Proceedings of the 36th WIC Symposium on Information Theory in the Benelux

and

The 5th Joint WIC/IEEE Symposium on Information Theory and Signal Processing in the Benelux

Université libre de Bruxelles, Brussels, Belgium
May 06–07, 2015

ULB

IEEE

Previous symposia

1. 1980 Zoetermeer, The Netherlands, Delft University of Technology
2. 1981 Zoetermeer, The Netherlands, Delft University of Technology
3. 1982 Zoetermeer, The Netherlands, Delft University of Technology
4. 1983 Haasrode, Belgium ISBN 90-334-0690-X
5. 1984 Aalten, The Netherlands ISBN 90-71048-01-2
6. 1985 Mierlo, The Netherlands ISBN 90-71048-02-0
7. 1986 Noordwijkerhout, The Netherlands ISBN 90-6275-272-1
8. 1987 Deventer, The Netherlands ISBN 90-71048-03-9
9. 1988 Mierlo, The Netherlands ISBN 90-71048-04-7
10. 1989 Houthalen, Belgium ISBN 90-71048-05-5
11. 1990 Noordwijkerhout, The Netherlands ISBN 90-71048-06-3
12. 1991 Veldhoven, The Netherlands ISBN 90-71048-07-1
13. 1992 Enschede, The Netherlands ISBN 90-71048-08-X
14. 1993 Veldhoven, The Netherlands ISBN 90-71048-09-8
15. 1994 Louvain-la-Neuve, Belgium ISBN 90-71048-10-1
16. 1995 Nieuwekerk a/d IJssel, The Netherlands ISBN 90-71048-11-X
17. 1996 Enschede, The Netherlands ISBN 90-365-0812-6
18. 1997 Veldhoven, The Netherlands ISBN 90-71048-12-8
19. 1998 Veldhoven, The Netherlands ISBN 90-71048-13-6
20. 1999 Haasrode, Belgium ISBN 90-71048-14-4
21. 2000 Wassenaar, The Netherlands ISBN 90-71048-15-2
22. 2001 Enschede, The Netherlands ISBN 90-365-1598-X
23. 2002 Louvain-la-Neuve, Belgium ISBN 90-71048-16-0
24. 2003 Veldhoven, The Netherlands ISBN 90-71048-18-7
25. 2004 Kerkrade, The Netherlands ISBN 90-71048-20-9
26. 2005 Brussels, Belgium ISBN 90-71048-21-7
27. 2006 Noordwijk, The Netherlands ISBN 90-71048-22-7
28. 2007 Enschede, The Netherlands ISBN 978-90-365-2509-1
29. 2008 Leuven, Belgium ISBN 978-90-9023135-8
30. 2009 Eindhoven, The Netherlands ISBN 978-90-386-1852-4
31. 2010 Rotterdam, The Netherlands ISBN 978-90-710-4823-4
32. 2011 Brussels, Belgium ISBN 978-90-817-2190-5
33. 2012 Enschede, The Netherlands ISBN 978-90-365-3383-6
34. 2013 Leuven, Belgium ISBN 978-90-365-0000-5
35. 2014 Eindhoven, The Netherlands ISBN 978-90-386-3646-7

Proceedings

Proceedings of the 36th Symposium on Information Theory in the Benelux and the 5th Joint WIC/IEEE Symposium on Information Theory and Signal Processing in the Benelux.
Edited by Jérémie Roland and François Horlin.
A catalogue record is available from the Université libre de Bruxelles Library.
ISBN: to come

The 36th Symposium on Information Theory in the Benelux and The 5th JointWIC/IEEE Symposium on Information Theory and Signal Processing in the Benelux have been organized by

Université libre de Bruxelles, Brussels, Belgium

http://sitb2015.ulb.ac.be

on behalf of the

Werkgemeenschap voor Informatie- en Communicatietheorie, the IEEE Benelux Information Theory Chapter and the IEEE Benelux Signal Processing Chapter.

Organizing committee
Jérémie Roland (ULB)
François Horlin (ULB)

Table of contents

Evidence-based discounting rule in Subjective Logic (extended abstract)

Boris Škorić[1], Sebastiaan J.A. de Hoogh[2], Nicola Zannone[1]
1) TU Eindhoven. {b.skoric, n.zannone}@tue.nl
2) Philips Research. sebastiaan.de.hoogh@philips.com

Abstract

We identify an inconsistency in Subjective Logic caused by the discounting operator '$\otimes$'. We propose a new operator, '$\boxtimes$', which resolves all the consistency problems. The new algebra makes it possible to compute Subjective Logic trust values (reputations) in arbitrarily connected trust networks. The material presented here is an excerpt of [3].

1 Subjective Logic

Subjective Logic (SL) [1] is a kind of 'fuzzy' logic that explicitly keeps track of uncertainties. The central concepts in SL are *evidence* and *opinions*. Let P be a proposition. Evidence about P is denoted as a vector (p, n), where p is the amount of evidence supporting P, and n the amount of evidence supporting $\neg P$. An *opinion* is a triplet $(b, d, u) \in [0, 1]^3$ satisfying $b + d + u = 1$. The b component is the 'belief' in proposition P, and it can be interpreted as the probability that P is provably true given the evidence. Likewise, the d is disbelief (belief in $\neg P$). The u is the uncertainty, the probability that nothing can be proven about P. There is a simple bijection between the evidence vector and the opinion based on it,

$$(b, d, u) = \frac{(p, n, 2)}{p + n + 2}; \qquad (p, n) = 2\frac{(b, d)}{u}. \tag{1}$$

This relation is based on an analysis of a posteriori probability distributions (beta distributions) [1]. Special points are Belief $B = (1, 0, 0)$, Disbelief $D = (0, 1, 0)$ and Uncertainty $U = (0, 0, 1)$. Triplets with $u = 0$ can only be reached with infinite amounts of evidence and are therefore often excluded from opinion space.

There are two important operations for combining opinions about the same proposition: *consensus* and *discounting*. The consensus operation simply adds up evidence vectors. Let $x = (x_\mathrm{b}, x_\mathrm{d}, x_\mathrm{u})$ be an opinion based on evidence (p_x, n_x) and $y = (y_\mathrm{b}, y_\mathrm{d}, y_\mathrm{u})$ an opinion based on (p_y, n_y). Then the combined evidence is $(p_x + p_y, n_x + n_y)$ and the corresponding opinion is given by

$$x \oplus y \stackrel{\mathrm{def}}{=} \frac{(x_\mathrm{u}y_\mathrm{b} + y_\mathrm{u}x_\mathrm{b}, x_\mathrm{u}y_\mathrm{d} + y_\mathrm{u}x_\mathrm{d}, x_\mathrm{u}y_\mathrm{u})}{x_\mathrm{u} + y_\mathrm{u} - x_\mathrm{u}y_\mathrm{u}}. \tag{2}$$

The consensus operation $\oplus$ is allowed only if the evidence in x and y is independent, otherwise 'double counting' of evidence occurs.

Discounting describes trust transitivity. Let Bob publish opinion y about proposition P. Let Alice have opinion x about Bob's trustworthiness. Then Alice's opinion about P is 'y discounted through x', which is denoted as $x \otimes y$ and defined as

$$x \otimes y \stackrel{\mathrm{def}}{=} (x_\mathrm{b}y_\mathrm{b},\ x_\mathrm{b}y_\mathrm{d},\ x_\mathrm{d} + x_\mathrm{u} + x_\mathrm{b}y_\mathrm{u}). \tag{3}$$

2 Problems with the $\otimes$ operator

The definition of $\otimes$ lacks a natural interpretation in evidence space. Let $z = x \otimes y$ in the Alice & Bob example above. The evidence vector (p_z, n_z) obtained using (1) is a messy function of p_x, p_y, n_x and n_y which under certain circumstances yields downright weird results. For instance, if $n_x = 0$, $n_y = 0$ and $p_y \gg p_x$, then $p_z \approx p_x/4$, which seems to imply that Alice's opinion about P is fully determined by x (which is not even an opinion about P), independent of y.

Furthermore, consider the following case. Alice has trust x in Bob. Bob gathers two independent evidence vectors, (p_1, n_1) and (p_2, n_2), about proposition P.
Scenario I: Bob forms two independent opinions, y_1 and y_2, based on the evidence. He publishes first y_1 and later y_2. Alice forms opinion $x \otimes y_1$ about P and later updates this to $(x \otimes y_1) \oplus (x \otimes y_2)$.
Scenario II: Bob combines his evidence and forms opinion $y_1 \oplus y_2$, which he publishes. Alice forms opinion $x \otimes (y_1 \oplus y_2)$ about P.
It is obvious that these scenarios should yield the same result for Alice. Yet the traditional discounting rule gives $x \otimes (y_1 \oplus y_2) \neq (x \otimes y_1) \oplus (x \otimes y_2)$. In SL the only correct expression is $x \otimes (y_1 \oplus y_2)$. We consider this to be a grave inconsistency in SL.

Next consider the trust network in Fig. 1. Due to the complicated mixup of evidence components in expressions of the form $x \otimes y$, combined with the prohibition on combining dependent evidence in $\oplus$ operations, it is impossible to write down a consistent SL result ('canonical expression' [2]) expressing the trust that node 1 has in node 6.

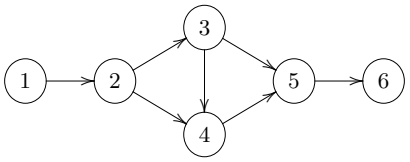


Figure 1: *Example of a trust network that is problematic for Subjective Logic.*

3 Bijection between evidence and opinion: Simplified derivation

We have found a simple way to obtain a bijection between evidence (p, n) and opinion $x = (b, d, u)$. Instead of looking at a posteriori probability distributions, we ask ourselves which natural constraints should be satisfied by such a bijection. If we impose the following conditions,

1. $b/d = p/n$
2. $b + d + u = 1$
3. $p + n = 0 \;\Rightarrow\; u = 1$
4. $p + n \to \infty \;\Rightarrow\; u \to 0$

then the relation between x and (p, n) can only be

$$x = (b, d, u) = \frac{(p, n, c)}{p + n + c} \quad ; \qquad (p, n) = c\frac{(b, d)}{u} \tag{4}$$

where $c > 0$ is a constant. Eq. (4) is precisely of the form (1), except for the constant '2' versus c. We make two important remarks: (i) The more generic mapping (4) is consistent with the $\oplus$ definition (2), i.e. the value of c does not matter, as long as all entities use the same c. (ii) The analysis of [1] can be re-done using a general constant c, and then still yields a consistent result. We see no reason to set $c = 2$.

4 New discounting operator: $\boxtimes$

We first define a new operation in Subjective Logic, *multiplication of a scalar and an opinion*. Let $x = (b, d, u)$ be an opinion based on evidence (p, n). Let $\lambda \geq 0$ be a scalar. In evidence space the product $\lambda \cdot x$ is defined as $(\lambda p, \lambda n)$. In opinion space this corresponds to the definition

$$\lambda \cdot x \stackrel{\text{def}}{=} \frac{(\lambda b, \lambda d, u)}{\lambda(b + d) + u}. \tag{5}$$

Next we define our new discounting operator '$\boxtimes$'. Let g be a function that maps opinions to $[0, 1]$, satisfying $g(B) = 1$ and $g(D) = 0$. We define

$$x \boxtimes y \stackrel{\text{def}}{=} g(x) \cdot y. \tag{6}$$

The function g can be chosen at will, depending on the context.
We refer to {SL with the new discounting operator} as Evidence-Based Subjective Logic (EBSL). EBSL avoids all the inconsistencies of the $\otimes$ operation,

- The expression $x \boxtimes y$ has a very simple interpretation in evidence space: Due to the disbelief and uncertainty present in x, only a fraction $g(x)$ of the evidence in y is accepted by the recipient.
- It holds that $x \boxtimes (y_1 \oplus y_2) = (x \boxtimes y_1) \oplus (x \boxtimes y_2)$, which is what is intuitively expected of a discounting operation.
- Due to the cleanness of the $\boxtimes$ operation, there is a strict separation between evidence on the one hand and the way it is carried over trust links on the other hand. Consequently EBSL can handle any trust network, no matter how complicated the graph. (See the next section.)

5 Arbitrary trust networks

Let opinion A_{ij} be the amount of trust that a node i has in node j, based on *direct* evidence, e.g. past interaction between i and j. We set the diagonal to $A_{ii} = U$. All nodes publish these direct opinions. Every node wants to know how much the other nodes can be trusted, and is willing to make use of the opinions published by others ('indirect evidence'). The mathematical problem is now to compute a meaningful reputation matrix R from A, giving proper weights to all the direct and indirect evidence. The diagonal of R is undefined, so we are free to set it arbitrarily. We set it to $B\mathbf{1}$, where $\mathbf{1}$ is the unit matrix. The following relation must be satisfied,

$$R = B\mathbf{1} \oplus (R \boxtimes A), \tag{7}$$

where the 'matrix multiplication' $R \boxtimes A$ is defined as $(R \boxtimes A)_{ij} = \oplus_k (R_{ik} \boxtimes A_{kj})$. Eq. (7) says that a reputation R_{ij} consists of a weighted sum of direct opinions A_{kj}, where the weights are determined by the reputations R_{ik}. Eq. (7) is a fixed-point equation. It can be solved e.g. by iterative methods such as repeatedly substituting (7) into itself. Experiments on synthetic as well as real data show fast convergence.

References

[1] A. Jøsang. A logic for uncertain probabilities. *International Journal of Uncertainty, Fuzziness and Knowledge-Based Systems*, 9(3):279–311, 2001.

[2] A. Jøsang, E. Gray, and M. Kinateder. Simplification and Analysis of Transitive Trust Networks. *Web Intelligence and Agent Systems*, 4:139–161, 2006.

[3] B. Škorić, S.J.A. de Hoogh, and N. Zannone. Flow-based reputation with uncertainty: Evidence-Based Subjective Logic, 2014. `http://arxiv.org/abs/1402.3319`.

Caching (a pair of) Gaussians

Giel J. Op 't Veld Michael C. Gastpar
School of Computer and Communication Sciences, EPFL
Lausanne, Switzerland
giel.optveld@epfl.ch michael.gastpar@epfl.ch

Abstract

A source produces i.i.d. vector samples from a Gaussian distribution, but the user is interested in only one component. In the cache phase, not knowing which component the user is interested in, a first compressed description is produced. Upon learning the user's choice, a second message is provided in the update phase so as to attain the desired fidelity on that component. We aim to find the cache strategy that minimizes the average update rate. We show that for Gaussian codebooks, the optimal strategy depends on whether or not the cache is large enough to make the vector conditionally independent. If it is, infinitely many equally optimal strategies exist. If it is not, we show that the encoder should project the source onto some subspace prior to coding. For a pair of Gaussians, we exactly characterize this projection vector.

1 Introduction

Nowadays, streaming-services draw a huge chunk of the available bandwidth. The *on-demand* aspect of video-on-demand results in an overload of individual requests at slightly different times of the day, albeit concentrated during peak hours. Caching is a strategy to move part of that load to off-peak times. During the night, a service could pre-load data onto your hard drive, taking an estimated guess of the content you might ask for during the day. If a user has a limited cache budget on his drive, what should the server put there in order to minimize traffic during the day? In this paper, we study these applied questions in a theoretical context of Gaussian vector sources.

A source produces length K vector samples of a Gaussian distribution, but the user is only interested in one of the components. In the cache phase, the encoder can code a first message up to cache rate R_c, without knowing the user's desired component. In the update phase, the user chooses component k uniformly, i.e., $p_Y(Y = k) = 1/K$ and reveals it to the encoder, who then sends an update at a rate R_u. The decoder then uses the cache and this update to construct a lossy representation of the k'th component at the desired fidelity. A schematic of this is depicted in Figure 1. Our goal is to find the caching strategy that minimizes the average update rate.

We show in this paper that for Gaussian codebooks the optimal coding strategy depends on whether or not the cache is sufficiently large so as to make the source components conditionally independent when conditioned on the cache. If that is so, Section 3.1 explains that there are infinitely many coding strategies that are equally optimal. If not, we argue in Section 3.2 that the encoder must project the source vector to a shorter vector. For a pair of Gaussians, we find this projection exactly; it turns out to be solely defined by the source's covariance and it does not change for different values of R_c.

All that we discuss in this paper relies on the successive refinability of Gaussian sources to connect the cache and update phase. For a general discussion we refer the reader to [1, 2]. For the (Gaussian) vector case, one should read [3] and [4] as its precursor. It describes the refinability of $\mathbf{X}_1$ to $\mathbf{X}_2$ as being possible if and only if their

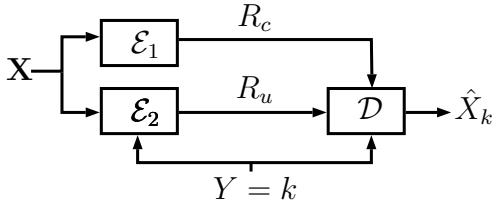


Figure 1: The caching scheme with cache rate R_c. After revealing $Y = k$ as the selection variable, $\mathcal{E}_2$ sends an update at rate R_u s.t. the decoder can retrieve $\hat{X}_k$.

covariances admit a semidefinite ordering $\Sigma_{X_1} \succeq \Sigma_{X_2}$. A discussion on the general rate-distortion function for Gaussian vectors was discussed in [5]. An attentive reader might also notice that our problem shows close resemblance to the Gray-Wyner system [6]. Namely, one could draw all the events of the user asking for one of the K components as K different decoders; the cache would then be their shared link and the required update their individual ones [7].

2 Definitions and Cache Rate-Distortion Function

Let $\mathbf{X}$ be an i.i.d. Gaussian vector source of dimension K, following the distribution $\mathcal{N}(0, \Sigma_{\mathbf{X}})$ with some potentially correlated covariance $\Sigma_{\mathbf{X}}$. That is, at each time instant, the source independently produces a vector sampled from this fixed distribution. We denote the source sample at time n by $\mathbf{X}(n)$, and we denote its kth component by $X_k(n)$, for $k = 1, 2, \ldots, K$. Independently of $\mathbf{X}$, a single random variable Y is drawn from the set $\{1, 2, \ldots K\}$ uniformly at random; we call it the selection variable.

We consider block coding of length N with two encoders. The first, referred to as the *cache encoder,* observes only $\{\mathbf{X}(n)\}_{n=1}^N$ and produces a description using NR_c bits, where R_c is called the *cache rate.* The second, referred to as the *update encoder,* gets to observe $\{\mathbf{X}(n)\}_{n=1}^N$ as well as the value of the random variable $Y = k$ and produces a description using $NR_u(k)$ bits, where $R_u(k)$ is called the *update rate* for the case $Y = k$. Hence, the average update rate of the encoder is given by $\overline{R}_u = \frac{1}{K}\sum_{k=1}^K R_u(k)$. Notation-wise, the sub- or superscript c stands for cache, while u stands for update.

Upon observing the realization y and both compressed descriptions, the decoder must output a sequence of estimates $\hat{X}_y(n)$ in such a way as to satisfy

$$\frac{1}{N}\sum_{n=1}^{N}\Big(X_y(n) - \hat{X}_y(n)\Big)^2 \leq D_u.$$

The question addressed in this paper is to characterize, for a fixed caching rate R_c, the smallest average update rate $\overline{R}_u$ for which the distortion constraint can be satisfied (irrespective of the value of Y).

For the cache phase, we allow the server to code any $\mathbf{Z}$ that is jointly Gaussian with the source. For large R_c, one can easily argue that Gaussian codebooks are optimal; for small R_c, it remains a difficult question that we unfortunately can not yet address in this article. In the update phase, one can compute $\hat{\mathbf{X}}^c = \mathbb{E}[\mathbf{X}|\mathbf{Z}]$ as the MSE-estimate of the source and subsequently the error as:

$$\mathbf{D}^c = \mathbb{E}[(\mathbf{X} - \hat{\mathbf{X}}^c)(\mathbf{X} - \hat{\mathbf{X}}^c)^T] \preceq \Sigma_{\mathbf{X}}. \tag{1}$$

The semidefinite ordering $\mathbf{D}^c \preceq \Sigma_{\mathbf{X}}$ means that $\Sigma_{\mathbf{X}} - \mathbf{D}^c$ is positive semidefinite. It yields an engineering perspective: any real symmetric matrix $\mathbf{D}^c$ that satisfies this ordering is an achievable Gaussian codebook.

At this point, there is no operational interest for a first estimate $\hat{\mathbf{X}}^c$ or its error $\mathbf{D}^c$. However, $\mathbf{D}^c$ has theoretical value. Namely, for any $\hat{\mathbf{X}}$ jointly (not necessarily Gaussian) distributed with $\mathbf{X}$, the mutual information satisfies

$$I(\mathbf{X};\hat{\mathbf{X}}) \geq \frac{1}{2}\log\frac{|\Sigma_{\mathbf{X}}|}{|\mathbf{D}|} = R(\mathbf{D}). \tag{2}$$

The last step in (2) is met with equality if indeed we use Gaussians codebooks, i.e., $\hat{\mathbf{X}}^c = \mathbb{E}[\mathbf{X}|\mathbf{Z}]$ with $\mathbf{Z} = \mathbf{X} + \mathbf{W}$ where $\mathbf{W}$ is independent from $\mathbf{X}$ and Gaussian as well [5, Lemma 2]. Thus, we may not need $\mathbf{D}^c$, but we can use it to characterize the rate associated with the cache phase. Therefore, a cache strategy that yields a particular error covariance $\mathbf{D}^c$ must have had a rate satisfying

$$R_c \geq \frac{1}{2}\log\frac{|\Sigma_{\mathbf{X}}|}{|\mathbf{D}^c|}. \tag{3}$$

Since our goal is to minimize $\overline{R}_u$ for a *fixed* R_c we can reverse (3) and state the following:

Definition 1. *A (valid)* caching strategy *is any real symmetric matrix* $\mathbf{D}^c$ *that satisifies the following two conditions:*

1. $|\mathbf{D}^c| = |\Sigma_{\mathbf{X}}|e^{-2R_c}$*, the rate-constraint* (3)*.*
2. $0 \preceq \mathbf{D}^c \preceq \Sigma_{\mathbf{X}}$*, the semidefinite ordering constraint* (1)*.*

In the update phase, $Y = k$ is revealed and consequently only an interest for X_k remains. Both the encoder and decoder have access to the side information presented by the cache. The MSE-estimator $\mathbb{E}[X_k|\mathbf{Z}]$ forms the first step to an estimate $\hat{X}_k$ and since $p(X_k|\mathbf{Z})$ is also a normal distribution, the update rate is lower bounded by the Gaussian rate distortion function. Namely, Gaussians are successively refinable [1, 3], which allows to combine the messages from the first and second phase. The variance of $p(X_k|\mathbf{Z})$ is simply the k'th diagonal entry of $\mathbf{D}^c$ and thus we have:

$$R_u(k) \geq \frac{1}{2}\log^+\frac{D^c_{kk}}{D_u},$$

which yields an *average* update rate for this construction:

$$\overline{R}_{u,\mathbf{D}^c}(D_u) = \frac{1}{K}\sum_{k=1}^{K}\frac{1}{2}\log^+\frac{D^c_{kk}}{D_u}. \tag{4}$$

The subscript $\mathbf{D}^c$ emphasizes that $\overline{R}_u$ depends on a particular cache strategy $\mathbf{D}^c$.

Definition 2. *The* cache rate-distortion function *is the average update rate needed to attain distortion* D_u *on any component, minimized over all caching strategies:*

$$\overline{R}_u(D_u) = \min_{\mathbf{D}^c} \quad \overline{R}_{u,\mathbf{D}^c}(D_u) \quad s.t. \begin{cases} 0 \preceq \mathbf{D}^c \preceq \Sigma_{\mathbf{X}} \\ |\mathbf{D}^c| = |\Sigma_{\mathbf{X}}|e^{-2R_c} \end{cases} \tag{5}$$

Our search for the best caching strategy thus translates to: What choice of $\mathbf{D}^c$ minimizes (4) given the search space of matrices set by Definition 1? What distortion profile for the cache phase minimizes that rate needed for the update phase?

Unfortunately, the cache rate-distortion function (5) is a minimization over a concave function. In many Gaussian source coding problems, the optimization variable $\mathbf{D}$ is found in the denominator, which is convex. It is now found in the numerator, which makes it concave and thus hard to solve. In the next section, we will argue on the different optimal caching strategies for small and large R_c.

3 Optimal Caching Strategies

A different way of writing (5) is to pull the sum of (4) inside the log:

$$\overline{R}_u(D_u) = \min_{\mathbf{D}^c} \frac{1}{2K} \log^+ \frac{\prod_k D^c_{kk}}{D_u^K} \quad \text{s.t.} \quad \begin{cases} 0 \preceq \mathbf{D}^c \preceq \Sigma_{\mathbf{X}} \\ |\mathbf{D}^c| = |\Sigma_{\mathbf{X}}| e^{-2R_c} \end{cases} \tag{6}$$

which leads to the insight that the numerator is lower bounded by the Hadamard inequality $\prod_k \mathbf{D}^c_{k,k} \geq |\mathbf{D}^c|$, hence

$$\overline{R}_u(D_u) \geq \frac{1}{2K} \log^+ \frac{|\mathbf{D}^c|}{D_u^K}, \tag{7}$$

and in turn $|\mathbf{D}^c|$ is bounded (or fixed even) by R_c, see again Definition 1. An interesting read on the relationship between the Hadamard inequality and Gaussians was presented in for example [8, Chapter 17]. The difference between a product of the diagonal entries of a covariance and its determinant stems from $h(\mathbf{X}) \leq \sum_k h(X_k)$. The mutual exclusiveness of the update phase, where the encoder only refines the one component the decoder asked for, combined with an objective to minimize the *average* update rate is the reason for why this product $\prod_k D^c_{kk}$ popped up instead of $|\mathbf{D}^c|$.

Interestingly, there are two distinct coding strategies depending on whether the lower bound (7) can be met or not. The Hadamard Inequality is met with equality if and only if the matrix $\mathbf{D}^c$ is diagonal. Algebraically, this is not trivial as one cannot have a diagonal $\mathbf{D}^c$ and satisfy $\mathbf{D}^c \preceq \Sigma_{\mathbf{X}}$ at the same time if $|\mathbf{D}^c|$ is too large. In terms of information theory, a diagonal cache distortion implies that the components of $\mathbf{X}$ become independent when conditioned on the cache. This is impossible if R_c is too small. These algebraic and information theoretic arguments are the same. In the next subsection, we elaborate on a threshold R^* on R_c and show that there are infinitely many equally optimal cache strategies if the rate is larger than R^*. In the subsection thereafter, we show that for smaller rates, the optimal strategy requires a dimensionality reduction. The cache should be a particular projection of the source components to some space. For a pair of Gaussians, we derive this projection exactly.

3.1 Large cache rates

The Hadamard inequality that was the lower bound in (7) hits equality if and only if a matrix is diagonal. Hence there must exist a decomposition $\Sigma_{\mathbf{X}} = \mathbf{D}^c + \Sigma_{\hat{\mathbf{X}}}$ where $\Sigma_{\hat{\mathbf{X}}}$ and $\mathbf{D}^c$ are both positive semidefinite* and $\mathbf{D}$ is diagonal. For this we derive:

Theorem 1. *For any cache rate $R_c \geq R^*$, there exists a caching strategy $\mathbf{D}^c$ that achieves the lower bound on the average update rate* (7)*, where R^* is the solution to*

$$\min_{\mathbf{D}^c} \ \frac{1}{2} \log \frac{|\Sigma_{\mathbf{X}}|}{|\mathbf{D}^c|} \quad \textit{s.t.} \quad \begin{cases} 0 \preceq \mathbf{D}^c \preceq \Sigma_{\mathbf{X}}, \\ \mathbf{D}^c \textit{ is diagonal.} \end{cases} \tag{8}$$

Proof. Recall that $R_c = \frac{1}{2} \log \frac{|\Sigma_{\mathbf{X}}|}{|\mathbf{D}^c|}$ implies the reverse relation on the determinant, $|\mathbf{D}^c| = |\Sigma_{\mathbf{X}}| e^{-2R_c}$. Suppose that $\mathbf{D}^*$ is the distortion matrix that minimizes (8) and let R^* be the cache rate associated to this point. Evidently, there cannot be a $\mathbf{D}'$ that is

*Demanding that $\Sigma_{\mathbf{X}}$ decomposes into a sum of two positive semidefinite matrices is equivalent to demanding $\mathbf{D}^c \preceq \Sigma_{\mathbf{X}}$ like we did before.

diagonal and has a determinant larger than that of $\mathbf{D}^*$ (or equivalenty, an R_c smaller than R^*), otherwise $\mathbf{D}'$ would have been the minimizer of (8). On the other end, for all $R_c \geq R^*$ there does exist a diagonal candidate caching strategy $\mathbf{D}'$. Namely, for diagonal matrices $\mathbf{D}' \preceq \mathbf{D}^*$ holds if and only if $D'_{i,i} \leq D^*_{i,i}$ $\forall i$. So one can construct another distortion matrix $\mathbf{D}'$ by decreasing the values on some arbitrary subset of the diagonal entries of $\mathbf{D}^*$. In doing so, any determinant $|\mathbf{D}'| \leq |\mathbf{D}^*|$ (and thus any $R_c \geq R^*$) can be achieved by a matrix that satisfies the chain $\mathbf{D}' \preceq \mathbf{D}^* \preceq \Sigma_{\mathbf{X}}$ and is thus both diagonal and achievable. □

In the proof we constructed matrices $\mathbf{D}'$ with any particular determinant in the region $R_c > R^*$ by decreasing some diagonal entries of $\mathbf{D}^*$, the solution to (8). It does not matter which entries we use for this or by what amount we decrease them, as long as the resulting determinant has the value we are after. Hence, in this high-R_c regime, there exists infinitely many diagonal $\mathbf{D}'$ with the same determinant that thus all achieve the same lower bound on $\overline{R}_u$ (7); they are equally optimal.

The minimization of (8) is simply a MaxDet problem, which can be solved efficiently numerically. The constraint that $\mathbf{D}^c$ must be diagonal is also a simple linear constraint, namely one can replace it by $\mathbf{D}^c - \operatorname{diag}(\mathbf{D}^c) \preceq 0$ and we already had $\mathbf{D}^c \succeq 0$. This brings about an interesting contrast with the original problem: Finding the general optimal distortion profile for our problem was a hard-to-solve concave minimization. The high-rate regime, however, now appears to be characterizable by a convex problem which is easily solvable. To our knowledge, we do not know of any analytical expression for $\mathbf{D}^c$ that minimizes (8), except for some special cases, one of which we will explain now.

3.1.1 Example: a Pair of Gaussians

Theorem 2. *For a pair of Gaussians the minimizer of* (8) *is* $R^* = \frac{1}{2}\log\frac{1+|\rho|}{1-|\rho|}$*, which is achieved by a distortion matrix*

$$\mathbf{D}^* = \begin{bmatrix} \sigma_1^2(1-|\rho|) & 0 \\ 0 & \sigma_2^2(1-|\rho|) \end{bmatrix}. \tag{9}$$

Proof. Let us find a decomposition of $\Sigma_{\mathbf{X}} = \mathbf{D}^c + \Sigma_{\hat{\mathbf{X}}}$ of a diagonal $\mathbf{D}^c$ by setting $\mathbf{D}^c = \operatorname{diag}(\alpha^2, \beta^2)$ and work out:

$$\begin{bmatrix} \sigma_1^2 & \rho\sigma_1\sigma_2 \\ \rho\sigma_1\sigma_2 & \sigma_2^2 \end{bmatrix} = \begin{bmatrix} \alpha^2 & 0 \\ 0 & \beta^2 \end{bmatrix} + \begin{bmatrix} \sigma_1^2-\alpha^2 & \rho\sigma_1\sigma_2 \\ \rho\sigma_1\sigma_2 & \sigma_2^2-\beta^2 \end{bmatrix}.$$

Such a decomposition yields positive semidefinite matrices (which is equivalent to $0 \preceq \mathbf{D}^c \preceq \Sigma_{\mathbf{X}}$) and is a valid caching strategy if and only if the following conditions are met:

1. $0 \leq \alpha^2 \leq \sigma_1^2$ and $0 \leq \beta^2 \leq \sigma_2^2$ ($\mathbf{D}^c$ is PSD).
2. $\frac{\rho^2\sigma_1^2\sigma_2^2}{(\sigma_1^2-\alpha^2)(\sigma_2^2-\beta^2)} \leq 1$ ($\Sigma_{\hat{\mathbf{X}}}$ is PSD).
3. $\alpha^2\beta^2 = |\mathbf{D}^c| =$ fixed (cache rate constraint).

Let us start by evaluating the 2nd condition:

$$\begin{aligned} 0 &\leq (\sigma_1^2-\alpha^2)(\sigma_2^2-\beta^2) - \rho^2\sigma_1^2\sigma_2^2 \\ &= -\alpha^2\sigma_2^2 - \frac{|\mathbf{D}^c|}{\alpha^2}\sigma_1^2 + |\Sigma_{\mathbf{X}}| + |\mathbf{D}^c|, \end{aligned} \tag{10}$$

where we replaced $\alpha^2\beta^2 = |\mathbf{D}^c|$ and $\beta^2 = \frac{|\mathbf{D}^c|}{\alpha^2}$ by condition 3. Moreover, note that for 2×2 matrices we have $|\Sigma_\mathbf{X}| = \sigma_1^2\sigma_2^2(1-\rho^2)$. The right hand side is convex. If it has two roots α_-^2 and α_+^2, all $\alpha^2 \in [\alpha_-^2, \alpha_+^2]$ are valid solutions, given that condition 1 is satisfied. Hence, if there are two roots then there exist infinitely many $\mathbf{D}^c$ that are equally optimal.

Equation (10) has only one root -and thus yields only one optimal $\mathbf{D}^c$- if the minimum of the right hand side is exactly at zero. By setting its derivative to zero, one finds that $\alpha_{min}^2 = \sqrt{|\mathbf{D}^c|\frac{\sigma_1^2}{\sigma_2^2}}$. Substituting α_{min}^2 into (10) and demanding equality:

$$\begin{aligned} 0 &= |\Sigma_\mathbf{X}| + |\mathbf{D}^c| - 2\sqrt{|\mathbf{D}^c|\sigma_1^2\sigma_2^2} \\ &= |\mathbf{D}^c|^2 - 2|\mathbf{D}^c|\sigma_1^2\sigma_2^2(1+\rho^2) + (\sigma_1^2\sigma_2^2(1-\rho^2))^2. \end{aligned}$$

This follows from pulling $-2\sqrt{|\mathbf{D}^c|\sigma_1^2\sigma_2^2}$ to the left hand side, squaring both sides and then pulling it back, while at the same time filling in $|\Sigma_\mathbf{X}| = \sigma_1^2\sigma_2^2(1-\rho^2)$. This is a new quadratic equation, which now revolves around $|\mathbf{D}^c|$ instead of α^2. Its roots are

$$\begin{aligned} |\mathbf{D}^c|_\pm^* &= \sigma_1^2\sigma_2^2(1+\rho^2) \pm \sqrt{\sigma_1^4\sigma_2^4(1+\rho^2)^2 - \sigma_1^4\sigma_2^4(1-\rho^2)^2} \\ &= \begin{cases} \sigma_1^2\sigma_2^2(1-|\rho|)^2 & \text{valid,} \\ \sigma_1^2\sigma_2^2(1+|\rho|)^2 & \text{invalid (since } |\mathbf{D}^c| > |\Sigma_\mathbf{X}| \text{ cannot be).} \end{cases} \end{aligned}$$

This bifurcation point $|\mathbf{D}^c|^*$ corresponds to a cache rate

$$R^* = \frac{1}{2}\log\frac{|\Sigma_\mathbf{X}|}{|\mathbf{D}^c|^*} = \frac{1}{2}\log\frac{\sigma_1^2\sigma_2^2(1-\rho^2)}{\sigma_1^2\sigma_2^2(1-|\rho|)^2} = \frac{1}{2}\log\frac{1+|\rho|}{1-|\rho|},$$

and marks the transition from having no to one and then to infinitely many $\mathbf{D}^c$ that have no correlation. We denote the actual distortion profile that achieves this rate $\mathbf{D}^*$ (9) and find it by filling $|\mathbf{D}^c|^* = \sigma_1^2\sigma_2^2(1-|\rho|)^2$ into $\alpha_{min}^2 = \sqrt{|\mathbf{D}^c|\frac{\sigma_1^2}{\sigma_2^2}}$ and $\beta^2 = \frac{|\mathbf{D}^c|}{\alpha^2}$. □

The value $R^* = \frac{1}{2}\log\frac{1+|\rho|}{1-|\rho|}$ also came forward in [7] as Wyner's Common Information for a pair of Gaussians.

3.2 Small Cache Rates for a Pair of Gaussians

For R_c smaller than the R^* of Theorem 1, no $\mathbf{D}^c$ can close the Hadamard inequality, but perhaps we can find another achievable lower bound. Here, we find this optimal strategy for a pair of Gaussians, which shows a strong connection to Theorem 2. For general dimensions, the problem remains open. One thing that is clear is the following:

Lemma 1. *If $R_c \leq R^*$, the $\mathbf{D}^c$ that minimizes* (6) *yields $\mathbf{D}^c \preceq \Sigma_\mathbf{X}$, but not $\mathbf{D}^c \prec \Sigma_\mathbf{X}$.*

We will not fully prove this here, but imagine that $\bar{\mathbf{D}}$ is some candidate strategy that yields $\bar{\mathbf{D}} \prec \Sigma_\mathbf{X}$. Since the ordering is not strict, we have room to rotate the matrix. Determinants are rotation-invariant, hence the R_c required for this rotated distortion profile is the same (3). The key insight is that rotation can always further minimize the product of the main diagonal, e.g., by bringing the matrix closer to eigendecomposition.

The difference between $\mathbf{D}^c \prec \Sigma_\mathbf{X}$ and $\mathbf{D}^c \preceq \Sigma_\mathbf{X}$ is that the latter implies $\exists v$ such that $v^T(\Sigma_\mathbf{X} - \mathbf{D})v = 0$; there exists a direction of which one learns nothing by

observing the cache. In other words, the cache encoder must have projected $\mathbf{X}$ onto some subspace prior to coding. For a *pair* of Gaussians, a lower-dimensional coding strategy simply means one codes a representation of $v^T\mathbf{X}$ for any (normalized) vector v in the cache, rather than $\mathbf{X}$ itself. The code in the cache can be represented as $v^T\mathbf{X}+W$ where W is a Gaussian noise and independent of $\mathbf{X}$. Then the error is found as $\mathbf{D}^c = [||\mathbf{X} - \mathbb{E}[\mathbf{X}|v^T\mathbf{X}+W]||^2]$, which can be worked out completely using channel models for lossy representations, e.g., [8, Chapter 10]. In short, any caching strategy featuring a projection to a vector v leads to a Schur complement:

$$\mathbf{D}^c(R_c) = \Sigma_{\mathbf{X}} - (1-e^{-2R_c})\frac{1}{v^T\Sigma_{\mathbf{x}}v}\Sigma_{\mathbf{X}}vv^T\Sigma_{\mathbf{X}}. \tag{11}$$

We specifically express this matrix as a function of R_c. Even the optimal choice of a vector v could in principle be different for different R_c, but for a pair of Gaussians we will prove that this is actually not the case. Note that (11) always satisfies both conditions of Definition 1 by construction. The border case $\mathbf{D}^*$ is also still a one-dimensional coding operation. We derived $\mathbf{D}^*$ algebraically, we can plug it into (11) and solve for the vector v that could have led us to it. The particular vector associated to $\mathbf{D}^*$ turns out to be of more importance than simply the border case:

Theorem 3. *If for a pair of Gaussians $R_c \leq \frac{1}{2}\log\frac{1+|\rho|}{1-|\rho|}$, then the caching strategy that uniquely minimizes* (6) *requires one to code $v^{*T}\mathbf{X}$ with*

$$v^* = \frac{1}{\sqrt{\mathrm{tr}\left(\Sigma_{\mathbf{X}}\right)}}\begin{bmatrix}\sigma_2\\ \mathrm{sign}(\rho)\cdot\sigma_1\end{bmatrix}. \tag{12}$$

Proof. By Lemma 1 we know it suffices to constrain the search space of $\mathbf{D}^c$ to those we can describe by means of (11). Hence, we can plug (11) into (6) and minimize over all v such that $v^Tv=1$. To find the optimal v it suffices to look at $\arg\min\prod_{k=1,2}D^c_{kk}$:

$$\arg\min_{\substack{v\\ :v^Tv=1}}\left(\sigma_1^2 - \frac{1-e^{-2R_c}}{v^T\Sigma_{\mathbf{x}}v}\left(\Sigma_{\mathbf{X}}vv^T\Sigma_{\mathbf{X}}\right)_{1,1}\right)\cdot\left(\sigma_2^2 - \frac{1-e^{-2R_c}}{v^T\Sigma_{\mathbf{x}}v}\left(\Sigma_{\mathbf{X}}vv^T\Sigma_{\mathbf{X}}\right)_{2,2}\right)$$

For a 2×2 matrix, one can work out the expression above by hand; it is not hard, but for length constraints we choose to omit this from this paper. Its derivative with respect to v has a clear root at (12), regardless of R_c. The $\mathrm{sign}(\rho)$ then ensures one picks the minimum rather than a maximum. □

As a closing comment, let us briefly explain where v^* comes from and what it entails. The vector can be found at the border case of $R_c = R^*$ by setting (11) equal to (9) and solve for v. As for the intuition, every positive semidefinite matrix can be uniquely represented by the ellipsoid $\mathcal{E}_{\mathbf{A}} = \{v : v^T\mathbf{A}^{-1}v = 1\}$. Its semiprincipal axes match the eigenvectors of $\mathbf{A}$ and have lengths equal to $\sqrt{\lambda_i}$. In Figure 2 we plot both $\mathcal{E}_{\Sigma_{\mathbf{X}}}$ and $\mathcal{E}_{\mathbf{D}^*}$. Recall that $\mathbf{D}^*$ is the covariance matrix with the largest possible determinant that still satisfies $\mathbf{D}^c \preceq \Sigma_{\mathbf{X}}$ without having any correlation. Since it is the border case, $\mathcal{E}_{\Sigma_{\mathbf{X}}}$ and $\mathcal{E}_{\mathbf{D}^*}$ touch (that is the impact of having $\mathbf{D}^c \preceq \Sigma_{\mathbf{X}}$ rather than $\mathbf{D}^c \prec \Sigma_{\mathbf{X}}$). Even more so, the vector where these ellipses touch is the orthogonal complement to our coding vector v^*; the cache provides information on all directions spanned by the source, except the one orthogonal to the one we coded.

A second consequence is that, since one should use the same vector to code $v^{*T}\mathbf{X}$ for all $R_c \leq R^*$, all resulting $\mathcal{E}_{\mathbf{D}(R_c)}$ touch $\mathcal{E}_{\Sigma_{\mathbf{X}}}$ at this same orthogonal complement. In other words, $\mathcal{E}_{\mathbf{D}^c(R_c)}$ is sandwiched between $\mathcal{E}_{\mathbf{D}^*}$ and $\mathcal{E}_{\Sigma_{\mathbf{X}}}$. The result is that for a

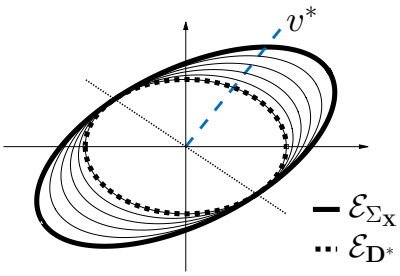


Figure 2: Ellipse of $\Sigma_{\mathbf{X}}$ and $\mathbf{D}^*$, together with the coding transform vector v^* (dashed) and its orthogonal complement ${}^1/\sqrt{\mathrm{tr}\,(\Sigma_{\mathbf{X}})}[\sigma_1,\ -\mathrm{sign}(\rho)\sigma_2]^T$ (dotted) that intersects the points where both ellipses touch. The thinner ellipses in between are the optimal $\mathbf{D}^c(R_c)$ for increasing R_c, coded with the same v^*, showing the ordering of (13).

sequence of cache rates $0 \leq R_1 \leq R_2 \leq \cdots \leq R_\ell \leq R^*$, the caching strategies that minimize (6) admit a semidefinite ordering:

$$\Sigma_{\mathbf{X}} \succeq \mathbf{D}^c(R_1) \succeq \mathbf{D}^c(R_2) \succeq \cdots \succeq \mathbf{D}^c(R_\ell) \succeq \mathbf{D}^*. \tag{13}$$

Hence, as a conclusion that stands apart from the goal of this paper, the Gaussian coding strategies that minimize the gap on the Hadamard Inequality for increasing rates form a Markov chain and are because of this successively refinable.

References

[1] W. Equitz and T. Cover, "Successive refinement of information," *Information Theory, IEEE Transactions on*, vol. 37, no. 2, pp. 269–275, 1991.

[2] B. Rimoldi, "Successive refinement of information: characterization of the achievable rates," *Information Theory, IEEE Transactions on*, vol. 40, no. 1, pp. 253–259, 1994.

[3] J. Nayak, E. Tuncel, D. Gunduz, and E. Erkip, "Successive refinement of vector sources under individual distortion criteria," *Information Theory, IEEE Transactions on*, vol. 56, no. 4, pp. 1769–1781, April 2010.

[4] H. Wang and P. Viswanath, "Vector gaussian multiple description with individual and central receivers," *Information Theory, IEEE Transactions on*, vol. 53, no. 6, pp. 2133–2153, June 2007.

[5] J. Xiao and Q. Luo, "Compression of correlated gaussian sources under individual distortion criteria," in *43rd Allerton Conference on Communication, Control, and Computing*, 2005, pp. 438–447.

[6] R. Gray and A. Wyner, "Source coding for a simple network," *Bell System Technical Journal, The*, vol. 53, no. 9, pp. 1681–1721, Nov 1974.

[7] G. Xu, W. Liu, and B. Chen, "Wyner's common information: Generalizations and A new lossy source coding interpretation," *CoRR*, vol. abs/1301.2237, 2013. [Online]. Available: http://arxiv.org/abs/1301.2237

[8] T. M. Cover and J. A. Thomas, *Elements of information theory (2. ed.)*. Wiley, 2006.

Automated Tissue Microarray Image Processing in Digital Pathology

Yves-Rémi Van Eycke[1,3], Olivier Debeir[1], Laurine Verset[2],
Pieter Demetter[2], Isabelle Salmon[2,3] Christine Decaestecker[1,3]
Université Libre de Bruxelles (U.L.B.)
[1]Laboratories of Image, Signal processing and Acoustics
Avenue Franklin Roosevel 50, 1050 Brussel, Belgium
[2]Pathology Department, Erasme Hospital
Route de Lennik 808, 1070 Bruxelles, Belgium
[3]DiaPath, Center for Microscopy and Molecular Imaging (CMMI)*
Rue Adrienne Bolland, 8, 6041 Gosselies, Belgium
yveycke@ulb.ac.be odebeir@ulb.ac.be laurine.verset@erasme.ulb.be
pieter.demetter@erasme.ulb.ac.be isabelle.salmon@erasme.ulb.ac.be cdecaes@ulb.ac.be

Abstract

Tissue microarray (TMA) is a widely used histological technology to analyze protein expression in various tissue samples at once. Combined with slide scanning and image analysis, this approach enables accurate quantitative analyses of protein biomarkers. To take into account the clinical data related to the tissue samples, it is necessary to store and to link meta-information with each sample included in the TMA. In this paper, we propose a method to make this specific linking between such information and TMA slide images. This method is then combined with a registration process to merge information from images of consecutive TMA slides.

1 Introduction

Tissue microarray (TMA) is a widely used histological technology to analyze various tissue samples at once [1]. It consists of a paraffin block in which hundreds of small cylindrical tissue samples (cores), extracted with a needle from larger tissue blocks, are arranged into arrays. Submitting TMA slices to immunohistochemistry (IHC) makes possible to evidence protein expression patterns in large collections of tissue samples in a standardized way. Combined with slide scanning and image analysis, this approach enables fast and accurate quantitative analyses of protein biomarkers, which are useful indicators for diagnosis, prognosis and therapeutic decisions [2]. As detailed below, TMAs are grids of tissue cores. Each core consists of a specific tissue sample, which is described by means of metadata (such as the patient ID, the lesion diagnosis, etc.) held in a "Design" file (required to carry out the TMA block). Depending of the tissue sample origins and characteristics, the data extracted from the cores can be processed differently. That is why each core is considered as a region of interest (ROI) in the image and its correct identification is critical for further analyses. Linking manually the metadata to each core is time consuming and error prone. Most of the commercially available solutions provide interactive grid fitting tools with limited tolerance to deformation [3]. These solutions rapidly require numerous manual interactions when TMA slides present defects, in particular when cores are shifted or lost. Thus we developed a tool which automatically identifies or interpolates the (present or absent) cores and locates each of them in the TMA grid to correctly link it with its own metadata. Characterizing complex cellular mechanisms, such as those involved in cancers, often requires to evaluate the colocalization of several biomarkers. This information can be collected by imaging and registering adjacent sections from the same TMA block and

*This research unit was funded by the Fonds Yvonne Boël (Brussels, Belgium), the European Regional Development Fund and the Walloon Region.

which evidence different proteins revealed by IHC. Algorithms for registration of whole-slide images have recently been proposed [4–7]. However, a TMA image significantly differs from a whole-slide image (grid of very similar circular samples with geometrical distorsions and sample losses, see Figure 1). We thus developed a tool which takes into account the particular layout of a TMA to ease and to accelerate the registration process. A short version of this study has been submitted to EMBC2015.

2 Material

To test our tool we used 7 TMA slides consisting of lymph node and rectal tissue cores organized in several subgrids. These 4-µm-thick slides were processed using IHC to evidence in brown (using DAB) expression of different proteins (α-SMA, IGF1, IGFBP2, IGF1R, Ki67, BAX and BCL2, see Table 2) whereas the tissue is counterstained in blue (HEM). The slides were then scanned using a Hamamatsu Nanozoomer 2.0-HT. The TMA coordinate system is also very specific. Each TMA grid is divided into subgrids containing the cores. Each core is associated to a alphanumeric coordinate as shown in Figure 1. The TMA organisation into subgrids is mentioned in the design file (required for constructing the TMA block).

Shortname	**Long name**	**Location**
α-SMA	Alpha smooth muscle actin	Myofibroblasts and smooth muscle cells in vessel walls, gut wall.
IGF1	Insulin-like growth factor 1	Extracellular
IGFBP2	Insulin-like growth factor-binding protein 2	Extracellular
IGF1R	Insulin-like growth factor 1 receptor	Present on the membrane of epithelial cells.
Ki67	/	Nuclear staining in proliferating cells
BAX	BCL2-associated X protein	Widely distributed cytoplasmic staining. Present in B-lymphocytes
BCL2	B-cell CLL/lymphoma 2	Present on the membrane of B-lymphocytes.

Table 1: Detailed description of the proteins targeted by IHC.

3 Method

The method used to extract the TMA core positions and to register core images is organized into 6 steps:

1. Tissue detection and core labeling
2. Identification of the 2D orientation of the TMA grid and average inter-core distance computation
3. Subgrid clustering

4. TMA core addressing
5. Missing core identification
6. Color deconvolution and fine registration of TMA image pairs

3.1 Tissue detection and core labeling

The low-resolution image (typically 2000x1200 pixels) is smoothed using a median filter and segmented in BLOBs (Binary Large Objects) using an Otsus threshold [8]. Labels are assigned to each BLOB, which is then filtered using its convex area. The bounding box and the centroid (mean position) of each filtered blob are then extracted.

3.2 TMA orientation Identification and inter-core distance computation

To detect the main directions of the whole grid (and so identifying its 2D orientation), we first compute the pairwise distances between the blobs using the properties extracted in the previous step (Figure 2A). We then pick the 4-nearest neighbours (Figure 2B-C).

Those neighbours are clustered into two categories according to their angles: the neighbours aligned horizontally and the neighbours aligned vertically. The median orientation is then extracted for each of the two groups. Those two values are considered as the main directions of the TMA grid.

When working with TMAs, the orientation matters. In particular, undetected section flipping leads to incorrect linking between the cores and the metadata. This is the reason why the last subgrid is let partially or totally empty (see Figure 1). To detect the (partially) empty subgrid the BLOB centroids are projected on the main directions computed in the previous step. Local density of the core is then approximated (see Figure 3). By comparing the mean density in the first half and the mean density in the second half of each chart we can efficiently detect in which quadrant of the image the (semi-)empty subgrid is.

3.3 Subgrid clustering

Subgrid clustering is done using the two charts obtained from the previous step (see Figure 3) for the two main directions. The TMA design file provides the number of subgrids in the image and their position relative to each other. If there are x subgrids

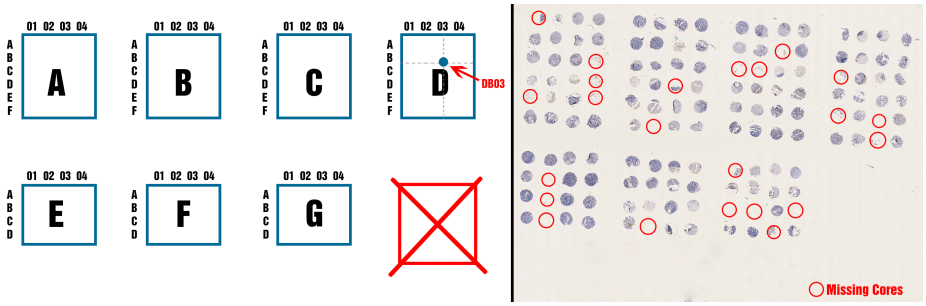


Figure 1: The coordinate system used in TMAs and a corresponding TMA image (where missing cores are located). A letter is assigned to each subgrid starting from the upper left corner. Each row is assigned to a letter and each column to a number. The last subgrid is partially or totally empty and acts as mistake proofing.

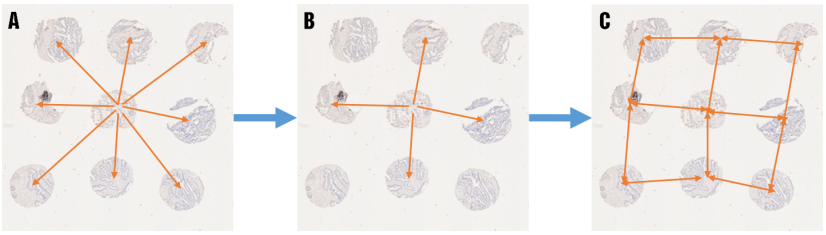


Figure 2: A. Pairwise distances between all the BLOB centroids are computed. B. The 4-nearest neighbours of one core. C. The nearest neighbours for all the cores.

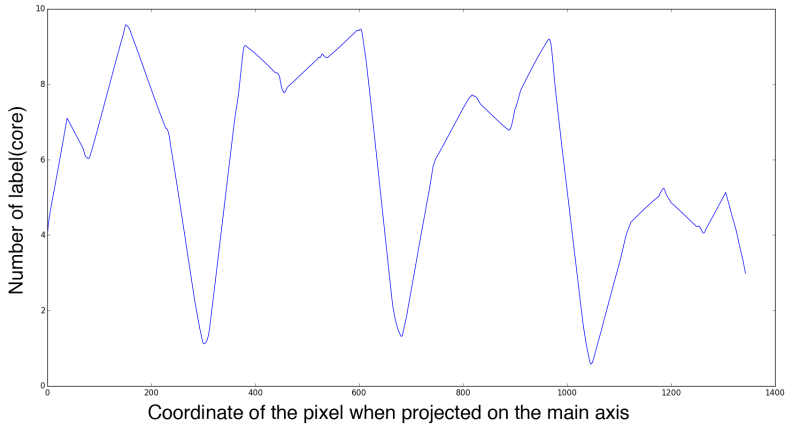


Figure 3: The approximation of the local density for the grid in one direction.

per row, we search the $x-1$ lowest local minima in the chart. Those minima correspond to the spaces between the subgrids. In Figure 3, the three local minima (corresponding to 4 subgrids per row) are easy-to-identify.

3.4 TMA core addressing and missing core identification

Once the subgrids have been detected, each core needs to be addressed according to the coordinate system. This step is straightforward when all the cores are present since the position of each subgrid and the direct neighbours of each core are known.

However, TMA sections are fragile and some cores can move or even completely disappear from the section. In such case, it is important to be able to track these (moved or missing) cores because false linking between a core and its metadata can lead to incorrect analyses (see Figure 1). In order to detect such problems, our method looks at the 4 nearest neighbours of each core. If the nearest neighbour in one direction is too far or not enough (vertically or horizontally) aligned with the core (to be identified without any ambiguity), this neighbour is considered as being missing. Each missing core so identified is placed in a reasonable interpolated position with respect to its detected neighbors, subgrid orientation and inter-core average distance.

3.5 Color deconvolution and fine registration

As detailed in section 2, 7 different IHC markers are used to reveal in brown protein expression in different cells or structures in the TMA sections, which are counterstained

in blue. Since image registration efficiency depends on the similarity of the input images, removing the brown staining from the images should improve the results. Stain separation is obtained using color deconvolution. Color deconvolution consist in projection of the original RGB colorspace, into a new space where two axes represent respectively the brown (DAB) and the blue (HEM) chromaticities, as described in [9], complemented with a third normal axis. Because the intensities in each of the RGB channels do not depend linearly on the concentration of the stain, the pixel value is converted into optical density (OD), which also normalizes the pixel value with regard to the glass slide background, using this formula:

$$OD = -log(\frac{I}{I_0}) \tag{1}$$

with I the intensity of the pixel located in the tissue area and I_0 the background intensity. Each pure stain needs to be defined by its own specific unit vector in the OD-converted RGB space. For this purpose, we designed and imaged two tissue slides, one stained with DAB and the other with HEM only, from which we extracted two stain-specific 3D scatter plots in the OD space. For each stain, the specific unit vector constituting the deconvolution matrix should be on an axis going through the origin of the coordinate system and fitting at best the main direction of the corresponding scatter plot in the OD space. One usual way is to point at the centroid. We also extracted the axis passing by the origin and which ensures the best representation of the data according to the least-square criterion. A projection can be written as:

$$\boldsymbol{x}_i' = a_i\boldsymbol{v} = \boldsymbol{v}^T\boldsymbol{x}_i\boldsymbol{v} \tag{2}$$

where $\mathbf{x}$ is a point in the OD space and $\boldsymbol{v}$ is the unit vector of the straight line on which we project the point. The least-square criterion to minimise can be written:

$$\begin{aligned}
J &= \sum_{i=1}^{n} ||a_i\boldsymbol{v} - \boldsymbol{x}_i||^2 \\
&= \sum_{i=1}^{n} a_i^2||\boldsymbol{v}||^2 - 2\sum_{i=1}^{n} a_i\boldsymbol{v}^T\boldsymbol{x}_i + \sum_{i=1}^{n} ||\boldsymbol{x}_i||^2 \\
&= \sum_{i=1}^{n} a_i^2 - 2\sum_{i=1}^{n} a_i^2 + \sum_{i=1}^{n} ||\boldsymbol{x}_i||^2 \\
&= -\sum_{i=1}^{n} [\boldsymbol{v}^T\boldsymbol{x}_i]^2 + \sum_{i=1}^{n} ||\boldsymbol{x}_i||^2 \\
&= -\sum_{i=1}^{n} \boldsymbol{v}^T\boldsymbol{x}_i\boldsymbol{x}_i^T\boldsymbol{v} + \sum_{i=1}^{n} ||\boldsymbol{x}_i||^2 \\
&= -\boldsymbol{v}^T\boldsymbol{M}\boldsymbol{v} + \sum_{i=1}^{n} ||\boldsymbol{x}_i||^2
\end{aligned} \tag{3}$$

where $\boldsymbol{M}_{kj} = \sum_i x_{ki}x_{ji}$. Hence, the solution is the unit vector $\boldsymbol{v}$ which maximizes $\boldsymbol{v}^T\boldsymbol{M}\boldsymbol{v}$. This is a constrained optimization problem:

$$max_{\boldsymbol{v}}(\boldsymbol{v}^T\boldsymbol{M}\boldsymbol{v}) \text{ subject to } (\boldsymbol{v}^T\boldsymbol{v} = 1) \tag{4}$$

The Lagrange function is:

$$\mathcal{L} = \boldsymbol{v}^T\boldsymbol{M}\boldsymbol{v} + \lambda(1 - \boldsymbol{v}^T\boldsymbol{v}) \tag{5}$$

and the condition needed for optimality is:

$$\partial_{\boldsymbol{v}} \mathcal{L} = 0 \tag{6}$$

Hence we get:

$$\boldsymbol{M}\boldsymbol{v} = \lambda \boldsymbol{v} \tag{7}$$

With $\boldsymbol{v}^T \boldsymbol{M} \boldsymbol{v} = \boldsymbol{v}^T \lambda \boldsymbol{v} = \lambda$ to maximize. So the solution is the eigenvector $\boldsymbol{v}$ of $\boldsymbol{M}$ with the highest eigenvalue, λ. In practice, we observed that this method results in vectors close to the ones obtained using the centroid. Figure 4 shows the deconvolution results obtained with pure staining.

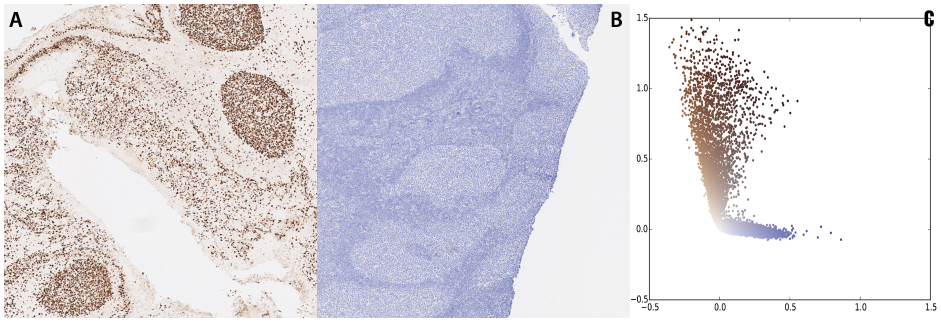


Figure 4: A. Sample of the pure DAB stained section. B. Sample of the pure HEM stained section. C. The pixel distribution in the deconvoluted color space.

Pairs of corresponding core images (i.e. labeled with the same grid location) from adjacent TMA slices are extracted and their orientation is matched using the TMA grid orientation (extracted in Section 3.2) to initialize the registration step. Those core images are then registered using the open-source Elastix framework [10] applied on the HEM (deconvoluted) channel of these images. The Elastix parameters are those determined in previous developments leading to successful registration of high-resolution fields of view from whole tissue sections [7]. To validate the registration method control points (CP) were manually placed and matched on different cores in the acquired TMA slide images. We evaluated registration accuracy per core using the root mean square error (RMSE) computed on the CPs.

4 Results and discussion

The automatic TMA grid fitting was successfully applied on the 7 slides without requiring any manual adjustment. The registration results obtained for pairs of consecutive slides show RMSE values corresponding to the diameter of a cell nucleus, i.e. about 5 µm. These results are illustrated in Figure 5. Intermediary results also confirm the efficiency of the Elastix-based registration step. Indeed, after matching the TMA grid orientations and before fine registration, the RMSE values are of 35 µm, i.e. slightly higher than the cell diameter (about 20 µm). The above data show that our image registration approach applied to consecutive TMA slides exhibiting different IHC markers has performances that enable accurate colocalization of protein expression. This approach is much more flexible than the one which consists in carrying out multiple staining IHC on one slide, using different chromogens to reveal the expression of different proteins. Indeed, this latter approach requires that the antibodies used in IHC come from different species (to avoid cross-reaction) and that the targeted proteins are expressed in different cells or cell compartments to avoid color overlay, which is

difficult to distinguish in brightfield microscopy. In addition, multiple staining is technically more difficult to set up and may cause staining artifacts as those we observed experimentally.

In our future works, we will use TMA core registration to match histologically relevant structures from an image to another, by using either a specific marker of the targeted structures or an annotation (made manually by an expert or resulting from a pattern recognition algorithm). Registered with a consecutive slide exhibiting the expression of another protein, e.g. specific of a cellular function (such as proliferation), this approach enables a compartmentalized analysis of the expression of the functional protein (e.g. inside and/or outside of the histological structures of interest).

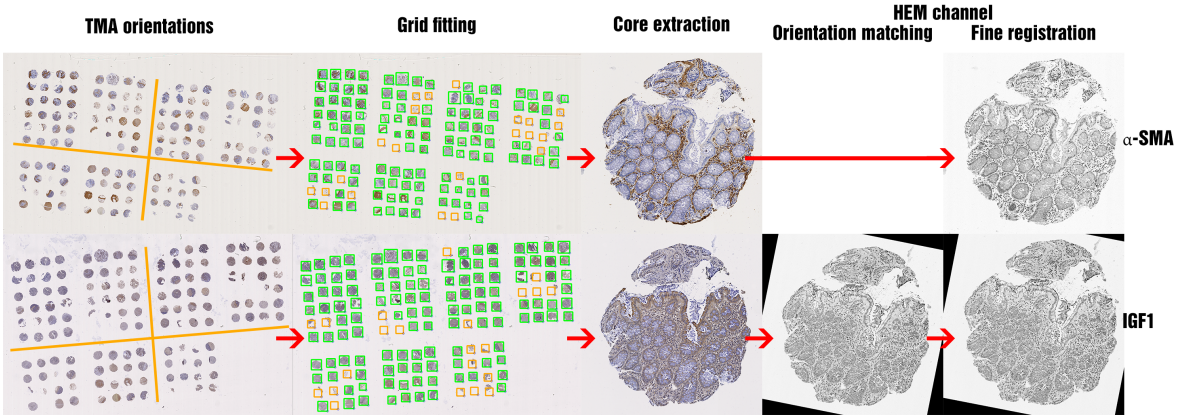


Figure 5: From left to right: original slide images where orange lines represent the orientation detected by the algorithm; results of the grid fitting on the images of two consecutive TMA slides (showing α-SMA and IGF1 expression in brown and negative tissue in blue) where absent vs. present cores are shown in orange vs. green; extraction of a pair of corresponding core images ; the HEM channel of the deconvoluted IGF1 core image after grid orientation matching; registration result of the blue (HEM) channel of the IGF1 image showing structure matching with the deconvoluted α-SMA core image.

References

[1] S. Hassan, C. Ferrario, A. Mamo, and M. Basik, "Tissue microarrays: emerging standard for biomarker validation," *Current Opinion in Biotechnology*, vol. 19, no. 1, pp. 19–25, 2008.

[2] R. L. Camp, G. G. Chung, and D. L. Rimm, "Automated subcellular localization and quantification of protein expression in tissue microarrays." *Nature medicine*, vol. 8, no. 11, pp. 1323–1327, 2002.

[3] "Arrayimager," http://www.visiopharm.com/solutions–deployed-image-analysis–arrayimager.shtml, 2012, [Online; accessed 24-April-2015].

[4] G. Nir, R. S. Sahebjavaher, P. Kozlowski, S. D. Chang, E. C. Jones, S. L. Goldenberg, and S. E. Salcudean, "Registration of whole-mount histology and volumetric imaging of the prostate using particle filtering," *IEEE Transactions on Medical Imaging*, vol. 33, no. 8, pp. 1601–1613, 2014.

[5] A. Sarkar, Q. Yuan, and C. Srinivas, "A robust method for inter-marker whole slide registration of digital pathology images using lines based features," in *2014 IEEE 11th International Symposium on Biomedical Imaging (ISBI)*, no. iii, 2014, pp. 762–765.

[6] Y. Song, D. Treanor, a. J. Bulpitt, N. Wijayathunga, N. Roberts, R. Wilcox, and D. R. Magee, "Unsupervised content classification based nonrigid registration of differently stained histology images," *IEEE Transactions on Biomedical Engineering*, vol. 61, no. 1, pp. 96–108, 2014.

[7] X. Moles Lopez, P. Barbot, Y.-R. Van Eycke, L. Verset, A.-L. Trépant, L. Larbanoix, I. Salmon, and C. Decaestecker, "Registration of whole immunohistochemical slide images: An efficient way to characterize biomarker colocalization," *Journal of the American Medical Informatics Association*, pp. 1–11, 2015.

[8] N. Otsu, "A threshold selection method from Gray-level," *IEEE Transactions on Systems, Man, and Cybernetics*, vol. SMC-9, no. 1, pp. 62–66, 1979.

[9] A. C. Ruifrok and D. A. Johnston, "Quantification of histochemical staining by color deconvolution." *Analytical and quantitative cytology and histology / the International Academy of Cytology [and] American Society of Cytology*, vol. 23, no. 4, pp. 291–9, Aug. 2001.

[10] S. Klein, M. Staring, K. Murphy, M. A. Viergever, and J. P. W. Pluim, "Elastix: A toolbox for intensity-based medical image registration," *IEEE Transactions on Medical Imaging*, vol. 29, no. 1, pp. 196–205, 2010.

Real-time Fullscale Model Colorization and Global Color Quality Evaluation for Rapid Scanning in Uncontrolled Environment

Arnaud Schenkel Olivier Debeir
Universit Libre de Bruxelles (U.L.B.)
Laboratories of Image, Signal processing and Acoustics
50, Av. F.Roosevelt. 1050 Bruxelles. Belgium
aschenke@ulb.ac.be odebeir@ulb.ac.be

(a) (b) (c)

Figure 1: Results from an unadjusted geometrically model: (a) reference colorization given by the device software (with visible artifacts), (b) our colorization, and (c) a combined color quality maps of (b) (in red the poorest quality, in green the best).

Abstract

The acquisition of a complete architectural or archeological model with a correct textures is still a problem. The fieldwork is often spread out over sufficiently large periods of time, to result in brightness or lighting changes in captured pictures (due to the inability to completely control the environment). Our promoted approach is defined to take into account all candidate color source to obtain homogeneous colorization. The process pipeline is designed to easily integrate new data (new photographic and geometric acquisitions) and to minimize memory footprint. Image stitching and weighting method are adapted to perform fine per vertex depth maps colorizations and to obtain a global quality evaluation. Each contribution is weighted using a quality measure of the information. A serie of quality maps are defined based on the distance between the geometry and the source, the relative orientation of the photographs to the surface, the visibility of the model from the point of view, the internal silhouette or the surface boundaries, and the image vignettings. We define and evaluate a set of transfer functions, relying on the best visual results to combine these maps.

1 Introduction

Preservation of cultural heritage benefits of the nowadays increase in interest for three-dimensional models as an information and communication medium. The digitization of existing objects and sites is a really fast and efficient way to obtain a very large amount of data of different kinds (shape, geometric size, appearance, surface reflectivity...). But obtaining a visually correct rendering of the appearance for a digital models remains an open issue in this field.

In order to get a complete model of a wide site and to minimize the so called geometric shadows, it is necessary to perform several geometric and colorimetric acquisitions during the field work from a variety of viewpoints. In natural uncontrolled climatic and light conditions, such as archaeological and architectural outdoor sites, the spacing in space of the colors measurements leads to consider multiple data sources whose quality depends on the surface distance and relative orientation, while the spacing in time induces risks of changes in the natural lighting conditions (i.e. effects due the suns positions or the weather changes). The variations of the brightness, the lightning and the acquisition parameters can induce significant differences in the visual appearance of a given scene. Without additional processing, such a multiview rendering models leads to a poor appearance, including artifacts such as color discontinuities (see Fig. 1a).

We promote in our work a method that produces visual representations of architectural sites with less visual artifacts (errors due to noise in the geometry or aberrant colorization...) and with a consistent and homogenous colorization. A real-time approach, compatible with fieldwork, enables to preview the results during the scanning procedure itself and thus allows to refine the acquisition parameters and to select new interesting viewpoints to improve the geometry and / or the colorization.

For models composed of large scans series, the outcomes exhibit a uniform coloring and no significant visual artifacts. The result also delivers a global measure of the color quality, also taking into account the number of contributing sources. The method allows to identify weaknesses in the model and the possible new interesting viewpoints, from which the acquisition should be completed.

2 Approach

In previous work [8], we have presented the more relevant approaches and identified several existing methods. Algorithms that require specific data or hardware were discarded. We also limit the methods to those suitable for point cloud on full-scale scene model and a reasonable computation time for real time application. We consider the class of weighing methods [1, 2, 3, 7] as the best existing candidate. We adapt this method in order to take into account several important features impacting the colorization quality of the produced model.

A object digitizing consists two types of data:

- a point cloud that is built by the successive addition of new scans, acquired from different viewpoints and registered into a single coordinate system;
- a series of pictures acquired by a 6M pixel camera.

Data are acquired in parallel to the colorization process by iterative and alternating phases of geometry and pictures capturing. In this work, three-dimensional registration of the scans (i.e. sensor's position and orientation matrix) and pictures registration relative to the point cloud (extrinsic and intrinsic parameters for each image) are assumed known. A large litterature covered these both themes [4, 6].

Our main goal is to quickly obtain a rendering, with a coherent colorization, of architectural or archaeological site models, which can be geometrically very complex and

have significant volumes, like the "Hôtel de Ville de Bruxelles", used for illustrations in this paper. We need to acquire data (three-dimensional scannings and photographies) from a large number of viewpoints for such acquisitions (for our model, 43 scans composed of 716M points, coupled with 211 pictures). Textures artifact-free are difficult to obtain during the fieldwork, it is thus important to work with pictures taken in arbitrary conditions (i.e. variable lighting and positioning conditions).

Our approach must meet three conditions:

- a homogeneous colorization calculated from pictures taken under varying conditions;
- a real-time computation relating to the acquisition time on the field;
- a manipulation of a large and growing amount of data (three-dimensional models and HD pictures).

We define a color source like the color value of a pixel in an image. For a 3D point, we could readily determine the related potential color sources in each image, based on the knowledge of the extrinsic and intrinsic parameters of the camera and on the perspective projection principles.

Our colorization approach consists of computing for each 3D points three information by considering all available colors sources :

- the evaluation of the global color quality of the sources relative to the geometry;
- the number of truly contributory pictures to the final color;
- the color, calculated as an average of the available sources weighted according to its respective quality.

Our notion of real-time is related to the fieldwork process. The computation time to colorize the complete model considering all the picture is lower than the time required to perform all measurements. To achieve this, we update the model information incrementaly with the new acquired data (3D geometry and pictures). Indeed, the knowledge of the previous information is used to calculate the new ones without reprocessed without any recalculation.

The manipulation of a large and growing amount of data is also a major difficulty. Our proposition is thus to keep the scans division of the entire model without merging all the three-dimensional data, and to adapt the colorization pipeline to consider independently each scan during the process.

A scan is basically a depth map and can therefore be seen as an "image". The depth map has the property to define a simple neighborhood notion without requiring heavy computation (i.e. the 2D neighboring element in the map can be seen as a neighboring point in the 3D space). This facilitates and speeds up calculations such as normals estimation, surface boundaries extraction,... by considering large neighborhood (defined by a radius around the point) in the image-space. These calculations can be refined using a variable radius (depending on the point depth) to take into account the geometry.

The proposed method consists of per vertex depth maps colorizations taking into account all available colors sources. Each contribution is weighted using a quality measure of the information. With this consideration, we compute score for each color sources (i.e. pictures).

3 Scores

A serie of scores is defined to weight each contribution. Inspired by Callieri et al. [3], we suggest to use a geometric mean of the normalized measures to combine these metrics. Each normalization is based on a transfer functions described bellow with the definition of his respective metrics.

The combination approach allows to obtain eventually better realistic visual results by extending the method to consider specialized weighting factors depending to the images content (focus [3], contrast, or saturation [7]) or the amplitude images [5].

We limits our scores evaluation to the more relevant ones (see Fig. 3) :

- Visibility. A portion of the depth map is invalid due to the limitations of the scanner, such as a limited range of use and difficulties with shiny, reflective or transparent surfaces. Pictures and geometry are acquired from different viewpoints. We compute the visibility mask by simulating and merging several z-buffers with decreasing resolutions.

- Distance.The distance between a color source and the geometry is an important score to reflect :

  - the amount of light that reaches the geometry
  - the pixel/surface ratio
  - the illumination due to nearby light such as flash

  We promoted the use of a normal function, centered in a parameter, defined both to mitigate the effects of overexposure due to flash light for close geometry and of remote sources. In general, the remote sources contribute less to the final color than other elements.

- Silhouette. Silhouette correspond to part in geometry located near a sharp variation in distance (i.e. border). Color is usually ill defined on these part since parallax problem can cause important error even for small aligment discrepancies. Silhouette score reflects the characteristics of complexity and uncertainty, both in terms of position and for the consistency of the colorization where the visibility of a surface can change.

  The internal silhouettes are defined as the set of points of the surface whose normal is perpendicular to the view vector. We used a cosine function to weight this value; sources value approaches or exceeds 90 degrees have a bad weight and do not contribute to the colorization. We use a hyperbolic tangent to obtain a rapid decreasing in range values near 90 degrees, and to avoid abrupt discontinuities.

- Orientation. The orientation is evaluated by computing the angle between the surface normal and the optical axis. This reflects that a picture oriented parallely to a surface should have a better score. Similarly to the Lambert's cosine law, the normalization of the orientation is based on a cosine function.

- Vignetting. An important roll-off in the photograph quality is measurable towards the borders and the corners, depending on the internal structure of the lens. We approximates this with a normalized distance map of the picture.

4 Global Color Quality Evaluation

In each 3D points, as illustrated in Fig. 4, we evaluate a global color quality as the square root of the product of :

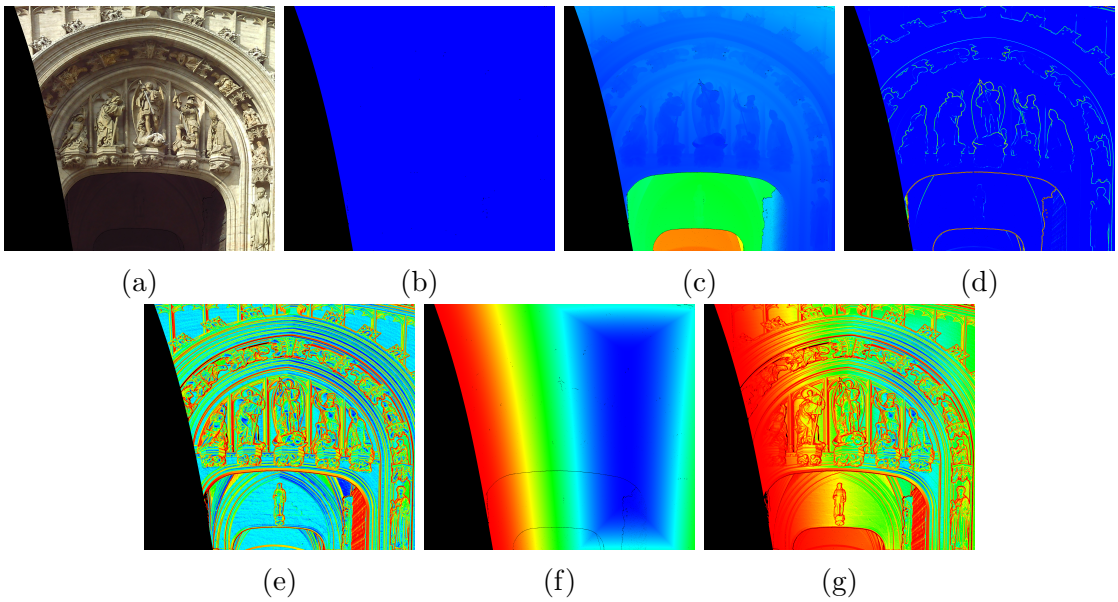


Figure 2: Normalized data corresponding to one picture: (a) color, (b) visibility, (c) distance, (d) silhouette, (e) orientation, (f) vignetting mask, and (g) obtained combination. The masks are standardized to use the color wheel defined in the HSV color space.

- the overall quality score, computed by combining the scores of all color sources;
- the number of contributive sources, defined as the number of non null scores.

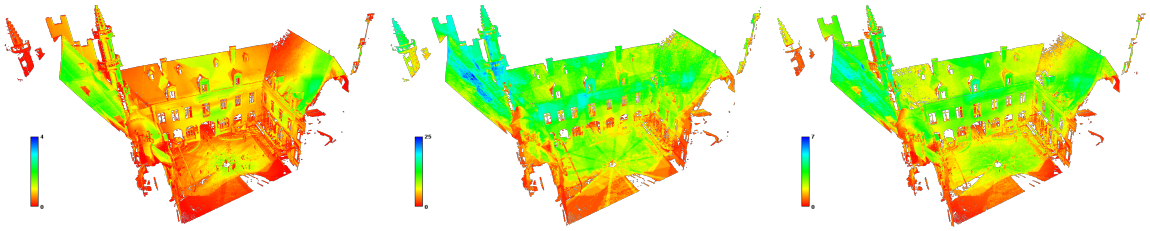


Figure 3: Global color quality evaluation steps: (a) the overall quality, (b) the number of contributive sources, and (c) the quality evaluation. No geometry pre-processing was applied on the architectural model

This representation illustrate the weakness in the colorization (in red). New interesting viewpoints and camera orientation can be identified. In order to identify a possible new acquisition place, we can simulate the expected quality before the acquisition itself. Indeed, we can produce the quality map for the simulated viewpoint based on the already acquired geometry and estimate the number and the quality of contributing color sources.

5 Conclusion

The rendering computation times, necessary to our solution to obtain good visual quality, is compatible with fieldwork. The integration of a new image in the process is near to one seconds, which allows us to test potential new captures.

For models composed of large scans series, the produced results exhibit a uniform coloring and no significant visual artifacts. The system also delivers a global measure of the color quality, also taking into account the number of contributing sources. The method allows to identify weaknesses in the model and can simulate the contribution of a new acquisition, and therefore enables to select better viewpoint during the acquisition campaign.

References

[1] A. Baumberg, "Blending Images for Texturing 3D Models," in Procedings of the British Machine Vision Conference 2002, 2002, pp. 404–413.

[2] F. Bernardini, I. M. Martin, and H. Rushmeier, "High-quality texture reconstruction from multiple scans," IEEE Transactions on Visualization and Computer Graphics, vol. 7, no. 4, pp. 318–332, 2001.

[3] M. Callieri, P. Cignoni, M. Corsini, and R. Scopigno, "Masked photo blending: Mapping dense photographic data set on high-resolution sampled 3D models," Computers & Graphics, vol. 32, no. 4, pp. 464–473, 2008.

[4] M. Dellepiane and R. Scopigno, "Global refinement of image-togeometry registration for color projection," in 2013 Digital Heritage International Congress (Digital Heritage). IEEE, Oct. 2013, pp. 39–46.

[5] U. Hahne and M. Alexa, "Exposure Fusion for Time-Of-Flight Imaging," Computer Graphics Forum, vol. 30, no. 7, pp. 18871894, 2011.

[6] N. Mellado, D. Aiger, and N. J. Mitra, "Super 4pcs fast global pointcloud registration via smart indexing," Computer Graphics Forum, vol. 33, no. 5, pp. 205–215, 2014.

[7] T. Mertens, J. Kautz, and F. Reeth, "Exposure Fusion," in 15th Pacific Conference on Computer Graphics and Applications, 2007, p. 382.

[8] A. Schenkel, N. Warzee, and O. Debeir, "Fast color correction for rapid scanning in uncontrolled environment," in 2013 Digital Heritage International Congress (Digital Heritage). IEEE, Oct. 2013, pp. 413–416.

Semiparametric Score Level Fusion: Gaussian Copula Approach

N. Susyanto[1], C.A.J. Klaassen[1], R.N.J. Veldhuis[2], and L.J. Spreeuwers[2]
[1]University of Amsterdam
Korteweg-de Vries Institute for Mathematics
P.O. Box 94248, 1090 GE Amsterdam, The Netherlands
[2]University of Twente
Faculty of EEMCS
P.O. Box 217, 7500 AE Enschede, The Netherlands
{n.susyanto,c.a.j.klaasen}@uva.nl
{r.n.j.veldhuis, l.j.spreeuwers}@utwente.nl

Abstract

Score level fusion is an appealing method for combining multi-algorithms, multi-representations, and multi-modality biometrics due to its simplicity. Often, scores are assumed to be independent, but even for dependent scores, according to the Neyman-Pearson lemma, the likelihood ratio is the optimal score level fusion if the underlying distributions are known. However, in reality, the distributions have to be estimated. The common approaches are using parametric and nonparametric models. The disadvantage of the parametric method is that sometimes it is very difficult to choose the appropriate underlying distribution, while the nonparametric method is computationally expensive when the dimensionality increases. Therefore, it is natural to relax the distributional assumption and make the computation cheaper using a semiparametric approach.

In this paper, we will discuss the semiparametric score level fusion using Gaussian copula. The theory how this method improves the recognition performance of the individual systems is presented and the performance using synthetic data will be shown. We also apply our fusion method to some public biometric databases (NIST and XMVTS) and compare the thus obtained recognition performance with that of several common score level fusion rules such as sum, weighted sum, logistic regression, and Gaussian Mixture Model.

1 Introduction

Multi-biometric system or biometric fusion is a combination of several biometric systems or algorithms in order to enhance the performance of the individual system or algorithm. In general, it can be characterized into six categories [15]: multi-sensor, multi-algorithm, multi-instance, multi-sample, multi-modal and hybrid. Several studies [7, 14, 15, 18] show that combining information from multiple traits or algorithms can provide better performance. For example, Lu et al. [7] combining three different feature extractions (Principle Component Analysis, Independent Component Analysis and Linear Discriminant Analysis) which is related to the multi-algorithm biometric fusion. In the fingerprint biometric field, Prabhakar and Jain [13] use the left and right index fingers to verify an individual's identity which is an example of the multi-instance biometric fusion.

Biometric fusion can be done at the sensor, feature, match score, rank and decision levels either for verification or identification. In this paper, we will focus on the match score level for person verification. This means that scores from multiple biometric

matchers for every pair of two subjects (user and enrollment) are transformed to a new score (a scalar) as a combined score. Once the new score has been generated, one has to decide whether the user and enrollment are from the same person or not. To do this, a threshold has to be set such that a score greater than or equal to the threshold is recognized as *genuine score* which means that the user and enrollment are the same subject while a score less than the threshold will lead to the conclusion that the user and enrollment are different people which will be called by *impostor score.* This threshold is determined using a set which is called the *training set* and is evaluated using a disjoint set which is called the *testing set*

There are three categories in biometric fusion: transformation-based [5], classifier-based [8], and density-based. The last category would be optimal if the underlying densities were known. However, in practice, such densities have to be estimated from the training set so that the performance relies on how well these two densities are estimated. The parametric models suffers from the limitation in choosing the appropriate parametric model to the data. The most successful parametric approach is the Gaussian Mixture Model (GMM) [10]. However, the number of the mixture components which is the most important part in estimating GMM is very hard to be determined. The author in his paper used GMM fitting algorithm proposed in [3] that automatically estimates the number of the mixture components using an EM algorithm and the minimum message length criterion. However, the computational cost is time consuming when the sample size is big or the the number of mixture components increases. On the other hands, the nonparametric models have a problem in choosing bandwidth and computational cost when working in the multidimensional space.

This paper focuses on the fusion strategy for dependent matchers. Using synthetic data, we will show that our approach is robust in handling the dependent classifiers even with an extremely high dependence structure. We will also apply our method on the public databases NIST-BSSR1 and XM2VTS. The rest of this paper is organized as follows. In Section 2, we will review the theory of Gaussian copula, why it is suitable to be chosen and how to do Gaussian copula based fusion. Some experimental results on the synthetic data are presented in Section 3 to show the robustness of our method in handling the dependence issues and the results on the public database will be provided to show the applicability of our method in the real world. Finally, this paper will be closed by our conclusions.

2 Gaussian Copula Fusion

2.1 Likelihood ratio based fusion

Suppose we have d matchers and let $\mathbf{X} = (X_1, \cdots, X_d)$ denote the d components of the matching(similarity or distance) scores where X_i is the random variable corresponding to the i-th match score where $\mathbf{X}$ takes its values in $\Omega \subset \mathbb{R}^d$. The decision function is a map $\psi : \mathbb{R}^d \mapsto \{0, 1\}$ where 0 and 1 corresponds to negative and positive decisions which are denoted by H_0 and H_1, respectively. A system can make two types of error(false): accepting an impostor score or rejecting a genuine score. The probability of accepting impostor score $P(\psi(\mathbf{X}) = 1|H_0)$ is called by *False Acceptance Rate (FAR)* while the probability of rejecting genuine score $P(\psi(\mathbf{X}) = 0|H_1)$ is called by *False Rejection Rate (FRR).* From the definition of FRR, it can be understood that the probability of accepting genuine score that will be called by *True Positive Rate (TPR)* is $TPR = 1 - FFR$. In application, the FAR has to be set very small since the cost of accepting an impostor may be much more expensive than the cost of rejecting a genuine user. For example, in security, allowing a forbidden person to access a secret place is much more dangerous that rejecting a "nice" person to access it. Therefore, for every given FAR, our fusion has to maximize the TPR.

Neyman and Pearson established the most powerful test based on the likelihood ratio [11]. Let f_{gen} and f_{imp} be the density of genuine and impostor scores, respectively. The likelihood ratio at a point $\mathbf{x} = (x_1, \cdots, x_d)$ is defined by

$$LR(\mathbf{x}) = \frac{f_{gen}(\mathbf{x})}{f_{imp}(\mathbf{x})}. \tag{2.1}$$

According to the Neyman-Pearson theorem, in order to get the maximum TPR for every fixed FAR, say α, we have to decide

$$\psi(X) = 1 \iff LR(\mathbf{x}) \geq \eta \tag{2.2}$$

where η is implicitly defined by

$$P(LR(\mathbf{X}) \geq \eta) = \alpha. \tag{2.3}$$

As a consequence, the optimal performance can be reached by defining the fused score as the likelihood ratio of the vector consisting of all matching scores.

2.2 Gaussian copula

Computing (2.1) means that the estimation of f_{gen} and f_{imp} is a must. Let H be any distribution function on $\mathbb{R}^d$ with density h. A classical result of Sklar [17] shows that H can be uniquely factorized into its univariate marginal distributions and a distribution function on the unit cube $[0,1]^d$ in $\mathbb{R}^d$ with uniform marginal distributions which is called by *copula:*

Theorem 2.1 (Sklar (1959))**.** *Let $d \geq 2$ and suppose H is a distribution function on $\mathbb{R}^d$ with one dimensional continuous marginal distribution functions $F_1, \cdots, F_d$. Then there is a unique copula C so that*

$$H(x_1, \ldots, x_d) = C(F_1(x_1), \ldots, F_d(x_d)) \ \ \forall (x_1, \ldots, x_d) \in \mathbb{R}^d. \tag{2.4}$$

This paper assumes that C is determined by a multivariate normal distribution with standard normal marginals and correlation matrix R. Note that this assumption is more flexible than assuming H to be multivariate normally distributed. The main difference is that each marginal of the multivariate normal has to be normally distributed while each marginal of a Gaussian copula can be any continuous distribution function. In section 3, we will see that our generated data follow a Gaussian copula distribution with normal and weibull marginal.

The key concept of the Gaussian copula is the assumption of the existence of a componentwise transformation $\tau : \mathbb{R}^d \mapsto \mathbb{R}^d$ such that $\tau(\mathbf{X}) \sim N(0, R)$. Here, each component τ_i of τ is a monotone continuous function. One can show that

$$\tau_i(x_i) = \Phi^{-1}(H_i(x_i)) \tag{2.5}$$

for $i = 1, \ldots, d$ where Φ and H_i denote the standard normal distribution function and the marginal distribution of the $i-$th component.

This means that (2.4) can be rewritten as

$$H(x_1, \ldots, x_d) = \Phi_R(\Phi^{-1}(u_1), \ldots, \Phi^{-1}(u_d)), \tag{2.6}$$

where $u_i = F(x_i)$, Φ the one-dimensional standard normal distribution function, and Φ_R the d-dimensional standard normal distribution function with correlation matrix R. Consequently, the density function of H is

$$h(x_1, \ldots, x_d) = \frac{1}{|R|^{1/2}} \exp\left(-\frac{1}{2}\mathbf{u}^T(R^{-1} - I)\mathbf{u}\right) \prod_{i=1}^{d} f_i(x_i), \tag{2.7}$$

with $\mathbf{u} = (\Phi^{-1}(F_1(x_1)), \cdots, \Phi^{-1}(F_d(x_d)))^T$.

2.3 Gaussian copula based fusion

Our fused score using the Gaussian copula approach is defined by (2.1) with the numerator f_{imp} and the denominator f_{gen} as in (2.7), i.e.,

$$LR(x_1,\ldots,x_d) = \frac{|R_{imp}|^{1/2} \times \exp\left(\frac{1}{2}\mathbf{u_{gen}}^T(R_{gen}^{-1}-I)\mathbf{u_{gen}}\right) \times \prod_{i=1}^{d} f_{gen,i}(x_i)}{|R_{gen}|^{1/2} \times \exp\left(\frac{1}{2}\mathbf{u_{imp}}^T(R_{imp}^{-1}-I)\mathbf{u_{imp}}\right) \times \prod_{i=1}^{d} f_{imp,i}(x_i)}. \quad (2.8)$$

Here, R_{gen} and R_{imp} denote the correlation matrices of transformed genuine and impostor scores, respectively, $\mathbf{u_{gen}}$ and $\mathbf{u_{imp}}$ are given by

$$\mathbf{u_{gen}} = (\Phi^{-1}(F_{gen,1}(x_1)),\cdots,\Phi^{-1}(F_{gen,d}(x_d)))^T$$

and

$$\mathbf{u_{imp}} = (\Phi^{-1}(F_{imp,1}(x_1)),\cdots,\Phi^{-1}(F_{imp,d}(x_d)))^T,$$

respectively. To obtain the LR value as given by (2.8), we need to estimate the correlation matrices $R_{gen}(R_{imp})$, the marginal densities $f_{gen,i}(f_{imp,i})$ and marginal distribution functions $F_{gen,i}(F_{imp,i})$ using a training set. Given a training set, we can extract to the genuine and impostor scores. Note that the scores often are dependent within the group of genuine scores, within the group of impostor scores, and between these two groups. However, we shall proceed as if all scores are independent. The resulting estimators are still reliable because most scores will be independent.

Let $W_1,\ldots,W_{n_{gen}}$ and $B_1,\ldots,B_{n_{imp}}$ be the two samples representing the genuine and impostor scores, respectively.

2.3.1 Matchers dependence

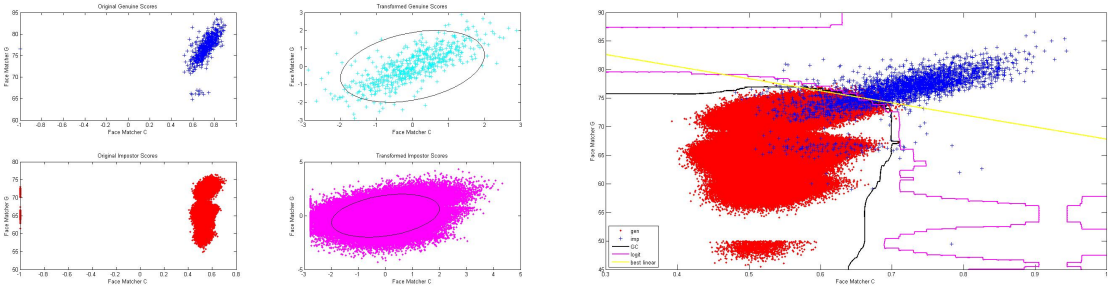
As stated above, some genuine and impostor scores are dependent. However, we are interested in the correlation matrices of the match scores, which we will assume to be the same, $R_{gen} = R_{imp} = R$. We shall estimate R using the combined sample, i.e,

$$(X_1,\ldots,X_n) = (W_1,\ldots,W_{n_{gen}},B_1,\ldots,B_{n_{imp}})$$

with $n = n_{gen} + n_{imp}$. Our experiments show that such restriction will improve the performance of the fused score. It is reasonable since we are estimating the matchers dependence not only the genuine or impostor scores dependence. Klaasen and Wellner [6] give an explicit formula to obtain an optimal estimator for the correlation matrix R via normal rank correlation by taking $\hat{R} = \left(\hat{\rho}_{rs}^{(n)}\right)$ where

$$\hat{\rho}_{rs}^{(n)} = \frac{\frac{1}{n}\sum_{j=1}^{n} \Phi^{-1}\left(\frac{n}{n+1}\mathbb{F}_r^{(n)}(X_{rj})\right)\Phi^{-1}\left(\frac{n}{n+1}\mathbb{F}_s^{(n)}(X_{sj})\right)}{\frac{1}{n}\sum_{j=1}^{n}\left[\Phi^{-1}\left(\frac{j}{n+1}\right)\right]^2} \quad (2.9)$$

where Φ denotes the one-dimensional standard normal distribution function while $\mathbb{F}_r^{(n)}$ and $\mathbb{F}_s^{(n)}$ are the marginal empirical distributions of F_r and F_s, respectively, is an efficient estimator for ρ_{rs} for every $1 \le r < s \le d$.



(a) Two face matchers scores from NIST-Multimodal

(b) Two face matchers scores from NIST-Face

Figure 1: Score transformation and Boundary decisions at 0.01% FAR

2.3.2 Marginal density estimation

To estimate the marginal density functions, we use the kernel bandwidth optimization as studied by Shimazaki and Shinomoto [16]. This method has two different kinds of choosing the optimal bandwidth. The first bandwidth choice is similar with the regular bandwidth selection but it performs much faster than the built-in *ksdensity* matlab. The second one is a local bandwidth optimization. This approach works very well in handling the data that have "spikes".

2.3.3 Marginal distribution function estimation

The empirical distribution function is an optimal estimator for the marginal distribution function and very easy to be implemented and very fast to be computed (see Figure 1a for example in biometric). The empirical distribution function, $\hat{F}$, is the distribution function that puts mass $1/n$ at each data point x_i where n is the number of the observation. In this paper, since we need to compute the quantile of the standard normal, then to avoid singularity, we prefer to put mass $1/(n+1)$. Explicitly, the empirical distribution function of genuine and impostor scores are given by

$$\hat{F}_{gen}(x) = \frac{1}{n_{gen}+1}\sum_{i=1}^{n_{gen}}(1)_{[W_i \le x]} \text{ and } \hat{F}_{imp}(x) = \frac{1}{n_{imp}+1}\sum_{i=1}^{n_{imp}}(1)_{[B_i \le x]}. \tag{2.10}$$

3 Experimental Results

To study the robustness of our method in fusing biometrics scores related to the classifiers dependence, genuine and impostor scores are generated that follow three different distribution functions and have three different dependence levels. Here, we assume that there are 1000 subjects with 2 biometric specimens for each subject, one is put as user and the other for enrollment. We also assume that we have 2 different biometric systems. Therefore, the size of genuine and impostor scores are 2×1000 and 2×9999000, respectively which we will use as training data. The testing data are obtained in the same way. The parameters for generating the data are:

- multivariate normal scores with correlations $0.99, 0.5$ and 0.1 with genuine means $[1,3]^T, [5,3]^T$ and $[5,3]^T$, respectively. All impostor means are set to be $[0,0]^T$.
- Gaussian copula with correlation value $0.9, 0.5$ and 0.1. The genuine and impostor marginals of the first matcher are set to follow weibull distribution with the shape

parameters 3 and 1, respectively, and the common scale parameter 4. For the second matcher, the genuine and impostor marginals follow normal distribution with parameter $(5, 1)$ and $(1, 0)$, respectively.

Once all data have been generated, for every pair of training and testing set, the exact likelihood ratio is computed which is called by true fusion. The next step is performing the sum rule with min-max and z-norm normalization and also the weighted sum using Fisher criterion [2]. Subsequently, we pick the best results. For the logit fusion, we use nonlinear logistic regression as given by W. Chen and Y. Chen[1]. The performance of several methods compared with the true fusion is provided in Table 1. The bold value is the best non-true fusion which indicated the TPR (%) at 0.01% FAR. We can see that our method is the most robust approach especially for the data with high dependence.

Table 1: Influence of Dependence in Biometric Fusion

Methods	**High**			**Moderate**			**Low**		
	MV	GC	Gu	MV	GC	Gu	MV	GC	Gu
True Fusion	90.70	93.20	99.90	91.00	90.70	97.40	96.90	90.90	84.70
Best Linear	89.80	90.40	94.00	**91.00**	90.20	90.90	**96.90**	89.90	83.50
Logistic Regression	00.10	88.20	87.60	90.60	90.50	87.40	**96.90**	**90.80**	82.80
Gaussian Copula	**90.10**	**92.80**	**99.70**	89.80	**90.70**	**93.50**	96.50	90.60	**84.70**

*MV: Multivariate Normal, GC: Gaussian Copula, Gu: Gumbel Copula.

We will also apply our method on the public databases: NIST-BSSR1 [9] and XM2VTS [12]. The NIST-BSSR1 database has three different set:

- NIST-Multimodal: Two fingerprints and Two face matchers applied to 517 subjects,
- NIST-Face: Two face matchers applied to 3000 subjects,
- NIST-Finger: Two fingerprints applied to 6000 subjects.

For every experiment, each set is split up randomly into two subsets, one is used for training and the other is used for testing. Then the naive sum rule with min-max normalization, naive sum with Z-normalization, weighted sum with Fisher criterion, nonlinear logistic regression, and our method are performed and the TPR at 0.01% is computed for every fusion strategy. This procedure is repeated 20 times and the average of all TPR at 0.01% for each fusion strategy is provided in the Table 2. We do not include the Gaussian Mixture Model (GMM) fusion strategy because the computation is very time consuming when it is done on a normal computer. However, we also provide the result of the GMM strategy as reported in [10] and we compare the 95% Confidence Interval on increase in TPR at 0.01% as given by Table 3. We can see that our approach outperforms all other fusion strategies (the bold value is the best one) even with GMM fusion which is computationally expensive. Also for the XM2VTS database that contains match scores from five face matchers and three speech matchers applied to 295 subjects with the partition of the training and testing set have been defined in [12], our method is the highest among all reported TPR at 0.01% FAR.

Table 2: TPR (%) values for different methods at 0.01% FAR on the public databases

Method	**NIST Multi modal**	**NIST Face**	**NIST Finger print**	**XM2VTS**
Naive Sum min-max	97.97	76.47	91.33	97.50
Naive Sum Z-norm	97.87	76.48	91.33	97.50
Weighted Sum	97.97	76.48	91.40	97.50
Logistic Regression	98.74	76.48	91.46	98.50
Gaussian Mixture Model[10]	99.10	**77.20**	91.40	98.70
This paper	**99.48**	**77.21**	**91.60**	**99.00**

Table 3: Comparison with LR fusion using Gaussian Mixture Model on the NIST-BSSR1 database

Database	**Mean TPR (%) at 0.01% FAR**			**95% Confidence Interval on increase in TPR (%) at 0.01% FAR**	
	BSM	GMM	GC	GMM	GC
NIST-Multimodal	85.30	99.10	**99.48**	[13.50,14.00]	[**13.51,14.84**]
NIST-Face	71.20	**77.20**	**77.21**	[4.70, 7.30]	[4.69, 7.32]
NIST-Fingerprint	83.50	91.40	**91.60**	[7.60, 8.20]	[**7.63, 8.57**]

*BSM: Best Single Matcher, GMM: Gaussian Mixture Model, GC: Gaussian Copula (used in this paper).

4 Conclusion

The Gaussian copula is a semiparametric model which is easy to be implemented, cheap in computation, and able to handle the dependence structure that usually appears in multi-algorithm fusion. Using several synthetic data, we have shown that our approach performs very well in dependent classifiers fusion even for extreme dependence structures when the performance of other approaches drops dramatically. We also see that our method works well when it is applied on the NIST-BSSR1 database (see Figure 1b for the comparison of the boundary decision with another approaches on this database) and even on the XM2VTS it reaches the highest TPR at 0.01% FAR among all reported results. However, it has limitations in estimating the tail density because estimation is based on the kernel density method. Our experiments show that although our approach works well at 0.01% FAR, it is sometimes much worse than individual classifiers at 0.001% FAR.

References

[1] W. Chen and Y. Chen, "DLR-b: Density-based Logistic Regression with bins for Large-scale Nonlinear Learning," *Technical report Department of Computer Science and Engineering, Washington University,* August 2013.

[2] C. M. Bishop, "Pattern Recognition and Machine Learning", *Springer-Verlag New York, Inc* 2006.

[3] M. Figueiredo and A. K. Jain, "Unsupervised Learning of Finite Mixture Models," *IEEE Transactions on Pattern Analysis and Machine Intelligence,* vol. 24, no. 3, pp. 381396, March 2002.

[4] J.P. Hube, "Neyman-Pearson Biometric Score Fusion as an Extension of the Sum Rule" *Proc. SPIE 6539*, Biometric Technology for Human Identification IV, 65390M, April 2007.

[5] J. Kittler,M. Hatef, R. Duin, and J. Matas, On combining classifiers, *IEEE Trans. Pattern Anal. Mach. Intell.,* vol. 20, no. 3, pp. 226239, Mar. 1998.

[6] C. A. J. Klaassen and J. A. Wellner, "Effcient estimation in the bivariate normal copula model: normal margins are least favourable," *Bernoulli* **3**, 55-77, March 1997.

[7] X. Lu, Y. Wang, and A. K. Jain. "Combining Classifiers for Face Recognition," *proc. IEEE International Conference on Multimedia and Expo (ICME),* volume 3, pages 1316, July 2003.

[8] Y. Ma, B. Cukic, and H. Singh, A Classification Approach to Multi-biometric Score Fusion, *Proc. Fifth International Conference on AVBPA, Rye Brook, USA,*, pp. 484493, July 2005.

[9] National Institute of Standards and Technology, "NIST Biometric Scores Set - release 1," 2004, Available at http://www. itl.nist.gov/iad/894.03/biometricscores.

[10] K. Nandakumar, Y. Chen, S. C. Dass, and A. K. Jain, "Likelihood ratio based biometric score fusion," *IEEE Trans on Pattern Analysis and Machine Intelligence,* Vol. 30, No. 2, Feb 2008

[11] J. Neyman and E. Pearson, "On the problem of the most efficient tests of statistical hypotheses," *Phil. Trans.Roy. Soc. London, Series A*, January 1933.

[12] N. Poh and S. Bengio, "Database, Protocol and Tools for Evaluating Score-Level Fusion Algorithms in Biometric Authentication," *Pattern Recognition,* vol. 39, no. 2, pp. 223233, February 2006.

[13] S. Prabhakar and A. K. Jain, "Decision-level Fusion in Fingerprint Verification," *Technical Report MSU-CSE-00-24,* October 2000.

[14] A. Ross, A. K. Jain, and J. Reisman, "A Hybrid Fingerprint Matcher," *Pattern Recognition,* 36(7):16611673, July 2003.

[15] A. Ross, K. Nandakumar, and A. K. Jain, *Handbook of Multibiometrics.* Springer-Verlag, 2006.

[16] H. Shimazaki and S. Shinomoto, "Kernel bandwidth optimization in spike rate estimation", *J Comput Neurosci* vol. 29, pp. 171182, August 2009.

[17] A. Sklar, "Fonctions de repartition a n dimensions et leurs marges," *Publ. Inst. Statist. Univ. Paris,* 8, 229-231, 1959.

[18] B. Ulery, A. Hicklin, C. Watson, W. Fellner, P. Hallinan, "Studies of Biometric Fusion" *-Executive Summary, NISTIR 7346, National Institute of Standards and Technology,* September 2006.

Facial recognition using new LBP representations

Alireza Akoushideh [1, 2], R.N.J. Veldhuis [1], L.J. Spreeuwers [1], and Babak M.-N. Maybodi [2]

[1] Dept. of Electrical Engineering, University of Twente, Enschede, the Netherlands.

[2] Dept. of Electrical and Computer, University of Shahid-Beheshti, G.C., Tehran, Iran.

{a.akoushideh, r.n.j.veldhuis, l.j.spreeuwers}@utwente.nl and b-mazloom@sbu.ac.ir

Abstract

In this paper, we propose a facial recognition based on the LBP operator. We divide the face into non-overlapped regions. After that, we classify a training set using each region at a time under different configurations of the LBP operator. Regarding to the best recognition rate, we consider a weight and specific LBP configuration to the regions. To represent the face image, we extract LBP histograms with the specific configuration (radius and neighbors) and concatenate them into feature histogram. We propose a multi-resolution approach, to gather local and global information and improve the recognition rate. To evaluate our proposed approach, we considered the FERET data set, which includes different facial expressions, lighting, and aging of the subjects. In addition, weighted *Chi-2* is considered as a dissimilarity measure. The experimental results show a considerable improvement against the original idea.

1 Introduction

Because of a wide range of applications in security, safety and access control, biometric pattern recognition has been a great challenge for researchers and scientists [1, 2]. In recent years, Local Binary Patterns (LBP) and its extensions, as a texture feature extractor, have been one of the most popular and successful applications. LBP [3] is the most widely used for the face detection, face recognition, facial expression analysis, and other related applications. Numerous approaches have been proposed based on LBP. For example, Ahonen *et al*. [4] proposed an LBP-based facial image analysis by dividing the face into some non-overlapping regions and concatenation of the LBP features that are extracted from each region. They assigned a weight to each region based on the importance of the information it contains. Zhang *et al*. [5] proposed a non-statistics based face representation approach, Local Gabor Binary Pattern Histogram Sequence (LGBPHS), with no training procedure to construct the face model. Chan *et al*. [6] proposed a face representation approach derived by the Linear Discriminant Analysis (LDA) of multi-scale local binary pattern histograms. Shan *et al*. [7] introduced Fisher Discriminant Analysis (FDA) of the LBP (LGBP) spatial histogram. Zhang *et al*. [8] encode Gabor phase through LBP and local histograms in addition to magnitudes of Gabor coefficients. Tan *et al*. [9] introduced local ternary patterns (LTP) to generalize of the LBP descriptor. Nikisins [10] proposed a

face recognition methodology, which is based on the combination of the texture operator, namely Multi-scale Local Binary Pattern (MSLBP), face image filtering and feature weighting algorithms. Ishraque *et al*. [11] proposed local directional pattern Variance (LDPv), to represent facial components.

In this paper, we have two contributions. First, we propose a new approach to assign a weight to the sub-regions of a face image. Second, a multi-resolution approach, to capture the local and global information of the face, is proposed. In addition, we have a discussion on the non-overlapped and overlapped regions. Finally, we compare our method with other state-of-the-art methods. In the experiments, *Chi-2* similarity measurement is implemented for unsupervised classification. In addition, the FERET data set [12, 13] is considered to generalize our experimental results. The rest of this paper is organized as follows: Section 2 is related to local binary pattern operator. Section 3 presents the proposed approaches. Experimental results and discussions are given in section 4. Section 5 concludes the paper.

2 Local Binary Pattern Operator (LBP) [14]

Due to impressive computational efficiency and good texture discriminative property of LBP operator[3], it has gained considerable attention since its publication. The LBP has already been used in many other applications including visual tracking, texture-based segmentation, image retrieval, face recognition, and texture classification. The LBP operator works in a 3×3 neighborhood, using the center value as a threshold. An LBP code is produced by multiplying the threshold values with weights given by the corresponding pixels. After that, the binary LBP code is converted to decimal number by using equations (1) and (2) to represent a unique spatial pattern:

$$\mathrm{LBP}_{\mathrm{P,R}} = \sum_{\mathrm{p}=0}^{\mathrm{P}-1} \mathrm{s}(\mathrm{g_p} - \mathrm{g_c}) 2^{\mathrm{p}} \tag{1}$$

$$\mathrm{s(x)} = \begin{cases} 1, & \mathrm{x} \geq 0 \\ 0, & \mathrm{x} \geq 0 \end{cases} \tag{2}$$

Where the gray value of the central pixel is g_c and g_p is the value of its neighbors, P is the number of neighbors and R is the radius of the neighborhood. $LBP\ {}^{ri}_{P,R}$ operator [3], defined by Equ. (3), removes effect of rotation. $ROR(x, j)$ performs a circular bit-wise right rotation j times on the P-bit number x. A majority of LBP patterns in a texture is termed "uniform" binary patterns that have limited number of transitions between zero and one. U value of an LBP ($LBP\ {}^{u2}_{P,R}$ operator) as shown by Equ. (4). $LBP\ {}^{riu2}_{P,R}$ operator is based on a circular symmetric neighborhood. In theory, it is invariant to any monotonic grey-scale transformation[15].

$$LBP\ {}^{ri}_{P,R} = min\{ROR(\mathrm{LBP_{P,R}}, j) \mid j = 0,1, \dots P-1\} \tag{3}$$

$$LBP\ {}^{u2}_{P,R} = U(\mathrm{LBP_{P,R}}) = |s(g_{P-1} - g_c) - s(g_0 - g_c)| + \sum_{p=1}^{P-1} |s(g_P - g_c) - s(g_{p-1} - g_c)| \tag{4}$$

Where the gray value of the central pixel is g_c and g_p is the value of its neighbors, R refers to the distance to the center, P stands for the number of sampling pixel in the neighborhood, and together they form the circularly symmetric neighborhood.

3 Proposed Approach

3.1 Fusing weight

We divide the face into some non-overlapping regions and calculate the LBP information of the regions as a feature vector. Regarding the recognition rate of each region on a training set, we define a weight for each region. We calculate the recognition rate of the regions in different LBP configurations (*R* and *P*). The best recognition rate under the specific LBP configuration makes the weight of each region. Figure 1 depicts the weights of regions under some configurations. The block size of the regions was considered 21×18 similar to [4]. Figure 2 depicts the best recognition rate of the regions along with the LBP configurations (fusing weight matrix).

2.06	1.55	2.06	1.03	2.06	1.55	2.06
1.55	3.61	8.76	2.06	8.76	3.61	1.55
1.03	0.52	0.52	0.00	0.52	0.52	1.03
0.52	0.52	0.00	0.00	0.00	0.52	0.52
0.52	0.00	0.52	0.00	0.52	0.00	0.52
0.52	0.52	0.52	0.00	0.52	0.52	0.52
0.00	2.06	1.55	0.00	1.55	2.06	0.00

(a) WP4R1

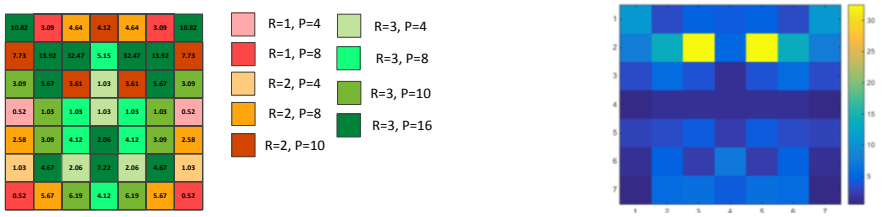
7.73	0.52	4.64	3.09	4.64	0.52	7.73
5.15	11.34	23.71	2.06	23.71	11.34	5.15
3.09	3.09	2.58	1.03	2.58	3.09	3.09
0.00	0.00	0.52	1.03	0.52	0.00	0.00
2.58	2.06	3.61	0.00	3.61	2.06	2.58
0.00	2.58	2.06	3.09	2.06	2.58	0.00
0.00	5.67	4.64	3.09	4.64	5.67	0.00

(b) WP8R2

10.82	1.55	1.55	2.06	1.55	1.55	10.82
5.67	13.92	32.47	3.09	32.47	13.92	5.67
2.58	5.67	3.09	1.03	3.09	5.67	2.58
0.00	1.03	1.03	1.03	1.03	1.03	0.00
1.55	3.09	3.09	2.06	3.09	3.09	1.55
0.00	4.64	2.06	7.22	2.06	4.64	0.00
0.00	2.58	5.15	4.12	5.15	2.58	0.00

(c) WP16R3

Figure 1 Weight values of the regions in some LBP configurations (*P* and *R*) with block size 21×18. (a) Weight under R=1 and P=4. (b) Weight under R=2 and P=8. (c) Weight under R=3 and P=16.



(a) (b)

Figure 2 Fusing weight matrix: the best recognition rate of the regions in different LBP configurations. (a) Fusion of region weights according to *P* and *R* configurations. (b) Weight values presented by the color bar.

3.2 Feature calculation method

For calculation of the feature vector in fusing weight approach, the face is first divided to sub regions. Regarding to the best LBP configurations in the fusing weigh matrix, the LBP information is calculated from all regions and concatenated together to make the feature vector. To reduce the feature dimensionality, we can eliminate the regions with low recognition rate. For example, the edge regions (in the bottom, in the left and right side of Figure 2.b) have very low efficiency and can be eliminated.

4 Experimental result and Discussion

4.1 Data set

The CSU Face Identification Evaluation System [12] was used to test the performance

of the proposed approaches. The CSU follows the procedure of the FERET test for semi-automatic face recognition algorithms [13] with slight modifications. Each set contains at most one image per person. These sets are *fa*, *fb*, *fc*, *dupI*, and *dupII*. The *fa* set (gallery set) contains frontal images of 1196 people. The *fb* set contains 1195 images. The subjects were asked for an alternative facial expression than in *fa* photograph. The *fc* set with 194 images was taken under different lighting conditions. The *dup I* set (722 images) were taken later in time and *dup II* set contains 234 images. This is a subset of the *dup I* set containing those images that were taken at least a year after the corresponding gallery image.

4.2 Face recognition results

The recognition rate of the FERET data set under different LBP configurations (*P* and *R*) and their respective weight of the regions are shown in Table I. As we can see, under fix LBP configuration, the best result can be obtained using *P*=10 and *R*=3. While, fusing weight approach improves recognition rate slightly better than the fix configuration approach.

Table I Recognition rate on the FERET data set under different weight calculation and LBP configuration with block size (21×18). The best scores are marked in bold.

		Fb	*fc***	*dupI*	*dupII*	*Weight*
R=1	P=4	94.06	55.67	45.43	31.20	*WP4R1*
	P=8	96.07	74.23	54.99	41.88	*WP8R1*
	P=4	94.64	69.07	57.76	55.56	*WP4R2*
R=2	P=8	96.99 (96.82*)	79.38	64.40 (65.7*)	63.68	*WP8R2*
	P=10	97.32	84.54	64.27	64.53	*WP8R2*
R=3	P=4	94.23	71.13	54.85	55.13	*WP4R3*
	P=8	96.90	83.51	63.99	63.25	*WP8R3*
	P=10	97.66	**87.63**	66.48	65.38	*WP10R3*
	P=16	96.65	87.63	63.16	66.67	*WP16R3*
Fusing Weight Approach		**97.74**	86.60	**66.62**	**67.09**	

* Ahonen's symmetric weighting approach taken from original paper [4]
** *fc* images from 1110 to 1206

4.3 Pyramid representation approach

In recent researches, pyramid representation has been used for achieving an effective local binary patterns texture descriptor [16], improving the scene categorization [17, 18], semantic concept retrieval [18], robustness against noise [19], and image classifying by the object categories they contain [20]. In the results, it can be seen that the feature selection from more pyramid representations can gather more information from an image. Further, more accuracy is attained by more pyramid representations. Therefore, apart from better accuracy we will have higher dimensionality. Of course, the researchers have shown peculiar effect that as the number of variables (feature dimensionality) is increased, the classification performance of the resulting decision surface initially improved, but then

began to deteriorate [21]. In this experiment we combine features of the first and the second pyramid representations with feature of original image. To keep the region size (21×18), we consider no down-sampling in pyramid representations. The weights of regions have been calculated similarly to before. The experimental results are shown in Table II. Regarding to [21], it can be seen that there is a slight improvement in recognition of some sets.

Table II Recognition rate under pyramid representation approach. There is no down-sampling in representations. The region size 21×18.

	fb	*fc*	*dupI*	*dupII*
$L0_{10,3}$+$L1_{16,3}$ (1)	98.08	81.44	68.14	67.09
$L0_{10,3}$+$L1_{16,3}$+$L2_{16,3}$ (2)	97.91	77.32	**69.39**	66.24
$L0_{FW}$+$L1_{FW}$ (3)	**98.07**	82.47	68.70	67.52
$L0_{FW}$+$L1_{FW}$+$L2_{FW}$ (4)	97.91	76.29	69.25	67.52

(1) Feature Vector = < Original Image under P=10;R=3, Image @ PL=1 under P=16;R=3 >. (2) Feature Vector = < Original Image under P=10;R=3, Image @ PL=1 under P=16;R=3, Image @ PL=2 under P=16;R=3 >. (3) Feature Vector = < Original Image under fused weight1, Image @ PL=1 under fused weight PL1 >. (4) Feature Vector = < Original Image under fused weight1, Image @ PL=1 under fused weight PL1, Image @ PL=2 fused weight PL2 >

4.4 Effect of using weights calculated in different configurations (Alternative weight)

In this experiment, we use the weight of the regions calculated under different values of *R* and *P* (e.g. WP4R1, WP8R1, …, WP16R3) for all LBP configurations. As we can see in the Table III, there is not a very huge variation in the recognition rate. The *fb* gallery with 1195 samples and *fc* gallery with the smallest samples (97 faces) have the lowest and the highest variance of recognition rate, respectively. Using alternative weight, the most recognition rate for *fb*, *fc*, *dupI*, and *dupII* have been obtained 98.16%, 89.69%, 66.48%, and 70.09%, respectively.

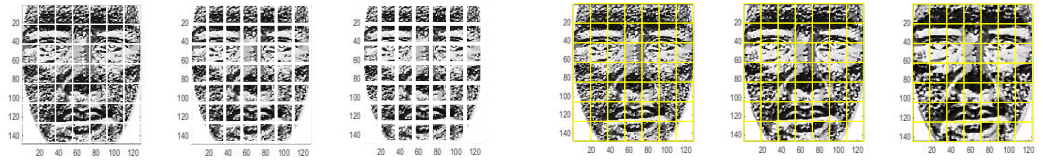
Table III The Effect of other region weights (Alternative weight).

Operator	*fb*	*fc*	*dupI*	*dupII*
$LBP_{4,1}$	92.62±0.83	64.26±4.19	47.71±1.00	36.75±2.70
$LBP_{8,1}$	95.23±0.56	73.88±1.93	55.02±1.07	46.72±3.15
$LBP_{4,2}$	94.82±0.56	67.92±4.23	55.36±1.50	52.90±3.15
$LBP_{8,2}$	96.76±0.61	79.38±3.18	63.05±1.21	63.87±3.39
$LBP_{10,2}$	97.31±0.56	83.05±2.78	63.90±1.13	63.68±2.76
$LBP_{4,3}$	95.04±0.56	65.06±4.30	55.32±1.46	51.47±3.00
$LBP_{8,3}$	96.93±0.61	82.13±2.82	63.28±1.30	60.45±2.89
$LBP_{10,3}$	97.31±0.53	86.03±2.07	64.68±1.46	62.68±2.79
$LBP_{16,3}$	96.90±0.54	87.97±1.26	62.94±1.45	66.05±3.60

4.5 Overlapped regions against non-overlapped regions

Ahonen *et al*. [4] split the image into some non-overlapped regions and extract the LBP images from each sub-images. Because of the small size of the regions (21×18), in their

approach we cannot use a large radius (R) for extraction of the LBP information. In the overlapped approach, we first calculate the LBP information of the image. After that, we split the image into the sub-regions and extract the LBP feature (histogram) from each sub-region. In the non-overlapped approach, shown in Figure 3.a, the dimension of the original image is 147×126 (21×18×7×7). The number of patterns under R=1, 2, and 3 will be 14896, 11662 and 8820. While, in the overlapped approach, shown in Figure 3.b, the number of patterns are 18944, 18396, and 17856 that are 27%, 58%, 102% more than non-overlapped approach. Table IV depicts the recognition rate of the overlapped approach. Better results than non-overlapped approach are marked in bold and improvement value have been shown in the parentheses. As we can see, regardless to have a more patterns in overlapped approach and improvement in the weight of the regions, there is not significant improvement in accuracy. We calculated the number of non-uniform and uniform patterns in non-overlapped and overlapped approaches. We saw that the relation on non-uniform to uniform patterns in both approaches are almost the same. It means that using the uniform LBP operator (LBP^{u2}), we will not have a significant improvement in the recognition rate.



R=1, P=8 R=2, P=8 R=3, P=8 R=1, P=8 R=2, P=8 R=3, P=8

(a) (b)

Figure 3 Non-overlapped and overlapped approaches. (a) Non-overlapped approach: A region with size 21×18 reduce to 19×16, 17×14, and 15×12 under R = 1, 2, and 3, respectively. (b) Overlapped approach: The number of patterns for R=1, 2, and 3 is 148×128, 146×126, and 144×124, respectively.

Table IV Recognition rate for overlapped regions approach on the FERET data set. Better scores than non-overlapped approach are marked in bold.

		fb	*fc*	*dupI*	*dupII*
R=1	P=4	92.80	**58.76**(+1.26)	**46.54** (+1.11)	**32.48** (+1.28)
	P=8	94.98	71.13	**55.26** (+0.27)	41.45
R=2	P=4	91.97	50.52	50.97	42.31
	P=8	**97.32** (+0.33)	71.13	59.83	55.13
	P=10	**97.82** (+0.50)	73.20	62.47	58.12
R=3	P=4	92.47	48.45	**57.06** (+2.23)	50.00
	P=8	**96.99** (+0.09)	70.10	63.30	60.26
	P=10	**97.74** (+0.08)	78.35	66.48	64.10
	P=16	**97.91** (+1.26)	82.47	**66.90** (+3.74)	66.24
Fusing Weight Approach		**98.33** (+0.59)	83.51	**68.56** (+1.94)	**70.51** (+3.42)
$L0_{16,3}$+$L1_{16,3}$ [(1)]		**98.16** (+0.08)	**83.51** (+2.07)	**69.67** (+1.53)	**69.23** (+2.14)
$L0_{16,3}$+$L1_{16,3}$+$L2_{16,3}$ [(2)]		**98.16** (+0.25)	**79.38** (+2.06)	**69.81** (+0.42)	**67.09** (+0.85)

[(1)] Feature Vector = < Original Image under P=16;R=3, Image @ PL=1 under P=16;R=3 >. [(2)] Feature Vector = < Original Image under P=16;R=3, Image @ PL=1 under P=16;R=3, Image @ PL=2 under P=16;R=3 >

Comparison of the best verification rate of the proposed approaches with other state-of-the-art methods have been shown in Table V. The methods use different cores of similarity measure and the comparison is not fair. However, we can see that the proposed approaches with simplest similarity measurement (*Chi-2*) have acceptable scores.

Table V Comparison of the best verification rate of the proposed approaches with other state-of-the-art methods.

Method	*fb*	*fc*	*dupI*	*dupII*	Core of Similarity measure
LBP [4]	93.39	50.5	61.4	49.6	*Chi-2* (LBP^{u2})
LBP Weigthed [4]	96.82	79.4	65.7	63.7	*Chi-2* (LBP^{u2})
LGBPHS [5]	98	97	74	71	WHIS
Multi-Scale LBP [6]	98.6	71.1	72.2	47.4	LDA
LGBP [7]	99.6	99	92	88.9	EPFDA
ELGBP (Mag+Pha) [8]	99	96	78	77	WHIS
Gabor+LBP [9]	98	98	90	85	KDCV
MSLBP (+ mean filter) [10]	96.8 (97.8)				MBDFW+WNNC (LBP)
MSLBP + MF + FW (+ BW) [10]	98.1 (99.2)	-	-	-	MBDFW+WNNC (LBP)
MSLBP + MF + FW + BW + PCA [10]	99.1	-	-	-	MBDFW+WNNC (LBP)
LDPv weighted [11]	0.97	0.74	0.64	0.59	*Chi-2* (LDP)
LDPv un-weighted [11]	0.97	0.71	0.6	0.57	*Chi-2* (LDP)
The best result in our approaches	98.33	89.69	69.39	70.51	*Chi-2* (LBP^{u2})

5 Conclusion

Experimental results have been shown that the most recognition rate on the FERET data set can be obtained under LBP configuration P=10 and R=3 and weighted region approach. The experimental result depicted that fusing weight approach improves the recognition rate. Pyramid approach makes a slight improvement in recognition rate with huge computation cost. In addition, with the overlapped regions, partial recognition rate (or weight of regions) has been improved and we significantly save image patterns.

For the future works, we should use other similarity measurements that have been used in other state-of-the-art methods to compare fairly with the proposed approaches. To reduce the feature dimensionality we can eliminate ineffective regions and use feature reduction methods like PCA.

6 References

[1] L. Spreeuwers, "Fast and Accurate 3D Face Recognition," *International Journal of Computer Vision*, vol. 93, pp. 389-414, 2011/07/01 2011.

[2] B. J. Boom, G. M. Beumer, L. J. Spreeuwers, and R. N. J. Veldhuis, "The Effect of Image Resolution on the Performance of a Face Recognition System," in *Control, Automation, Robotics and Vision, 2006. ICARCV '06. 9th International Conference on*, 2006, pp. 1-6.

[3] T. Ojala, M. Pietikainen, and T. Maenpaa, "Multiresolution gray-scale and rotation invariant texture classification with local binary patterns," *Pattern Analysis and Machine*

Intelligence, IEEE Transactions on, vol. 24, pp. 971-987, 2002.

[4] T. Ahonen, A. Hadid, and M. Pietikäinen, "Face description with local binary patterns: application to face recognition," *IEEE Trans. Pattern Anal. Mach. Intell.,* vol. 28 pp. 2037-2041, 2006.

[5] Z. Wenchao, S. Shiguang, G. Wen, C. Xilin, and Z. Hongming, "Local Gabor binary pattern histogram sequence (LGBPHS): a novel non-statistical model for face representation and recognition," in *Computer Vision, 2005. ICCV 2005. Tenth IEEE International Conference on*, 2005, pp. 786-791 Vol. 1.

[6] C.-H. Chan, J. Kittler, and K. Messer, "Multi-scale Local Binary Pattern Histograms for Face Recognition," in *Advances in Biometrics*. vol. 4642, S.-W. Lee and S. Li, Eds., ed: Springer Berlin Heidelberg, 2007, pp. 809-818.

[7] S. Shiguang, Z. Wenchao, S. Yu, C. Xilin, and G. Wen, "Ensemble of Piecewise FDA Based on Spatial Histograms of Local (Gabor) Binary Patterns for Face Recognition," in *Pattern Recognition, 2006. ICPR 2006. 18th International Conference on*, 2006.

[8] W. Zhang, S. Shan, L. Qing, X. Chen, and W. Gao, "Are Gabor phases really useless for face recognition?," *Pattern Anal Applic,* vol. 12, pp. 301-307, 2009.

[9] Xiaoyang Tan and B. Triggs, "Enhanced Local Texture Feature Sets for Face Recognition Under Difficult Lighting Conditions," *IEEE Transactions on Image Processing,* vol. 19, pp. 1635-1650, 2010.

[10] O. Nikisins, " Weighted Multi-scale Local Binary Pattern Histograms for Face Recognition," presented at the The 2013 International Conference on Applied Mathematics and Computational Methods, AMCM Venice, Italia, 2013.

[11] S. Z. Ishraque, T. Jabid, and O. Chae, "Face Recognition Based on Local Directional Pattern Variance (LDPv)," 2012.

[12] M. Topi, P. Matti, and O. Timo, "Texture classification by multi-predicate local binary pattern operators," in *Pattern Recognition, 2000. Proceedings. 15th International Conference on*, 2000, pp. 939-942 vol.3.

[13] T. Randen and J. H. Husoy, "Texture segmentation using filters with optimized energy separation," *Image Processing, IEEE Transactions on*, vol. 8, pp. 571-582, 1999.

[14] M. Pietikäinen, A. Hadid, G. Zhao, and T. Ahonen, *Computer Vision Using Local Binary Patterns* vol. 40: Springer, 2011.

[15] F. Xiao, Y. Liang, and X. Qu, "Learning Local Binary Patterns with Enhanced Boosting for Face Recognition " presented at the Seventh International Conference on Computational Intelligence and Security (CIS), , 2011.

[16] X. Qian, X.-S. Hua, P. Chen, and L. Ke, "PLBP: An effective local binary patterns texture descriptor with pyramid representation," *Pattern Recognition,* vol. 44, pp. 2502-2515, 2011.

[17] S. Lazebnik, C. Schmid, and J. Ponce, "Beyond Bags of Features: Spatial Pyramid Matching for Recognizing Natural Scene Categories," in *Computer Vision and Pattern Recognition, 2006 IEEE Computer Society Conference on*, 2006, pp. 2169-2178.

[18] X. Qian, G. Liu , D. Guo, Z. Li, Z. Wang, and H. Wang, "HWVP: hierarchical wavelet packet descriptors and their applications in scene categorization and semantic concept retrieval," *Multimed Tools Appl,* June 2012.

[19] A. Akoushideh and B. Mazloom-Nezhad Maybodi, "High-accurate and noise-tolerant texture descriptor," 2015, pp. 94450V-94450V-6.

[20] A. Bosch, A. Zisserman, and X. Muñoz, "Representing shape with a spatial pyramid kernel," presented at the CIVR, 2007.

[21] D. J. Hand, *Discrimination and Classification*: John Wiley & Sons, 1981.

Decoding delay in network coded multipath transmissions

Berksan Serbetci, Jasper Goseling,
Jan-Kees van Ommeren and Richard J. Boucherie

Stochastic Operations Research, University of Twente, The Netherlands

{b.serbetci, j.goseling,
j.c.w.vanommeren, r.j.boucherie}@utwente.nl

Abstract

We investigate the decoding delay performance of a communication network in which a single source is transmitting data packets to a single receiver via multiple routers. Network coding is applied to all data packets at the source at each transmission opportunity. Receiver receives network coded packets from routers and decodes them. We define the delay as the time between arrival of a data packet at the source and decoding of all the packets served in the busy period of the source queue starting from the arrival of that data packet. We show that the delay can be expressed in closed-form.

1 Introduction

In modern communication networks, data packets are transmitted from the gateway to user equipments via base stations. In principle, each base station is responsible of transmitting data packets to users that are present in its coverage. In practice, there are many areas that are covered by multiple base stations. Depending on channel conditions, it may be more viable for the user equipment to receive data packets from different base stations at different transmission opportunities. At some occasions, it is possible that same data packets are requested by multiple users. Then, we can come up with an alternative way of transmitting data packets to these users as in [1], [2], [3]. In these works, it is shown that sending random linear combinations of all data packets is another way of transmitting all data packets and this alternative data transmission scheme is called network coding.

The system consists of a single source transmitting data packets to a single receiver via multiple routers. The source refers to gateway, routers refer to base stations and receiver refers to user equipment. New data packets arrive at the source according to a Poisson process. The intermediate network consists of two routers that receive packets from the source and forward these to the receiver. The source and the routers have exponential service rates. The source transmits network coded packets through the network. In particular, at each transmission opportunity, the source transmits a random linear combination over all data packets that are present at the source at that time. Each network coded packet is then transmitted to one of the routers with probabilistic routing. Once a network coded packet is transmitted to one of the routers, the source drops the data packet that is located at the head of the queue as proposed in [4], [5], [6].

As the receiver obtains network coded packets from multiple routers, it is necessary to decode these network coded packets in order to retrieve the data packets. Decoding is only possible when the number of network coded packets is at least equal to the number of data packets involved in the received linear combinations. We show that once the source queue becomes empty and all network coded packets that have been generated so far have been received by the receiver, it decodes all these network coded packets and retrieve data packets.

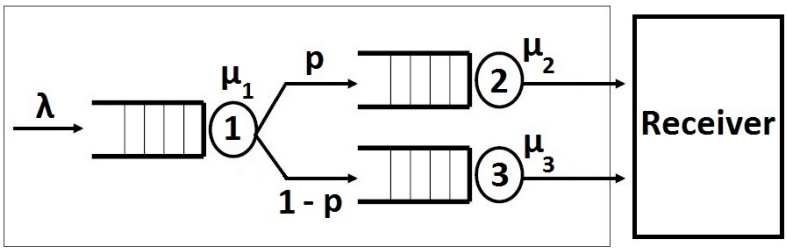


Figure 1: Queueing network for the system.

This work mainly focuses on analyzing the delay where the delay is defined as the time between arrival of a data packet at the source and decoding of all the packets served in the busy period of the source queue starting from the arrival of that data packet. Note that even though sending network coded packets do not save any resource over sending data packets for unicast transmissions, it is still useful to analyze the delay for the unicast system in order to prepare a baseline for future research.

2 Model and Problem Statement

We define the source and the routers as independent $M/M/1$ queues. Data packets arrive at the source according to a Poisson process with rate λ. The source queue is called *Queue* 1 ($Q1$) and has an exponential service rate μ_1. At each transmission opportunity, network coding is applied to all data packets at the source, namely each data packet that is present at the source at that transmission opportunity is multiplied with a random coefficient and the sum of them forms a network coded packet. Then the network coded packet is routed to one of the two routers called *Queue* 2 ($Q2$) and *Queue* 3 ($Q3$) with probabilistic routing with parameter p. Namely, the network coded packet that is ready to be transmitted from the source is routed to $Q2$ with probability p, and to $Q3$ with probability $1-p$. The system is shown in Figure 1.
As the routers transmit network coded packets to the receiver, the receiver must decode these network coded packets in order to retrieve the data packets. $Q2$ and $Q3$ have exponential service rates μ_2 and μ_3 respectively. The receiver can decode the data packets when it receives as many network coded packets as at least equal to the number of data packets involved in the received linear combinations. Once a network coded packet is transmitted to one of the routers, the source drops the data packet that is located at the head of the queue. Then, when the source queue becomes empty and all network coded packets that have been generated so far have been received by the receiver, this condition is satisfied. We assume that all linear combinations that have been generated are independent. Probability of receiving identical network coded packets is neglected. This probability can be made arbitrarily small by making the field size over which network coding is performed sufficiently large. In this work, the delay which is defined as the time between arrival of a data packet at the source and decoding of all the packets served in the busy period of the source queue starting from the arrival of that data packet is analyzed.

3 Preliminaries

At this section, we will list the specifications and necessary tools that we will use to analyze the system that is the shown in Figure 1. The source and the routers form a Jackson network. From the traffic equations of the Jackson network, it follows from [7]

that arrival rates to the routers are $\lambda_2 = p\lambda$ and $\lambda_3 = (1-p)\lambda$. We denote $\rho_1 = \lambda/\mu_1$, $\rho_2 = p\lambda/\mu_2$ and $\rho_3 = (1-p)\lambda/\mu_3$. Throughout the paper, we assume $\rho_1 < 1$, $\rho_2 < 1$ and $\rho_3 < 1$ for stability.

Lemma 1. *It follows directly from [7] that equilibrium distribution of the system is defined as*

$$\pi(n_1, n_2, n_3) = \prod_{i=1}^{3} (1-\rho_i)\rho_i^{n_i}$$

where n_1, n_2 and n_3 are the number of packets located at $Q1$, $Q2$ and $Q3$ respectively.

Based on the proposed performance parameter, we need to define the probability of a departure from $Q1$ when $n_1 = 1$ so that $Q1$ becomes empty after this departure and then the receiver can decode the packets that it has received from the source in $Q1$'s last busy period. In order to do so, we need to define *Palm probabilities*. *Palm probability* is used on defining a specific transition by characterizing the past and future of a Continuous Time Markov Chain (CTMC) at such a transition. The issue deals with how to evaluate any probability for a CTMC conditioned that a specific transition occurs. Since the occurrence of any specific transition at any time has probability 0, conventional conditional probabilities cannot be used. Instead these conditional probabilities must be formulated as *Palm probabilities*. Then, the *Palm probability* of a stationary CTMC conditioned that a specific transition occurs at any time is the ratio of the expected number of that specific transition at which that specific transition occurs in a fixed time interval divided by the expected number of all possible transitions in the interval.

Theorem 1. *Palm probability* $P_H(C)$ of event C given that H occurs for an $M/M/1$ queue follows directly from [8] as:

$$P_H(C) = \frac{\sum_{(n,n')\in C} \pi(n)q(n,n')}{\sum_{(n,n')\in H} \pi(n)q(n,n')}, \quad C \subseteq H$$

where n is the current state, n' is the next state, $\pi(n)$ is the equilibrium distribution and $q(n,n')$ is the transition rate.

4 Analysis

The system can be defined as a three dimensional Markov chain with state space $S = (n_1, n_2, n_3)$ where each non-negative value corresponds to number of customers in $Q1$, $Q2$ and $Q3$ respectively. $Q1$, $Q2$ and $Q3$ become busy and idle sequentially as shown in Figure 2. Transitions between these states are the crucial moments as specified earlier. At time A_1, busy period of $Q1$ started. At B_1, last packet is served from $Q1$ and it becomes empty again. Hence, $B_1 - A_1$ is a busy period duration for $Q1$. At B_1, all coded packets that have been served in the busy period $[A_1, B_1]$ are routed to $Q2$ and $Q3$. Then, $Q2$ becomes empty for the first time after $Q1$ finishes its busy period at B_2' and $Q3$ becomes empty at B_1''. Hence, we define two time parameters defined as $T_2 = B_2' - B_1$ and $T_3 = B_1'' - B_1$. When $Q1$ finishes its busy period, $Q2$ becomes empty after T_2 and $Q3$ becomes empty after T_3. To conclude, the receiver can decode all packets served in $[A_1, B_1]$ in $T_{dec} = B_1 - A_1 + \max\{T_2, T_3\}$. Once the distribution of the number of packets at $Q2$ and $Q3$ is known at the end of the busy period of $Q1$, we can find the maximum of T_2 and T_3.

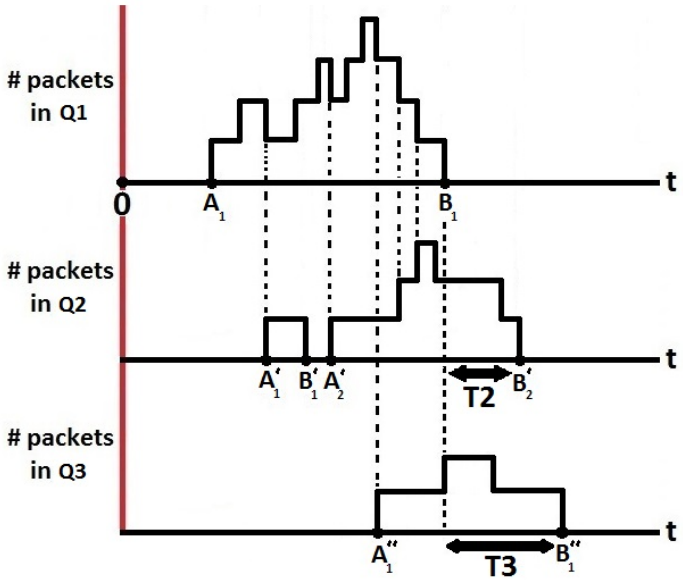


Figure 2: Timeline for the system.

Lemma 2. *The distribution of the number of packets at Q2 and Q3 at the end of the busy period of Q1 is equal to*

$$P(N_2 = n_2, N_3 = n_3 | N_1 \to 0) = \begin{cases} R\left(p\rho_2^{n_2-1}\rho_3^{n_3} + (1-p)\rho_2^{n_2}\rho_3^{n_3-1}\right) & \text{if } n_2 \geq 1, n_3 \geq 1 \\ Rp\rho_2^{n_2-1} & \text{if } n_2 \geq 1, n_3 = 0 \\ R(1-p)\rho_3^{n_3-1} & \text{if } n_2 = 0, n_3 \geq 1 \end{cases} \tag{1}$$

where $R = (1-\rho_2)(1-\rho_3)$.

Proof. We will use Theorem 1 here. We define the event C as one packet from the source is transmitted to one of the two routers when $n_1 = 1$ and the event H as all possible transitions from $Q1$ when $n_1 = 1$.
For $n_2 \geq 1, n_3 \geq 1$,

$$\begin{aligned} P(N_2 = n_2, N_3 = n_3 | N_1 \to 0) &= \frac{\pi(1, n_2-1, n_3)p\mu_1}{\sum_{j=0}^{\infty}\sum_{k=0}^{\infty}\pi(1,j,k)\mu_1} + \frac{\pi(1, n_2, n_3-1)(1-p)\mu_1}{\sum_{j=0}^{\infty}\sum_{k=0}^{\infty}\pi(1,j,k)\mu_1} \\ &= \frac{\rho_1\rho_2^{n_2-1}\rho_3^{n_3}p}{\sum_{j=0}^{\infty}\sum_{k=0}^{\infty}\rho_1\rho_2^j\rho_3^k} + \frac{\rho_1\rho_2^{n_2}\rho_3^{n_3-1}(1-p)}{\sum_{j=0}^{\infty}\sum_{k=0}^{\infty}\rho_1\rho_2^j\rho_3^k} \\ &= \frac{\rho_2^{n_2-1}\rho_3^{n_3}p}{\frac{1}{1-\rho_2}\frac{1}{1-\rho_3}} + \frac{\rho_2^{n_2}\rho_3^{n_3-1}(1-p)}{\frac{1}{1-\rho_2}\frac{1}{1-\rho_3}} \\ &= R\left(p\rho_2^{n_2-1}\rho_3^{n_3} + (1-p)\rho_2^{n_2}\rho_3^{n_3-1}\right). \end{aligned}$$

For $n_2 \geq 1, n_3 = 0$,

$$\begin{aligned} P(N_2 = n_2, N_3 = 0 | N_1 \to 0) &= \frac{\pi(1, n_2-1, 0)p\mu_1}{\sum_{j=0}^{\infty}\sum_{k=0}^{\infty}\pi(1,j,k)\mu_1} \\ &= \frac{p\rho_1\rho_2^{n_2-1}}{\sum_{j=0}^{\infty}\sum_{k=0}^{\infty}\rho_1\rho_2^j\rho_3^k} \end{aligned}$$

$$= \frac{p\rho_2^{n_2-1}}{\frac{1}{1-\rho_2}\frac{1}{1-\rho_3}}$$
$$= Rp\rho_2^{n_2-1}.$$

And similarly, for $n_2 = 0, n_3 \geq 1$,

$$P(N_2 = 0, N_3 = n_3 | N_1 \to 0) = R(1-p)\rho_3^{n_3-1}.$$

□

Theorem 2. *We know that for each queue every single service time is exponentially distributed and service times between n occurrences are Erlang-distributed with the number of packets n, rate parameter λ_{Er} with mean $\mu_{Er} = n/\lambda_{Er}$ and cdf*

$$F(t) = 1 - \sum_{j=0}^{n-1} \frac{(\lambda_{Er}t)^j}{j!} e^{-\lambda_{Er}t}, \quad t \geq 0.$$

We have $T_2 \sim Er\{n_2\}$ and $T_3 \sim Er\{n_3\}$ where n_2 and n_3 are number of customers at any moment in $Q2$ and $Q3$ respectively. We need to determine the expected value of $T_{max} = \max\{T_2, T_3\}$.

Lemma 3. *We have*

$$E[T_{max}] = \frac{p\left(\mu_2 + (1-p)\lambda\right)}{\mu_2\left(\mu_2 - p\lambda\right)} + \frac{(1-p)\left(\mu_3 + p\lambda\right)}{\mu_3\left(\mu_3 - (1-p)\lambda\right)} - \frac{p(1-p)\lambda(\mu_2+\mu_3)}{\mu_2\mu_3(\mu_2+\mu_3-\lambda)}.$$

Proof.

$$\begin{aligned}
E[T_{max}] &= E\left[E[T_{max}|N_2, N_3]\right] \\
&= E\left[E[\max\{T_2, T_3\}|N_2, N_3]\right. \\
&= \sum_{n_2=0}^{\infty}\sum_{n_3=0}^{\infty} E[\max\{T_2, T_3\}|N_2 = n_2, N_3 = n_3] P(N_2 = n_2, N_3 = n_3 | N_1 \to 0) \\
&= \underbrace{\sum_{n_2=1}^{\infty} pR\rho_2^{n_2-1} \int_0^{\infty} P(T_2 > m)dm}_{S_1} + \underbrace{\sum_{n_3=1}^{\infty}(1-p)R\rho_3^{n_3-1} \int_0^{\infty} P(T_3 > m)dm}_{S_2} \\
&+ \underbrace{\sum_{n_2=1}^{\infty}\sum_{n_3=1}^{\infty} R\left(p\rho_2^{n_2-1}\rho_3^{n_3} + (1-p)\rho_2^{n_2}\rho_3^{n_3-1}\right) \int_0^{\infty} P(T_2 > m)dm}_{S_3} \\
&+ \underbrace{\sum_{n_2=1}^{\infty}\sum_{n_3=1}^{\infty} R\left(p\rho_2^{n_2-1}\rho_3^{n_3} + (1-p)\rho_2^{n_2}\rho_3^{n_3-1}\right) \int_0^{\infty} P(T_3 > m)dm}_{S_4} \\
&- \underbrace{\sum_{n_2=1}^{\infty}\sum_{n_3=1}^{\infty} R\left(p\rho_2^{n_2-1}\rho_3^{n_3} + (1-p)\rho_2^{n_2}\rho_3^{n_3-1}\right) \int_0^{\infty} P(T_2 > m)P(T_3 > m)dm}_{S_5}
\end{aligned} \tag{2}$$

We split (2) into 5 pieces and compute these terms separately. We start with computing S_1 as follows

$$\begin{aligned}
S_1 &= \sum_{n_2=1}^{\infty} pR\rho_2^{n_2-1} \int_0^{\infty} P(T_2 > m)dm \\
&= \sum_{n_2=1}^{\infty} pR\rho_2^{n_2-1} \int_0^{\infty} \left[\sum_{i=0}^{n_2-1} \frac{1}{i_1!} e^{-\mu_2 m} (\mu_2 m)^i \right] dm \\
&= \sum_{n_2=1}^{\infty} pR\rho_2^{n_2-1} \sum_{i=0}^{n_2-1} \frac{1}{i_1!} \left[\int_0^{\infty} e^{-\mu_2 m} (\mu_2 m)^i \right] dm \\
&= \sum_{n_2=1}^{\infty} pR\rho_2^{n_2-1} \sum_{i=0}^{n_2-1} \frac{1}{\mu_2} \\
&= \sum_{n_2=1}^{\infty} pR\rho_2^{n_2-1} \frac{n_2}{\mu_2} \\
&= \frac{p(1-\rho_3)}{\mu_2(1-\rho_2)}.
\end{aligned}$$

Similarly,

$$S_2 = \frac{(1-p)(1-\rho_2)}{\mu_3(1-\rho_3)}.$$

Due to space constraints, computation steps of S_3 and S_4 are skipped and the results of the two parameters are given as

$$S_3 = \frac{(1-p)\rho_2 + p\rho_3}{\mu_2(1-\rho_2)},$$

and

$$S_4 = \frac{(1-p)\rho_2 + p\rho_3}{\mu_3(1-\rho_3)}.$$

Finally S_5 can be computed as follows

$$\begin{aligned}
S_5 &= \sum_{n_2=1}^{\infty} \sum_{n_3=1}^{\infty} R \underbrace{\left(p\rho_2^{n_2-1}\rho_3^{n_3} + (1-p)\rho_2^{n_2}\rho_3^{n_3-1} \right)}_{h(n_2,n_3)} \int_0^{\infty} P(T_2 > m) P(T_3 > m) dm \\
&= R \sum_{n_2=1}^{\infty} \sum_{n_3=1}^{\infty} h(n_2, n_3) \sum_{i=0}^{n_2-1} \sum_{j=0}^{n_3-1} \frac{1}{i!j!} \int_0^{\infty} e^{-(\mu_2+\mu_3)m} (\mu_2 m)^i (\mu_3 m)^j dm \\
&= \underbrace{\frac{R}{(\mu_2+\mu_3)}}_{K} \sum_{n_2=1}^{\infty} \sum_{n_3=1}^{\infty} h(n_2, n_3) \sum_{i=0}^{n_2-1} \sum_{j=0}^{n_3-1} \frac{(i+j)!}{i!j!} \frac{\mu_2^i \mu_3^j}{(\mu_2+\mu_3)^{i+j}} \\
&= K \sum_{i=0}^{\infty} \sum_{j=0}^{\infty} \frac{(i+j)!}{i!j!} \frac{\mu_2^i \mu_3^j}{(\mu_2+\mu_3)^{i+j}} \sum_{n_2=i+1}^{\infty} \sum_{n_3=j+1}^{\infty} \left(p\rho_2^{n_2-1}\rho_3^{n_3} + (1-p)\rho_2^{n_2}\rho_3^{n_3-1} \right) \\
&= K \sum_{i=0}^{\infty} \sum_{j=0}^{\infty} \frac{(i+j)!}{i!j!} \left(\frac{\mu_2}{\mu_2+\mu_3} \right)^i \left(\frac{\mu_3}{\mu_2+\mu_3} \right)^j \left(\frac{\rho_2^i \rho_3^j \left[(1-p)\rho_2 + p\rho_3 \right]}{(1-\rho_2)(1-\rho_3)} \right)
\end{aligned}$$

$$= \frac{(1-p)\rho_2 + p\rho_3}{\mu_2+\mu_3} \sum_{i=0}^{\infty}\sum_{j=0}^{\infty} \frac{(i+j)!}{i!j!} \left(\frac{p\lambda}{\mu_2+\mu_3}\right)^i \left(\frac{(1-p)\lambda}{\mu_2+\mu_3}\right)^j$$
$$= \frac{(1-p)\rho_2 + p\rho_3}{\mu_2+\mu_3} \sum_{i=0}^{\infty} \frac{1}{i!} \left(\frac{p\lambda}{\mu_2+\mu_3}\right)^i \left(1-\frac{(1-p)\lambda}{\mu_2+\mu_3}\right)^{-1-i} i!$$
$$= \frac{(1-p)\rho_2 + p\rho_3}{\mu_2+\mu_3} \left(\frac{\mu_2+\mu_3}{\mu_2+\mu_3-\lambda}\right)$$
$$= \frac{p(1-p)\lambda(\mu_2+\mu_3)}{\mu_2\mu_3(\mu_2+\mu_3-\lambda)}.$$

Then we use (2) to compute $E[T_{max}]$ as follows

$$E[T_{max}] = S_1 + S_2 + S_3 + S_4 - S_5$$
$$= \frac{p\left(\mu_2+(1-p)\lambda\right)}{\mu_2\left(\mu_2-p\lambda\right)} + \frac{(1-p)\left(\mu_3+p\lambda\right)}{\mu_3\left(\mu_3-(1-p)\lambda\right)} - \frac{p(1-p)\lambda(\mu_2+\mu_3)}{\mu_2\mu_3(\mu_2+\mu_3-\lambda)}.$$

□

Now we are ready to state the main result.

Theorem 3. *The total expected delay for the time that is needed to decode all the packets served in a single busy period of $Q1$ is equal to*

$$E[T_{dec}] = \frac{1}{\mu_1-\lambda} + \frac{p\left(\mu_2+(1-p)\lambda\right)}{\mu_2\left(\mu_2-p\lambda\right)} + \frac{(1-p)\left(\mu_3+p\lambda\right)}{\mu_3\left(\mu_3-(1-p)\lambda\right)} - \frac{p(1-p)\lambda(\mu_2+\mu_3)}{\mu_2\mu_3(\mu_2+\mu_3-\lambda)}.$$

Proof. All packets served in a busy period of $Q1$ will certainly be decoded in $E[T_{dec}]$ which is computed as follows:

$$E[T_{dec}] = E[BP_{Q1}] + E[T_{max}]$$
$$= \frac{1}{\mu_1-\lambda} + \frac{p\left(\mu_2+(1-p)\lambda\right)}{\mu_2\left(\mu_2-p\lambda\right)} + \frac{(1-p)\left(\mu_3+p\lambda\right)}{\mu_3\left(\mu_3-(1-p)\lambda\right)} - \frac{p(1-p)\lambda(\mu_2+\mu_3)}{\mu_2\mu_3(\mu_2+\mu_3-\lambda)}.$$

□

For $\lambda = 1$, $\mu_1 = 4$, $\mu_2 = 2$, $\mu_3 = 0.5$, expected delay vs. probabilistic routing parameter p graph is shown in Figure 3. Delay can be computed only for the case when all queues are stable and it is infinity otherwise. For this specific example, $Q3$ has a lower service rate. $Q3$ receives more packets as p decreases. When $(1-p)\lambda > \mu_3$, $Q3$ is not stable anymore and the queue is exploded. This means that receiver will not be able to decode data packets. As p increases, $Q2$ starts receiving more packets and delay decreases since $Q3$ has a lower service rate compared to $Q2$.

5 Discussion & Conclusion

In this work, we have presented a network scenario containing a source transmitting network coded packets via multiple routers to a receiver. The receiver must receive enough number of packets to decode network coded packets and retrieve data packets. We define the delay as the time between arrival of a data packet at the source and decoding of all the packets served in the busy period of the source queue starting from the arrival of that data packet. We show that for the proposed network scenario, the

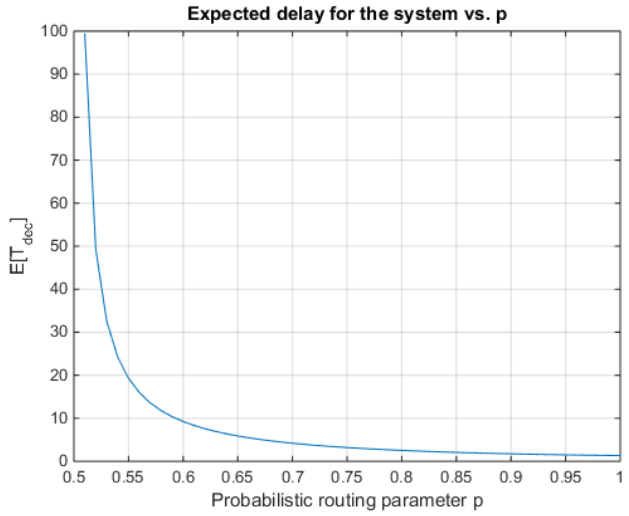


Figure 3: Expected delay vs. p for $\lambda = 1$, $\mu_1 = 4$, $\mu_2 = 2$, $\mu_3 = 0.5$.

delay can be expressed in closed-form. In practice, as the service rates of the routers change due to channel conditions, it is possible to minimize the delay by changing the probabilistic routing parameter p.
Even though sending network coded packets do not save any resources over sending data packets for unicast transmission scenarios, it is still useful to analyze the delay for the unicast system in order to prepare a baseline for future research. Various systems with different network coding techniques and comparisons between coded and non-coded systems will be analyzed in the future.

References

[1] R. Ahlswede, N. Cai, S.-Y. R. Li, and R. W. Yeung, "Network information flow", *IEEE Trans, Inform. Theory*, vol. 46, no. 4, pp. 1204-1216, July 2000.

[2] P. A. Chou, Y. Wu, and K. Jain, "Practical network coding", *Proc. 41st Allerton Conf. Communication, Control and Computing*, Monticello, IL, Oct. 2003.

[3] T. Ho, R. Koetter, M. Medard, M. Effros, J. Shi, and D. Karger, "A random linear network coding approach to multicast", *IEEE Trans, on Information Theory*, vol. 52, iss. 10, pp. 4413-4430, October 2006.

[4] C. Fragouli, D. Lun, M. Medard, and P. Pakzad, "On feedback for network coding", *CISS 2007*, March 2007.

[5] J. K. Sundararajan, D. Shah, and M. Medard, "ARQ for Network Coding", *ISIT 2008*, Toronto, Canada, July 2008.

[6] J. K. Sundararajan, D. Shah, M. Medard, M. Mitzenmacher, and J. Barros, "Network coding meets TCP", *INFOCOM 2009*, April 2009.

[7] L. Lipsky, "Queueing Theory: A Linear Algebraic Approach", Second Edition, Springer, 2009.

[8] R. Serfozo, "Basics of Applied Stochastic Processes", Springer, Berlin, 2009.

Video analysis for acute pain detection in infants

B.P.S. Slaats, S. Zinger, P. H.N. de With, W. Tjon a Ten, S. Bambang Oetomo

TU/e, Dept. Signal Processing, Group VCA
Den Dolech 2, Eindhoven
biancaslaats@gmail.com s.zinger@tue.nl

Abstract

Detecting pain in infants is of vital importance in healthcare. This work investigates two different systems for automated continuous facial analysis for detection of stress and pain in infants. The first system uses an Active Appearance Model (AAM) and a three-class SVM classifier. The second system detects three Regions Of Interest (ROI), aiming at detecting the presence of the brow bulge, eye squeeze and the nasolabial furrow. For this system, the resulting pain/stress level is detected with an accuracy of 67%. The second system follows directly the PIPP pain scale form and is able to detect the facial regions, even with occlusions like feeding tubes. The first (AAM-based) system is not able to handle occlusions, but has an accuracy of 92%, classifying the facial expressions into comfort, discomfort and the Primal Face of Pain (PFP).

1 Introduction

Infants cannot communicate verbally and are therefore unable to report about their discomfort and pain. Frequent and long-term pain can cause severe complications, such as a delay in development or a change in the nervous system. Therefore, continuous monitoring of infants for possible signs of pain and discomfort is necessary.

Research on detection of acute pain in infants by analyzing the facial expressions is reported in [1] and [2]. However, these studies do not automatically find the face and do not exploit the properties of a video sequence but use photographs, on which the user has to manually indicate the region of interest. Another study focuses on discomfort detection for infants or small children [3]. Unfortunately, the method in this study is only tested for discomfort and not for acute pain. Furthermore, the face detection in this study is not always robust [3].

Many challenges arise when developing an automated pain detection system for infants. The systems available for adults ([4] [5]) cannot be used because these systems are trained on adults and the image frames only show the face without any other objects near the face such as hands, pacifiers or toys. The reported systems also need specific features like eyebrows or pupils to be present, which is not always the case for infants. Infants often do not have visible eyebrows and have their eyes closed, which makes the eyes and face harder to detect. This means that the system needs to be trained and designed specifically for infants in order to work properly.

This study contributes in several ways. We show that it is possible to create an automated pain and stress detection system for infants, adopting information from the PFP. Also, we are the first to translate a part of a clinical pain scale form, the Premature Infant Pain Profile (PIPP) [6], to a computer system and we select features that are able to extract necessary information from the face or from specific parts of the face.

In this paper, we present and discuss two novel systems for pain detection. The AAM-based system of earlier work [3] is extended to detect the PFP as a whole,

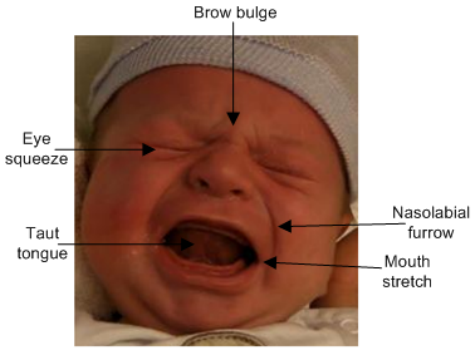


Figure 1: Primal Face of Pain.

while a second ROI-based system detects individual facial expressions according to the PIPP pain scale and combines those to obtain the intensity of the pain. Both systems automatically find the face and the position of the eyes and mouth, even with closed eyes, and analyze the facial expression with respect to pain and stress.

This paper contains the following sections. Section 2 explains and discusses the design of our systems. The performed experiments and the comparison of the two systems are shown in Section 3. Section 4 contains the overall conclusions and a discussion of possible future work.

2 Methodology

2.1 Face detection

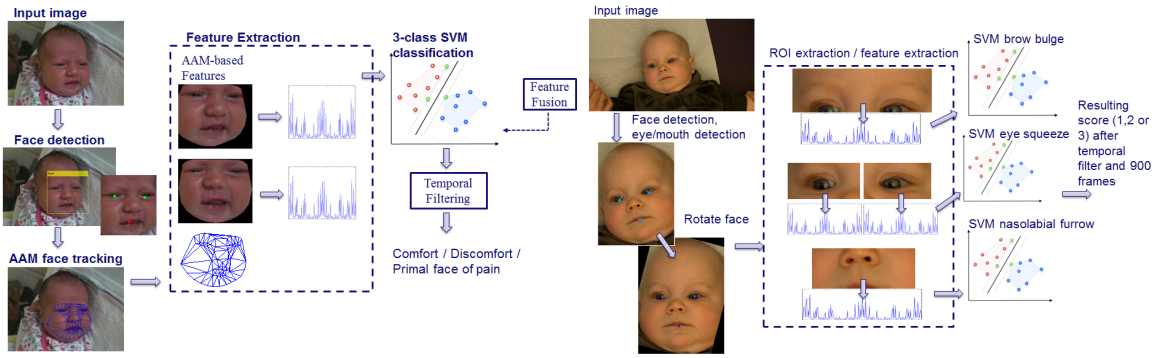
Both systems apply the same face detection algorithm, which is explained first. Face detection is an important part of an automatic pain recognition system. First, the position of the face is detected. After this, the eye and mouth positions are identified inside the face area. To find the eye and mouth positions, we use the method of Fotiadou *et al.* [3]. The face detection technique combines the Gaussian skin detector with a Viola-Jones algorithm. If a face is found with the Viola-Jones algorithm and its color does not coincide with the skin-color region, the face is discarded. This reduces false positives of the Viola-Jones algorithm.

2.2 AAM-based system

The AAM-based system aims to split facial expressions into three groups: normal, stress and the primal face of pain. The block scheme of this system is shown in Figure 2a. This system consists of three parts: the face detection/tracking, feature extraction and classification. We address these parts briefly.

2.2.1 Face detection/tracking

For the AAM-based system, the tracking of the face is achieved using an AAM (Active Appearance Model) [4]. During a training phase, a statistical model of the object shape and the appearance is created. For each new infant, this training phase has to be repeated and the landmark points have to be manually annotated for the different facial positions. This AAM face tracking method is also used for infant discomfort detection in [3] with the exception of the initialization, where the eye and mouth positions are found as described in the previous section.



(a) AAM-based system. (b) ROI-based system.

Figure 2: Proposed pain detection systems.

2.2.2 Feature extraction/classification

For feature extraction, we apply the same features as investigated and explained in [3]. The first feature is the similarity-normalized shape obtained from the AAM algorithm. The second feature consists of Elongated Local Binary Patterns (ELBP) with similarity-normalized appearance. The third feature also uses ELBP, but here the canonical-normalized appearance is used.

Using these features, a 3-class SVM classifier is generated with the following classes: comfort, discomfort and the Primal Face of Pain (PFP). If the input features are calculated, a one-versus-all SVM classifier (first step) is applied, in order to distinguish comfort from pain and stress. The images that do not belong to the comfort class are inserted to the next SVM classifier (second step) which distinguishes discomfort from the PFP. For both steps, all three features are used. Each feature is trained with an individual SVM classifier and the outputs of these classifiers are combined.

2.3 ROI-based system

In order to create the ROI-based system, we have analyzed several pain scale methods in the hospital. The PIPP scale form [6], used to find pain in neonates and small infants, is of high interest for our research because every parameter of this scale can be observed by the monitor, or is visible in the facial expressions. Our ROI-based system interprets several PIPP scale form parameters by means of video analysis. These parameters are the brow bulge, eye squeeze and the nasolabial furrow. This system is shown in Figure 2b.

2.3.1 Facial region extraction

To interpret the facial analysis part of the PIPP score, three different facial actions have to be recognized. Corresponding parts are called the Regions Of Interest (ROI). To extract the ROI, we start with the eye and mouth positions and the detected face. Firstly, the face has to be rotated using the positions of the eyes. Now we calculate the distance between the eyes D_{eyes} in the x direction and the distance between the eyes and mouth $D_{e/m}$ in the y direction. We employ this information to calculate the upper and lower borders of the brow bulge area using the relation

$$bb_{up} = P_{eye,y} - \frac{D_{e/m}}{\alpha}, \tag{1}$$

for the upper border, where the scaling parameter α is determined empirically and with $P_{eye,y}$ being the position of the eyes in the y direction. For the lower border of the brow bulge, we add the upper and lower borders of the eye area, devide them by

2 and substract $P_{eye,y}$ with the resulting value. For the left and right borders of the brow bulge area, we define the same borders as used for the eye area.

For the upper bound of the eye area, we specify that

$$eye_{up} = P_{eye,y} - \frac{D_{e/m}}{\beta} - \kappa - (\frac{P_{eye,y} - \frac{D_{e/m}}{\beta} - \kappa - bb_{up}}{\gamma}) \tag{2}$$

with $P_{eye,y}$ denoting the position of the eyes in the y direction, κ a number of pixels which have to be determined experimentally and β and γ scaling parameters which have to be evaluated empirically. For the lower bound of the eye, we define

$$eye_{down} = P_{eye,y} + \frac{D_{e/m}}{\beta}. \tag{3}$$

For the left border, we specify

$$eye_{left} = P_{eyeL,x} - \frac{D_{eyes}}{2}, \tag{4}$$

where $P_{eyeL,x}$ defines the positions of the left eye in the x direction. For the right border, we use the same equation but with a positive sign and the position of the right eye.

The eyes are split into the left and right eye using the middle point of the mouth. This point is calculated as follows. We first compute the area where the mouth is located based on the previously computed mouth point. For the upper or lower border, we subtract or add $D_{e/m}/\delta$ to the y position of the mouth. Parameter δ denotes a scaling value, which is determined empirically. For the left and right border, we subtract or add $D_{e/m}/\alpha$ to the x position of the mouth. From this region, the saturation is calculated and thresholded to find mouth points. The middle of the mouth is found by taking the mean of all mouth points.

The nasolabial furrow area uses the bottom border of the eyes as upper border and the middle of the mouth in y direction as lower border. The left and right border of the nasolabial furrow region NF_z is then calculated by

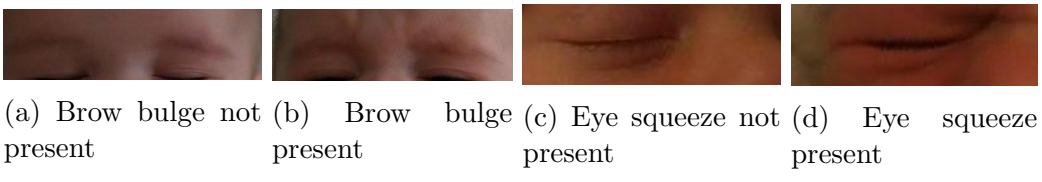
$$NF_z = eye_z\frac{3}{4} + mouth_x\frac{1}{4}, \tag{5}$$

with z referring to either left or right and $mouth_x$ representing the middle of the mouth in the x direction.

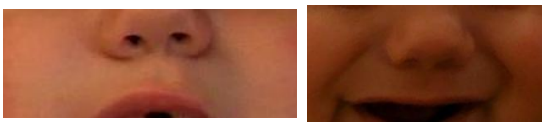
For each region, different features are evaluated (but not discussed here) to select the best performing feature and are used as an input for the SVM classification. In the sequel, we only elaborate on the best feature for each region.

2.3.2 Detecting the ROIs

First, we detect the presence of the brow bulge. In Figure 3b and 3a, an example is shown of the brow bulge region, both with and without the presence of the corresponding facial expression. When the brow bulge appears, the texture between the eyes changes. Therefore, we describe this ROI with HOG features. Prior to feature extraction, a 5×5 median filter is applied, so that the image is optimally pre-processed for further analysis.



(a) Brow bulge not present (b) Brow bulge present (c) Eye squeeze not present (d) Eye squeeze present



(e) Nasolabial furrow not present (f) Nasolabial furrow present

Figure 3: Examples of facial actions

The second ROI, the eye squeeze, is the most difficult facial expression to detect. An example of an eye with and without eye squeeze is shown in Figure 3d and 3c. In order to facilitate the detection, we first consider if the eye is open or closed using HOG features. If the eye is detected as a closed eye, HOG and ELBP features are extracted to determine the presence of the eye squeeze. We concentrate on HOG and ELBP because HOG gives texture information and the eyes have an elliptic shape which corresponds to the ELBP neighborhood. For both features, separate SVM classifiers are trained which are combined using a simple summation.

Lastly, we analyze the nasolabial furrow. When present, two lines between the nose and the corners of the mouth appear, as shown in Figure 3f. HOG features are applied to the total nasolabial furrow region.

2.3.3 Scoring of ROIs

In this part, we show how the different facial sections are combined in a score from 1 to 3 with 1 being minimal pain or stress and 3 being maximal pain or stress. We first analyze the separate regions for presence of a facial action for 30 seconds. After this, we score the separate sections from 0 to 3. If the facial action occurs less than 10% of the time, we score a 0, less than 40% yields a score of 1, less than 70% a score of 2 and if the facial action is present for 70% or more, the score is 3. We now have three scores and add them to obtain a score from 0 to 9. This is equal to the extraction of facial action scores in the PIPP pain scale form. Finally, the score is assigned to an interval corresponding with the strength of the facial expression. For example, the first strength level corresponds to a score between 0 and 2, the second level to a score between 3 and 6 and the third strength level to a score between 7 and 9.

3 Experimental results for both systems

3.1 Database

For testing, a database is created using videos recorded at the Maxima Medical Center (MMC) in Veldhoven. Videos are recorded of the faces of infants, experiencing pain from an invasive procedure (prick) or from post-operative pain. Videos are also captured of the same infants in a relaxed state. For additional information about the difference between stress and pain, infants who experience stress but no pain, are also recorded. The age of the 50 recorded infants ranges between 2 days and 17 months. Out of the 50 infants, 31 experience a painful stimulus. Our database consists of 177

videos with a frame rate of 30 frames per second and a spatial resolution of 1280×720 pixels. For training, we also apply the database obtained from and used by E. Fotiadou [3]. All tests are performed using Matlab on a 2.5-GHz dual-core processor.

3.2 AAM-based system performance

In this section, we discuss the performance of the AAM-based system. Face detection is discussed first, after this, the classification performance is shown.

3.2.1 Face detection

Here, face detection is examined based on the combination of Viola-Jones with skin color. A face is considered to be successfully detected if the AAM algorithm is able to track it correctly. The Viola-Jones algorithm consists of 10 stages with a false alarm rate of 0.3 and is trained using 352 images of 28 different infants. For detection performance testing, 20 videos from 8 different infants with a total of 17,132 frames are used. The combination with Viola Jones can correctly track 61.6% of the frames. This is higher than when using the skin color method only, where 40.1% of the frames would be tracked correctly.

3.2.2 Classification

The performance of the 3-class SVM classifier system is analyzed. For training, 24 videos of 16 infants with 12,552 correctly tracked frames are used, all showing one or more states in one video. Of these videos, 22 show comfort, 17 show stress and 12 of the movies have images with the PFP. Testing is performed on the same infants using leave-one-out cross validation. The accuracy of this system is 91.9%, the specificity is 94.3% and the sensitivity is 68.4%.

3.3 ROI-based system performance

In this section, we discuss the performance of the ROI-based system in a similar way as the AAM-based system. However, specific tests for infants with occlusions are performed and results are shown.

3.3.1 Face detection

Here we present the performance of the face detection using the ROI-based system. For the parameters mentioned in the equations from Section 2.3.1, we have determined the following optimal parameter settings: $\alpha = 1.2$, $\beta = 5$, $\gamma = 3$, $\kappa = 8$ and $\delta = 4$. In the system, a frame can only be used if the Viola-Jones algorithm is able to identify a face. A face is detected properly if all three regions shown in Section 2.3.1 are detected correctly. We use visual inspection to determine the percentage of correctly tracked frames. From the 400 tested frames, 77.5% are found correctly and 11% false positives occur.

3.3.2 Classification

The performances of the three facial region detectors are investigated for different features. For the experiments concerning the separate regions, 400 clearly different test samples are used. First, we evaluate the performance of detecting the brow bulge using the optimal brow bulge size, as shown in Section 2.2.2. The accuracy of the brow bulge detection is 73.3%. The eye squeeze is detected with an accuracy of 76.5%. The

Table 1: Performance of the ROI-based system

	Accuracy	Sensitivity	Specificity
Nurse and non-medical scores compared	91.7%	85.7%	100%
System using nurse scores as ground truth	56.0%	53.9%	83.3%
System using non-medical scores as ground truth	64.0%	63.6%	85.7%
Modified system using nurse scores as ground truth	62.5%	57.1%	90.0%
Modified system using non-medical scores as ground truth	**66.7%**	**66.7%**	**91.7%**

last facial region is the nasolabial furrow. This ROI performs best with an accuracy of 87.0%.

Table 1 presents the performance of the total ROI-based system as shown in Figure 2b, using 26 unaltered videos of 30 seconds from 9 different infants. For ground truth, we have used scores from a healthcare professional and from a non-medical, but algorithm expert who looked at the video frames separately.

The nasolabial furrow detection has the highest performance. Therefore, we give this region a two times higher weight. We call this approach the modified system. The modified system offers an accuracy of 66.7%. The brow bulge region analysis leads to the lowest performance. The infants and lighting conditions differ and this causes changes in the texture of the brow bulge. By manually splitting the infants into three groups, based on lighting conditions and appearance of the infant, the overall accuracy is improved to 73.9%.

3.3.3 Occlusions

In order to apply a pain detection system in a clinical setting, occlusions such as hands, pacifiers and breathing tubes have to be dealt with. In majority, we focus on infants with feeding tubes. Frames from 5 different infants are used with a total of 1,394 frames and the infants vary in gestational age from 34 to 37 weeks. The system performs best for infants with open eyes (81.8% correctly tracked) and with the feeding tube between the eyes (59.4% correctly tracked). If the eyes are closed, the system has a low tracking score (16.0% correctly tracked). Furthermore, the system performs best when the infant is seen in frontal view. Moreover, as expected, the lighting conditions have a clear impact on the image quality and thus on the detection performance.

3.4 Comparison of ROI- and AAM-based systems

In this section, we compare the performance of the two systems, using the combination of skin color and Viola-Jones for face detection. We first look at the percentage of correctly tracked frames. For the AAM system, this means that the AAM mask is created correctly. For the ROI system, it means that all regions are correctly detected. Both systems are tested using 1,100 frames from 9 different infants. From this set, the ROI system is able to find 77.5% of the frames while the AAM-based system finds 71.0% of the frames.

When comparing the performance of the pain/stress classification, we see that the accuracy of the AAM-based system is 25.3% higher than the accuracy of the ROI-based system but it cannot handle occlusions. We note that the ROI-based system currently includes only the facial expression analysis, which is only a fraction of the PIPP Parameters. This enables further improvement, which is addressed below.

4 Conclusions and discussion

In this paper, we have presented two innovative systems for stress and pain detection for infants, which are capable to automatically detect facial expressions belonging to pain. The proposed algorithms and the results bring us closer to the objective of continuous and fully automatic monitoring of facial expressions of infants for possible signs of pain.

The two investigated systems are an AAM-based system using the conventional AAM model with ELBP features, and a ROI-based system exploiting individual facial expression in a direct way. The ROI-based system is able to handle occlusions such as hands near the face, feeding tubes and pacifiers. The three strength levels of the facial expressions are detected with an accuracy of 66.7%. In contrast, the AAM-based system is not able to handle occlusions but detects the Primal Face of Pain with an accuracy of 91.7%. This makes the AAM-based system better for pain classification provided that occlusions are absent. While the AAM-based model has a high accuracy for pain detection, the system requires a tedious manual annotation of the AAM shape parameters for each head position and various lighting conditions. This infant-dependent AAM model is therefore less desirable for an application in a clinical setting. The ROI-based model is a better choice here because it is designed to be completely infant-independent.

Although its lower score, the ROI-based system can still be further improved when the classification of the separate regions is optimized. For example, it is possible to split the infant videos in smaller groups based on age or lighting conditions, or changing and optimizing the input features or classification algorithm. An alternative for improving the pain detection is to include other parameters from the PIPP pain scale form to the ROI-based system. We can also convert the whole PIPP scale form to an unobtrusive automated monitoring system, which continuously checks the facial expressions, heart rate and oxygen saturation for signs of pain.

References

[1] S. Brahnam, C.-F. Chuang, F. Y. Shih, and M. R. Slack, "Machine recognition and representation of neonatal facial displays of acute pain." *Artificial intelligence in medicine*, vol. 36, no. 3, pp. 211–22, Mar. 2006.

[2] B. Gholami, W. M. Haddad, and A. R. Tannenbaum, "Relevance Vector Machine Learning for Neonate Pain Intensity Assessment Using Digital Imaging," *Biomedical Engineering*, vol. 57, no. 6, pp. 1457–1466, 2010.

[3] E. Fotiadou, S. Zinger, W. E. Tjon a Ten, S. Bambang Oetomo, and P. H. N. de With, "Video-based facial discomfort analysis for infants," in *Proc. SPIE 9029, Visual Information Processing and Communication V, 90290F*, 2014.

[4] P. Lucey, J. F. Cohn, I. Matthews, S. Lucey, S. Sridharan, J. Howlett, and K. M. Prkachin, "Automatically Detecting Pain in Video Through Facial Action Units," *Systems, Man, and Cybernetics, Part B: Cybernetics, IEEE Transactions*, vol. 41, no. 3, pp. 664–674, 2011.

[5] S. Kaltwang, O. Rudovic, and M. Pantic, "Continuous Pain Intensity Estimation from Facial Expressions," *Advances in Visual Computing. Springer Berlin Heidelberg*, pp. 368–377, 2012.

[6] B. Stevens, C. Johnston, P. Petryshen, and A. Taddio, "Premature Infant Pain Profile: development and initial validation." *The Clinical journal of pain*, vol. 12, pp. 13–22, 1996.

Analysis of an Arbitrated Quantum Authentication Scheme

Helena Bruyninckx Dirk Van Heule
Royal Military Academy, Brussels, Belgium
Department MWMW
Avenue de la Renaissance, 30, B-1000, Brussels
helena.bruyninckx@rma.ac.be dirk.van.heule@rma.ac.be

Abstract

In this paper, we review an unconditionally secure quantum authentication scheme for authenticating a classical message in the presence of an online, semi-trusted arbiter. This scheme is classified as a quantum protocol among two distrustful parties. The scheme respects the non-repudiation property (by both sender and receiver), is secure against a malicious receiver and offers dispute resolution. We present an analysis of the scheme and propose an amelioration for it by emphasizing on the issue of reusing the shared keys.

1 Introduction

Authentication of information is of vital concern in several applications of information exchange. The receiver of a message must be able to verify that it was deliberately created by the sender and that it has not been substituted or altered during transmission. Authentication schemes offer a solution to this problem, but when they are based on symmetric key principles, and the transmitter and receiver do not trust each other, we need an authentication scheme with arbitration. In such schemes, a trusted third party called an arbiter can help resolve disputes between the sender and the receiver. Moreover, the scheme must provide a solution in case the third party wants to deceive either the transmitter or receiver.

Unconditionally secure authentication schemes with an arbiter have been proposed in the literature (e.g., [1, 2]) and this idea was extended in the context of quantum authentication [3]. Compared to other arbitrated quantum schemes, the scheme respects the non-repudiation property (by both sender and receiver) and is also secure against a malicious receiver. Moreover, the arbitrator is not an inline-party (in contrast to other arbitrated quantum schemes).

Unconditionally secure authentication schemes are secure against any adversary, on the condition that the participants keep their shared keys completely secret. Such schemes (of classical and quantum nature) normally have a high demand for fresh secret key material, but in this scheme, this requirement is not needed.

Objectives of the paper. In this paper, we describe the set-up phase from the scheme from [3], and analyze the issue of reusing the shared keys. Details on appropriate hash functions will be given. We use different hash functions for each pair of participants, together with quantum encryption. We compare the scheme with the unconditionally secure authentication schemes from Wegman-Carter [4] and from Brassard [5], both intended for classical authentication without the help of an arbitrator.

Paper outline. The paper is organized as follows. In section 2, we give a brief description of the model of authentication with arbitration, and an informal description of the quantum authentication scheme with arbitration from [3]. We focuss on the set-up phase from [3] and present its analysis in section 3. Finally, we draw our conclusions in section 4.

2 Review of the quantum authentication scheme

2.1 Authentication with arbitration

We introduce the authentication model by given a brief description. The quantum authentication scheme (see Section 2.2) is based upon the authentication model with arbitration [2, 6], and translated to the context of quantum cryptography. In this authentication model, not only attacks from an outsider (opponent) are considered, but also attacks from the insiders (transmitter, receiver and arbiter). Figure 1 represents the main components of this model.

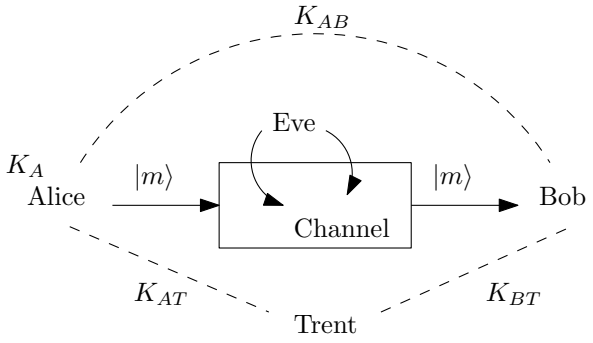


Figure 1: General model of authentication with arbitration

Participants. In this model, we consider four different participants in the scheme: Alice the sender (A), Bob the receiver (B) , Trent the arbiter (T) , and Eve the outside opponent (E) . The different parties taking part in the authentication protocol (Alice, Bob and Trent) are guaranteed to follow the protocol honestly, but are interested in breaking (some of) the requirements of the scheme. This means that they will try to learn as much as possible during the execution of the protocol and that they will use that information later on. Such participants are called honest-but-curious participants. It is obvious that we do not allow the different participants to collude with each other.

Keys. Three shared keys are used in this model: the arbitrator will share two secret keys K_{AT} and K_{BT} with Alice and Bob, respectively. Alice and Bob also share a secret key K_{AB}, unknown to the arbiter.

The arbiter. Unconditionally secure authentication schemes that permit arbitration involve an arbiter that possesses some privileged information, i.e., information not available to one or more of the other participants. He has however no access to the information shared between Alice and Bob.

The arbiter in this scheme is a trusted party as far as arbitration between the parties is concerned. However, he may have an interest in breaking some of the other requirements of the scheme, e.g., impersonating a party. He is thus semi-trusted; this means that a great trust is placed in the arbiter (e.g., concerning his arbitration), but that his cheating can be proven (i.e., a verifiable third party [6]).

The arbiter is needed to settle a dispute, but the receiver should be able to verify the validity of the messages on his own. This requirement means that the arbiter is not an inline-party, but an on-line party.

Requirements. The requirements for our scheme are the same as for other authentication schemes with arbitration, more precisely: Verifiability, Unforgeability, Non-repudiation and Resolution for disputes. The requirement for non-repudiation poses the biggest problem, because we want this requirement to be respected by both the sender and the receiver. This means that it should be impossible for Alice to disavow an authenticated message created by herself. But secondly, we do not allow Bob to falsely claim that the received message was not authentic. In case such disputes occur among the participants, the arbiter can resolve the dispute [3].

Attacks. There are different kind of attacks which are possible in this model and we refer to [2] for a detailed discussion. We can classify these attacks with respect to the attacker: security against an external opponent (Eve) and security against a dishonest participant (the arbiter, Alice or Bob). The attacks can be related to the quantum nature of the scheme (thus performing quantum attacks), or to the classical parts of the scheme. An informal security analysis of the scheme from Section 2.2 can be found in [3].

2.2 Informal description of the scheme

In this section, we give a short description of the arbitrated quantum authentication scheme from [3]. The problem of authenticating a classical message in the presence of a semi-trusted arbiter is based upon a non-repudiation technique due to Asokan et al. [6]. Following this technique, the authentication consists of two parts and they must be used together in order to create the complete authentication for a certain message.

The scheme consists of four stages: (1) The initialization phase in which Alice, Bob and Trent obtain the necessary shared keys and agree on two hash functions. Alice will also construct her hash chain; (2) The setup phase in which a non-repudiation token for Alice is created; (3) The authentication phase in which Bob receives a message and wants to verify that the message is authentic; (4) The dispute phase in which Trent is requested to resolve a dispute between Alice and Bob.

The different steps in the setup phase and authentication phase are as follows. Firstly, Alice sends the message and other information to be authenticated together with a one-time ticket to the arbiter. The arbiter verifies the received information and computes a non-repudiation token over the received data. This token constitutes the first part of the authentication. Alice receives the token from the arbiter and produces her one-time signature (second part of the authentication). Finally, Bob receives the information from Alice and verifies the correctness of the non-repudiation token. Then, he verifies the second part of the authentication, i.e., Alice's one-time signature.

3 Analysis of the scheme

During the initialization phase, all participants agree on two collision resistant one-way hash functions. Those functions will remain unchanged during the complete execution of the scheme. In [3], no details were given on the appropriate hash functions. Secondly, quantum encryption is used whenever a message is sent from one participant to another in order to hide the quantum information securely. Encoding a message into a quantum string was done by using the shared keys K_{AT}, K_{BT} and K_{AB}.

In this section, we discuss the hash functions together with the quantum encryption and propose a method for reusing the shared keys.

3.1 Universal hashing

Authentication is usually performed by sending a message together with a tag t. Verification is done by computing the tag of the received message, and comparing it with the received tag.

We denote $\mathcal{M}$ the set of classical messages, $\mathcal{T}$ the set of tags and $\mathcal{K}$ the set of shared keys. Let $\mathcal{H} = \{h_k : \mathcal{M} \to \mathcal{T}\}_{k \in \mathcal{K}}$ be a family of hash functions. From this family, a hash function is chosen uniformly at random by means of the shared key $k \in \mathcal{K}$.

The authentication scheme commonly used in Quantum Key Distribution (QKD) is the Wegman-Carter authentication scheme [4], because it offers information-theoretic security. The hash function is selected by the shared key from a family of $\epsilon-\mathrm{ASU}_2$ hash functions [8] and for each message, a new hash function will be chosen from this family. The demand of fresh key material is thus very high.

When several messages must be authenticated, the same hash function can be used each time, but the tag should be encrypted with a one-time pad (OTP). Alice thus appends $h_{k_1}(m) \oplus r$ to her message m, where k_1 is used for all messages and r is a one-time pad. This second scheme uses less key material and is called an authentication scheme with key recycling. In this case, we use a weaker family of hash functions, more precisely an $\epsilon-$almost XOR universal$_2$ ($\epsilon-\mathrm{AXU}_2$) family of hash functions [9].

Even though the second authentication scheme consumes less key material, we still need a OTP in order to provide information-theoretic security. This requirement is impractical due to the need of synchronization between Alice and Bob. Therefore, Brassard [5] proposed a modification to the Wegman-Carter scheme by replacing the OTP with a pseudo random number generator (PRNG). The seed for this PRNG consists of a shared secret key between Alice and Bob and Brassard showed that this authentication scheme offers computationally secure authentication. Remark that Brassard's scheme uses two secret keys: one key will be used as the seed for the PRNG and the other key will select a hash function from an $\epsilon-\mathrm{ASU}_2$ family of hash functions. This hash function remains unchanged during the execution of the protocol.

3.2 Discussion

In this section, we first discuss the choice of the two hash functions as explained in [3] and then, provide more detailed information on a more suitable choice of the functions.

On the hash chain. During the initialization phase [3], the participants agree on a one-way, collision resistant hash function: $h(\cdot) : \{0,1\}^* \to \{0,1\}^{n_1}$ which will be used by Alice for generating a hash chain (see [3] and [6] for a detailed discussion). The notation $h_A^i(\cdot)$ can be seen as a keyed hash where the known key is the concatenation of Alice's identity (ID_A or A) together with the counter, i ($i = n-1, n-2, \ldots, 1$).

The signer Alice will select a random secret key K_A and construct her hash chain, starting with her secret key as seed: $K_A^0, K_A^1, K_A^2, \ldots, K_A^t$, where

$$K_A^0 = K_A,\ K_A^i = h_A^i(K_A) = h_A\left(K_A^{i-1}\right)$$

and t denotes the number of messages Alice wants to authenticate.

The hash function $h(\cdot)$ must be selected at random, but needs to be known to all participants. The security of the scheme is based on the one-wayness of the chosen hash function and its collision-resistance, more precisely its second-preimage resistance.

On the second hash function. During the initialization phase, the participants also agree on a second one-way, collision resistant hash function: $H(\cdot) : \{0,1\}^* \rightarrow \{0,1\}^{n_2}$. Besides transforming the information into a string of fixed length, the generated hash indicates if the information sent from one participant to another was received correctly.

We will now elaborate on this second hash function $H(\cdot)$ and explain that it does not have to be the same function for Alice, Bob and Trent. They can mutually agree on different hash functions, only known to two participants each time.

The authentication schemes from Section 3.1 are based on symmetric key principles, and therefore, they repose on the mutual trust between the participants. Since we consider an authentication scheme for mutually distrusting and deceitful parties, we will adapt the described authentication schemes and include the non-repudiation property.

In the key recycling authentication scheme from Wegman-Carter, the shared key does not to be renewed for each message, but we still need to use a OTP for each message (which can be seen as a new key for each message). Therefore, we will use the authentication scheme from Brassard as starting point.

Assume Alice wants to send a message m to the receiver Bob. First, she needs to contact Trent with the following information: the data to be authenticated, i.e., $m_p = (ID_A, ID_B, m, i)$, and a one-time ticket, K_A^i. This one-time ticket is defined by the hash function and therefore only known to Alice. m_p stands for the public classical message, which is known to all participants. The different steps in the set-up phase are as follows.

1. Alice computes the information she wants to send to the arbiter:

$$m_T = K_A^i || H_{k_1}\left(ID_A, ID_B, m, i, K_A^i\right),$$

   where $H_{k_1}(\cdot)$ is taken from a strongly universal family of hash functions. The key k_1 is a shared fixed key between Alice and Trent.

2. Instead of using the XOR operation from Brassard's authentication scheme, we encode the classical message m_T into a quantum string by using the quantum encryption for classical messages (see Table 1). Alice and Trent need to agree on two orthonormal bases and a shared secret key will denote which basis is used during the encryption. The shared key K, is obtained from the outcome of a PRNG, where the seed was specified by a shared fixed key k_2. K^i and m_T^i denote the $i-$th bit of the key K and the message m_T respectively. For each bit of the message, m_T^i, the quantum state is determined by its value and the value of the bit K^i. Note that for each bit of the message, another bit from the output sequence of the PRNG is used and that the key K is completely independent of the message m_T. The key K used for the quantum encryption will thus be renewed for every message. The length of K must be the same as the length of the message m_T, more precisely $n_1 + n_2$.
   This means that the key K_{AT}, shared between Alice and Trent, consists of the concatenation of k_1 and k_2, i.e., $K_{AT} = k_1 || k_2$.

	$m_T^i = 0$	$m_T^i = 1$
$K^i = 0$	$\|0\rangle$	$\|1\rangle$
$K^i = 1$	$\|+\rangle = \frac{1}{2}(\|0\rangle + \|1\rangle)$	$\|-\rangle = \frac{1}{2}(\|0\rangle - \|1\rangle)$

Table 1: Quantum encryption of classical messages

After applying quantum encryption on the message m_T, Alice sends the quantum message $|m_T\rangle = E_K(m_T)$, together with $m_p = (ID_A, ID_B, m, i)$ to the arbiter. The classical information is sent through an unauthenticated public channel.

After reception of the quantum and classical information, the arbiter measures the received qubits $|m_T\rangle$ with the bases determined by the outcome of his PRNG and he obtains $m_T^{'}$. The arbiter is interested in the one-time ticket K_A^i and therefore, he computes $H_{k_1}(ID_A, ID_B, m, i, K_A^i)$ and verifies that this corresponds to the received information from Alice.

We now analyze the security of this authentication method. For every message exchanged between Alice and Trent, the hash function $H_{k_1}(\cdot)$ will remain the same, but the encryption key K will change, according to the sequence of bits generated by the PRNG. The PRNG used by Alice and Trent is a deterministic (or cryptographic) PRNG that produces a sequence of bits which distribution is indistinguishable from the uniform distribution. The security of a PRNG is defined as the hardness to tell the difference between its pseudo-random sequence output and truly random sequences (i.e., *distinguishing attacks*). A subclass of distinguishing attacks consists of *state recovery attacks*. A state recovery attack on a pseudo-random generator is an algorithm that, given a pseudo-random sequence, recovers the seed [11]. In order to perform such attacks, the adversary needs to know a subsequence of generated bits from the PRNG (Direct Cryptanalytic Attack), or he needs to use some knowledge on the PRNG input (Replayed-input attacks or Known-input attacks). Sometimes, the attacker can even choose or manipulate the input to the PRNG (Chosen input attacks). Moreover, quantum computing can be used for implementing efficient algorithms to certain problems where an efficient classical algorithm is not known [12].

To attack a PRNG in quantum or probabilistic polynomial time, the adversary needs to know some information on the sequence of output bits produced by the generator under attack. In our scheme, the output of the PRNG is hidden by the quantum encryption. An attacker can perform some attack strategies, e.g., the intercept and resend attack. The idea is to capture all or a proportion of the states sent by Alice to Trent, and then prepare new quantum states (based on the measurement outcomes). The attacker may learn something about the bases she chose by observing Trent's reaction.

The problem for Eve is that she can only access the public message m_p and the quantum encoded version of the message m_T. By looking at the outcome from her measurements on the quantum states, she cannot obtain enough information. The probability that her interfering will not be detected is 75% on each qubit. But if Trent accepts the message from Alice (intercepted by Eve), Eve doesn't know if she selected the correct basis or not. In 25% of the cases, Eve obtained the wrong classical message outcome, without knowing. In order to verify this, she can try to compute $H_{k_1}(ID_A, ID_B, m, i, K_A^i)$ and compare her outcome with the corresponding second part of her intercepted message m_T'. However, Eve doesn't know the key k_1, needed to select the correct hash function, and she doesn't know the correct value of K_A^i. Therefore, she can not verify her intercepted message and be certain of the outcome of the PRNG. Attacking the PRNG in order to retrieve the secret key k_2 is thus impossible.

If Eve wants to attack the authentication method of this round, she can use the same strategies as for attacking the authentication method of Brassard, which was proven to be computationally secure. Including quantum encryption of the classical message will enhance the security of the protocol.

Once the arbiter accepted the received message from Alice, he checks that $h_A^{t-i}(K_A^i) = K_A^t$ in order to verify if K_A^i was indeed correctly created by Alice. If a message from Alice containing K_A^i has not been received by him, he considers K_A^i as consumed and

will update i by $i-1$.

As the last step in the set-up phase, Trent sends back a message that links m to K_A^i such that it can only be verified by Bob, R_T. This message is in fact the non-repudiation token and constitutes the first part of the authentication. Trent creates a message $R_T||H_{k_1}(ID_A, ID_B, m, i, K_A^i, R_T)$ by using the same hash function $H_{k_1}(\cdot)$. This message is encoded by quantum encryption and sent to Alice. The key used for this second encryption is a completely different key than the one for the first encryption.

Alice verifies the received information and retains R_T as the first part of her non-repudiation token to Bob. This concludes the set-up phase.

For the rest of the protocol, we apply the same reasoning leading to a shared fixed hash function $H_2(\cdot)$ between Alice and Bob. This function will be used for every message sent between Alice and Bob. The encryption key, on the other hand, will change for every message.

Since we use different hash functions for each pair of participants, and we encode the classical information by quantum encryption, the scheme offers more security than the scheme from [3]. The shared keys can be reused, thanks to the construction of Brassard.

4 Conclusion

In this paper, we reviewed an arbitrated quantum authentication scheme. The scheme offers authentication of classical messages and does not use entangled states. Just like classical authentication schemes with arbitration, the scheme satisfies the property of non-repudiation (by both the sender and receiver). Details on the scheme were given with an emphasize on the hash functions and the issue of reusing the shared keys. We compared the scheme with the unconditionally secure authentication schemes from Wegman-Carter and from Brassard. Since we use different hash functions for each pair of participants, and we encode the classical information by quantum encryption, the scheme offers more security than the analyzed scheme.

References

[1] T. Johansson, "Lower bounds on the probability of deception in authentication with arbitration," IEEE Transactions on Information Theory, vol. 40, no. 5, pp. 1573–1585, 1994.

[2] G. Simmons, "A cartesian product construction for unconditionally secure authentication codes that permit arbitration," Journal of Cryptology, vol. 2, no. 2, May 1990.

[3] H. Bruyninckx, D. Van Heule, "Arbitrated Secure Authentication realized by using quantum principles," IEEE ICC 2015, London, UK, June 2015.

[4] M. N. Wegman and J. Carter, "New hash functions and their use in authentication and set equality," Journal of Computer and System Sciences, vol. 22, no. 3, pp. 265–279, 1981.

[5] G. Brassard, "On computationally secure authentication tags requiring short secret shared keys," in Advances in Cryptology, Springer US, pp. 79–86, 1983.

[6] N. Asokan, G. Tsudik, and M. Waidner, "Server-supported signatures," Journal of Computer Security, vol. 5, no. 1, pp. 91–108, Jan 1997.

[7] L. Lamport, "Password Authentication with Insecure Communication," Communications of the ACM, vol. 24, no. 11, pp. 770–772, Nov 1981.

[8] D. R. Stinson, "Universal hashing and authentication codes," Designs, Codes and Cryptography, vol. 4, no. 4, pp. 369–380, Oct. 1994.

[9] P. Rogaway, "Bucket hashing and its application to fast message authentication," Journal of Cryptology, vol. 12, no. 2, pp. 91–115, 1999.

[10] H. Krawczyk, "LFSR-based hashing and authentication," Advances in Cryptology - CRYPTO '94, ser. LNCS, vol. 839, pp. 129–139., 1994.

[11] A. Sidorenko and B. Shoenmakers, "State Recovery Attacks on Pseudorandom Generators," WEWoRC 2005, LNI, vol. P-74, pp. 53–63, Leuven, Belgium, July 2005.

[12] E. Guedes, F. de Assis, B. Lula Jr., "Quantum Permanent Compromise Attack to Blum-Micali Pseudorandom Generator," IEEE International Telecommunications Symposium 2010 (ITS 2010), Manaus, Amazonas, Brazil, Sept 2010

Towards a home video monitoring system for patients with Parkinson's disease

B. Abramiuc, S. Zinger, P.H.N. de With

Eindhoven University of Technology

Faculty of Electrical Engineering, Video Coding and Architectures group
P.O. Box 513, 5600 MB Eindhoven, the Netherlands

(b.abramiuc, s.zinger, p.h.n.de.with)@tue.nl

N. de Vries-Farrouh, M.M. van Gilst, B. Bloem, S. Overeem

Donders Institute for Brain Cognition and Behavior
Department of Neurology

Nijmegen, The Netherlands

(nienke.devries-farrouh, merel.vangilst, bas.bloem, sebastiaan.overeem)@radboudumc.nl

Abstract

Parkinson's disease (PD) is a chronic disorder that is characterized by severe joint cognitive-motor impairments, which are difficult to evaluate on a frequent basis. Home monitoring of PD extends and enhances the diagnosis process and can lead to better treatment adjustment. In this paper we propose a video monitoring system that measures the quantity and quality of two clinically relevant motor symptoms. Our system separates the patients silhouette from the background exploiting the HSV color-space properties. Further, the silhouette is split into anatomical regions, which enables detection of the upper and lower limb extremities. Finally, step length and arm-swing angles are measured to assess Parkinson's disease severity. The experiments show that the system is able to accurately measure the parameters (found tolerance 2-5%) related to PD severity. Additionally, the results suggest that the system can be improved by analyzing the dynamic behavior/patterns of the key parameters.

1 Introduction

Parkinson's disease is a progressive neurodegenerative disorder that has many motor, as well as non-motor consequences. Amongst the various symptoms, slow movement, postural instability, tremor and freezing of gait are characteristic to PD [1]. As the disease progresses, it usually leads to deterioration of the physical functioning. In advanced disease states, patients with PD are rather dependent on daily home assistance. Moreover, non-motor symptoms, including sleep disorders like disorder of REM sleep behavior, cognitive problems and depression, are also frequently reported [2]. Because of our ageing society, the absolute amount of PD patients is increasing. In addition, the healthcare costs are 6-7 times higher for PD patients than for matched non-PD individuals [3] and spending increases with disease severity. These factors have a major social and financial impact on society.

Treatment with medicine can improve the motor symptoms, but after a few years most patients experience a wear-off effect [4]. Monitoring clinical performance of patients on the long term and finding a balanced treatment is difficult for various reasons. One of the reasons is the inability to evaluate symptom severity on a frequent

basis. In the current clinical practice, the patient visits the neurologist once or at most a few times per year. During these visits, medical treatment is adjusted based on patients experience and clinical assessment. This method does neither take into account day-to-day changes in symptoms severity, nor is it able to register changes in different moments of the day. Proposed solutions based on wearable accelerometers or pressure-sensitive mats require proper placement on the body or floor [6]-[8]. Additionally, maintenance is needed and this can become a burden for the patient. A more elegant solution is an unobtrusive video-based system. There are several visual monitoring systems proposed for home surveillance with clinical purpose, but they either monitor only one component of the human gait [9], or specially colored costumes have to be worn [10].

In order to solve the current challenges in PD motor symptom assessment, we are designing an unobtrusive video-based system for PD movement analysis. The system registers a set of body motion parameters that are clinically relevant and the movement behavior of body parts is then automatically analyzed according to the current clinical practice. Our approach aims at considering the quantity and quality of motor symptoms for later detailed evaluation, which improves the sensitivity for changes from day-to-day and for changes during different moments of the day.

In this paper, we present an innovative algorithm that is able to provide robust measurements of motor symptoms dynamics for PD patients. Section 2 of this paper explains the details of the processing stages applied to the video sequences and Section 3 presents our initial experimental results. The conclusions are presented in Section 4.

2 System for PD patient monitoring

In order to quantify the degree of impairment caused by the PD symptoms, we develop a system that detects and registers clinically relevant movement parameters. Due to their biomechanical nature, accurate location of the limbs is necessary. The parameters we extract are step length and arm-swing angle. We choose these two as a starting step, because they are associated with disease severity and are also commonly evaluated in current clinical practice [11]-[13]. Therefore, measuring these parameters properly is crucial for both aspects.

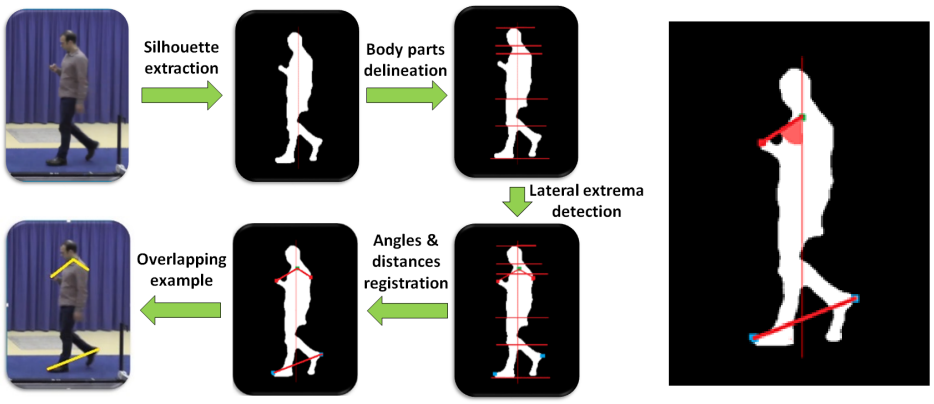


Figure 1: Diagram of processing steps of the monitoring system, on the left. The input image to the system is a patient video (first image) which goes through the processing steps. The output consists of the input video combined with graphic elements indicating the measurements (last image). Example of the measured parameters (arm-swing angle and step length), on the right.

Let us now briefly discuss the video monitoring system in more detail. Our video analysis algorithm consists of several steps (Figure 1, left). In the first step, we employ a *silhouette extraction* technique based on background subtraction. This separates the human silhouette from the background in the raw images. The background is obtained by averaging video frames in the beginning of the video, where the human is not present, and then subtraction is performed in the HSV color space. This color choice is motivated by the challenges that arise in the home-monitoring scenario. These challenges include similar colors in the foreground and background, uneven illumination and shadows. For better foreground segmentation, our technique exploits the ability to independently differentiate between color, saturation and brightness of the foreground object and background, which are illuminated by the same light source. These are favorable properties which are specific to the HSV color space [14]. The differences in each channel are normalized and then summed up to give a grayscale image. Then, by applying a threshold on the newly formed image, a binary mask is obtained. Noise and other possible movements of object that are part of the background also appear in the mask. Therefore, using the *a-priori* knowledge that video monitoring is exploited in a home environment and that there is only one human in the frame, the largest connected component is selected and preserved. Further, the algorithm verifies the height and orientation of the selected component. If the height is above a predefined value and the orientation is within the interval between -60 to +60 degrees with respect to the vertical axis, then the blob is considered to represent a standing human.

The following step is *body-part delineation*, where the silhouette is split into regions of corresponding human anatomy. To achieve this, we apply proportions that are derived from the height of the silhouette. The proportions are defined according to the study of Plagenhoef *et al.* [15], who have determined relationships between the height of a human and the average length of the other body parts. By applying these findings, the silhouette is separated into head, shoulders, trunk, thighs and feet regions.

Once the anatomical regions are known, in the next stage several points of interest are determined. These points are the center of the shoulder region and the *lateral extremities* of the trunk and feet regions. The lateral extremities represent the hands and feet during the walking motion. The center of the shoulder region is later used as a reference point to draw the angle between the vertical axis and the line that connects the hand and shoulder. This represents the arm-swing angle. The Euclidian distance between the two extremities of the feet is expressed in pixels and represents a measure of the step length. The examples of these measurements are presented in Figure 1, on the right.

During the *angles and distances registration* step, these parameter values are registered in a list from which plots can be drawn for later clinical evaluation. Additionally, we visualize the detected points of interest and the measured parameters on the input video. This helps the clinician to have an intuitive view over the measured parameters and their temporal dynamics.

3 Experimental results

Our initial experiments are performed in an environment with various background colors, where we record videos of a healthy volunteer. The parameters are evaluated in two types of scenarios and they are registered only if certain conditions are satisfied. The obtained data is then compared with the ground truth, by performing a Mean Absolute Error (MAE) analysis.

The video sequences are captured with an HD Panasonic HC-V720 camera at 50 fps. The room contains background objects that are similar in color to the volunteers clothes and the volunteer is asked to act and walk naturally in two different scenarios. The first scenario assumes the individual is walking into the scene (Figure 2). In this

case, the registration starts only when the human silhouette is not connected to the edge of the image and when it is above a predefined height. The height threshold prevents registration of wrong measurements, as the body proportions change when for e.g. the individual is bent. In the second scenario, called timed up and go test, the individual is sitting on a chair in the scene and then he stands up, walks a certain distance within the scene, comes back and sits in the chair again (Figure 3). This is a standard clinical test for PD patients in the gait laboratories. The action is usually timed by the clinician between the moments that the patient stands up at the beginning, and sits down again at the end. To reproduce this, our algorithm starts timing and registering the parameters only when the silhouette height is above the predefined threshold, while it is not connected to the edges of the image. The registered parameters are saved and then used to obtain diagrams (Figure 4), from which gait patterns can be later extracted.

Figure 2: System output for three different moments during monitoring. Example for walking in the scene scenario with ground truth annotations of: step length (middle) and arm-swing angle (left).

The camera is positioned such that the walking actions will be captured from a lateral view, enabling correct detection and registration of the parameters. The field of view of the camera covers, in this case, an approximate distance of 7 meters. For a healthy individual, this results in an average of 6 steps, 4 arm swings and the time per one direction walk is approximately 9 seconds.

To verify the accuracy of our system, the angles and distances are manually annotated on the video sequence by a human observer at key moments of the human gait (Figure 5). This is done with the help of Kinovea software, which is a component from an open source library, commonly used for biomechanical analysis of human body movement [16]. The annotated values are considered ground truth and further compared with the values from our system, obtained for the same frames. We make use of the MAE, which gives an indication of how close the measurements are to the ground truth.

In this set-up, a total of 23 valid step lengths and 10 arm-swing angles have been registered, for both scenarios. The remaining measurements, during stopping or turning, are ignored because, for our investigation, these phases of human gait do not possess any clinical value. The MAE comparison results in mean step-lengths of 241.9 and 240.3 pixels with an MAE of 8.6 and 5.3 pixels, respectively. The arm-swing angles comparison yields a mean of 39.5 and 33.8 degrees with an MAE of 1.8 degrees, for

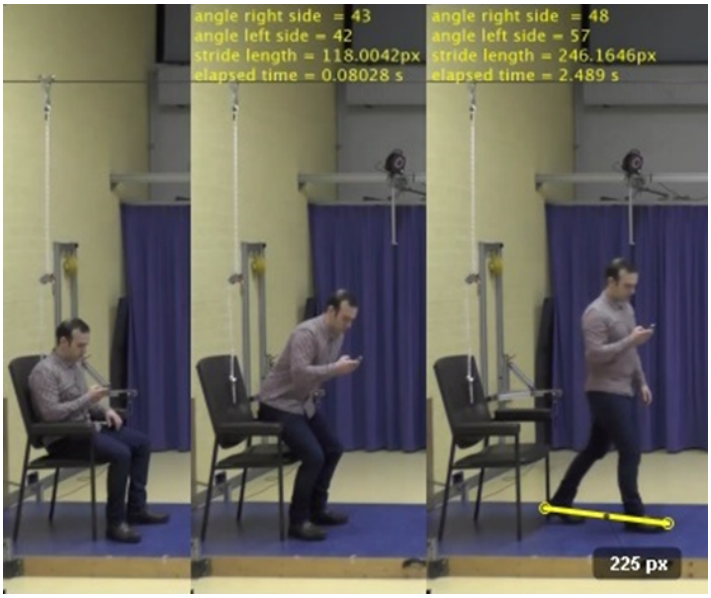


Figure 3: System output for three different moments during monitoring. Example for timed up and go scenario with ground truth annotation of: step length (right).

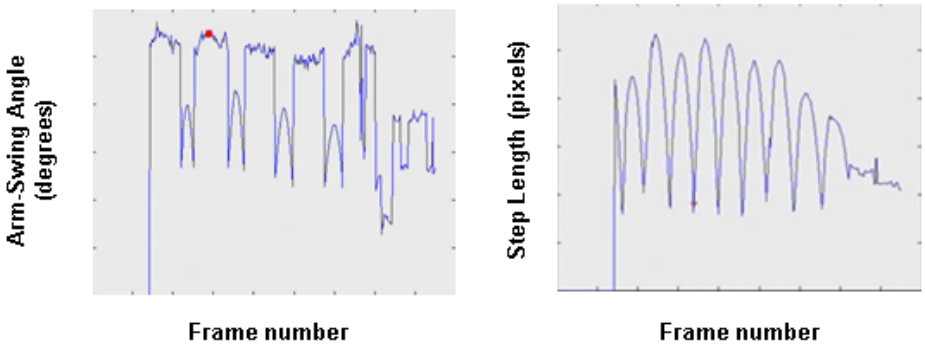


Figure 4: Diagrams of arm-swing angles in degrees (left) and step length in pixels (right).

both scenarios.

4 Conclusions

In this paper, we have studied a home-monitoring system for PD patients in order to quantify the degree of impairments caused by PD symptoms. The system is based on an algorithm that first separates the human silhouette from the background, then splits the body into anatomical parts and finally measures and registers clinically relevant parameters. Our monitoring system requires relatively static background environment, for deriving a mean background image. We have implemented a novel human body segmentation method that approximates anatomical regions, which are further used to detect parts of the upper and lower limbs.

The results show good compliance with the ground-truth measurements and they suggest that the system can objectively measure symptoms in patients with PD in the home environment. However, several technical challenges arise regarding the

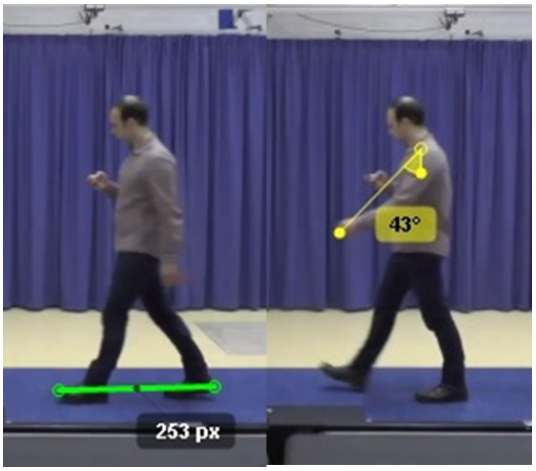


Figure 5: Examples of ground-truth annotations: step length (left) and arm-swing angle (right).

possible motion interruptions and irregularities. Example challenges are an interrupted walk, arm swing due to performing a task in daily routine, walking irregularities or unequal arm swing. In these situations, analyzing the patterns associated with the registered parameters will enhance the system. This improvement will allow the reliable detection and analysis of more complex PD symptoms like arm-swing asymmetries and freezing of gait, which is left for future work.

This work is only an initial study with positive implications. Our final objective is to create a monitoring system that can help clinical evaluation of PD symptoms by bringing a diagnostic tool into the home environment. This will allow to further refine and balance the decision-making process in the diagnosis and treatment of PD.

References

[1] A. J. Lees, J. Hardy, and T. Revesz, Parkinsons disease., Lancet, vol. 373, no. 9680, pp. 205566, 2009.

[2] K. R. Chaudhuri and A. H. V. Schapira, Non-motor symptoms of Parkinsons disease: dopaminergic pathophysiology and treatment., Lancet Neurol., vol. 8, no. 5, pp. 46474, 2009.

[3] S. J. Johnson, A. Kaltenboeck, M. Diener, H. G. Birnbaum, E. Grubb, J. Castelli-Haley, and A. D. Siderowf, Costs of Parkinsons disease in a privately insured population, Pharmacoeconomics, vol. 31, no. 9, pp. 799806, 2013.

[4] W. H. Poewe, Clinical aspects of motor fluctuations in Parkinsons disease., Neurology, vol. 44, no. 7 Suppl 6, pp. S69, 1994.

[5] M. M. van Gilst, B. R. Bloem, and S. Overeem, Sleep benefit in Parkinsons disease: a systematic review., Parkinsonism Relat. Disord., vol. 19, no. 7, pp. 6549, 2013.

[6] A. Salarian, H. Russmann, F. J. G. Vingerhoets, P. R. Burkhard, and K. Aminian, Ambulatory Monitoring of Physical Activities in Patients With Parkinsons Disease, IEEE Trans. Biomed. Eng., vol. 54, no. 12, pp. 22962299, 2007.

[7] N. L. W. Keijsers, M. W. I. M. Horstink, and S. C. A. M. Gielen, Ambulatory motor assessment in Parkinsons disease., Mov. Disord., vol. 21, no. 1, pp. 3444, 2006.

[8] B. Bilney, M. Morris, and K. Webster, Concurrent related validity of the GAITRite walkway system for quantification of the spatial and temporal parameters of gait, Gait Posture, vol. 17, no. 1, pp. 6874, 2003.

[9] E. Stone and M. Skubic, Passive, In-Home Gait Measurement Using an Inexpensive Depth Camera: Initial Results, Proc. 6th Int. Conf. Pervasive Comput. Technol. Healthc., pp. 183186, 2012.

[10] R. Chang, L. Guan, and J. A. Burne, An automated form of video image analysis applied to classification of movement disorders., Disabil. Rehabil., vol. 22, no. 12, pp. 97108, 2000.

[11] D. C. Dewey, S. Miocinovic, I. Bernstein, P. Khemani, R. B. Dewey, R. Querry, and S. Chitnis, Automated gait and balance parameters diagnose and correlate with severity in Parkinson disease., J. Neurol. Sci., vol. 345, no. 12, pp. 1318, 2014.

[12] A. Salarian, F. B. Horak, C. Zampieri, P. Carlson-Kuhta, J. G. Nutt, and K. Aminian, iTUG, a sensitive and reliable measure of mobility., IEEE Trans. Neural Syst. Rehabil. Eng., vol. 18, no. 3, pp. 30310, 2010.

[13] C. Zampieri, A. Salarian, P. Carlson-Kuhta, K. Aminian, J. G. Nutt, and F. B. Horak, The instrumented timed up and go test: potential outcome measure for disease modifying therapies in Parkinsons disease., J. Neurol. Neurosurg. Psychiatry, vol. 81, no. 2, pp. 1716, 2010.

[14] R. Cucchiara, C. Grana, M. Piccardi, A. Prati, and S. Sirotti, Improving shadow suppression in moving object detection with HSV color information., IEEE Intell. Transp. Syst. Proc., pp. 334339, 2001.

[15] S. Plagenhoef, F. G. Evans, and T. Abdelnour,Anatomical data for analyzing human motion., Research Quarterly for Exercise and Sport, 54, 169178. 1983.

[16] C. H. Guzmn-Valdivia, A. Blanco-Ortega, M. A. Oliver-Salazar, and J. L. Carrera-Escobedo, Therapeutic Motion Analysis of Lower Limbs Using Kinovea, Int. J. Soft Comput. Eng., vol. 3, no. 2, pp. 359365, 2013.

Energy-efficient user scheduling algorithm for LTE networks

Marcos Rubio del Olmo[1], Rodolfo Torrea-Duran[1],
Aldo G. Orozco-Lugo[2], and Marc Moonen[1]

[1]STADIUS Center of Dynamical Systems, Signal Processing and Data Analytics, KU Leuven Department of Electrical Engineering (ESAT), Belgium

[2]Center for Research and Advanced Studies of IPN, Communications Section, Av. IPN No.2508, Col. San Pedro Zacatenco, Mexico

marcos.rubiod@gmail.com, aorozco@cinvestav.mx,
{Rodolfo.TorreaDuran, Marc.Moonen}@esat.kuleuven.be

Abstract

Long Term Evolution (LTE) has become one of the main cellular technologies to cope with the tremendous demand for higher data rates in mobile communications. Typical LTE base stations are designed to operate continuously at full power to meet this demand, while the diversity of the network load allows however a large flexibility in the assignment of resources. Most schedulers are designed with the goal of having benefits in terms of total throughput or fairness. Nevertheless, the impact on energy consumption is rarely considered. Therefore, it is of paramount importance to develop algorithms that allow base stations to reduce their energy consumption by considering changes in the network load. To tackle this problem, we propose two new scheduling algorithms that exploit the users' channel conditions to reduce the energy consumption. Using a state-of-the-art base station power model, we show that by serving users at time slots when they have favorable channel conditions, and delaying transmissions when they have unfavorable channel conditions, we can use higher modulation and coding schemes that increase the energy efficiency of the base station. Still, we guarantee a minimum QoS by setting a maximum delay time.

1 Introduction

Long Term Evolution (LTE) has become one of the main cellular technologies to cope with the tremendous demand for higher data rates in mobile communications [1]. For this purpose, typical LTE base stations are designed to operate continuously at full power in order to meet the quality-of-service (QoS) of all the potentially connected users, regardless of the operating conditions. This increases the operational expenses for the network providers as well as the environmental impact, which has been a growing concern in the ICT industry in the last years [2, 3].

In cellular networks there is a large variability in the network load due to the changing number of users connected to the network. Also, the traffic generated varies drastically from application to application. This allows the base stations a large flexibility in the assignment of resources.

Most schedulers are designed with the goal of having benefits in terms of total throughput or fairness. For instance, users that have favorable channel conditions or stringent QoS requirements are scheduled with higher priority since otherwise the important performance metrics could be degraded. However, the impact on energy

consumption is rarely considered. Therefore, it is of paramount importance to develop algorithms that allow base stations to reduce their energy consumption by considering changes in the network load without sacrificing the minimum QoS requirements. As the base station downlink transmissions are amongst the largest contributors of energy expenditure in cellular networks [4], we focus on the downlink transmissions in this paper.

To tackle this problem, we propose two new scheduling algorithms that exploit the users' channel conditions to reduce the energy consumption. Using a state-of-the-art base station power model, we show that by serving users at time slots when they have favorable channel conditions, and delaying transmissions when they have unfavorable channel conditions, we can use higher modulation and coding schemes that increase the energy efficiency of the base station. Still, we guarantee a minimum QoS by setting a maximum delay time.

Our first proposed algorithm is based on proportional fairness, which favors a high instantaneous data rate and penalizes a high average data rate. However, the proposed algorithm penalizes a high average energy consumption instead, hence rewarding energy efficiency. The second proposed algorithm considers the period of time that a user has not been served, hence rewarding users that are able to postpone their transmission. Both of the proposed algorithms have been evaluated in a LTE standard-compliant framework and compared against the most well-known scheduling algorithms for LTE. Simulations show that our proposed algorithms have a lower energy consumption at the cost of a small degradation in throughput, while guaranteeing minimum QoS constraints.

This paper is organized as follows. Section 2 describes the system model including LTE radio interface characteristics and the baseline and proposed approaches. Section 3 shows the performance evaluation. Finally section 4 draws the conclusions.

2 System Model

2.1 LTE radio interface

In LTE, a physical resource block (PRB) is the minimum resource allocation unit for a user. It is formed by 12 consecutive subcarriers, or 180 kHz, with a duration of one time slot (1ms*). Each scheduled user is assigned by the base station a transmission scheme composed of a certain modulation and coding rate. This transmission scheme is based on the channel quality indicator (CQI) fed back by each user, which is computed based on measurements of the reference signals (RS) transmitted by the base station over the whole bandwidth. This value is computed for the whole bandwidth or per PRB, as assumed in this paper.

Based on the RS, users compute the CQI as the index corresponding to the highest modulation and coding transmission scheme that supports a block error rate not exceeding 10%. The CQI is hence a measure of both the signal-to-interference-and-noise ratio and the receiver capabilities in a certain PRB [1]. The CQI can take one of 15 possible values as described in Table 1 [5]. After receiving the CQIs of all users, the base station assigns the PRBs to each user with the modulation and coding rate indicated by the CQI. Although the CQI feedback procedure is standardized, the assignment of PRBs to users is the manufacturer's choice of implementation and it is done at each base station independently.

*Although 3GPP defines the duration of a PRB as 0.5ms, the minimum resource allocation unit for a user is 1ms. Hence, in this paper we assume the duration of a PRB to be 1ms without loss of generality.

SINR (dB)	CQI	Modulation	Coding rate for a 1024 size block	Energy per information bit (μJ/bit)
-6.937	1	QPSK	78	5.43
-5.148	2	QPSK	120	3.53
-3.181	3	QPSK	193	2.20
-1.253	4	QPSK	308	1.38
0.761	5	QPSK	449	0.95
2.699	6	QPSK	602	0.71
4.694	7	16QAM	378	0.57
6.525	8	16QAM	490	0.44
8.573	9	16QAM	616	0.35
10.366	10	64QAM	466	0.31
12.289	11	64QAM	567	0.26
14.173	12	64QAM	666	0.22
15.887	13	64QAM	772	0.19
17.813	14	64QAM	873	0.17
19.828	15	64QAM	948	0.16

Table 1: SINR and CQI mapping with energy consumption.

2.2 Base station power model

The PRB assignment has an impact on the energy efficiency of the base station. Intuitively, a high modulation order and a large coding rate can transmit more information bits in each subcarrier, hence minimizing the energy consumed per information bit. This can be seen in the last column of Table 1 [6].

To compute the energy per information bit, we use the EARTH power model [4, 7], which is based on the individual power consumption of each of the components forming the transceiver of a base station. These components are grouped as power amplifier (PA), radio frequency (RF), baseband processor (BB), digital converter (DC), cooling system (CO), and main supply (MS). The power consumption of these components varies according to the network load, the bandwidth, the number of antennas, and the type of base station: macro, pico, or femto. For a macro base station with 2 transmit antennas and 10 MHz bandwidth, the power consumption of the different components of the base station is as shown in Fig. 1.

One of the main features of this power model is the scalability it provides based on the network load, which allows to reduce the energy consumption when the base station is not operating in fully loaded conditions.

2.3 Baseline scheduling algorithms

In this section, we present the most common scheduling algorithms used in cellular networks. We assume that each PRB can be assigned to a different user and that the time slot duration is 1ms.

2.3.1 Round Robin [8]

The PRBs are assigned in a circular fashion without taking into account the users' channel conditions. Every PRB is assigned to a different user until all the PRBs are

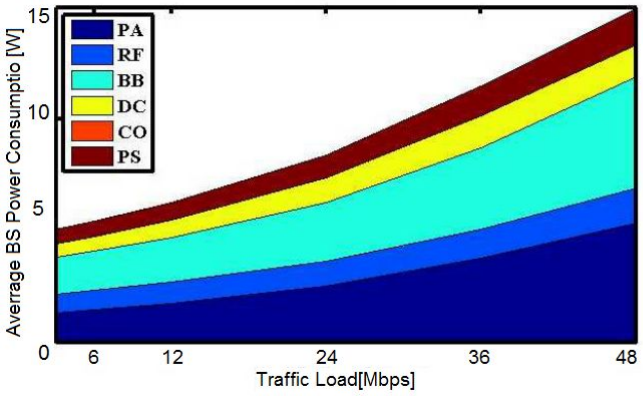


Figure 1: Average power consumption of a macro base station with 2 transmit antennas using 10 MHz bandwidth as a function of the network load [6].

assigned. If more PRBs remain to be scheduled, every user is assigned another PRB, and so on.

2.3.2 Maximum CQI (Max-CQI) [8]

Each PRB is assigned to the user which reported the highest CQI in each PRB, maximizing in this way the total throughput. This means that at time slot t, PRB k is assigned to the user i that has the highest ratio:

$$i = \arg\max r_i^k(t), \tag{1}$$

where $r_i^k(t)$ is the number of information bits per time slot of user i when using the CQI computed for PRB k. This value can be obtained from Table 1 using the corresponding modulation and coding rate.

2.3.3 Proportional fairness (PF)

This scheduling algorithm presented in [9] sacrifices maximum throughput at the cost of achieving fairness for all the users. With this algorithm, at time slot t, PRB k is assigned to the user i that has the highest ratio:

$$i = \arg\max \frac{r_i^k(t)}{(R_i)^{\beta_i(t)}}, \tag{2}$$

where R_i represents the average number of information bits of user i from the assignment of PRBs from previous time slots and

$$\beta_i(t) = m^{\text{sign}[(m-1)(r_i^k(t)-r_i^k(t-1))]}, \tag{3}$$

where

$$m = \frac{r_i^k(t) - r_i^k(t-1)}{\bar{r}_i(t) - \bar{r}_i(t-1)}, \tag{4}$$

and where $\bar{r}_i(t)$ is the average number of information bits from all the PRBs assigned to user i in time slot t.

2.3.4 Minimum CQI (Min-CQI)

This scheduling algorithm is used only as a bottom line and we assume that it assigns the PRBs to the user which has the lowest non-zero CQI. This results in achieving the minimum throughput and corresponds to assigning PRB k to the user i with the lowest ratio:

$$i = \arg\min r_i^k(t). \tag{5}$$

2.4 Energy-efficient scheduling algorithms

In order to increase the energy efficiency of the base station, we propose two algorithms. The main idea of both is to exploit the users' channel conditions to reduce the energy consumption. Specifically, by serving users at time slots when they have favorable channel conditions, and delaying transmissions when they have unfavorable channel conditions, we can use mostly higher modulation and coding schemes that increase the energy efficiency of the base station. In our simulations we assume favourable conditions if the CQI reported by a particular user for each PRB is above 5.

Whenever the transmission to a user is delayed, the EARTH power model allows the base station to reduce the energy consumption because of the reduced network load. To avoid starvation of resources for a user with continuously poor channel conditions and to avoid disturbance in the retransmissions, a transmission is delayed only if it is not a retransmission from a previously erroneous packet, and if the user has not delayed transmissions for more than 10 time slots. By setting this maximum delay time, we are able to guarantee a minimum QoS to each user.

2.4.1 Power-based proportional fariness (PPF)

This algorithm is based on the PF approach. However, instead of using the average number of information bits per user, we use the average consumed energy per user. In this way, this algorithm favors a high instantaneous rate of information bits and penalizes a high average energy consumption. This means that at time slot t, PRB k is assigned to the user i that has the highest ratio:

$$i = \arg\max \frac{r_i^k(t)}{(E_i)^{\alpha_i(t)}}, \tag{6}$$

where E_i is the average energy consumed from the assignment of PRBs from previous time slots to user i and

$$\alpha_i(t) = n^{\text{sign}[(n-1)(e_i^k(t) - e_i^k(t-1))]}, \tag{7}$$

where

$$n = \frac{e_i^k(t) - e_i^k(t-1)}{\bar{e}_i(t) - \bar{e}_i(t-1)}, \tag{8}$$

and where $e_i^k(t)$ is the energy consumed from the assignment of PRB k to user i in time slot t obtained from Table 1 and $\bar{e}_i(t)$ is the average energy consumed from all the PRBs assigned to user i in time slot t.

2.4.2 Window-based proportional fairness (WPF)

This algorithm considers the period of time that a user has not been served, hence rewarding users that are able to postpone their transmission. Similarly as PPF, it favors the instantaneous rate of information bits, but it rewards a long transmission delay. This means that user i is scheduled if

$$i = \arg\max r_i^k(t)(T^i)^n, \tag{9}$$

Parameter	Value
Number of users	15
Bandwidth	1.4 MHz
Simulation time	2s
Maximum speed	50 km/h
Minimum speed	3 km/h
Channel profile	ITU pedestrian B
Number of transmit antennas	1
Number of receive antennas	1

Table 2: Simulation parameters.

where T^i represents the window of time since the last time slot in which user i was scheduled. The variable n is a factor that tunes the weight of this period of time. In our simulations we set it to 2.

3 Performance evaluation

In this section we analyze the performance of the baseline and proposed scheduling algorithms in terms of power consumption and throughput using the parameters of Table 2. The power consumption is computed based on the energy per information bit of Table 1 for each assigned PRB, while the throughput is computed based on the rate of information bits obtained from the modulation and coding rate per PRB of Table 1. For our simulations, we use a LTE standard-compliant framework and the EARTH power model [4].

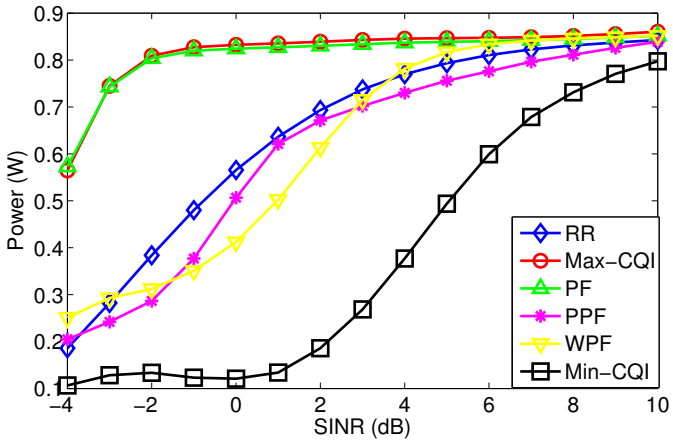


Figure 2: Power consumption of the baseline and proposed scheduling algorithms.

Fig. 2 shows the total power consumption of the base station. Max-CQI and PF lead to the largest power consumption as they both serve users with the highest instantaneous rate of information bits, regardless of the energy consumption. As expected Min-CQI leads to the lowest energy consumption, while RR shows an average performance.

The proposed energy-efficient algorithms, on the other hand, are able to drastically reduce the energy consumption of the base station compared to Max-CQI and PF. This comes at the cost of a decrease in the total throughput as seen in Fig. 3. As expected, Max-CQI offers the largest total throughput as it maximizes the instantaneous rate of information bits. PF offers a lower throughput as it achieves fairness for all the users by sacrificing maximum throughput. The proposed energy-efficient algorithms offer a throughput higher than RR for all SNR values and higher than PF for some SNR values.

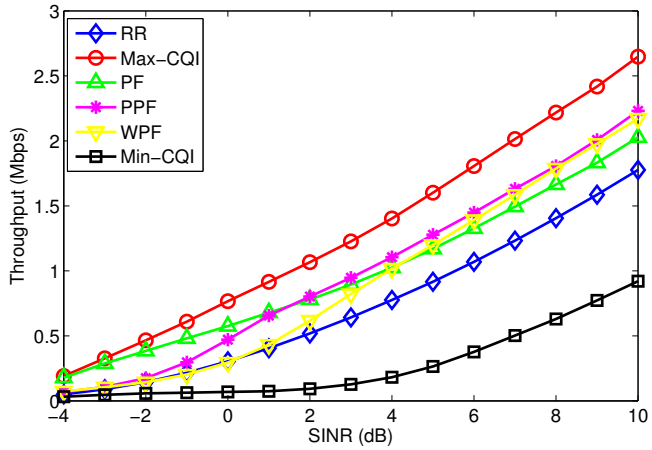


Figure 3: Throughput of the baseline and proposed scheduling algorithms.

Nevertheless, the previous analysis gives no indication of how the throughput is distributed among the users. For this purpose, we plot the cumulative distribution function (CDF) in Fig. 4.

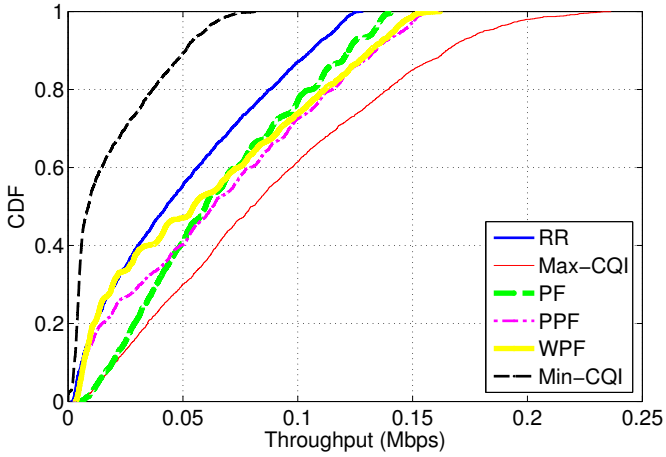


Figure 4: CDF of the baseline and proposed scheduling algorithms.

Evidently, Max-CQI offers the best performance in terms of the probability of achieving a certain throughput, while Min-CQI and RR offer the worst performance. PF offers the same performance as Max CQI for low throughput, which shows the fairness achieved by this approach. However, the proposed energy-efficient approaches are able to achieve a performance close to Max-CQI for low and high throughput and better than PF for high throughput.

4 Conclusions

In this paper we have proposed two energy-efficient scheduling algorithms for downlink transmission in LTE networks that exploit the users' channel conditions to reduce the energy consumption. Using the EARTH base station power model, we show that by serving users at time slots when they have favorable channel conditions, and delaying transmissions when they have unfavorable channel conditions, we can use higher modulation and coding schemes that increase the energy efficiency at the base station. Still, we guarantee a minimum QoS by setting a maximum delay time. The simulations show that we can drastically reduce the energy consumption with a small sacrifice in total throughput.

Acknowledgements

This research work was carried out at the ESAT Laboratory of KU Leuven, in the frame of KU Leuven Research Council PFV/10/002(OPTEC), the Belgian Programme on Interuniversity Attraction Poles initiated by the Belgian Federal Science Policy Office "Belgian network on stochastic modelling, analysis, design and optimization of communication systems (BESTCOM)" 2012-2017. The second author acknowledges the support of the Mexican National Council for Science and Technology (CONACYT). The scientific responsibility is assumed by its authors.

References

[1] 3GPP, "Evolved Universal Terrestrial Radio Access (E-UTRA) and Evolved Universal Radio Access Network (E-UTRAN): Overall Description", TS 36.300, v12.2.0, stage 2, release 12, June 2014.

[2] Parliament Office of Science and Technology, "ICT and CO2 Emmissions", no. 319. Dec. 2008.

[3] EARTH project deliverable 2.1, "Economical and ecological impact of ICT", https://www.ict-earth.eu/publications/deliverables/deliverables.html, Apr. 2011.

[4] EARTH project deliverable 2.3, INFSO-ICT-247733, "Energy efficiency analysis of the reference systems, areas of improvements and target breakdown", https://www.ictearth.eu/publications/deliverables/deliverables.html, Nov. 2010.

[5] 3GPP,"Physical layer procedures",Technical Specification Group Radio Access Network, TS 36.213, Release 12, v12.4.0, Dec. 2014.

[6] R. Torrea-Duran, C. Desset, S. Pollin, and A. Dejonghe, "Adaptive energy efficient scheduling algorithm for LTE pico base stations", en Future Network Mobile Summit (FutureNetw), 2012, pp. 1-8.

[7] EARTH project deliverable 4.1, "Most Promising Tracks of Green Radio Technologies", https://www.ictearth.eu/publications/deliverables/deliverables.html, Dec. 2010.

[8] E. Dahlman, S. Parkvall, and J. Skold, "4G: LTE/LTE-Advanced for Mobile Broadband", 1st edition. Academic Press, Oxford, Burlington, MA, 2011.

[9] X. Li, B. Li, B. Lan, M. Huang, and G. Yu, "Adaptive pf scheduling algorithm in LTE cellular system", en 2010 International Conference on Information and Communication Technology Convergence (ICTC), 2010, pp. 501-504.

Performance of multihop CSMA unicast under intermittent interference

Chara Papatsimpa[1] Jean-Paul Linnartz[1,2] Peiliang Dong[3]
Eindhoven University of Technology[1]
Philips Research, Eindhoven[2]
Philips Research China,Shanghai[3]

Abstract

We study the effect of Wi-Fi or other forms of interference on unicast multihop 802.15.4 traffic in a sensor network, based on a layered state-driven Markov chain. The probability of a successful transmission attempt is treated as a conditional probability that depends on the state, that is, on the previous experience that the packet had with the presence of harmful interference. This allows us to evaluate end-to-end success probabilities.

1 Introduction

The increasing popularity of wireless sensor networks has led to a rapid growth in the number of devices that use the IEEE 802.15.4 standard. To allow for fast deployment, the IEEE 802.15.4 opted for the 2.4 GHz ISM band. However, the presence of other wireless technologies like Wi-Fi (IEEE 802.11) or Bluetooth (IEEE 802.15.1) across the same band potentially causes coexistence issues, leading to loss of reliability for the network or inefficient use of the radio spectrum. Wi-Fi transmitters are the more concerning since they are commonly used in office or residential environments.The coexistence of 802.15.4 and Wi-Fi has been a subject of many previous papers. Most papers focus on the power differences and the large differences in time constants between the slow 802.15.4 and the fast 802.11 in accessing the channel [1, 2]. Publications on interference within an 802.15.4 network are also relevant. In [3], the back-off state of a node has been modeled as a Markov Chain. We adopt a similar model, but with a number of differences: We include the impact of Wi-Fi interference but initially neglect interference from other ZigBee nodes. We consider a unicast multihop network, in which packets follow a particular route, that is, we extend the model to include more than one hop. The probability of sensing the channel busy is assumed in [3] to be a constant, that is, independent of the history of the packet in the network. We extend this by considering a conditional probability of sensing the channel busy, which depends on whether it has experienced idle or busy channels in the past. The rest of the paper is structured as follows. The network interference is modeled in section 2. Section 3 provides an overview of the IEEE 802.15.4 protocol with the aim to derive an accurate model for our analysis. A layered Markov model is formulated in section 4. Each-subsection includes a layer describing a different protocol operation. Finally, concluding remarks are given in section 5.

2 The Interference Model

The network interference, for instance from Wi-Fi, will be modeled as a stochastic process $C'(t)$ that is independent of the packet network. This implies the assumption that the 802.15.4 network does not affect Wi-Fi devices, which is true for a specific distance range (region R3 in [2]). We will compare two different models in the following sub-sections.

2.1 Markovian Interference Model

Here, we assume that the interference is a binary on-off process alternating between active and idle periods. Let $\{C_{th}(t) : t \geq 0\}$ represent a stochastic process with discrete state space $S = \{1, 0\}$, 1 representing a clear channel (idle state) and 0 the busy state. The process transitions from idle to busy state, independent of the past, according to a continuous-time Markov chain. This process can be fully described by the finite state space $S = \{1, 0\}$, the transition matrix $P_I(t)$ and the holding-time rates $\alpha_k, k \in S$. Every time that state k is visited, the chain spends on average $\bar{t_k} = 1/a_k$ units of time there before moving on. Once we choose particular $\bar{t_1}$ and $\bar{t_0}$, that is, the average inter-frame idle time between two consecutive IEEE 802.11 packets and the average time that the Wi-Fi traffic is on respectively, not only we fix the holding rates, but also the probability γ_0 is defined as $\gamma_0 = P_r(C_{th}(t) = 1) = \bar{t_1}/(\bar{t_0} + \bar{t_1})$.
The transition probabilities can be calculated by solving the Kolmogorov backward equations for the CTMC with transition matrix $P_I(t)$. For the given two state CTMC the transition matrix are of the following form:

$$P_I(t) = \begin{pmatrix} \frac{a_0}{a_0+a_1} & \frac{a_1}{a_0+a_1} \\ \frac{a_0}{a_0+a_1} & \frac{a_1}{a_0+a_1} \end{pmatrix} + \begin{pmatrix} \frac{a_1}{a_0+a_1} & \frac{-a_1}{a_0+a_1} \\ \frac{-a_0}{a_0+a_1} & \frac{a_0}{a_0+a_1} \end{pmatrix} e^{-(a_0+a_1)t} \tag{1}$$

2.2 Interference Traffic Model based on measured data

In this traffic model, we assume that the interference pattern is characterized by measurements at a survey site. Let $C_m(t)$ be a stochastic process that describes the channel status at instant time t, $t \in [0, T]$.

$$C_m(t) = \begin{cases} 1 & \text{if } E(t) < E_{th} \\ 0 & \text{if } E(t) \geq E_{th} \end{cases} \tag{2}$$

In the above equation, $E(t)$ is the measured interference power level at time t and E_{th} is the threshold value above which we assume that 802.15.4 communication is disrupted. The clear channel rate ($\gamma_0 = P_r(C_m(t) = 1)$ is obtained directly from measured realizations $c_m(t), t \in T$ of the stochastic process $C_m(t)$ in a sampling window of length T.

3 IEEE 802.15.4 Standard Overview

We review the IEEE 802.15.4 protocol in order to derive a simplified but sufficiently accurate model. IEEE 802.15.4 employs CSMA/CA for medium access control. When a node has a packet to transmit, it backs off for a random number of backoff slots (each slot lasts for $T_{bs} = 0.32$ msec) chosen uniformly between 0 and $2^{BE} - 1$. After the backoff, the channel is checked using a clear channel assessment (CCA). If the channel is sensed idle, the node starts transmitting its packet. This transmission can be successful or run into a collision, for instance because Wi-Fi is ignorant of 802.15.4 traffic. These transmission failures can be remedied by a positive acknowledgment scheme (ACK), that is, the packet is retransmitted up to a maximum number of retries R, if no acknowledgment packet is received. If the the maximum number R is exceeded, the protocol terminates with a communications failure. On the other hand, if the channel is found to be busy, the backoff exponent (BE) is incremented by one and the node waits for a new random number of back off slots until the channel can be sensed again. This procedure continues up to a maximum number N of allowed back-offs and the protocol terminates with a channel access failure.

4 Formulation of the protocol model

The protocol operations, as described in section 3, are modeled in the form of a layered Markov Chain presented in Fig.1. A detailed description of each layer is provided in the following sub-sections.

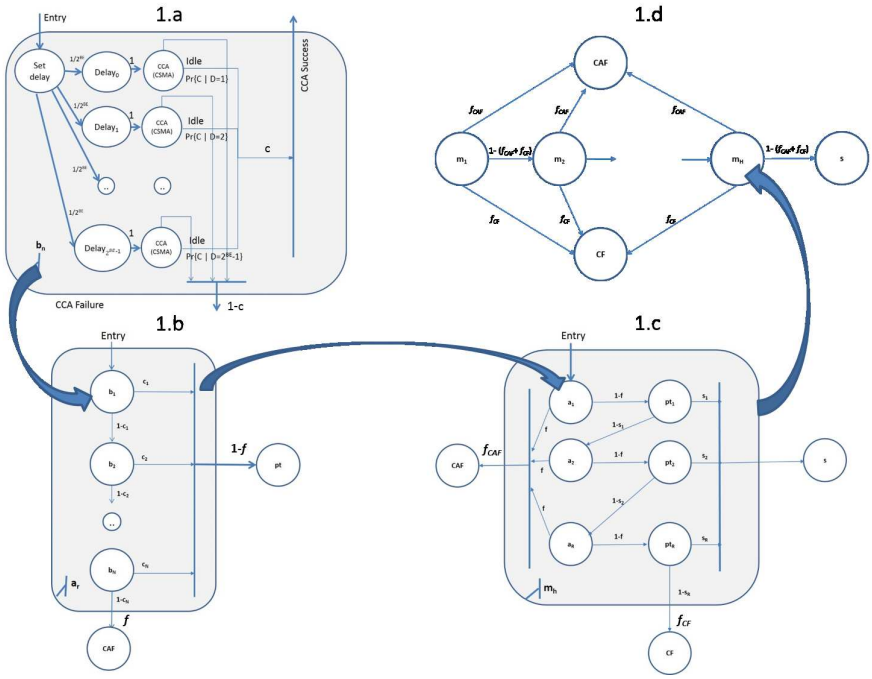


Figure 1: Layered state transition diagram for protocol operations

4.1 The Random Delay Mechanism

The random delay mechanism in CSMA/CA can be approximated by means of a discrete-time transition chain as shown in Fig. 1a. The random back off period before attempting a CCA is represented by a transition to one of the Delay_d states, $d \in [0, 2^{BE} - 1]$. We assume that the conditional idle probability $P_r(C|D = d)$ for a clear channel assessment depends on previous observations and the observation delay $\tau = dT_{bs}$, i.e., on how long ago these previous observation have been done. In this paper, we are not (yet) interested in latency. Thus we have no need to consider time-driven Markov chains and we will further collapse this into an event-based Markov chain. We use the model of Fig. 1a to calculate the probability of sensing the channel idle during CCA.

$$P_r(C) = \sum_{d=0}^{2^{BE}-1} P_r(C|D = d)P_r(D = d) = \frac{1}{2^{BE}} \sum_{d=0}^{2^{BE}-1} Pr(C|D = d) \tag{3}$$

We can distinguish two different cases:

Initial back-off: No evidence about channel status
In the case of the first attempt to access the channel, we have no previous knowledge about the channel status. Thus, the probability $Pr(C|D = d))$ to sense the channel

idle after the backoff time $\tau = dT_{bs}$ is assumed independent on time τ and equal to the clear channel rate for both interference models.

$$P_r(C) = \frac{1}{2^{BE}} \sum_{d=0}^{2^{BE}-1} Pr(C|D=d) = \frac{1}{2^{BE}} \sum_{d=0}^{2^{BE}-1} \gamma_0 = \gamma_0 \forall d \in [0, 2^{BE}-1] \quad (4)$$

Channel assessment following a busy detection

When a node is attempting to access the channel given that it was sensed busy in the previous attempt, we know that there was an on-going Wi-Fi transmission in the medium. We assume that probability $Pr(C|D=d)$ to sense the channel idle given a previous busy detection after the back off waiting time $\tau = dT_{bs}$ depends on the result of the previous CCA attempt. This probability will be estimated for the two traffic models:

i. Markovian Interference traffic model

The probability to sense the channel idle (state 1) after time $\tau = dT_{bs}$ given that now is in state 0 is given by the (2, 1) entry of the transition matrix $P_I(\tau)$. Thus according to Eq.3:

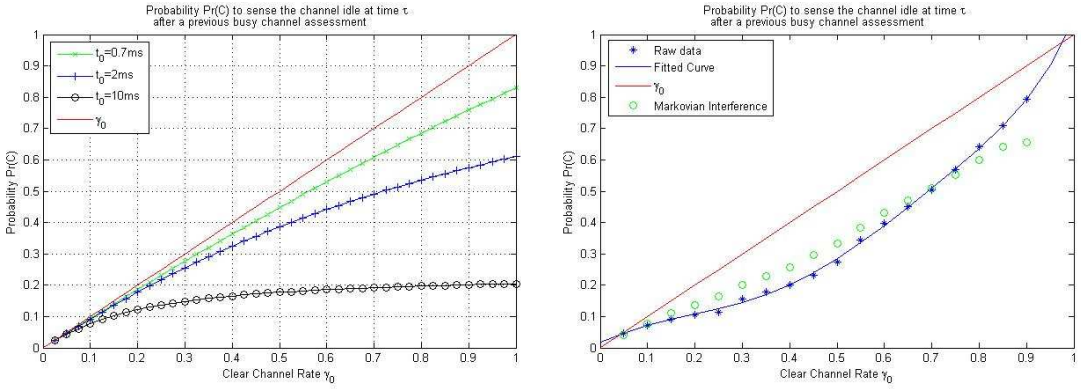
$$P_r(C) = \frac{1}{2^{BE}} \sum_{d=0}^{2^{BE}-1} Pr(C|D=d) = \frac{1}{2^{BE}} \sum_{d=0}^{2^{BE}-1} P_{I_{2,1}}(dT_{bs}), d \in [0, 2^{BE}-1] \quad (5)$$

ii. Interference traffic model based on measured data:

Similarly, we are interested in the conditional probability of an idle channel assessment at time $t + \tau$ given that the previous CCA at time t indicated a busy channel.

$$P_r(C) = \frac{1}{2^{BE}} \sum_{d=0}^{2^{BE}-1} P_r\{C_m(t + dT_{bs}) = 1 | C_m(t) = 0\} \quad (6)$$

Fig.2 shows the probability $P_r(C)$ of a channel idle assessment after backoff time τ given a previous busy channel detection for the two interference models. The calculations in



(a) Markovian Interference traffic model (b) Measured data traffic model

Figure 2: Probability $P_r(C)$ of a channel idle assessment given a previous busy channel detection

Fig.2b are based on measured data collected from a survey site in an office environment in Shanghai [4]. In both models, we can observe that the previous busy channel detection provides us with information about the consequent channel assessment, i.e., the probability to sense the channel idle is lower than the unconditional probability γ_0 (red line). The green dots in Fig.2b present the results from the Markovian interference model assuming the same average busy time ($\bar{t_0}$).

4.2 Back-off state Markov Model

Extending the backoff state $\mathrm{b_n}$ as proposed in Fig.1b, a Markov Chain is constructed. We start from the first attempt to access the channel, till either the channel is sensed idle and the packet is transmitted or the maximum number of back-off stages N has been reached and the protocol terminates. We use the following notation: $\mathrm{b_n}$ denotes the n^{th} back-off attempt, $n \in [1, \ldots, N]$, pt a packet transmission and the CAF state denotes protocol termination with a channel access failure. Finally, $B(t)$ represents a stochastic process such that:

$$B(t) = \{\mathrm{b_n}, \mathrm{pt}, \mathrm{CAF}\}, n \in [1, \ldots, N] \tag{7}$$

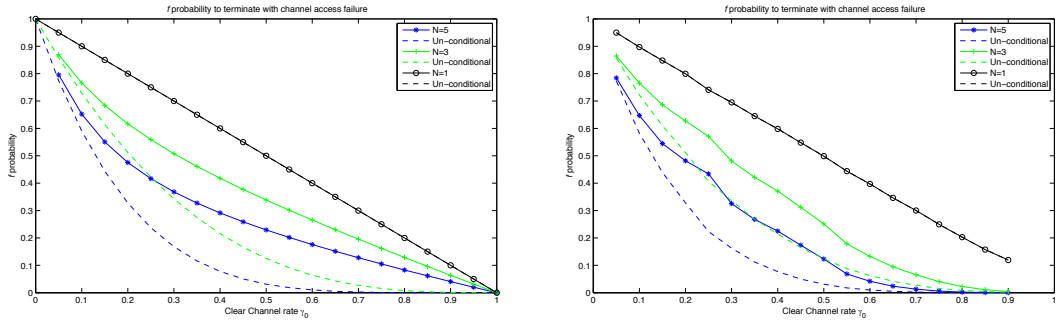
Let P_B be the transition matrix of the Markov chain with transition probabilities*:

$$P_B(\mathrm{b_n}|\mathrm{b_{n+1}}) = 1 - \mathrm{c_n} \tag{8}$$

$$P_B(\mathrm{b_n}|\mathrm{pt}) = \mathrm{c_n} \tag{9}$$

$$P_B(\mathrm{b_N}|\mathrm{CAF}) = 1 - \mathrm{c_N} \tag{10}$$

The transition probabilities $\mathrm{c_n}$ are the idle probabilities $P_r(C)$ as calculated in the previous sub-section ($\mathrm{c_1}$ from Eq.4 and $\mathrm{c_n}$ from Eq.5-6. Filling in the transition matrix P_B, enables us to determine the overall probability f to terminate the protocol with a channel access failure (reach CAF state in Fig.1b). This probability is presented in Fig.3. for a different number of allowed backoff attempts N. The dashed lines present probability f in the case that we do not consider conditional probabilities. That is, the probability to sense the channel idle (c_n) given a previous busy detection does not depend on history and is considered equal to the clear channel rate.



(a) Markovian Interference traffic model

(b) Measured data traffic model

Figure 3: Probability f to terminate the protocol with a channel access failure

*For sake of notational simplicity, we shorten $P_B(B(t) = a)|B(t+1) = b)$ as $P_B(a|b)$.

4.3 Single Hop Markov Model

In order to model the packet re-transmission attempts, we extend the backoff state $\mathrm{a_r}$ as proposed in Fig.1c, starting from the first transmission attempt, till either the packet has been successfully transmitted to the next hop or the maximum number of allowed retries has been reached. At any point in time, a packet is in one of the following states: state $\mathrm{a_r}$, during which a node is contending to access the channel, state $\mathrm{pt_r}$, where a packet is being transmitted in the current retry, the s state denoting a successful packet transmission or in the CAF or CF states that denote a channel access failure or a communications failure respectively. Finally $A(t)$ represents a stochastic process such that:

$$A(t) = \{\mathrm{a_r}, \mathrm{pt_r}, \mathrm{s}, \mathrm{CF}, \mathrm{CAF}\}, r \in [1, \ldots, R] \tag{11}$$

Let P_A be the transition matrix of the Markov chain with probabilities:

$$P_A(\mathrm{a_r}|\mathrm{pt_r}) = 1 - f \tag{12}$$

$$P_A(\mathrm{a_r}|\mathrm{CAF}) = f \tag{13}$$

$$P_A(\mathrm{pt_r}|\mathrm{s} = \mathrm{s_r} \tag{14}$$

$$P_A(\mathrm{pt_r}|\mathrm{a_{r+1}}) = 1 - \mathrm{s_r} \tag{15}$$

$$P_A(\mathrm{pt_R}|\mathrm{CF}) = 1 - \mathrm{s_R} \tag{16}$$

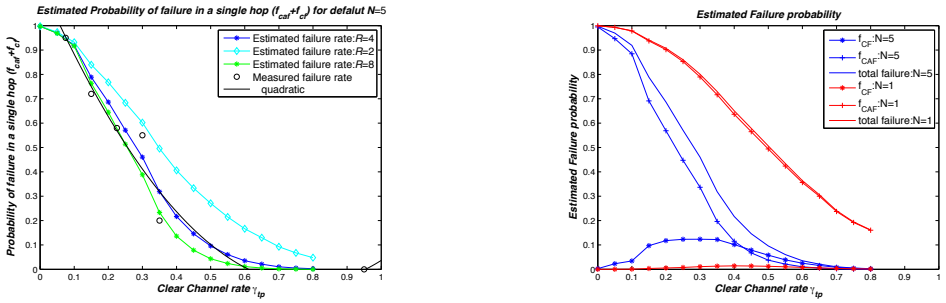
Filling in the transition matrix P_A, enables us to determine the probabilities f_{CAF} (reach CAF state in Fig.1c) and f_{CF} (reach CF state in Fig.1c). Transition probabilities f are calculated as in sub-section 4.2 and probability s_r (successful packet transmission after a clear CCA) will be estimated in the following sub-sections:

i. Markovian Interference traffic model
In this case, a lossless channel is assumed, i.e., we do not take into account any form of packet losses that may occur during the transmission ($s_r = 1$). Since a packet is always transmitted successfully, for the Markovian interference, we do not model the ACK packets and we do not consider any packet retransmissions ($R = 1$). Thus, the overall probability to fail to transmit the packet in a single hop is as presented in Fig.3a.

ii. Interference traffic model based on measured data:
We adopt a similar approach to [5], in order to consider a packet failure after a clear CCA. Let $S(t)$ be a stochastic process denoting the availability of the channel for a transmission of duration t_p starting at instant t, $t \in \{0, T - t_p\}$.

$$S(t) = \begin{cases} 1, & \text{if } \int_{t}^{t+t_p} C_m(t)d\tau = t_p \\ 0, & \text{if } \int_{t}^{t+t_p} C_m(t)d\tau < t_p \end{cases} \tag{17}$$

where t_p denotes the time duration that the channel needs to be free of interference. The clear channel rate γ_{t_p} is defined as $\gamma_{t_p} = P_r\{S(t) = 1\}$.
Surprisingly, experiments in [4] revealed that the probability to find a clear time slot for a successful packet transmission is virtually not correlated with the outcome of the preceding CCA. Using this as an approximation, the probability of a channel clear assessment and the probability of a successful packet transmission become independent. Thus the probability for a successful packet transmission (s_r) is equal to the clear channel rate γ_{t_p}.



(a) Default N=5 backoff attempts

(b) f_{CF} and f_{CAF}

Figure 4: Estimated failure rate of the 802.15.4 32-byte packet transmission in a single hop under Wi-Fi interference

Fig.4 shows the overall probability to fail to transmit the packet in a single hop ($f_{CAF}+f_{CF}$). A number of repetitive re-transmission attempts significantly reduces the failure rate. The circles in Fig.4a present measured failure rates from a lab test in Shanghai [4]. Our theoretical estimation of the failure rate in a single hop is very close to the fitted curve of the above mentioned measured rates, denoting the accuracy of the model. From Fig.4b we can observe that at low clear channel rate the protocol mostly fails during CCA. However, as the clear channel rate increases transmissions fail even if CCA denoted a clear channel.

4.4 MultiHop Markov Model

At the highest layer of abstraction, a Markov chain is constructed that follows the packet as it passes over multiple hops. At any point in time, a packet is in one of the following states: state $\mathrm{m_h}$, including the channel access mechanism and the packet transmission in the current hop h, the s state denoting a successful transmission or in the CAF or CF states that denote a channel access failure or a communications failure respectively. $M(t)$ represents a stochastic process such that:

$$M(t) = \{\mathrm{m_h}, \mathrm{s}, \mathrm{CF}, \mathrm{CAF}\}, h \in [1, \ldots, H] \tag{18}$$

Let P_M be the transition matrix of the Markov chain with transition probabilities:

$$P_M(\mathrm{m_h}|\mathrm{m_{h+1}}) = 1 - (f_{CAF} + f_{CF}) \tag{19}$$

$$P_M(\mathrm{m_m}|\mathrm{CAF}) = f_{CAF} \tag{20}$$

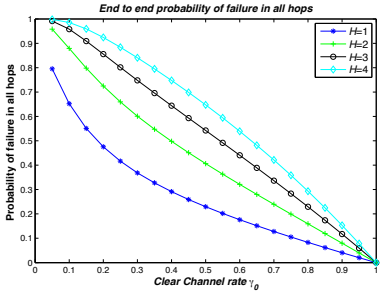
$$P_M(\mathrm{m_h}|\mathrm{CF} = f_{CF} \tag{21}$$

$$P_M(\mathrm{m_H}|\mathrm{s} = 1 - (f_{CAF} + f_{CF}) \tag{22}$$

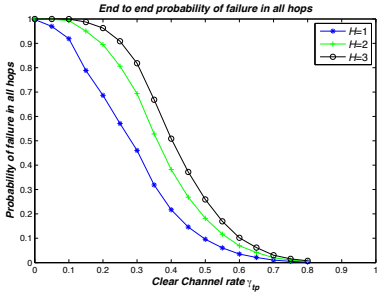
Filling in the transition matrix P_M, enables us to calculate the end to end probability of a successful packet transmission in all hops as presented in Fig.5. As expected, as the number of hops a packet has to pass through to reach the final destination increases, the probability of a successful transmission decreases. The difference in the two models is due to the fact that in the case of the Markovian Interference, we assume that the 802.15.4 packets are always transmitted successfully ($s_r = 1$).

5 Conclusions

Our model was motivated by a need to analyze and predict the performance of IEEE 802.15.4 under IEEE 802.11 (Wi-Fi) interference. We adapted and extended a Markov Chain analysis to account for Wi-Fi interference to a unicast multihop CSMA/CA network. The main strength and novelty of this model is that it accounts for a changing conditional probability of successful transmission, as the node makes successive attempts, given the history of earlier transmission attempts. Preliminary results in a single hop, showed good agreement with the measured failure rates, indicating that the model can be used in a multiple hops scenario. Results revealed that when the channel is very busy ($\gamma_{tp} < 0.4$) the failure rate increases dramatically. Increasing the re-transmission attempts R is not recommended,as most transmissions fail during CCA. Other measures, like frequency agility, should be performed. At high γ_{tp} values, transmissions fail even after a clear CCA, indicating that the channel should be free of interference for a gap length at least equal to the 802.15.4 packet for a successful transmission. Packet re-transmissions are now essential, however, increasing R too much does not have a huge impact (recommended value between 3-5).



(a) Markovian Interference traffic model



(b) Measured data traffic model

Figure 5: End-to-end failure rate of the 802.15.4 packet transmission in many hops for N=5 allowed back off attempts and R=4 packet re-transmissions

References

[1] Soo Young Shin; Hong Seong Park; Wook Hyun Kwon (2007). Mutual interference analysis of IEEE 802.15.4 and IEEE 802.11b, Computer Networks, Vol. 51, Issue 12,pp 3338-3353

[2] Wei Yuan; Xiangyu Wang; Linnartz, J.-P.M.G.. A Coexistence Model of IEEE 802.15.4 and IEEE 802.11b/g. In proceedings of the SCVT '07, Delft, the Netherlands.

[3] Pollin, S.; Ergen, M.; Ergen, S.; Bougard, B.; Der Perre, L.; Moerman, I.; Bahai, A.; Varaiya, P.; Catthoor, F.. Performance Analysis of Slotted Carrier Sense IEEE 802.15.4 Medium Access Layer. Wireless Communications, IEEE Transactions on , vol.7, no.9, pp.3359,3371, September 2008

[4] P. Dong; Z. Zhang; Jun Yao. Survey and Study of IEEE 802.15.4 network link quality under Wi-Fi interference.

[5] Z. Zhang; P. Dong; Jun Yao; Holtman, K.. A Performance Study of ZigBee Broadcasts in Coexistence with Wi-Fi. In proccedings of the CyberC'14.

A Dynamic Digital Signature Scheme Without Third Parties

Maarten van Elsas
Jan C.A. van der Lubbe
Jos H. Weber
Delft University of Technology
Fac. EEMCS, Cyber Security Group
2628 CD, Delft, the Netherlands
M.vanElsas@student.tudelft.nl
J.C.A.vanderLubbe@tudelft.nl
J.H.Weber@tudelft.nl

Abstract

We propose a digital signature scheme for dynamic coalitions. Particularly, we enable members to join and leave. Our scheme does not need trusted or oblivious third parties (TTPs or OTPs). In this distributed scheme there is a changing secret group key dependent on the members in the group. Each member's secret key remains the same for all group compositions. The downside of this approach is that we need to keep track of the signatures signed by each group composition to prevent backlogging. We use One-Way Accumulators to minimise the amount of information that needs to be saved for this.

1 Introduction

In various collaborative environments such as alliances for joint peacekeeping military operations or joint emergency responses to the spread of e.g. an infectious disease such as Ebola, coalitions are formed to achieve common objectives by resource sharing and joint decision making. In practice the coalition members are dissimilar with regards to their disposition. In the Ebola example they vary from private and non-governmental organisations (such as the Red Cross and Doctors Without Borders), to governmental organisations (such as the military, hospitals, local and federal government). In general each of the coalition partners has its own policies and will participate in the joint operation for a limited period of time. This makes the coalitions very dynamic. In peacekeeping operations partners will join and leave the coalition. It is clear that while they are participating in the coalition they should share information as efficiently as possible. However, if a partner leaves the coalition this partner should no longer have access to the shared information. On the other hand if a new partner joins the coalition it should have access to the shared information. In other words, access control is essential. Therefore it should be possible to authenticate users that log on to the combined information network. To authenticate a user it is essential that users have their own private key and that the public keys are certified by a Certification Authority (CA). The CA is a very important factor in the coalitions. It is more important still that all partners take part in it, as in practice the coalition partners do not accept one commonly trusted party that can be used to provide the coalition partners with their secret signing keys. When the coalition partners want to be able to place a signature jointly without addressing a Trusted Third Party (TTP), there are several distributed key generation and signature protocols available. In 2014 Van der Lubbe et al. [4] described a distributed (n, n) signature scheme for a dynamic coalition defence environment, which can be expanded to a $(n + 1, n + 1)$ signature scheme. A (k, n)

threshold signature scheme requires k out of n group members to sign a message. So in an (n, n) signature scheme all n group members participate in the signing process. Their scheme is based on the modified ElGamal type signature scheme described by Park and Kurosawa [3]. In order to avoid usage of one TTP they proposed the usage of two OTPs . But the usage of OTPs has its disadvantage; if they work together there are potential weaknesses in the scheme. In this paper we propose a dynamic (n, n) scheme where OTPs are not needed.

Our aim is to create an (n, n) signature scheme for our dynamic coalition. $\mathcal{N}$ is the current set of members of this coalition. Each member has an unique identifier o_i. We denote $\mathcal{O}$ as the set of unique identifiers of all members of $\mathcal{N}$. More formally $\mathcal{O} = \{o_i | i \in \mathcal{N}\}$. During every phase only the current set of members $\mathcal{N}$ can give out a signature. This requires only the cooperation of every member in $\mathcal{N}$. When the group composition changes, members keep their own secret key.

In Section 2 we give a static protocol and the challenges to make it dynamic. Then in Section 3 we present our dynamic protocol. Section 4 concludes the paper.

2 A static group $\mathcal{N}$

First we will give a static signature scheme as introduced by Park and Kurosawa [3]. We refer to this as the static scheme. We will then expand this scheme to a dynamic scheme, enabling members to leave and join the set $\mathcal{N}$.

The following applies to both schemes: p and q are large primes such that q divides $p - 1$. g generates the group G_q which is a subgroup of $\mathcal{Z}_p$, of order q. We assume p, q and g are publicly known. In every case m is the message that all parties agree to by signing. This can be e.g. a public key of an individual or information about the changes in the group composition. We use $h(m)$ to denote the hash value of m where h is a publicly known hash function with a range from 1 to $q - 1$.

2.1 Static Initialization Protocol

First we give the initialization protocol for the static scheme. The protocol reads as follows:

1. Each member $i \in \mathcal{N}$ chooses a random secret x_i from $\mathcal{Z}_q$.
2. Each member $i \in \mathcal{N}$ broadcasts $y_i = g^{x_i} \mod p$ to all other members.
3. Each member in $\mathcal{N}$ computes $y = \prod_{i \in \mathcal{N}} y_i = g^x \mod p$.

Hence, in the initialization we have each member pick a secret key x_i and share the corresponding y_i with each other. The shared secret x is defined as follows:

$$x \triangleq \sum_{i \in \mathcal{N}} x_i$$

This shared secret is not known by any member. However, the corresponding value of y is known by all members.

2.2 Signature Issuing Protocol for a static group $\mathcal{N}$

Secondly we give the signature issuing protocol in the static scheme. The protocol reads as follows:

1. Each $i \in \mathcal{N}$ chooses a random secret β_i from $\mathcal{Z}_q$.

$$\beta \triangleq \sum_{i \in \mathcal{N}} \beta_i$$

   Here β is the shared random secret, not known by any member.

2. Each $i \in \mathcal{N}$ broadcasts $c_i = g^{\beta_i} \mod p$ to all other members.

3. Each $i \in \mathcal{N}$ reveals $a_i = g^{\gamma_i}$ where $\gamma_i \triangleq wx_i + h(m)\beta_i \mod q$. Here w is equal to $v \mod q$ with $v = \prod_{i \in \mathcal{N}} c_i = g^{\beta} \mod p$.

4. Each member in $\mathcal{N}$ verifies that $\forall j, a_j = (y_j)^w (c_j)^{h(m)}$.

5. Each member in $\mathcal{N}$ computes $a = \prod_{i \in \mathcal{N}} a_i = \prod_{i \in \mathcal{N}} g^{\gamma_i} = g^{\sum_{i \in \mathcal{N}} \gamma_i} = g^t$ where $t = wx + h(m)\beta \mod q$.

The validity of the signature (t, w, y) is verified by

$$w \equiv (g^{t/h(m)} y^{-w/h(m)} \mod p) \mod q$$

We have altered the original protocol slightly by revealing g^{γ_i} instead of γ_i. We do this so that it becomes harder to derive x_i from this value.

2.3 Backlogging

Imagine we would use the static protocol for a dynamic group and have a group $\mathcal{N}$ as well as a group $\mathcal{N}' = \mathcal{N} \cup \{k\}$. Where k is the new member that is joining. The initial group was $\mathcal{N}$ which gave out several signatures before admitting member k. In this situation the group $\mathcal{N}$ can still give out signatures pretending k has not yet joined because k has no way of knowing rather a signature given out by $\mathcal{N}'$ was created before or after it joined. If a signature is given out by $\mathcal{N}$ after k joined it is backlogged. We will prevent this backlogging using One-Way Accumulators.

2.4 One-Way Acummulator

A One-Way Accumulator (OWA) is a one way membership function. Depending on the implementation only an identifier (such as the hash of a document) or and identifier and a witness value need to be provided by the party identifying itself.

For our implementation, we require that the OWA has no trapdoor. This is because if there were a trapdoor, there is no way for any group $\mathcal{N}$ to know rather the previous groups know this trapdoor and as such can backdate additional signatures. Trapdoorless OWAs have been given by Lipema [1] and Nyberg [2]. It is inefficient to have to send the witness values to the certificate holders after the OWA is no longer being updated. As such we choose to use an OWA that does not update the witness values when more values are added. We use an OWA as given by Nyberg [2] that does in fact not use a witness value at all but is instead based on bloom filters. This does mean that the amount of memory required is linear to the amount of elements saved. When the OWA is initialized the amount of items it can contain needs to be determined, the size can not be increased later.

3 A dynamic group $\mathcal{N}$

We define two kinds of signatures: regular signatures and group signatures. Regular signatures give out certificates to individuals using their identifying information e.g. their public key. Group signatures on the other hand are used to confirm changes to the composition of group $\mathcal{N}$. We use the very similar signature protocols for these but the message differs. For regular signatures the message m is the identifying information of the individual the certificate is issued to. For group signatures it contains the group composition, the y value of the composition and the OWA.

In the following paragraphs we outline the protocols for initializing the group, having a member join and having a member leave. In each of these cases a group signature is created. These group signatures and their messages are kept by each group member and past on to any new members. We will refer to the collection of group signatures as the memberlog. The period between changes in the group composition is a phase.

We assume we have a public hash function h', a security parameter τ and maximum number of items N for our OWA. Here $e^{-\tau}$ is the probability of forgery. From N and τ the required size of our boolean array follows. This boolean array together with the hash function h' forms the OWA. Every phase a new OWA z is made by every member by initializing a new (empty) boolean array. For every regular signature that is given out the value m is added to z. $h(z)$ is the hash value of the boolean array. Z is denoted as the collection of z values belonging to all previous groups.

3.1 Dynamic Initialization Protocol

In this subsection, we present the initialization protocol for the dynamic scheme. The proposed protocol reads as follows:

1. Each member $i \in \mathcal{N}$ chooses a random x_i from $\mathcal{Z}_q$.
2. Each member $i \in \mathcal{N}$ broadcasts o_i and $y_i = g^{x_i} \mod p$ to all other members.
3. Each member $i \in \mathcal{N}$ computes $y = \prod_{i \in \mathcal{N}} y_i$.
4. A group signature is issued by the members of $\mathcal{N}$ with $m := (\mathcal{O}, y)$.

Other than in the static protocol, we have the members initialize a memberlog with the signature containing $\mathcal{O}$. This information is later used to verify validity of the memberlog.

3.2 Joining Protocol

In this subsection, we consider the situation in which a new member k joins the group. The extended group $\mathcal{N} \cup \{k\}$ is denoted by $\mathcal{N}'$. Additionally, $\mathcal{O} \cup \{o_k\}$ is denoted by $\mathcal{O}'$. The proposed protocol reads as follows:

1. k requests Z and the memberlog from a member $i \in \mathcal{N}$.
2. The member sends Z and the memberlog to k.
3. k verifies that the memberlog consists of a valid group signature sequence starting with the initialization group signature.
4. k verifies that $\forall z \in Z$, z corresponds to the $h(z)$ value in the memberlog and that there is a z corresponding to each $h(z)$ value in the memberlog.
5. Each member $i \in \mathcal{N}$ sends o_i and $y_i = g^{x_i} \mod p$ to k.

6. k chooses a random x_k from $\mathcal{Z}_q$.
7. k broadcasts its o_k value and $y_k = g^{x_k} \mod p$ to each member $i \in \mathcal{N}$.
8. Each member in $\mathcal{N}'$ computes $y' = \prod_{i \in \mathcal{N}'} y_i$.
9. A group signature is issued by the members in $\mathcal{N}$ with $m := (\mathcal{O}', y', h(z))$. It is added to the memberlog.
10. A member $i \in \mathcal{N}$ sends this signature and z to k.
11. k checks whether the group signature is valid for z.
12. A new phase starts with the group $\mathcal{N}'$: $\mathcal{N} := \mathcal{N}'$ where $\mathcal{O} := \mathcal{O}'$, $y := y'$ and $Z := Z \cup \{z\}$.

By having the members in the group $\mathcal{N}$ sign the O', the new member k has proof it has been admitted by all members in $\mathcal{N}$. This prevents the old members from silently reverting back to before k joined. The $h(z)$ value is signed in order to lock in the signatures that have been signed in the phase by $\mathcal{N}$ preventing backlogging.

3.3 Leaving Protocol

Next we consider the situation in which a member j leaves the group. The reduced group $\mathcal{N} \setminus \{j\}$ is denoted by $\mathcal{N}'$. Additionally, $\mathcal{O} \setminus \{o_j\}$ is denoted by $\mathcal{O}'$. The proposed protocol reads as follows:

1. Each member in $\mathcal{N}'$ computes $y' = \prod_{i \in \mathcal{N}'} y_i$.
2. A group signature is issued by the members $i \in \mathcal{N}$ with $m := (\mathcal{O}', y', h(z))$. It is added to the memberlog.
3. A new phase starts with the group $\mathcal{N}'$: $\mathcal{N} := \mathcal{N}'$ where $\mathcal{O} := \mathcal{O}'$, $y := y'$ and $Z := Z \cup \{z\}$.

By having the group $\mathcal{N}$ sign a signature with $\mathcal{O}'$, this new group has proof that j agreed to leave.

3.4 New OWA Protocol

Because our OWA needs to be set up at the beginning of the phase and has a limited size, it might reach this size. If this happens we add the old OWA to the memberlog and start a new one.

1. A group signature is issued by the members $i \in \mathcal{N}$ with $m := (\mathcal{O}, h(z))$. It is added to the memberlog.
2. Z is updated by each member: $Z := Z \cup \{z\}$.

3.5 Regular Signature issuing for a dynamic group $\mathcal{N}$

Our signature issuing is similar to the static case. However, we add the value of $h(m)$ to z to prevent backlogging.

1. Each $i \in \mathcal{N}$ chooses a random β_i from $\mathcal{Z}_q$.

$$\beta \triangleq \sum_{i \in \mathcal{N}} \beta_i$$

   Here β is the shared random secret, not known by any member.

2. Each $i \in \mathcal{N}$ broadcasts $c_i = g^{\beta_i} \mod p$ to all other members.

3. Each $i \in \mathcal{N}$ reveals $a_i = g^{\gamma_i}$ where $\gamma_i \triangleq wx_i + h(m)\beta_i \mod q$. Here w is equal to $v \mod q$ with $v = \prod_{i \in \mathcal{N}} c_i = g^{\beta} \mod p$.

4. Each member in $\mathcal{N}$ verifies that $\forall j, a_j = (y_j)^w (c_j)^{h(m)}$.

5. Each member in $\mathcal{N}$ computes $a = \prod_{i \in \mathcal{N}} a_i = \prod_{i \in \mathcal{N}} g^{\gamma_i} = g^{\sum_{i \in \mathcal{N}} \gamma_i} = g^t$ where $t = wx + h(m)\beta \mod q$.

6. Each member in $\mathcal{N}$ adds m to z.

The validity of the signature (t, w, y) is verified by

$$w \equiv (g^{t/h(m)} y^{-w/h(m)} \mod p) \mod q$$

We only use this signature to authenticate the message if its been given out in the current phase. We check this by confirming that the most recent y value in the memberlog matches that of the signature. Because of the potential for backlogging using just the signature is only possible for the current group composition. If this is not the case, we need to use our OWA and check rather $h(m)$ is in Z. $h(m)$ being in Z suffices for authentication in itself. However, we do not use the OWA for the current group composition because it would require us to keep the OWA up to date at every location at which we check credentials. By using the signatures for the current phase we only have to update these locations when the group composition changes.

3.6 Group Signature issuing for a dynamic group $\mathcal{N}$

In this final subsection, we give the group signature protocol:

1. Each $i \in \mathcal{N}$ chooses a random β_i from $\mathcal{Z}_q$.

$$\beta \triangleq \sum_{i \in \mathcal{N}} \beta_i$$

   Here β is the shared random secret, not known by any member.

2. Each $i \in \mathcal{N}$ broadcasts $c_i = g^{\beta_i} \mod p$ to all other members.

3. Each $i \in \mathcal{N}$ reveals $a_i = g^{\gamma_i}$ where $\gamma_i \triangleq wx_i + h(m)\beta_i \mod q$. Here w is equal to $v \mod q$ with $v = \prod_{i \in \mathcal{N}} c_i = g^{\beta} \mod p$.

4. Each member in $\mathcal{N}$ verifies that $\forall j, a_j = (y_j)^w (c_j)^{h(m)}$.

5. Each member in $\mathcal{N}$ computes $a = \prod_{i \in \mathcal{N}} a_i = \prod_{i \in \mathcal{N}} g^{\gamma_i} = g^{\sum_{i \in \mathcal{N}} \gamma_i} = g^t$ where $t = wx + h(m)\beta \mod q$.

6. Each member in $\mathcal{N}$ adds (t, w, y) to the memberlog.

The validity of the signature (t, w, y) is verified by

$$w \equiv (g^{t/h(m)} y^{-w/h(m)} \mod p) \mod q$$

When we issue a group signature we do not need to use the OWA. This is because the memberlog contains successive signatures of different group compositions. In order to alter anything about the memberlog one would need the secret keys of all the members from the phase one wants to alter to the current phase to cooperate.

4 Conclusion

We have given a dynamic (n, n) signature scheme. There is no limit on the amount of members that can join and members can leave till none are left. The secret key pieces do not need to be updated when the group composition changes. In future work one might incorporate threshold cryptography as in Park and Kurosawa [3]. Additionally, the efficiency might be increased, the amount of data communicated is quite high and the OWA implementation requires memory linear to the amount of elements.

References

[1] Helger Lipmaa. Secure accumulators from euclidean rings without trusted setup. In *Applied Cryptography and Network Security*, pages 224–240. Springer, 2012.

[2] Kaisa Nyberg. Fast accumulated hashing. In *Fast Software Encryption*, pages 83–87. Springer, 1996.

[3] Choonsik Park and Kaoru Kurosawa. New elgamal type threshold digital signature scheme. *IEICE transactions on fundamentals of electronics, communications and computer sciences*, 79(1):86–93, 1996.

[4] Jan C.A. van der Lubbe, Merel J. de Boer, Zeki Erkin. A signature scheme for a dynamic coalition (defence) environment without a Trusted Third Party, *Balkan-CryptSec* 2014, Istanbul, Oct. 16-17, 2014 (In: *Lecture Notes in Computer Science*, B. Ors, and B. Preneel (eds.), Springer-Verlag, 2015)

DNA sequence modeling based on context trees

Lieneke Kusters Tanya Ignatenko
Eindhoven University of Technology
Dept. of Electrical Engineering, SPS group
Eindhoven, The Netherlands
c.j.kusters@tue.nl t.ignatenko@tue.nl

Abstract

Genomic sequences contain instructions for protein and cell production. Therefore understanding and identification of biologically and functionally meaningful patterns in DNA sequences is of paramount importance. Modeling of DNA sequences in its turn can help to better understand and identify such patterns and dependencies between them. It is well-known that genomic data contains various regions with distinct functionality and thus also statistical properties. In this work we focus on modeling of such individual regions of distinct functionalities. We apply the concept of context trees to model these DNA regions. Based on the Minimum Description Length principle, we use the estimated compression rate of a genomic region, given such models, as a similarity measure. We show that the constructed model can be used to distinguish specific genes within DNA sequences.

1 Introduction

The human genome contains information about human evolution and physiological properties. The genetic research community put a lot of effort in projects like the human genome project, the 1000 genomes project and the HapMap project, in order to collect, analyze and understand the human genome. These efforts resulted in the human reference genome sequence (that is a general representation of the human genome) and many new insights regarding population evolution, functional properties of the genome, as well as genetically inherited diseases and disease predispositions, and their treatment.
It is known that certain regions of the genome encode for proteins. In these regions, triplets of nucleotides (codons) encode for the amino-acids that together construct a protein of specific shape. Research on automatic detection of protein-coding regions in the genome, includes spectral analysis techniques [1],[2],[3] and Markov models [4],[5]. However, besides the protein coding regions, there are also regions in the genome with other functionalities, such as e.g. regulatory elements (control transcription of a nearby gene). To the best of our knowledge, there exists no general model that can be used to automatically identify and distinguish between various regions of different functionalities within genomic sequences.
It is our goal to construct a generic statistical model for genetic sequences. Since various regions in the genome have different functionality, their statistical properties also differ within the genome. Therefore, as a first step in constructing a generic model, we focus on determining individual models corresponding to the different functional regions in the genome. In [6] it was shown that context trees can be used to model and distinguish between the human chromosomes. We propose to use a similar approach, and construct models that can help discriminate between smaller regions of different functionality.We propose to use context trees [7] to model genetic sequences. We show that the context tree model can be used to distinguish regions of similar statistics within a sequence.

This paper is organized as follows. In the next section we first explain the proposed methods for constructing and evaluating the model. In Section 3 we present our experimental results for modeling of different types of sequences. Finally, we discuss our findings and future work in Section 4.

2 Methodology

In this work we propose to use a two-pass method, to construct the model that we can use for DNA modeling. With the two-pass method, we first construct the maximum a posteriori model corresponding to a given sequence, and then apply the constructed model to estimate the compression rate of a sequence given the model. We use the resulting compression rate as criteria to make a decision whether the sequence was generated by the given model, and thus is functionally similar to the sequence(s) used to estimate this model. In the following, we first summarize the properties of the DNA data. Next, in Sections 2.2 and 2.3 we introduce the context tree model and describe the two-pass method that we use to construct the maximum a posteriori tree model of a sequence. Finally, we describe the application of this two-pass method to DNA modeling.

2.1 DNA Sequences

Human genetic information is encoded in deoxyribonucleic acid (DNA) sequences. The DNA sequence is composed of four different symbols that correspond to the DNA building blocks, called nucleobases, i.e. Adenine (A), Cytosine (C), Guanine (G) and Thymine (T). DNA sequences vary across populations and generations. These variants occur due to mutations and generally occur once per thousand nucleotides in the sequence. Typical genetic variations include *substitution* of one nucleotide for another and *insertion* or *deletion* of a short subsequence of nucleotides.

2.2 Context Tree Model for DNA sequences

A DNA sequence is a string of concatenated quaternary symbols, where each symbol can take on a value from a quaternary alphabet $(A, C, G, T) \in \mathcal{A}$, corresponding to the four different nucleobases. Let a DNA sequence $x_0x_1x_2\ldots x_{N-1}$ of length N be denoted by x_1^N. We assume that the DNA sequence is generated by a tree source. For a tree source the probability $\Pr\{X_t = a\}$ of a symbol X_t in the sequence to take on a value $a \in \mathcal{A}$, is determined by its context, where the context is defined by at most D preceding symbols in the sequence. Such a tree source can be described by a context tree. A context tree is a set of nodes labeled with contexts s with $0 \leq len(s) < D$, and a set of leafs that correspond to the contexts of maximum depth, $len(s) = D$, with $len(s)$ the length of the context s. An example context tree is shown in Figure 1. Given such a context tree, we can determine the probability $\Pr\{X_t = a|x_{t-D}^{t-1}\}$, by starting at the root λ of the tree and moving along the nodes $x_{t-1}, x_{t-2}, \ldots$ until a leaf of the tree is reached. In this leaf, s, we find the corresponding parameter $\Theta_s = \left\{\theta_s^A, \theta_s^C, \theta_s^G, \theta_s^T\right\}$, that are the conditional probabilities of a symbol to take a certain value from the alphabet $\mathcal{A}$, given its context s. Therefore, using the context tree with parameters, we can find the conditional probability of our symbol as $\Pr\{X_t = a|x_{t-D}^{t-1}\} = \theta_s^a$. The suffix set, that represents the leafs of the tree, $\mathbf{S}$ is called the model of the source and the corresponding parameters are stored in its leafs and denoted by $\mathbf{\Theta}$. Furthermore, we define the mapping from the context of depth D, to a suffix s in the model $\mathbf{S}$ as $\omega^{\mathbf{S}}(x_{t-D}^{t-1}) = s \in \mathbf{S}$.

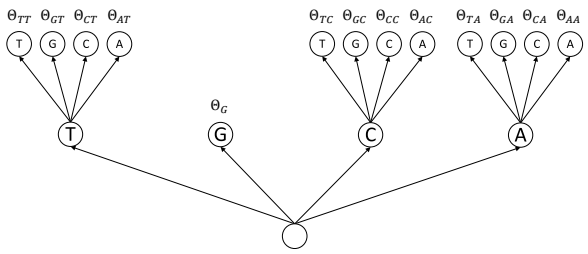


Figure 1: Context Tree $\mathbf{S}$ and model parameters $\mathbf{\Theta}$.

Now, given the example tree model with parameters, in Figure 1, the probability of a subsequence $ACGTC$, in $x_1^N = (\ldots CAA\mathbf{ACGTC}GG\ldots)$, can be estimated as follows:

$$\Pr\{ACGTC|\mathbf{S},\mathbf{\Theta}\} = \theta_{AA}^{A}\theta_{AA}^{C}\theta_{CA}^{G}\theta_{G}^{T}\theta_{TG}^{C} \tag{1}$$

In general we do not know the actual source model that corresponds to the DNA sequence. In [7] the context tree weighting algorithm (CTW) is proposed to estimate the unknown sequence distribution. In CTW they estimate a good coding distribution that can be used to compress data in the sequential way. Instead, we want to find the model that best describes the sequence and evaluate its performance. This can be achieved by the CTW two-pass method [8], which uses the techniques to determine the maximum a posteriori (MAP) model after observing the complete sequence. In [8] this MAP model is first estimated and then used for compression of the sequence.
Here we propose to use this two-pass method to first estimate an optimized statistical model corresponding to a given training sequence. Then we evaluate the model performance, based on the compression rate of a sequence, given this model.

2.3 Maximum a posteriori (MAP) model selection

In this section we summarize the algorithm that is used for the MAP model selection. Our implementation is based on the two-pass method as proposed by Willems *et al.* in [8]. First, we construct a context tree where we assume a maximum depth D. Then, we process the training sequence sequentially, and estimate the probabilities of the subsequences that correspond to each of the contexts s in the tree. Finally, we evaluate the estimated sequence probabilities at different nodes in the tree to find the MAP model, selecting the nodes as either leafs or nodes based on the estimated probabilities.
First of all, for each context that corresponds to a node in the tree, we count the symbols that occur with this context in the training sequence. We store the counts c_s^a that correspond to symbol $a \in \mathcal{A}$, occurring with context $s \in \mathbf{S}$ in the corresponding node. We use the KT-estimator from [9] to estimate the probability of the subsequence P_e^s with context s, given the counts in the corresponding node, as follows:

$$P_e^s\left(c_s^A, c_s^C, c_s^G, c_s^T\right) = \frac{\prod_{a\in\mathcal{A}}(c_s^a - 1/2)!}{((\sum_{a\in\mathcal{A}} c_s^a) - |\mathcal{A}|/2)!}. \tag{2}$$

Where the probability of the next symbol is estimated as,

$$\Pr\{X_t = a|c_s^A, c_s^C, c_s^G, c_s^T\} = \frac{c_s^a + 1/2}{\sum_{a'\in\mathcal{A}} c_s^{a'} + |\mathcal{A}|/2}, \tag{3}$$

for the symbol X_t with value $a \in \mathcal{A}$ and given the counts of corresponding context s.

In the next step, we use the method proposed in [8] to find the maximum a posteriori tree model. That is, we estimate for each node the maximum a posteriori probability of the corresponding subsequence,

$$P_m^s = \begin{cases} \max\left(\alpha \cdot P_e^s, (1-\alpha) \cdot \prod_{a \in \mathcal{A}} P_m^{as}\right) & \text{if depth}(s) < D, \\ P_e^s & \text{otherwise.} \end{cases} \tag{4}$$

Where $\alpha = \frac{|\mathcal{A}|-1}{|\mathcal{A}|}$, is a penalty for the model complexity, increasing with the depth of the tree, see also [9]. We find the nodes corresponding to the MAP model, by tracking the above maximization procedure, starting from the root. If in a node s in the context tree $\alpha \cdot P_e^s \geq (1-\alpha) \cdot \prod_{a \in \mathcal{A}} P_m^{as}$, this node is a leaf in the MAP model and the corresponding context s is added to the model $\mathbf{S}$. Otherwise this node is an internal node in the MAP model and we continue to evaluate the children that are deeper in the tree: $\{As, Cs, Gs, Ts\}$. In this way we find all the leafs corresponding to the MAP model $\mathbf{S}$. Finally, we can compute the parameters $\mathbf{\Theta}$ of our model using equation 3.

2.4 DNA sequence model evaluation

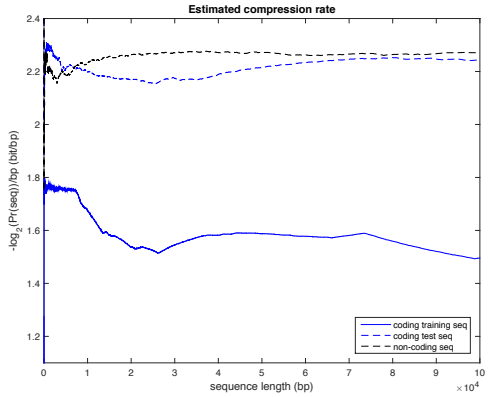
As explained in Section 2.2, the CTW two-pass algorithm for MAP model approximation, was originally developed for compression of the corresponding sequence. However, we would like to apply this model to evaluate or to detect sequences of similar functionality. The Minimal Description Length principle [10], states that the model that describes the data in the shortest possible way is the model that produced the data. Therefore, we use the estimated compression rate of a sequence, given the model, as a measure of the correctness of the model. We can estimate the compression rate, by using the constructed model $\mathbf{S}$ and corresponding probabilities $\mathbf{\Theta}$, to estimate the probability of the sequence given the model. We have shown in Section 2.2 how to estimate the probability of a sequence (or a single symbol) given the model.
We estimate the compression rate of the sequence x_1^t as follows:

$$R(x_1^t) = -\sum_{j=1}^{t} \log_2(\theta^{a_j}_{\omega^{\mathbf{S}}(x_{j-D}^{j-1})})/t, \tag{5}$$

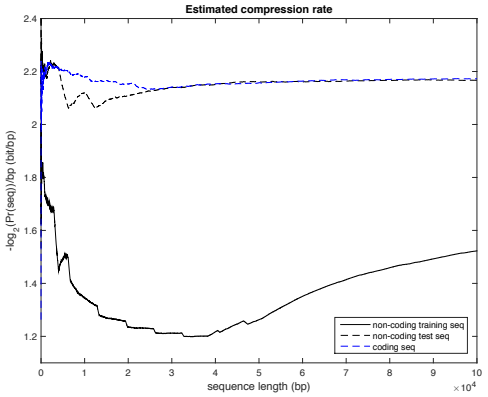
with $\omega^{\mathbf{S}}(x_{j-D}^{j-1}) \in \mathbf{S}$ the context of X_j (a symbol in the sequence) that corresponds with a leaf in the model $\mathbf{S}$, and $\theta^{a_j}_{\omega^{\mathbf{S}}(x_{j-D}^{j-1})}$ the corresponding probability of the symbol X_j having its corresponding value a_j. Furthermore, we can also estimate the contribution of a symbol X_t to the compression rate, as

$$R(X_t) = -\log_2(\Pr\{X_t = a | x_{t-1} \dots x_{t-D}\}) = -\log_2(\theta^{a}_{\omega^{\mathbf{S}}(x_{t-D}^{t-1})}), \tag{6}$$

with $\omega^{\mathbf{S}}(x_{t-D}^{t-1}) \in \mathbf{S}$ is the mapping of the context of X_t to a leaf in the model $\mathbf{S}$, and $\theta^{a}_{\omega^{\mathbf{S}}(x_{t-D}^{t-1})}$ the corresponding probability of the symbol $X_t = a$. Finally, we note that the compression rate is measured in bits per base-pair, which means that for our data, a compression rate smaller than 2, i.e. $\log_2(4)$, corresponds to actual compression of the sequence.



(a) MAP model of coding sequences.



(b) MAP model of non-coding sequences.

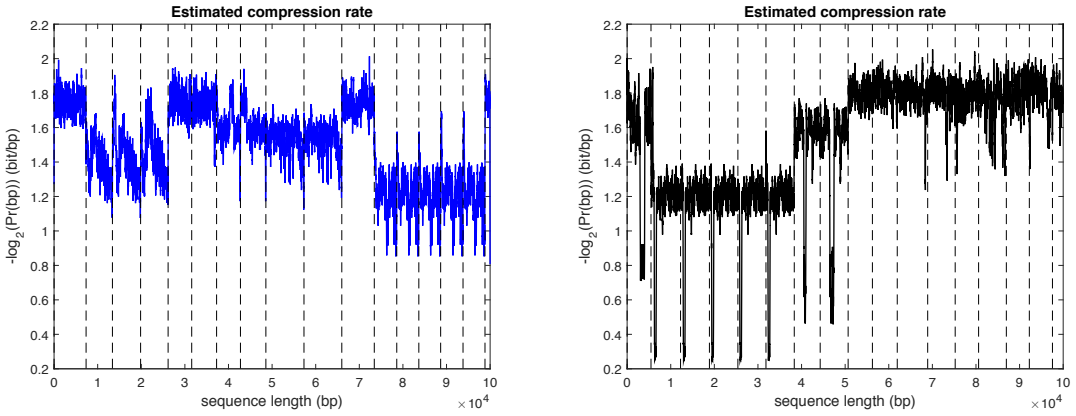
Figure 2: Achievable compression rates for coding and non-coding sequences, using MAP context tree model. Two models were trained, one on coding (2a) and one on non-coding (2a) DNA compound sequences. For both models the performance is shown, when applied to the sequence used for training, when applied to a sequence of similar functionality, and when applied to a sequence of the opposite functionality

3 Experimental results

We evaluate the performance of the maximum a posteriori tree model in two experiments. In each experiment we first use the techniques explained in Section 2.3 to construct the MAP model corresponding to the training sequence. Then we test the performance of the model, by estimating the resulting compression rate for various sequences.

3.1 Coding and non-coding sequences

In the first experiment we construct two MAP tree models for (protein) coding and non-coding sequences respectively. We use a set of sequences from the Human Reference Genome (build: GRCh38.p2) annotated as mRNA (coding) and ncRNA (non-coding) in the NCBI Homo sapiens Annotation Release 107 [11].
First, we construct one coding and one non-coding sequence of 10^5 base-pairs long, by compounding the subsequences of corresponding functionality from the annotated set. We estimate the MAP model for each of the constructed sequences, assuming maximum tree depth 7. We estimate the compression rate of each model on both sequences, the resulting compression rates are shown in Figure 2. We can see that both models have a good compression rate on the sequence that was used for the model training, and they can be used to distinguish between the coding and non-coding training sequences. Therefore, we may conclude that the resulting model provides a good estimate of the source model of the sequences.
Now, we construct two more ('test') sequences in a similar way as before, but from other sets of the annotated ncRNA and mRNA sequences. When we apply the previously constructed coding and non-coding models to the new sequence of corresponding functionality (Figure 2), the compression rate is above 2, which means that the sequences are not compressible at all (see Section 2.4).
In Figure 3, we estimate the compression rate per symbol for the models on their corresponding training sequence. Now we can see, that the rate varies for different regions in the sequence. We conclude that, though the constructed models do have a



(a) MAP model of coding sequences.

(b) MAP model of non-coding sequences.

Figure 3: Achievable compression rates per symbol, for coding (3a) and non-coding (2a) sequences using MAP context tree model. The dotted lines mark the regions corresponding to different subsequences that were used to construct the total sequence.

sufficient overall performance to represent the entire sequence (Figure 2), the model varies for different regions in the sequences and the overall model is actually a mixture of models. Furthermore, we find that the variations in the compression rate are related to the transitions between the subsequences (marked by the dotted lines) that jointly form the total sequence.

3.2 MAP model for gene in mitochondrion

Now we concentrate on construction of the model for a sequence with a more specific functionality than just coding or non-coding functionality. In this experiment we would like to detect COX1 gene in the mitochondrial DNA and use our model to detect the gene in a set of mitochondrial DNA variant sequences.
For this experiment we have used a set of mitochondrial DNA sequences from 20 individuals of various ethnicity (America, Africa, Europe, Asia)*. Between those sequences small variations occur in the form of substitutions, insertions and deletions of nucleotides (see also Section 2.1). We use the sequence from two persons to construct the MAP model, with initial context tree depth 5, for the COX1 gene (approx. 1500 bps long). Then we evaluate the performance of the model on the sequences that correspond to the other 18 individuals. The estimated compression rate per symbol is shown in Figure 4. We observe a very good compression rate of the subsequence corresponding to the COX1 gene, when the learned COX1 model is used for mitochondrion compression. On the other hand, in the other regions no compression is achieved. Therefore, we can clearly distinguish the COX1 gene in the sequences, using this model. Furthermore, the model is generic in the sense that its performance is similar for all sequences, despite the small variations that occur.

4 Discussion and Future Work

In this study we have shown that the context tree can be used to model the statistics of DNA sequences. Though a model can be constructed to represent sequences of variable length and functionality, it is not clear whether the model also implies information

*Sequences were downloaded from the mitochondrion database [12]

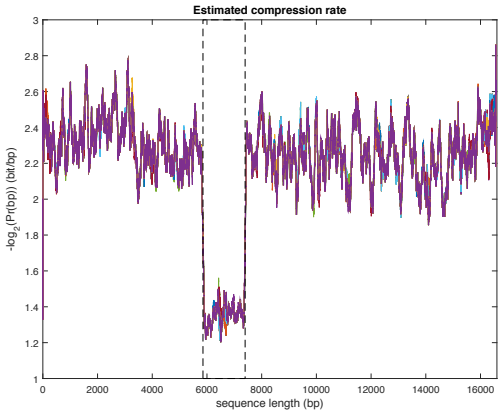


Figure 4: MAP model applied to detect COX1 gene in mitochondrion. The region corresponding to the COX1 gene is marked between the dotted lines.

about the functionality of the modeled sequence. The model can be used to recognize sequences that have similar statistics to the original sequence. Besides functionality analysis, other applications of such a model include read mapping and genome compression.
In this work we assumed that non-coding and coding regions in DNA sequences are stationary. However, our experiments imply that these regions have non-stationary statistics. As a future work we plan to develop an algorithm that automatically recognizes a change in the model and constructs multiple models to accurately represent the different regions in the source-sequence. These models can give more insight in the statistics of the different regions and can be related to the functionality of the region. As a final remark we state that the strength of the context tree model for DNA sequences, is that it has low sensitivity to variations in the sequence. We plan to further explore this property for application in privacy-sensitive modeling of DNA sequences, since variations are hidden in the model.

5 Acknowledgment

This work has been funded by the EC via grant agreement no. 611659 for the AU2EU project.

References

[1] M. Roy and S. Barman, "Effective gene prediction by high resolution frequency estimator based on least-norm solution technique," *EURASIP journal on bioinformatics & systems biology*, vol. 2014, no. 2, 2014.

[2] S. A. Marhon and S. C. Kremer, "Gene prediction based on DNA spectral analysis: A literature review," *Journal of Computational Biology*, vol. 18, no. 4, pp. 639–676, 2011.

[3] D. Kotlar and Y. Lavner, "Gene prediction by spectral rotation measure: A new method for identifying protein-coding regions," *Genome Research*, vol. 13, no. 8, pp. 1930–1937, 2003.

[4] A. M. Gupal and A. V. Ostrovsky, "Using compositions of Markov models to determine functional gene fragments," *Cybernetics and Systems Analysis*, vol. 49, no. 5, pp. 692–698, 2013.

[5] M. Stanke and S. Waack, "Gene prediction with a hidden Markov model and a new intron submodel," in *Bioinformatics*, vol. 19, no. Suppl. 2, 2003, pp. 215–225.

[6] T. Ignatenko and M. Petković, "AU2EU: Privacy-Preserving Matching of DNA Sequences," in *Information Security Theory and Practice. Securing the Internet of Things (WISTP 2014 Proceedings)*. Springer Berlin Heidelberg, 2014, pp. 180–189.

[7] F. Willems, Y. Shtarkov, and T. Tjalkens, "The Context-Tree Weighting Method: Basic Properties," *IEEE Transactions on Information Theory*, vol. 41, no. 3, pp. 653–664, May 1995.

[8] F. M. J. Willems, A. Nowbakht, and P. A. J. Volf, "Maximum a posteriori probability tree models," in *Proceedings of the 4th International ITG Conference on Source and Channel Coding*, Berlin, Germany, 2002, pp. 335–340.

[9] T. J. Tjalkens, Y. M. Shtarkov, and F. M. Willems, "Sequential weighting algorithms for multi-alphabet sources," 1993, pp. 22–27.

[10] J. Rissanen, "Modeling by shortest data description," pp. 465–471, 1978.

[11] *The NCBI handbook [Internet]*. Bethesda (MD): National Library of Medicine (US), National Center for Biotechnology Information, 2002.

[12] M. Ingman and U. Gyllensten, "mtDB: Human Mitochondrial Genome Database, a resource for population genetics and medical sciences," *Nucleic Acids Res*, vol. 34, pp. D749–D751, 2006.

Adaptive Channel Selection and Sensing based on Reinforcement Learning

Sreeraj Rajendran Sofie Pollin
KU Leuven
ESAT, TELEMIC
Kasteelpark Arenberg, 10 2444
sreeraj.rajendran@esat.kuleuven.be sofie.pollin@esat.kuleuven.be

Abstract

The rapid growth in the number of wireless devices like smart-phones, tablets and wireless sensors have resulted in scarcity of available spectral resources. As the availability of spectrum is limited, a huge interest is spawned in co-existence solutions where new users can share the available spectrum along with the legacy systems even in licensed bands. This demands development of new algorithms which can provide satisfactory throughput to secondary users with less interference to the legacy systems (Primary User). Open spectrum access is still an active research problem. New technical standards should be established which will account for transmitter and receiver based regulations instead of legacy transmitter centric spectrum regulation. A Spectrum Sharing Challenge [1] is organized by IEEE keeping these motivations in mind. In this paper a reinforcement learning based algorithm is proposed which will serve as a starting point for approaching the problem statement mentioned in the Spectrum Sharing challenge. The proposed algorithm does a single channel selection and adapts its transmission and sensing actions. Simulation results revealed that the proposed algorithm yields better throughput results over the Upper Confidence Bound [2] based strategies.

1 Introduction

Spectrum scarcity is a serious issue faced while devising new communication protocols and standards. The next generation wireless systems will possibly be equipped with algorithms which can make use of underutilized licensed spectrum assigned to a Primary User (PU). This allows coexistence of a Secondary User (SU) in a frequency band which is licensed to a PU. Many studies have revealed that existing wireless systems, for example 802.15.4 and 802.11, create interference resulting in reduced throughput when operated on the same frequency bands [3]. A survey of various schemes appeared in literature for solving coexistence issues are presented in [4]. The IEEE Spectrum Sharing Challenge is hosted to motivate researchers and developers to come up with innovative solutions for coexistence which will help in devising transmitter and receiver based spectrum usage restrictions instead of providing licensed spectrum explicitly.

There are various surveys [5,6] which briefs various machine learning techniques for dynamic spectrum sharing solutions. Some techniques make use of Reinforcement learning [8] under the assumption that the SU-PU interaction forms a Markov Decision Process (MDP). Other schemes studied in literature take care of scenarios when the assumption is non-Markovian [5]. There are also ad-hoc methods where the problem is modelled as a Multi-Armed Bandit problem and solutions are presented using strategies based on Upper Confidence Bound [2].This paper details the problem stated in the IEEE DySPAN 2015 Spectrum Sharing Challenge and discusses some of the approaches that can be used to solve the problem.

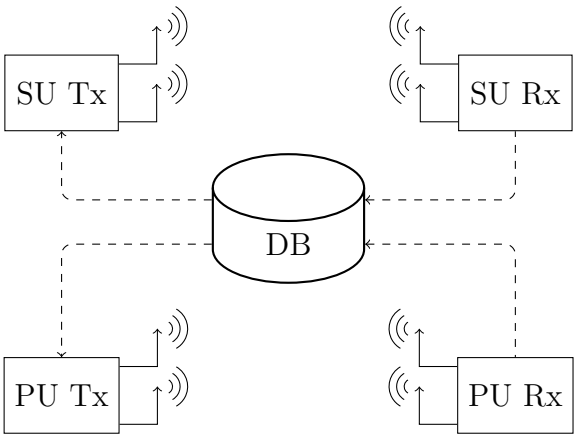


Figure 1: Spectrum challenge setup

The rest of the paper is organized as follows. Section 2 defines the problem statement and the winning parameters. A brief overview of the reinforcement learning framework is presented along with the proposed method is Section 3. Section 4 details the initial results from the proposed algorithm and comparisons with one of the algorithms mentioned in literature. Shortcomings of the proposed algorithm, possible enhancements and conclusions are presented in Section 5.

2 Problem Definition

The 5G spectrum challenge setup is detailed in Figure 1. A multi-channel radio which will implement IEEE 802.15.4 PHY will be used as the primary user (PU). The PU will be transmitting on four predefined frequency bands each of bandwidth 2 MHz with channel spacing of 5 MHz. Another pair of radios which act as the SU will try to achieve maximum throughput over the same 20 MHz which the PU is using. Both pair of radios will be connected to the database (DB) which is responsible for providing

- Fixed length layer-2 packets for SU
- PU receiver feedback
- Performance monitoring

The final challenge score will be calculated as a product of SU throughput (T_{SU}) and PU satisfaction. The PU satisfaction (S_{PU}) is calculated from the offered PU throughput ($\widehat{T}_{PU}$) and the delivered PU throughput (T_{PU}) as given in equation 1. A maximum throughput loss tolerance of 10% is admitted. More than 10% PU throughput loss will result in no PU satisfaction at all.

$$\begin{aligned} Score &= T_{SU} \times S_{PU} \\ S_{PU} &= max\Big(0, \frac{10}{9} T_{PU} - \widehat{T}_{PU}\Big) \end{aligned} \tag{1}$$

3 Reinforcement learning based score maximization

For maximizing the score defined in the problem statement, only single channel selection algorithms are explored in this paper. That is, the SU will only try use only one

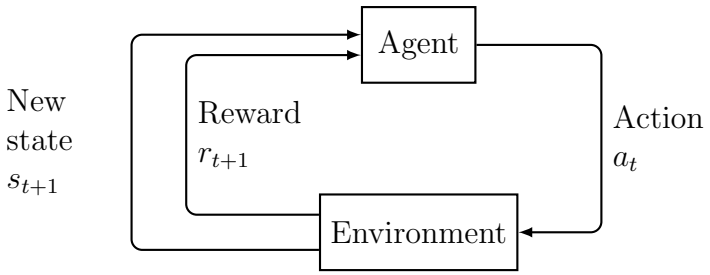


Figure 2: Reinforcement Learning framework

frequency band of 2 MHz at a time which is best suited for maximizing the defined score. In a multi-channel scenario, if the secondary user has information about the channel occupancy of each channel in advance, it is always best to select the channel with lower average occupancy which will effectively maximise the SU throughput. As the channel occupancy probabilities of the channels are typically unknown, the SU should explore the channels to estimate the channel occupancies. The more the exploration, the better will be the channel occupancy estimate. In the meantime SU should also exploit the channel to maximize its throughput and cannot keep on exploring the channel. This will result in an exploration-exploitation trade-off.

The aforementioned channel access problem is closely related to multi-armed bandit problems which are discussed in detail in literature [7]. There exists various formally justified techniques presented in literature which solve the bandit problem making the assumption that the PU-SU interaction forms an MDP [8,9]. In [2] authors use Upper Confidence Bound (UCB) based strategies to exploit the spectrum resources intelligently. A real world implementation of the some of these algorithms can be found in [10]. If the prefect system model is known in advance, then these kind of objective maximization problems can be solved using dynamic programming [8,11]. Since we have no priori information about the channel occupancies reinforcement learning framework can be used to approach the problem. Reinforcement Learning allows unsupervised learning by interacting with the environment.

In a reinforcement learning model, an agent interacts with the environment through actions as depicted in Figure 2. The action will change the state of the environment which is communicated to the agent as a delayed reinforcement or reward. A general reinforcement learning based framework consists of

- A discrete set of states, $\mathbb{S}$
- A discrete set of actions, $\mathbb{A}$
- A policy π that maximizes the expected reward

The agent's job is to come up with a policy that maximizes some measure of reinforcement. For example in the SU channel selection problem, a reinforcement framework can be used to select the best channel which maximises the SU throughput. The throughput feedback from the secondary user receiver can be used as a reinforcement, which tells how good or bad the last channel selection action was.

3.1 Q-Learning

Watkins' Q-learning algorithm [12], one of the most popular model free algorithms that learns from delayed rewards, is used to model the spectrum challenge problem. The basic QL update equation is given as

$$Q_{t+1}(s,a_t) = Q_t(s,a_t) + \alpha \Big(r(s,a_t) + \gamma \max_a Q_t(s,a) - Q_t(s,a_t) \Big) \tag{2}$$

where α is the learning rate and γ is the discount factor. The Q-value $Q(s,a)$ is the expected discounted reinforcement for taking an action a in state s and then continuing to act optimally by selecting suitable actions. The higher the value of α, the greater the agent relies on the delayed and future rewards when compared to the current Q value. Similarly the higher the value of γ the greater the agent relies on the discounted future reward compared to the delayed reward.

3.2 Proposed Method

The algorithm model proposed in this section is motivated from the algorithm presented in [13] where the authors use multiple channels as the states, and sensing, transmitting and channel switching as the actions. The proposed scheme tries to maximize the score mentioned in challenge by

- Selecting the most reliable channel for transmission
- Adapting the number of sensing and transmission actions in the selected channel

The action and state space of the RL algorithm are selected as given below.

- Action set: $\{sense, transmit, channel\_switch\}$
- States: $\{0, .., n\}$ where n is the number of available channels

The QL update equations for each action, the corresponding rewards and the policy selection are explained below.

3.2.1 Expected sensing reward: $Q(s, a_{se})$

$$Q_{t+1}(s,a_{se}) = Q_t(s,a_{se}) + \alpha \Big(r(s,a_{se}) + \gamma \max_a Q_t(s,a) - Q_t(s,a_{se}) \Big) \tag{3}$$

$$r(s,a_{se}) = \begin{cases} 0 & \text{if channel is occupied} \\ 1 & \text{if channel is free} \end{cases}$$

3.2.2 Expected transmission reward: $Q(s, a_{tx})$

To maximize the score mentioned in the problem statement, the algorithm should increase secondary throughput and primary user satisfaction. A reasonable assumption is made that the reduction is primary throughput is only caused by secondary user collisions. Based on the previous assumption we assign positive rewards based on the instantaneous secondary throughput (T_{SU}) and negative rewards (penalty) based on the collision count (T_{CO}) which will in turn maximize the score,

$$Q_{t+1}(s,a_{tx}) = Q_t(s,a_{tx}) + \alpha \Big(r(s,a_{tx}) + \gamma \max_a Q_t(s,a) - Q_t(s,a_{tx}) \Big) \tag{4}$$

$$r(s, a_{tx}) = T_{SU} - T_{CO}.$$

. For example if the secondary user is transmitting 1 bit each in 4 slots and if 3 of them resulted in collision, the reward is $1 - 3 = -2$

3.2.3 Expected switching reward: $Q(s, a_{cs})$

A state value $V(s)$ is defined as

$$V(s) = Q_t(s, a_{se}) + Q_t(s, a_{tx}). \tag{5}$$

If a switching action is selected then the secondary user switches from state s to $\widehat{s}$ such that

$$\widehat{s} = \arg\max_{h \in S} V(h) \tag{6}$$

The Q value for switching is calculated as a gain in terms of $V(s)$, as

$$Q_{t+1}(s, a_{cs}) = V(\widehat{s}) - V(s). \tag{7}$$

.

A higher $Q(s, a_{cs})$ for a state s indicates that there is another good state (channel) which can give higher transmission gains.

3.2.4 Action selection policy: $\pi_t(s, a)$

At any time t the SU action in a state (channel) s an is selected based on a *soft-max* selection policy. Here the parameter τ is called the temperature. High temperatures makes the selection of all actions to be nearly equi-probable. The soft-max selection becomes the same as the greedy selection as $\tau \to 0$.

$$\pi_t(s, a) = \frac{e^{Q_t(s,a)/\tau}}{e^{Q_t(s,a_{se})/\tau} + e^{Q_t(s,a_{tx})/\tau} + e^{Q_t(s,a_{cs})/\tau}} \tag{8}$$

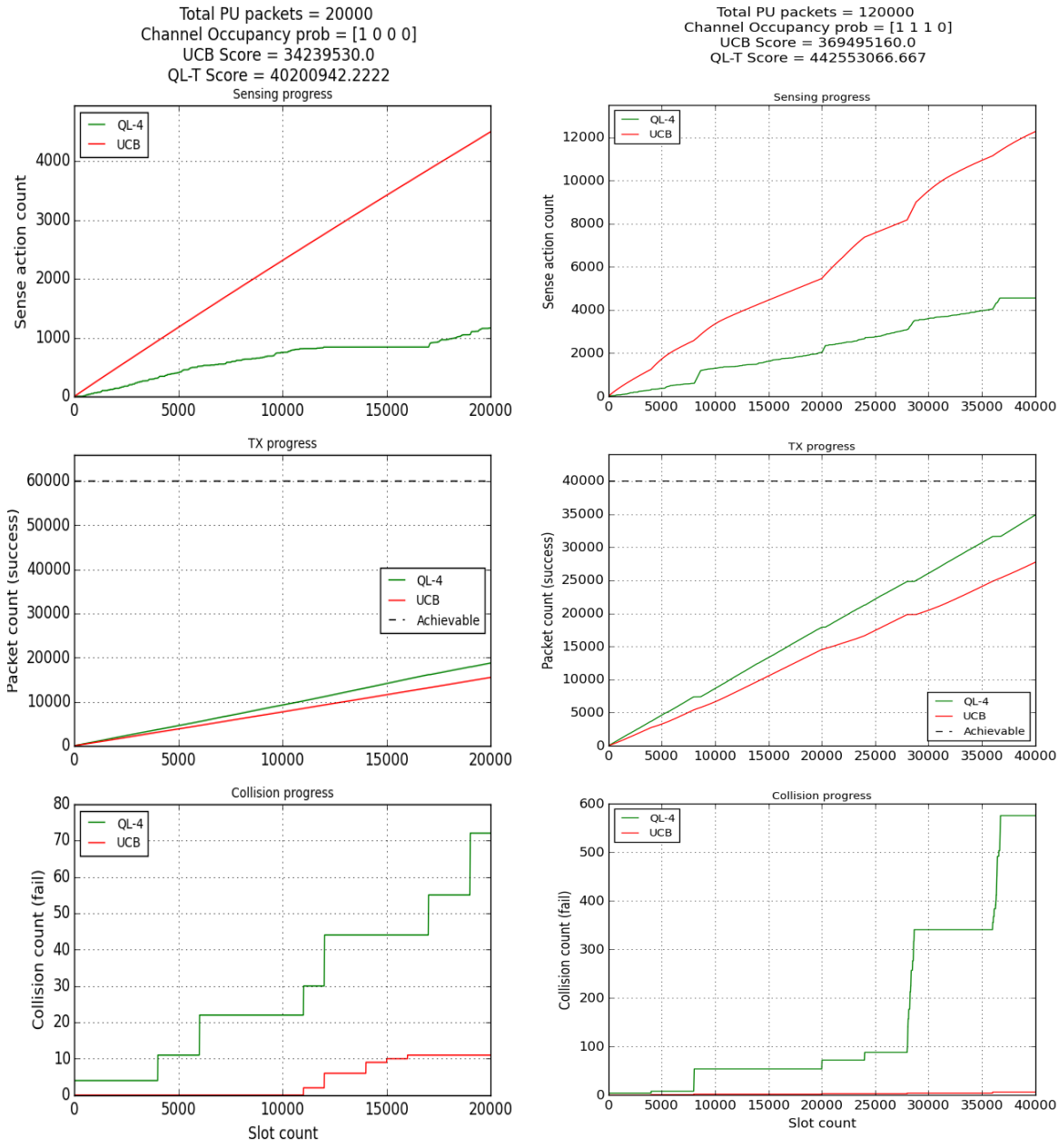
4 Experimental Results and Discussion

This section details some initial simulation results of the proposed algorithm. For simulations, the entire available time duration is split into slots, each having a slot duration of T_{slot}. An assumption is made that the channel switching time, $T_{cs} << T_{slot}$ and no time penalty is assigned for channel switching. For the QL algorithm, parameter values $\alpha = 0.8$, $\gamma = 0.1$, and $\tau = 1$ are used. For all the mentioned experiments a transmission duration, $T_{tx} = 4 \times T_{slot}$ and sensing duration $T_{se} = T_{slot}$ is used.

1. Exp1: A frequency hopping PU system is considered in this experiment. After every $1000 \times T_{slot}$ the primary user will hop to a random channel. Figures 3a and 4a show the simulation progress and action statistics respectively.

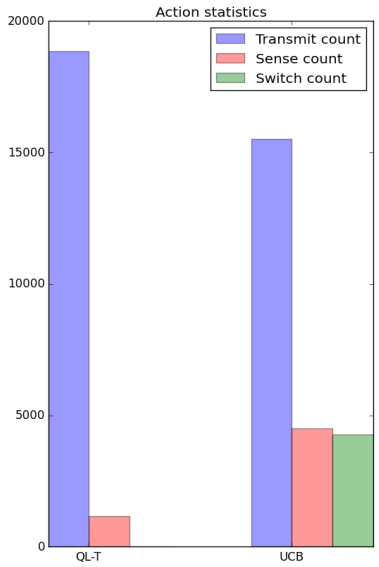
2. Exp2: A simulation where only a single channel is free at a time. After every $4000 \times T_{slot}$, a random channel is made free. Figures 3b and 4b show the simulation progress and action statistics respectively.

From the initial simulation results its clear that the proposed algorithm provides much improvement in the secondary user throughput which in-turn increases the spectrum challenge score. The improvement in throughput is due to the fact that QL algorithm is able to reduce its sensing actions when the channel is free and transmission action is giving positive rewards. It can be also noticed that the channel switching action is considerably reduced when compared to the UCB algorithm.

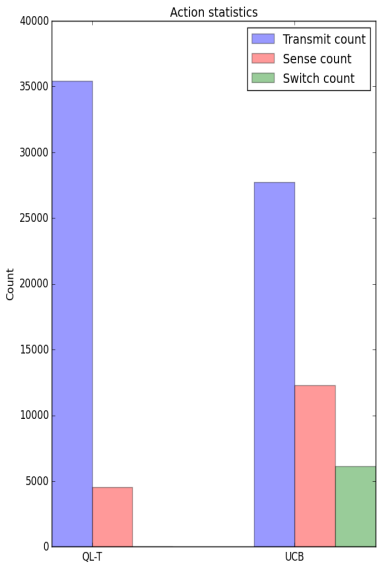


(a) Exp1: Frequency hopping scheme (b) Exp2: Single free channel scenario

Figure 3: Sensing, transmission and collision progress



(a) Exp1: Frequency hopping scheme



(b) Exp2: Single free channel scenario

Figure 4: Transmit, sense and channel switch action count

5 Conclusion

In this paper a RL based approach is presented to solve the Spectrum Sharing Challenge problem. An adaptive scheme is developed which yields better throughput when compared to the UCB strategies. It should be noted that the PU throughput is not yet considered in the QL algorithm. The PU throughput feedback will be crucial when PU is trying to transmit in a channel where SU transmission is in progress. The PU PHY will be equipped with some carrier sense mechanism which will prevent it from transmitting in the SU occupied channel. This will result in a reduced PU throughput and satisfaction. The proposed algorithm considers only single channel selection. This is obviously not an ideal solution for a frequency hopping PU as only one channel will be occupied by the PU at a time. These two directions will be investigated in future which may significantly improve the algorithm performance.

References

[1] IEEE DySPAN 2015 Spectrum Sharing Challenge, *http://dyspan2015.ieee-dyspan.org/content/5g-spectrum-sharing-challenge*

[2] W. Jouini, D. Ernst, C. Moy, and J. Palicot, "Upper confidence bound based decision making strategies and dynamic spectrum access," in *Proceedings of the IEEE International Conference on Communications (ICC 10)*, pp. 15, May 2010

[3] Sofie Pollin, Ian Tan, Bill Hodge, Carl Chun, Ahmad Bahai, "Harmful Coexistence Between 802.15.4 and 802.11: A Measurement-based Study" in *3rd International Conference on Cognitive Radio Oriented Wireless Networks and Communications*,May 2008

[4] Dong Yang, Youzhi Xu, and Mikael Gidlund, "Wireless Coexistence between IEEE 802.11- and IEEE 802.15.4-Based Networks: A Survey",in *International Journal*

of Distributed Sensor Networks, vol. 2011, Article ID 912152, 17 pages, 2011. doi:10.1155/2011/912152

[5] Mario Bkassiny, Yang Li and Sudharman K. Jayaweera, "A Survey on Machine Learning Techniques in Cognitive Radios" in *Communications Surveys & Tutorials, IEEE (Volum:15, Issue:3)*, pp. 1136 - 1159 , October 2012

[6] Kok-Lim Alvin Yau, Geong-Sen Poh, Su Fong Chien, and Hasan A. A. Al-Rawi, "Application of Reinforcement Learning in Cognitive Radio Networks: Models and Algorithms," in *The Scientific World Journal*, vol. 2014, Article ID 209810, 23 pages, 2014. doi:10.1155/2014/209810

[7] J. C. Gittins, "Bandit Processes and Dynamic Allocation Indices," in *ournal of the Royal Statistical Society*, Vol. 41, No. 2 (1979), pp. 148-177

[8] Leslie Pack Kaelbling and Michael L. Littman and Andrew W. Moore, "Reinforcement Learning: A Survey" in *Journal of Artificial Intelligence Research*,vol. 4, pages 237-285, 1996

[9] Ali Motamedi, and Sofie Pollin "Optimal Dynamic Channel Selection and Application to Open Spectrum Sharing" in *IEEE DySPAN*, March 2007

[10] Christophe Moy, "Reinforcement Learning Real Experiments for Opportunistic Spectrum Access" in *Eighth Karlsruhe Workshop on Software Radios*, March 2014

[11] K. Zheng and H. Li, "Achieving energy efficiency via drowsy transmission in cognitive radio," in *Proceedings of the 53rd IEEE Global Communications Conference (GLOBECOM 10)*, pp. 1-6, December 2010

[12] Richard Sutton and Andrew Barto, "Reinforcement Learning: An Introduction", *MIT Press*, 1998.

[13] Marco Di Felice, Kaushik Roy Chowdhury, Andreas Kassler and Luciano Bononi, "Adaptive Sensing Scheduling and Spectrum Selection" in *Proceedings of the 20th International Conference on Computer Communications and Networks (ICCCN 11)*, pp. 16, August 2011

From Minimal Distortion to Good Characterization: Perceptual Utility in Privacy-Preserving Data Publishing (Extended Abstract)

Raphaël Peschi, François-Xavier Standaert & Vincent Blondel
Université catholique de Louvain, ICTEAM Institute, Louvain-la-Neuve, Belgium.

Introduction

The collection of digital information by organizations has been an increasingly important trend over the last decade. While it creates new opportunities for knowledge-based decision making, it also raises new challenges regarding the privacy of the individuals whose personal information is collected. In this context, privacy-preserving data publishing aims at releasing data in a way that it is practically useful (e.g. allows data mining) while preserving individual privacy. In other words, it aims at trading utility and privacy. A wide literature has investigated metrics for privacy, including the well established k-anonymity [8] and its numerous refinements. By contrast, metrics for measuring the utility of a database are sparser, and generally face the difficulty of defining what is "useful data". In order to be independent of the type of data processing purposed, the typical solution is to follow the "principle of minimal distortion". This implies assuming that the database is useful anyway, and measuring a (pseudo) utility based on this a priori, by quantifying the damage caused by the anonymization of the data [5]. Quite naturally, it also means that any modification of the data is damaging by definition/assumption. In this paper, we aim to investigate an alternative track for measuring utility, based on recent advances in the certification of the information leakage in cryptographic implementations [4]. More precisely, we propose to quantify utility based on whether the statistical attributes of which the samples form a database are "well characterized". We further describe how to use the notion of Perceived Information (PI) for this purpose. Intuitively, the PI captures the amount of information that an adversary can extract from some observations, given a (possibly biased) model of the data. If the model is perfect, the PI correspond to Shannon's classical definition of Mutual Information (MI). If the model is imperfect (as usually the case in practice), the PI is the best approximation of the MI that is available to the adversary. Based on this notion and using the tools in [4], we can trade the speed of convergence of a model with its informativeness, and derive a perceptual utility metric for actual databases. We use this metric to illustrate concrete situations where the anonymization of the data does not have any utility cost. For example, the accuracy of an attribute's observations can be too high for being characterized with the number of available samples. In this case, reducing the accuracy of the collected data is beneficial to individual privacy, but does not reduce the perceptual utility. We also describe a couple of experiments that allow us to discuss the impact of grouping users from the k-anonymity and perceptual utility points-of-view, as well as the curse of dimensionality for the characterization of attributes. Eventually, we conclude the paper by discussing why perceptual utility could also be seen as a privacy metric (although not an anonymity one).

1 Definitions and framework

In this first section, we introduce the definitions and mathematical framework that allow us to reason formally about privacy and utility in databases. We use capital letters for random variables, small caps for their samples, calligraphic letters for sets and sans

serif fonts for functions. We start by defining a set of n *users* $\mathcal{U} = \{u_1, u_2, \ldots, u_n\}$, and m random variables $X_1, X_2, \ldots, X_m$ with (discrete or continuous) sample spaces $\mathcal{X}_1, \mathcal{X}_2, \ldots, \mathcal{X}_m$. We denote these random variables as *attributes*, that are specified by their probability function $\mathsf{p}(x)$ in the discrete case, and probability density function $\mathsf{f}(x)$ in the continuous case. We also use the more generic term probability distribution to denote both $\mathsf{p}(x)$ and $\mathsf{f}(x)$, when we do not want to distinguish between discrete and continuous random variables.

Deterministic data. In this context, we define a deterministic *Data Structure* (DS) as a set of m attributes together with their probability distributions $\mathsf{p}_i(x)$ or $\mathsf{f}_i(x)$, with $1 \leq i \leq m$. And we define a deterministic *Data Base* (DB) as the sampling of a deterministic DS, i.e. a set of $m \times n$ samples, one per attribute X_i and user u_j.

Probabilistic data. In this context, we define a probabilistic DS as a set of m attributes, each of them described by n probability distributions $\mathsf{p}_i^j(x)$ or $\mathsf{f}_i^j(x)$, with $1 \leq j \leq n$. And we define a probabilistic DB as the sampling of a probabilistic DS, where we denote the number fo samples per user as $m \times n_{u_j}$, so that the total number of samples in the DB equals $m \times \sum_{j=1}^{n} n_{u_j}$ (similarly to the deterministic case).

Both for deterministic and probabilistic data, we further denote any vector of m samples corresponding to a line of a DB (excluding the user) as an *observation*. We also define any subset of users that one may be interested to characterize as a *group*, and a set of q groups as $\mathcal{G} = \{g_1, g_2, \ldots, g_q\}$. Eventually, we call aggregation the process of replacing an attribute X_i by an *aggregated attribute* Y_i, such that the original sample space $\mathcal{X}_i$ is replaced by a set of events $\mathcal{Y}_i$, with $|\mathcal{Y}_i| < |\mathcal{X}_i|$ if the attribute was discrete, and $\mathcal{Y}_i$ a discretized version of $\mathcal{X}_i$ if the attribute was continuous. Note that in concrete case studies, the DS is always unknown, and the only thing that can be analyzed is the DB (i.e. a sampling of the DS). However, it is sometimes interesting to consider *simulated* DB where the samples are produced according to known distributions. The latter context typically allows a better theoretical understanding.

2 Privacy metric(s)

As mentioned in introduction, numerous metrics were introduced to quantify various aspects of privacy in databases, some of them being surveyed in [5]. Our goal is not to argue about which metric to use in which context. We rather recall the popular k-anonymity[8], in order to evaluate and discuss it in front of the utility notion in the next section. We first denote an observation o_j^k as the vector of k^{th} samples obtained for the m attributes of user u_j as:

$$o_j^k =: [x_1(u_j, k), x_2(u_j, k), \ldots, x_m(u_j, k)]\cdot$$

Secondly, we denote the set of observations $\mathcal{O}(u_j)$ found in a DB for a user u_j as:

$$\mathcal{O}(u_j) =: \{o_j^k \mid 1 \leq k \leq n_{u_j}\}\cdot$$

Thirdly, we denote the anonymity set $\mathcal{A}(o)$ as the set of users for which a given observation o is in the DB as:

$$\mathcal{A}(o) =: \{u_j \mid o \in \mathcal{O}(u_j)\}\cdot$$

Based on these notations, we say that a DB preserves k-anonymity if:

$$k =: \min_{\substack{1 \leq j \leq n \\ 1 \leq k \leq n_{u_j}}} |\mathcal{A}(o_j^k)|\cdot$$

Intuitively, the k-anonymity guarantees that an observation does not allow to (strictly) distinguish (i.e. with probability one) a user from at least $k-1$ other users in the DB.

Concretely, this metric is usually computed with respect to deterministic attributes that are supposed to be easier to collect for the adversary (e.g. the sex, ZIP code, ...) in order to obtain sensitive information (e.g. some medical data). Note that in probabilistic DB, the k-anonymity ignores the possibility that different users have different probabilities given an observation. By denoting the number of apparitions of an observation o in a DB as $\#o$ and its number of apparitions for user u_j as $\#o|u_j$, we can then define the *pseudo-probability* of a user u_j given an observation o as:

$$\tilde{\Pr}[U = u_j | O = o] =: \frac{\#o|u_j}{\#o}.$$

The term pseudo-probability reflects the fact that $\tilde{\Pr}[u_j|o]$ is defined based the sampled data of a DB, which does not mandatorily represents well the true distribution of the attributes. Such pseudo-probabilities could be used to estimate other anonymity metrics, such as the ones in [2]. Although our following discussions will focus on k-anonymity for quantifying privacy, we will use this notion of pseudo-probability in order to illustrate the conceptual differences between anonymity and (perceptual) utility.

3 Perceptual utility metric

Approaches to guarantee privacy in DB generally imply a number of anonymization operations, which include aggregation, noise addition, suppression, ... This leads to the complementary problem of determining if the sanitized data remains useful.

Both general purpose and specific metrics have been introduced to answer this question [1]. On the one hand, general purpose metrics rely on the goal of minimal distortion. That is, they start from the a priori that the DB is useful, and quantify utility by measuring the distance between the original and anonymized DB. To some extent, this approach resembles the one to quantify privacy in the previous section, since it is also based on the sampled data of a DB, independent of its DS. We use the term *pseudo-utility* to reflect this fact. Minimizing distortion does not guarantee that an anonymized DB is useful, it only guarantees that it is nearly as useful as originally. On the other hand, specific metrics aim at measuring utility based on the purpose of the data collected (e.g. estimating some statistical moment for an attribute, or classifying users based on this attribute). Compared to the previous case, such an approach suffers from the complementary drawback that it hardly allows comparing the utility of data collected for different purposes, and therefore requires knowing this purpose at the time the data is published. Strictly speaking, this last drawback is unavoidable: utility is indeed most accurately defined in function of a task to perform. However, we argue next that an intermediate path is possible, by quantifying (perceptual) utility based on whether the data collected represents well the DS (i.e. the true distribution of the attributes). We first define the perceived information metric that we will use for this purpose, and then provide the rationale behind our new approach.

3.1 The Perceived Information

The Perceived Information (PI) was introduced in the context of side-channel attacks against cryptographic devices, of which the goal is to recover some secret data (aka key) given some physical leakage [7, 10]. The PI aims at quantifying the amount of information about the secret key, independent of the adversary who will exploit this information. Informally, we will use this metric in a similar way, by just considering users as the data to recover, and observations as leakages. Using the previous notations, we can first define the Mutual Information (MI) between the users random variable U and the observation random variable O:

$$\mathrm{MI}(U; O) = \mathrm{H}[U] + \sum_u \Pr[u] \cdot \sum_o \mathsf{p}(o|u) \cdot \log_2 \Pr[u|o],$$

if the observations are discrete, and:

$$\mathrm{MI}(U;O) = \mathrm{H}[U] + \sum_u \Pr[u] \cdot \int \mathsf{f}(o|u) \cdot \log_2 \Pr[u|o]\ do,$$

if they are continuous. For conciseness, we use the notation $\Pr[X = x] =: \Pr[x]$ whenever clear from the context. In these equations, the last probability $\Pr[u|o]$ is derived via Bayes' theorem, e.g. $\Pr[u|o] = \frac{\mathsf{f}(o|u)}{\sum_{u^*} \mathsf{f}(o|u^*)}$ for the continuous case, and $\mathrm{H}[U]$ is computed based on the a priori distribution of the users (e.g. $\mathrm{H}[U] = \log_2(n)$ if it is uniform). Concretely though, and as previously discussed, the true distribution of the attributes is always unknown. Therefore, it is not possible to compute the MI directly (excepted in the case of simulated DB). In order to avoid this caveat, the approach in side-channel analysis, that we repeat here, is to split the DB that one wishes to evaluate in two parts: the first one, denoted as $\mathrm{DB_l}$ is used for learning a model, the second one, denoted as $\mathrm{DB_t}$ is used to test it.* The PI is then computed in two phases:

1. A probabilistic model $\hat{\mathsf{p}}^j_{\mathrm{model}}$ (resp. $\hat{\mathsf{f}}^j_{\mathrm{model}}$) is estimated for each user u_j, which we denote with the conditional distribution $\hat{\mathsf{p}}^j_{\mathrm{model}}(o|u_j) \leftarrow \mathrm{DB_l}$ (resp. $\hat{\mathsf{f}}^j_{\mathrm{model}}(o|u_j) \leftarrow \mathrm{DB_l}$). Note that in the discrete case, such a model can be quite close to the previously defined pseudo-probabilities. The main conceptual difference is that this model is only built from a (learning) part of the DB that will be tested on independent observations (in the second phase below), and can be "simplified" (see, e.g. the example in Subsection 4.4 taking advantage of an independence assumption). In the continuous case, differences are generally more explicit, since the model will be based on a continuous distribution.

2. The model is tested by computing the PI estimate:

$$\hat{\mathrm{PI}}(U;O) = \mathrm{H}[U] + \sum_{j=1}^{n} \Pr[u_j] \cdot \sum_{k=1}^{n^t_{u_j}} \frac{1}{n^t_{u_j}} \cdot \log_2 \hat{\Pr}_{\mathsf{model}}[u_j|o^k_j],$$

where $n^t_{u_j}$ is the number of observations for user u_j in $\mathrm{DB_t}$, and $\hat{\Pr}_{\mathsf{model}}[u_j|o^k_j]$ is derived from $\hat{\mathsf{p}}^j_{\mathrm{model}}$ (resp. $\hat{\mathsf{f}}^j_{\mathrm{model}}$) via Bayes' theorem, as in the classical MI computation. In the ideal case where the model is perfect, the PI is an estimate of the MI (i.e. its value tends towards the MI one as the number of samples in $\mathrm{DB_t}$ increases). In the practical cases where the model differs from the attributes' true distribution, the PI captures the amount of information that is extracted from the DB, biased by the model errors. That is, the PI becomes lower than the MI as the model errors increase, and can be negative in case the model does not approximate at all the attributes' true distribution.

3.2 Perceptual utility rationale

In the following, we will say that a DB is perceptually useful if it allows extracting a large amount of PI. Here, the word perceptual relates to the fact that the definition of the PI is based on a model for the DS attributes, which may be incorrect (because of estimation or assumption errors), or become incorrect at some point (since it is hard to perfectly characterize users in the long term, e.g. because of preference or habit changes). Intuitively, a perceptually useful DB implies that the collected data represents well the DS. Hence it is fundamentally different from metrics based on the pseudo-probabilities of a DB, and will be concretely different when the number of samples in a DB is too low for characterizing the attributes accurately. More precisely,

* One can possibly split the DB in more parts in order to take advantage of cross-validation [4].

the two main advantages of this definition of utility relate to two recent results in the field of side-channel attacks. (1) The perceived information can be used to bound the success rate of a Bayesian adversary trying to distinguish a user based on "new" samples of his DS (i.e. independent of the samples used to build the model) – which is in contrast with the k-anonymity game, where the goal is to identify a user based on an observation in a DB. Hence, it relates to the "best possible" characterization of the attributes that can be obtained thanks to statistical sampling [3]. (2) The perceived information can benefit from "leakage certification" tests [4], which aim to guarantee that the PI is "close enough" to the MI. In such cases, we have that the collected data is "useful for anything" since it nearly perfectly represents the true distribution of the attributes. We intentionally leave these advantages informal because of place constraints and refer to [3, 4] for more details. Note that that our notion of perceptual utility is based on whether some user distributions are well characterized. This directly corresponds to probabilistic DB. However, even in the case of deterministic DB, one will generally describe group features, in which case the deterministic user data also becomes probabilistic. Eventually, in the extreme case where a single group has to be characterized, the conditional distributions in the PI derivation become irrelevant, but one can still exploit leakage certification to verify that this single group is well described.

4 Simulated experiments

In this section, we analyse the evolution of the k-anonymity and perceptual utility in the context of a simulated database containing individuals' shopping lists. We first define our simulation settings. Next, we put forward a number of intuitions regarding the impact of aggregating atributes, grouping users and the list's curse of dimensionality.

4.1 Simulation settings

We consider a DS (corresponding to a shop) with n users (aka clients). The shop sells N_i different items. For simplicity, each item can be puchased in N_q (integer) quantities. Hence, we have a set of $N_l = N_q^{N_i}$ possible shopping lists:

$$\mathcal{L} = \{(q_{i_1}, q_{i_2}, \ldots, q_{i_{N_i}}) | 1 \leq q_{i_j} \leq N_q\},$$

that we will also denote as $\mathcal{L} = \{l_1, l_2, \ldots, l_{N_l}\}$. For example, a shop with $N_i = 2$ items and $N_q = 3$ quantities will lead to the following set of $3^2 = 9$ lists:

$$\{(0,0), (0,1), (0,2), (1,0), (1,1), (1,2), (2,0), (2,1), (2,2)\}.$$

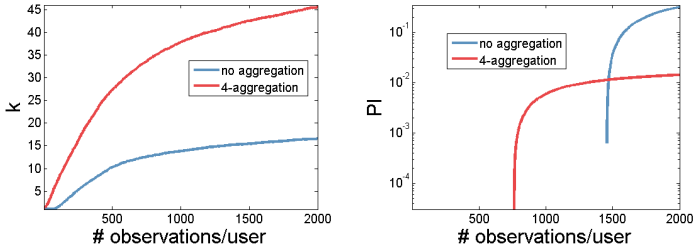
In this context, each user u_j has a single attribute. We define our simulated DS by selecting the user's probability functions $\mathsf{p}_1^j(o|u_j)$. For this purpose, for each user we pick up N_l values at random from a normal distribution $\mathcal{N}(\mu, \sigma)$ that we normalize (adapting the variance allows us to make users more or less different). Concretely, we analyzed a case study with $n = 100$ users, $N_i = 4$ items and $N_q = 5$ quantities. The number of observations per user will be variable in our experiments, but it is always identical for all users. Eventually, and taking advantage of our simulated context, we will report results averaged over 100 sampled DB (which allows obtaining smooth curves and gaining intuition about the average behavior of our metrics).

4.2 Impact of aggregation

As a first illustration of the tradeoff between k-anonymity and perceptual utility, we investigate the impact of a simple aggregation process for the previously defined shopping list attribute. Namely, we define N_a-aggregated shopping lists as lists where sets

of N_a original (consecutive) items are considered as single (aggregated) items that can be purchased in $N'_q = N_a \cdot (N_q - 1) + 1$ quantities. This reduces the cardinality of the set of lists from $N_q^{N_i}$ down to $N_q'^{N_i/N_a}$. For simplicity, we only consider cases wher N_a divides N_i. Figure 1 represents the evolution of the k-anonymity and the PI in function of the size of the DB (measured in number of observations per user), for the 100-users DB defined in the previous subsection, with and without aggregation. By making users more similar, the aggregation process *asymptotically* increases the k-anonymity and decreases the PI. But quite interestingly, we see that for DB with up to 1500 observations per user, the PI of the aggregated data is in fact larger than for the orginal one. This typically corresponds to the "win-win" scneario mentioned in introduction. That is, the amount of data collected is not sufficient to fully characterize the original lists. So aggregation allows improved k-anonymity without any loss of (perceptual) utility.

Figure 1: Average impact of aggregating items.

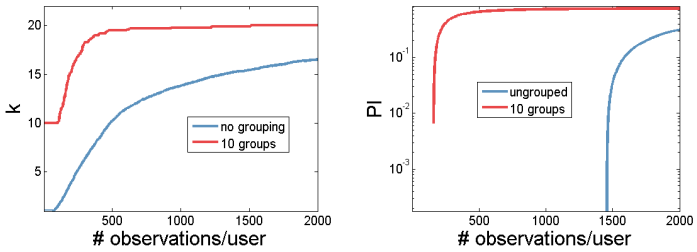


(a) k-anonymity. (b) Perceived Information.

4.3 Impact of grouping

We now study a complementary experiment in which the users are grouped in subsets. For this purpose, and in order for the grouping to make sense, our DS described in Subsection 4.1 actually embeds an additional feature that we now detail. Namely, we only created $q = 10$ user's probability functions p_1^j (with $1 \leq j \leq q$), and each of them was repeated 10 times to obtain $n = 100$ users. In this context, one can naturally group each subset of 10 identical users together. As illustrated in the right part of Figure 2, this significantly improves the convergence of the PI metric (since we have 10 times more observations per user). Furthermore, if the grouping is perfect (i.e. if the users of each group have identical distributions), there is no perceptual utility loss since, e.g. in our simulated case, we have:

Figure 2: Average impact of grouping.

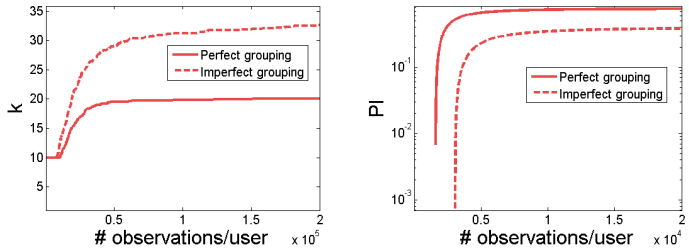


(a) User k-anonymity. (b) Perceived Information.

$$\begin{aligned}\mathrm{MI}(U;O) &= \log(100) + \sum_{u=1}^{100} \tfrac{1}{100} \sum_o \mathsf{p}(o|u) \log \Pr[u|o];\\ &= \log(100) + \sum_{g=1}^{10} \tfrac{10}{100} \sum_o \mathsf{p}(o|g) \log \tfrac{1}{10} \Pr(g|o);\\ &= \log(10) + \sum_{g=1}^{10} \tfrac{1}{10} \sum_o \mathsf{p}(o|g) \log \Pr(g|o);\\ &= \mathrm{MI}(G;O).\end{aligned}$$

So the gap between the PI curves in Figure 2 is only due to a lack of samples to characterize the ungrouped users. As for the k-anonymity in the left part of the figure, it is positively impacted by grouping as well. Indeed, whenever grouping, any observation recorded for a user u_j will be only be labeled as belonging to a group g_j. So in the simple case where groups have identical sizes, we can derive the user k-anonymity by multiplying the group k-anonymity by the group size. This implies a minimum k-anonymity of 10. Quite naturally, the situation substantially differs when the grouping is imperfect, as reflected in Figure 3. In this case, the characterization is still faster. However, it comes at the cost of a perceptual utility loss. This loss can be explained by the distributions of the groups that are becoming more similar, which is also reflected in a larger k-anonymity. Let us finally mention that grouping is an relevant option to

Figure 3: Impact of imperfect grouping.



(a) User k-anonymity. (b) Perceived Information.

preserve (generalizations of) k-anonymity when multiple observations for probabilistic attributes are leaked for a single user (since their combination usually allows better identification), which is an interesting scope for further research.

4.4 The curse of dimensionality

As clear from the previous discussions, the size of the sample space for shopping lists' distributions grows exponentially in the number items they contain. This suggests that exhaustively characterizing lists rapidly turns out to be infeasible (despite our toy examples made it possible by limiting N_i to 4). In this context, a last natural direction, that we investigate in this subsection, is to characterize items independently. As illustrated in the right part of Figure 4, this indeed allows making the collected data perceptually useful much faster. We further observe that the independence assumption was incorrect in our setting, since the characterization of four independent items is significanly less informative than the one of full lists (yet more informative than the characterization of a single item, as expected).

Interestingly, and assuming that only single items have to be characterized, it also becomes possible to sanitize a DB with "utility-preserving" operations that substantially increase the k-anonymity. In particular, it is easy to see that some types of data swapping (similar to the proposal in [6]) will not affect the utility of items considered independently. For example, let us assume 3 user observations o_1, o_2, o_3 made for 3 items i_1, i_2, i_3. In this case, any permutation of the lines below will lead to "potential user observations" o'_1, o'_2, o'_3 that do not affect the items' characterization:

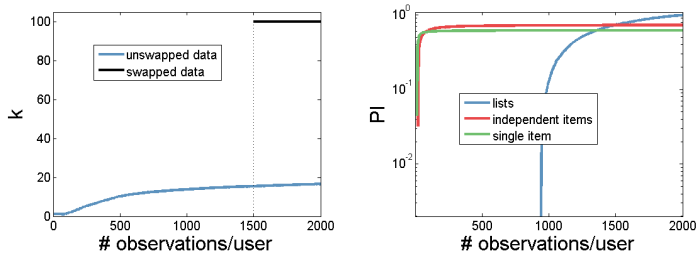
	o_1	o_2	o_3
i_1	1	0	1
i_2	0	0	1
i_3	2	1	1

swap $\rightarrow$

	o'_1	o'_2	o'_3
i_1	0	1	1
i_2	0	1	0
i_3	2	1	1

Since the permutations are unknown, such operations increase the number of potential user observations, and therefore the k-anonymity. Quite naturally, this also makes the computation of the k-anonymity more challenging, but it at least guarantees that as soon as every quantity appeared once for every user and item, the k-anonymity will be maximum. As illustrated in the left part of Figure 4, this condition was typically observed after 1500 observations per user in our example.[†] As the previous grouping, this type of anonymization will preserve k-anonymity even in contexts where multiple observations are leaked about a user. But contrary to grouping, it will not maintain the probabilistic anonymity metrics such as the privacy degree in [2].

Figure 4: Independent items characterization.



(a) k-anonymity. (b) Perceived Information.

Note that this last subsection also suggests that in most practical cases, the utility will be very perceptual (i.e. the PI will significantly differ from the MI), because of the difficulty to characterize large distributions in a non-parametric manner. However, making assumptions about a distribution (as in our last experiment) is still different than deciding in advance the goal for which some data is collected - which makes perceptual utility different than more specific metrics.

5 Conclusions

In the present state-of-the-art, privacy and utility are essentially seen as two different and conflicting goals. However, there are examples where privacy metrics such as the k-anonymity can be improved without loss of perceptual utility, as shown in this paper. And more fundamentally, the most striking conclusion of our experimental case studies is that, as the number of samples in a DB increases, both the k-anonymity and the characterization of its users generally increases. This suggests that anonymity (in general) and perceptual utility could in fact be two different facets of privacy. On the one hand, anonymity allows a user to deny allegations (i.e. claiming that he is not the only one having some attributes). On the other hand, useful data characterizes its users, which potentially allows identifying them based on observations that are not (yet) in the DB. In this respect, an important scope for further research will be to better connect anonymity metrics with perceptual utility. We anticipate that considering k-anonymity in front of multiple observations, and its extension towards (pseudo) probabilistic metrics such as the privacy degree in [2], could be an interesting

[†] By slightly biasing the DB with additional fake observations (which will then decrease its perceptual utility), we can easily enforce that this condition is met earlier.

direction for this purpose. Note that such connections can only exist asymptotically (i.e. for large enough DB), since conceptually, it always remains that perceptual utility requires that the samples in a DB represent well the true distribution of some attributes, while anonymity is defined based on the samples of the DB (i.e. independent of whether they are sufficient to characterize the attributes). Natural connections are also foreseen with the location privacy metrics in [9], for which characterization is indeed a central ingredient of the definition. For large enough DB, investigating the links between perceptual utility and the success rate, that have been proved useful in cryptographic contexts, is an important open problem as well (which could potentially improve the evaluation of location privacy). More generally, applying the tools in this paper to real case studies, and investigating the risks when combining multiple DB, is certainly needed to confirm their relevance and improve understanding. Eventually, developing tools to characterize the evolution of a user's profile over time (i.e. what is the impact of a change of preferences or habits on privacy and utility) is yet another challenge.

References

[1] J. Brickell and V. Shmatikov. The cost of privacy: destruction of data-mining utility in anonymized data publishing. In Y. Li, B. Liu, and S. Sarawagi, editors, *Proceedings of the 14th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, Las Vegas, Nevada, USA, August 24-27, 2008*, pages 70–78. ACM, 2008.

[2] C. Díaz, S. Seys, J. Claessens, and B. Preneel. Towards measuring anonymity. In R. Dingledine and P. F. Syverson, editors, *Privacy Enhancing Technologies, Second International Workshop, PET 2002, San Francisco, CA, USA, April 14-15, 2002, Revised Papers*, volume 2482 of *Lecture Notes in Computer Science*, pages 54–68. Springer, 2002.

[3] A. Duc, S. Faust, and F. Standaert. Making masking security proofs concrete - or how to evaluate the security of any leaking device. In E. Oswald and M. Fischlin, editors, *Advances in Cryptology - EUROCRYPT 2015 - 34th Annual International Conference on the Theory and Applications of Cryptographic Techniques, Sofia, Bulgaria, April 26-30, 2015, Proceedings, Part I*, volume 9056 of *Lecture Notes in Computer Science*, pages 401–429. Springer, 2015.

[4] F. Durvaux, F. Standaert, and N. Veyrat-Charvillon. How to certify the leakage of a chip? In P. Q. Nguyen and E. Oswald, editors, *Advances in Cryptology - EUROCRYPT 2014 - 33rd Annual International Conference on the Theory and Applications of Cryptographic Techniques, Copenhagen, Denmark, May 11-15, 2014. Proceedings*, volume 8441 of *Lecture Notes in Computer Science*, pages 459–476. Springer, 2014.

[5] B. C. M. Fung, K. Wang, R. Chen, and P. S. Yu. Privacy-preserving data publishing: A survey of recent developments. *ACM Comput. Surv.*, 42(4), 2010.

[6] S. P. Reiss. Practical data-swapping: The first steps. *ACM Trans. Database Syst.*, 9(1):20–37, 1984.

[7] M. Renauld, F. Standaert, N. Veyrat-Charvillon, D. Kamel, and D. Flandre. A formal study of power variability issues and side-channel attacks for nanoscale devices. In K. G. Paterson, editor, *Advances in Cryptology - EUROCRYPT 2011 - 30th Annual International Conference on the Theory and Applications of Cryptographic Techniques, Tallinn, Estonia, May 15-19, 2011. Proceedings*, volume 6632 of *Lecture Notes in Computer Science*, pages 109–128. Springer, 2011.

[8] P. Samarati and L. Sweeney. Generalizing data to provide anonymity when disclosing information (abstract). In A. O. Mendelzon and J. Paredaens, editors, *Proceedings of the Seventeenth ACM SIGACT-SIGMOD-SIGART Symposium on Principles of Database Systems, June 1-3, 1998, Seattle, Washington, USA*, page 188. ACM Press, 1998.

[9] R. Shokri, G. Theodorakopoulos, J. L. Boudec, and J. Hubaux. Quantifying location privacy. In *32nd IEEE Symposium on Security and Privacy, S&P 2011, 22-25 May 2011, Berkeley, California, USA*, pages 247–262. IEEE Computer Society, 2011.

[10] F. Standaert, T. Malkin, and M. Yung. A unified framework for the analysis of side-channel key recovery attacks. In A. Joux, editor, *Advances in Cryptology - EUROCRYPT 2009, 28th Annual International Conference on the Theory and Applications of Cryptographic Techniques, Cologne, Germany, April 26-30, 2009. Proceedings*, volume 5479 of *Lecture Notes in Computer Science*, pages 443–461. Springer, 2009.

On RAKE processing with estimation errors in an UWB system

Arie G.C. Koppelaar, Stefan Drude, Marinus van Splunter
Andries Hekstra, Frank Leong
NXP Semiconductors
High Tech Campus 46, Eindhoven, The Netherlands
arie.koppelaar@nxp.com stefan.drude@nxp.com

Abstract

In a Passive Keyless Entry (PKE) system the user does not have to use a physical key actively to get access to the system. The presence of the key in the vicinity of the system is sufficient to get access. PKE is a feature introduced by car manufacturers some time ago. PKE systems are prone to so-called relay attacks in which the attacker can relay messages between car and key fob such that the electronics in the car believe that the key fob is in the vicinity of the car. Distance-bounding protocols enable the verification that a legitimate key is not further away from the car than it should be to grant access to the car. An element of the distance bounding protocol is the ability to determine the range between car and key. Impulse Radio Ultra-Wideband (IR-UWB) modulation due to its short time pulses is suitable to do range measurement using Time-of-Flight techniques. Next to ranging, IR-UWB modulation is also suitable for realizing low-power low-cost communication. The pulse based communication offers possibilities to make the reception resilient to multipath propagation. RAKE processing is a method to coherently add up the energy of the individual reflected paths. For this purpose channel estimation has to be carried out. In this paper the influence of channel estimation inaccuracies on the receiver performance is investigated.

1 Introduction

Car manufacturers are gradually introducing Passive Key Entry (PKE) as a comfort feature in cars. The user can unlock the car without actively pushing a button on her or his key fob. The detection by the car electronics of the key fob in the vicinity of the car is sufficient to unlock the car. A PKE system without proper vicinity check is vulnerable to so-called relay attack. In Section 2 a relay attack in the context of a PKE system is explained. Distance bounding is a way to protect against a relay attack. It requires Time-of-Flight measurements, for which Impulse Radio Ultra-Wideband (IR-UWB) modulation is proposed. IR-UWB can also be used for low-power communication in a multipath environment. RAKE processing is an essential receiver operation to make the communication reliable. Some results from literature related to UWB channel models and RAKE receivers are presented in Section 3. For channels with a large delay spread it is known that it makes sense to have a large amount of RAKE fingers for collecting the energy of many multipath reflections. In Section 4 it is investigated whether it makes sense to implement a large amount of RAKE fingers when the receiver's channel estimate is suffering from inaccuracies. Finally, in Section 5 some conclusions are provided.

2 Passive Key Entry systems

A Passive Key Entry system is a system that provides a user access to a protected area (e.g. home, office or car) without the need of using its key fob actively. The only requirement is that the user has the key fob with him or her and that he or she is in the protected area.

In literature the setting above is often referred to as a location verification problem between a verifier (in our setting the car) and a prover (in our setting the car key fob). Moreover, the location verification in our setting is confined to a distance verification problem. When a user is in the proximity of the car, the car detects the presence of a key fob and can initiate communication with the key fob such that credentials of the key fob can be authenticated by the car.

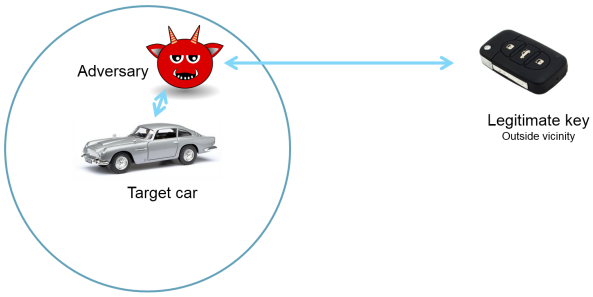


Figure 1: Typical scenario for a relay attack of a PKE system

However, this approach is vulnerable to a so-called relay attack [1]. A person with bad intentions can trigger the car short range detection means and relay the communication of messages between the car and the key fob, see Figure 1. A solution to the relay attack is that a key fob can prove its proximity to the car in a way that cannot be modified in favor to a malicious person. Distance bounding [2] is a method to defeat relay attacks and is based on the fact that information is local and cannot travel faster than light. By measuring Time-of-Flight (ToF) of messages the distance between key fob and car can be determined. An adversary can only relay the messages that are exchanged but cannot insert messages that shorten the ToF without knowing the secret.

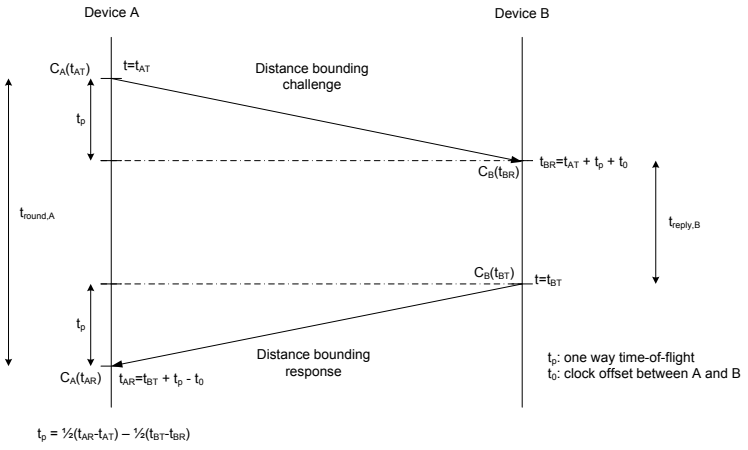


Figure 2: Determining the distance by two-way ToF measurement

In Figure 2 the principle of distance determination by two-way ToF measurement is shown. Device A wants to check whether device B is inside or outside the protected

zone. It sends a challenge message to device B and at the same time starts a timer. Device B computes the response and sends it back to Device A. At reception of the response, Device A stops its timer and verifies the response. Using its timer values, device A can determine the round trip delay and after compensation with the processing time of device B (either fixed or appended to the response in an encrypted way) it can calculate the ToF and therefore also the distance to device B.

IR-UWB is a modulation method that is very well suited for doing precise ToF measurements. The used pulse shape is narrow and provides therefore a good time resolution. Moreover, the short pulse shape gives an adversary little room to detect the leading edge of a pulse, analyze the modulation content and to advance the modulated pulse in time with a large amplitude. Moreover, IR-UWB modulation lends itself to making low-power implementations and can be made resilient to multipath propagation. The resilience to multipath propagation can be accomplished by using low duty cycling signalling. In combination with short pulses, reflected pulses will have less probability of overlapping each other. The dispersion of signal pulses in time can be compensated by using a RAKE receiver [3]. A RAKE receiver attempts to time align the reflected pulses and to add them up coherently. A RAKE finger is the processing part of a RAKE receiver that captures an individual received pulse, provides it with an appropriate delay and complex weight such that the pulse can be coherently added to the other pulses. The amount of fingers to implement is a trade-off between implementation complexity and receiver performance.

3 UWB Channel and RAKE receiver

UWB communication is standardised in IEEE802.15.3a and IEEE802.15.4a. Channel models with several profiles for evaluating PHY proposals were defined. The IEEE802.15.3a channel models [6] represent typical indoor channel environments and are based on a modified Saleh-Valenzuela model. The RMS delay spread varies from 5 till 25 ns. In Morche et. al. [4] these channel models are used to simulate the performance of a RAKE receiver. As a trade-off between performance and complexity it was decided to implement a RAKE receiver with only 4 fingers. For the channel with an RMS delay spread of ≤ 15 ns, less than 2 dB energy was lossed compared to a receiver that collected all multipath components. Using 4 fingers on a channel with an RMS delay spread of 25 ns results in a loss of more than 3 dB. The number of paths needed to recover 85% of all energy grows fast with increasing RMS delay spread. For an RMS delay spread of 15 ns more than 50 fingers are needed while for 25 ns over 100 finger are needed.

Similarly, based on measurements in several environments, a statistical channel model was developed and accepted as a standard model for evaluation of UWB system proposals by the IEEE802.15.4a Task Group. In the paper of Ahmadian et.al.[5] simulations are carried out with an IEEE802.15.4a receiver on some of the channel models defined by the Task Group. They investigated how the performance (in terms of BER and/or FER) depends on the number of used RAKE fingers. As a reference they compare to a so-called all-RAKE (ARAKE) configuration in which all channel impulse response contributions are used in the combining. Selective RAKE (SRAKE) is a combining method in which the L_s RAKE fingers are assigned to the channel impulse response contributions with the largest magnitude. The performance results presented in [5] show that for the Residential environment (NLOS) channel model (CM2) the performance of an SRAKE with $L = 20$ fingers approaches the performance of an ARAKE receiver. However, for the Outdoor NLOS channel model (CM6) many more fingers are needed in order to approach the ARAKE performance (See Figure 3).

Measurements conducted by Karedal et.al. [9] show that in an industrial environment with NLOS, more than 100 multipath components need to be combined in order to collect over 50% of the total energy. In the next section it is investigated how

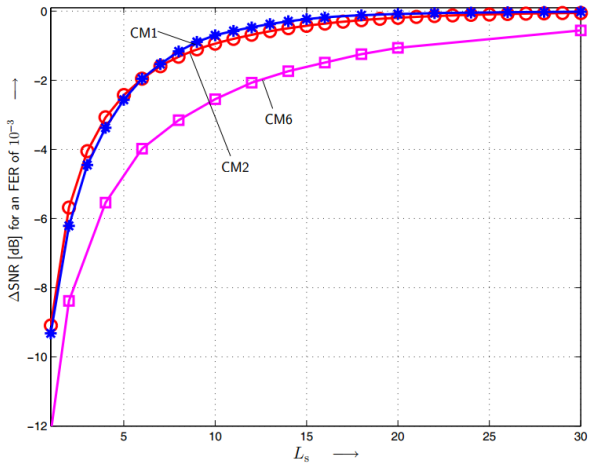


Figure 3: SNR loss of SRAKE compared to ARAKE as function of the number of RAKE fingers [5]

realistic a RAKE receiver with such a huge number of RAKE fingers is.

4 RAKE processing with estimation errors

For our analysis we assume a stationary channel with a channel impulse response consisting of L_{max} multipath components :

$$h(t) = \sum_{i=0}^{L_{max}-1} h_i \delta(t - \tau_i) \tag{1}$$

Furthermore we assume that BPSK or PAM modulation is employed using IR-UWB communication with a low pulse repetition rate. When transmitting a symbol X through the channel, the received signal is equal to :

$$r(t) = X \sum_{i=0}^{L_{max}-1} h_i \delta(t - \tau_i) + n(t) \tag{2}$$

In a RAKE finger i, $i \in \{0, \ldots, L\}$ for some $L \leq L_{max}$, we receive $R_i = Xh_i + n_i$, where n_i is AWGN. In an ideal RAKE receiver, the RAKE finger contributions are appropriately time aligned and added using RAKE finger coefficients $c_i = h_i^*$, leading to the decision variable U :

$$U = \sum_{i=0}^{L-1} R_i c_i = \sum_{i=0}^{L-1} R_i h_i^* = \sum_{i=0}^{L-1} X|h_i|^2 + n_i h_i^* \tag{3}$$

We observe that the decision variable U consists of a signal part $U_s = \sum_{i=0}^{L-1} X|h_i|^2$ and a noise part $U_n = \sum_{i=0}^{L-1} n_i h_i^*$. The realised signal to noise ratio for combining $L \leq L_{max}$ fingers is :

$$SNR_{ideal}(L) = \frac{E[U_s U_s^*]}{E[U_n U_n^*]} = X^2 \frac{\sum_{i=0}^{L-1} |h_i|^2}{\sigma_n^2}, \tag{4}$$

The realised SNR is a positive non-decreasing function of the number of applied RAKE fingers L and is a function of the channel realisation.

In order to implement a RAKE receiver, both the path amplitude and the path delay have to be estimated. Estimation errors in the path delay and the path amplitude will lead to performance degradation of the RAKE receiver [7, 8]. In the analysis below, we only assume estimation errors in the path amplitude. Due to estimation errors we assume that the RAKE coefficients are polluted with estimation noise, which we assume to be additive and i.i.d. Gaussian distributed, i.e. $c_i = h_i^* + m_i$, where m_i is a complex Gaussian variable with variance σ_m^2.

Now we want to determine the decision variable in case of non-perfect RAKE coefficients :

$$U = \sum_{i=0}^{L-1} R_i c_i = \sum_{i=0}^{L-1} R_i (h_i^* + m_i) = \sum_{i=0}^{L-1} X|h_i|^2 + X h_i m_i + n_i h_i^* + n_i m_i. \tag{5}$$

The signal part of the decision variable remains unchanged, but the noise part has now 3 contributions. The noise contribution of RAKE finger i is equal to $U_{n,i} = X h_i m_i + n_i h_i^* + n_i m_i$.

We assume that the coefficient noise, additive noise and the path amplitude are uncorrelated. In Sheng & Haimovich [8] it is shown that the variance of the coefficient noise due to path amplitude estimation errors can be written as :

$$\sigma_m^2 = \frac{1}{M} \frac{\sigma_n^2}{X^2}, \tag{6}$$

where M stands for the estimation effort to estimate the channel and is linear with the number of preamble or pilot symbols that are used for channel estimation.

Using this relation between noise variance and combiner tap variance we obtain :

$$SNR_{non-ideal}(L) = \frac{SNR_{ideal}(L)}{1 + \frac{1}{M}\left[1 + L/SNR_{ideal}(L)\right]} \tag{7}$$

For perfect channel estimates ($\lim M \to \infty$) $SNR_{non-ideal}(L)$ converges to the ideal RAKE receiver $SNR_{ideal}(L)$. In case of non-perfect channel estimates $SNR_{non-ideal}(L)$ can have a maximum. As function of L, $SNR_{ideal}(L)$ will converge to an end value and for some value of L $SNR_{non-ideal}(L)$ will not increase anymore but instead it will decrease. Figure 4 is an illustration of this behaviour, the green and blue SNR curves show the achieved SNR as function of the number of RAKE fingers for a channel realisation.

Simulations using the exponential channel model [10] are carried out to determine the average number of RAKE fingers needed to maximize the SNR and the SNR loss compared to an ideal RAKE receiver. The model is a multipath model from which the taps are independent complex Gaussian variables with average power profile that decays exponentially. For the exponential channel model the following set of RMS delay spreads were used: $\tau_{RMS} \in \{5, 10, 20, 50, 100\}$ [ns]. To investigate the estimation accuracy the parameter M is taken from the set $\{1, 2, 5, 10, 20, 50\}$. Two main concepts of RAKE receiver were investigated: SRAKE and partial RAKE (PRAKE). With PRAKE the L RAKE fingers are assigned to the multipath components according to their time arrival. Note that due to imperfect channel estimation the RAKE finger assignment in SRAKE becomes hypothetical. As an upperbound to the expected performance we assume for SRAKE an assignment of RAKE fingers according to the true impulse response. Our first interest is in assessing the number of RAKE fingers that maximize the SNR. This number is determined as a function of the RMS delay spread and as function of the estimation accuracy.

In Figure 5 and 6 the results are shown for both an SRAKE and a PRAKE receiver. Due to the unsorted channel impulse response, the PRAKE receiver needs many more RAKE fingers to optimize the SNR than an SRAKE receiver.

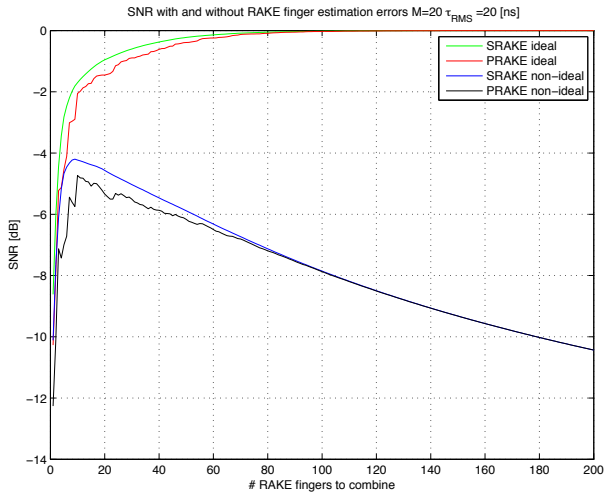


Figure 4: SNR gain as function of number of RAKE fingers for a channel realisation

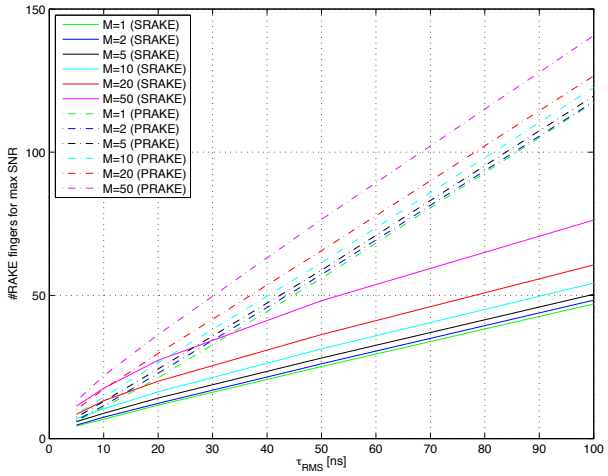


Figure 5: Average number of RAKE fingers to maximize SNR as function of RMS delay spread

The optimum number of RAKE fingers shows a linear relation with both the RMS delay spread and the estimation accuracy parameter M. Increasing the estimation effort (increase of M) allows to use more RAKE fingers without sacrificing SNR.

The results in Figure 7 show that the SNR loss as function of the RMS delay spread increases. The SNR loss for the SRAKE solution is in the order of 2 dB less than for the PRAKE receiver. The SNR loss as function of the estimation effort is shown in Figure 8. The graphs show that for an RMS delay spread of $\geq$ 50 ns, quite some estimation effort ($M \geq 50$) has to be made in order to bring the SNR loss below 5 dB.

5 Conclusions

In this paper the performance of RAKE receivers with imperfect channel estimates is investigated. Using the exponential channel model in simulations it is shown that the number of RAKE fingers to be used should be chosen carefully. The realized SNR as function of the number of RAKE fingers has a maximum, meaning that using

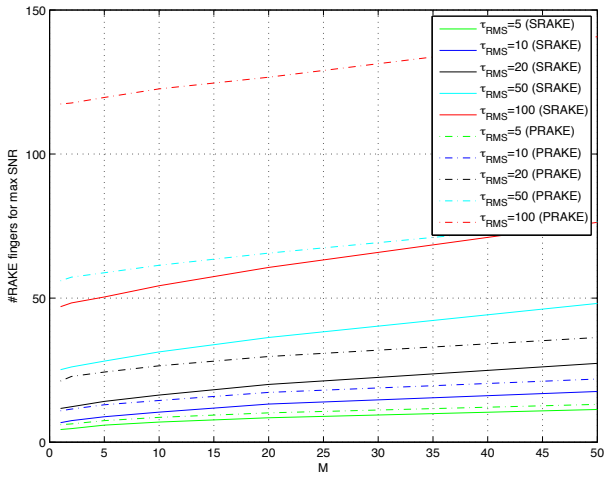


Figure 6: Average number of RAKE fingers to maximize SNR as function of M

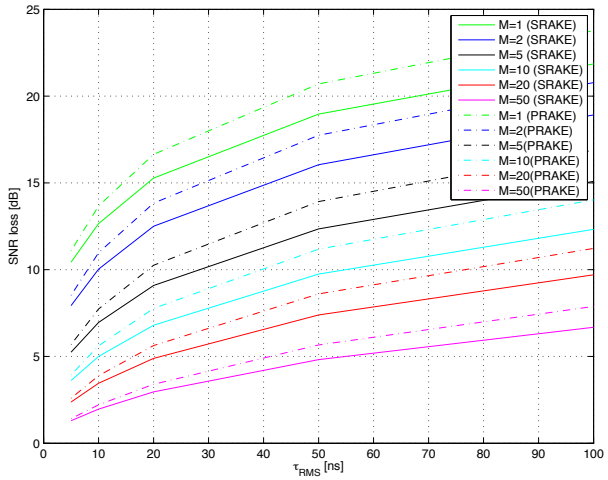


Figure 7: Average SNR loss compared to ideal RAKE as function of RMS delay spread

more fingers than this maximum will result in an SNR penalty. The consequences of non-perfect channel estimation are that less of the available multipath energy can be used and that the realized SNR will be smaller than in an ideal RAKE receiver. The maximum SNR and the corresponding optimal number of RAKE fingers depend on the channel realisation and extra acquisition effort is needed in the receiver to limit an extra SNR penalty due to using a wrong number of RAKE fingers.

Theoretically, an SRAKE receiver is a good trade-off between SNR performance and implementation complexity. However, due to imperfect channel estimates, the selection of the strongest multipath components cannot be done reliable. Therefore significant more RAKE fingers will be needed and an extra penalty in the realized SNR can be expected.

In future work we like to investigate possibilities to close the gap between SRAKE and PRAKE receivers by using a reliable selection criterion. Moreover, we like to make an implementation that is less sensitive to L variation.

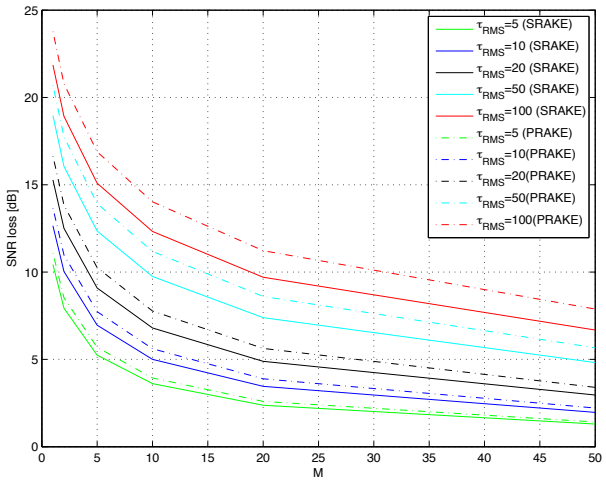


Figure 8: Average SNR loss compared to ideal RAKE as function of M

References

[1] Francillon, A., Danev, B., & Capkun, S. *Relay Attacks on Passive Keyless Entry and Start Systems in Modern Cars*. 18th Annual Network and Distributed System Security Symposium, February 6-9, 2011

[2] Brands, S. & Chaum, D. *Distance-Bounding Protocols (Extended Abstract)*. EUROCRYPT'93, Lecture Notes in Computer Science 765, pp. 344-359, Springer Verlag

[3] Proakis, John G., *Digital Communications*, 3rd edition, McGraw-Hill, 1995

[4] Morche, D. et. al., *Double-Quadrature UWB Receiver for Wide-Range Localization Applications With Sub-cm Ranging Precision*, IEEE Journal of Solid-State Circuits, Vol.48, No.10, pp.2351-2362, October 2013.

[5] Ahmadian, Z. and Lampe L., *Performance Analysis of the IEEE 802.15.4a UWB System*, IEEE transactions on Communications, Vol.57, No.5, pp. 1474-1485, May 2009.

[6] Molisch, A.F. et. al., *Channel Models for Ultrawideband Personal Area Networks*, IEEE Wireless Communications, pp14-21, December 2003.

[7] Choi, J.D. and Stark W.E., *Performance of Ultra-Wideband Communications With Suboptimal Receivers in Multipath Channels*, IEEE Journal on Selected Areas in Communications, Vol.20, No.9, pp.1754-1766, December 2002.

[8] Sheng H. and Haimovich A.M., *Impact of Channel Estimation on Ultra-Wideband System Design*, IEEE Journal of Selected Topics in Signal Processing, Vol.1, No.3, pp. 498-507, October 2007.

[9] Karedal, J., et.al., *A Measurement-Based Statistical Model for Industrial Ultra-Wideband Channels*, IEEE trans. on Wireless Communications, pp.3028-3037, Vol.6, No.8, August 2007

[10] Halford, S. et.al., *Evaluating the Performance of HRb Proposals in the Presence of Multipath*, IEEE802.11-00/282r1, Sept 2000.

Binary Puzzles as an Erasure Decoding Problem

Putranto Hadi Utomo Ruud Pellikaan
Eindhoven University of Technology
Dept. of Math. and Computer Science
PO Box 513. 5600 MB Eindhoven
p.h.utomo@tue.nl g.r.pellikaan@tue.nl

Abstract

Binary puzzles are interesting puzzles with certain rules. A solved binary puzzle is an $n \times n$ binary array such that there are no three consecutive ones and also no three consecutive zeros in each row and each column, the number of ones and zeros must be equal in each row and in each column, every two rows and every two columns must be distinct.
Binary puzzles can be seen as constrained arrays. Usually constrained codes and arrays are used for modulation purposes. In this paper we investigate these arrays from an erasure correcting point of view. We give lower and upper bound for the rate of these codes, the probability of correct erasure decoding and erasure decoding algorithms.

1 Introduction

Sudokus are nowadays very popular puzzles and they are studied for their mathematical structure [2, 5, 18]. For instance the minimal number of entries that can be specified in a single 9×9 puzzle to ensure a unique solution was in [14] conjectured to be 17, and this was proved by means of the chromatic polynomial of the Sudoku graph [7]. Furthermore the erasure correcting capabilities and decoding algorithms of the collection of $n \times n$ a Sudokus are considered [13, 16]. The asymptotic rate is still an open problem [1, 7]. Solving an $n \times n$ Sudoku puzzle is an NP-hard problem [17].

The binary puzzle is also an interesting puzzle with certain rules and is the focus of this paper. We look at the mathematical theory behind it. The solved binary puzzle is an $n \times n$ binary array that satisfies:

1. no three consecutive ones and also no three consecutive zeros in each row and each column,

2. every row and column is balanced, that is the number of ones and zeros must be equal in each row and in each column,

3. every two rows and every two columns must be distinct.

Figure 1 is an example of a binary puzzle. There is only one solution satisfying all three conditions. But there are 3 solutions satisfying (1) and (2). The solution satisfying three all conditions is given in Figure 2. Figure 3 and 4 are solved puzzles where the third constraint is excluded.

Binary and Sudoku puzzle can be seen as constrained arrays. Usually constrained codes and arrays are used for modulation purposes [8, 9]. We investigate these arrays from an erasure correcting point of view. We give lower and upper bound for the rate of these codes, the probability of correct erasure decoding and erasure decoding algorithms.

	0						
			1		1		0
		0					
	1						
				1			
	0				1		
	0			0			
				0		0	

Figure 1: Unsolved Puzzle

1	0	1	0	1	0	1	0
0	1	0	1	0	1	1	0
1	0	0	1	1	0	0	1
0	1	1	0	0	1	1	0
0	1	0	1	1	0	0	1
1	0	1	0	1	1	0	0
1	0	0	1	0	0	1	1
0	1	1	0	0	1	0	1

Figure 2: Solved Puzzle

1	0	1	0	1	0	0	1
0	1	0	1	0	1	1	0
1	0	0	1	1	0	1	0
0	1	1	0	0	1	0	1
0	1	0	1	1	0	1	0
1	0	1	0	1	1	0	0
1	0	0	1	0	0	1	1
0	1	1	0	0	1	0	1

Figure 3: Solved Binario Puzzle with repetition of column/row allowed

0	0	1	0	1	0	1	1
1	1	0	1	0	1	0	0
1	0	0	1	1	0	1	0
0	1	1	0	0	1	0	1
0	1	0	1	1	0	1	0
1	0	1	0	1	1	0	0
1	0	0	1	0	0	1	1
0	1	1	0	0	1	0	1

Figure 4: Solved Binario Puzzle with repetition of column/row allowed

2 Constrained sequences and constrained array

Let C be a code in Q^n, where the alphabet Q has q elements.
Recall that the *(information) rate* of C is defined by

$$R(C) = \frac{\log_q |C|}{n}.$$

In the following $Q = \mathbb{F}_2$, $n = lm$ and $\mathbb{F}_2^{l\times m}$ is the set of binary $l \times m$ arrays. Define:

$$\begin{aligned}
A_{l\times m} &= \{X \in \mathbb{F}_2^{l\times m} \mid X \text{ satisfies (1) }\};\\
B_{l\times m} &= \{X \in \mathbb{F}_2^{l\times m} \mid X \text{ satisfies (2) }\};\\
C_{l\times m} &= \{X \in \mathbb{F}_2^{l\times m} \mid X \text{ satisfies (3) }\};\\
D_{l\times m} &= \{X \in \mathbb{F}_2^{l\times m} \mid X \text{ satisfies (1), (2) and (3) }\}.
\end{aligned}$$

The theory of constrained sequences, that is for $l = 1$, is well established and uses the theory of graphs and the eigenvalues of the incidence matrix to give a linear recurrence. An explicit formula for the number of such sequences of a given length m can be expressed in terms of the eigenvalues. The asymptotical rate is equal to $\log_q(\lambda_{max})$, where λ_{max} is the largest eigenvalue. See [8, 9]. Shannon [15] showed already that the following relation holds for $m \geq 1$:

$$|A_{1\times(m+2)}| = |A_{1\times(m+1)}| + |A_{1\times m}|.$$

Asymptotically this gives

$$R(A_{1\times m}) \approx \log_2\left(\frac{1}{2} + \frac{1}{2}\sqrt{5}\right), \quad \text{for } m \to \infty$$

The number of balanced sequence is equal to a number of combination of ones, that is $B_{1\times 2m} = \binom{2m}{m}$ and asymptotically $R(B_{1\times 2m}) \approx 1$, for $m \to \infty$.
It was shown [6, 10, 11] that the balanced property does not influence the asymptotic rate of constrained sequences. So $R(A_{1\times 2m} \cap B_{1\times 2m}) \approx \log_2\left(\frac{1}{2} + \frac{1}{2}\sqrt{5}\right)$, for $m \to \infty$.
We expect that a similar result holds for balanced constrained arrays.
For arrays we know that $\binom{2l}{l}^m \leq |B_{2l\times 2m}| \leq \binom{2l}{l}^{2m}$. From these inequalities it is can be shown that, asymptotically:

$$\tfrac{1}{2} \lesssim\leq R(B_{2m\times 2m}) \leq 1, \quad \text{for } m \to \infty$$

Four arbitrary elements of $B_{2m\times 2m}$ gives an element of $B_{4m\times 4m}$. So $|B_{4m\times 4m}| \geq |B_{2m\times 2m}|^4$. Therefore $R(B_{2m\times 2m})$ is increasing in m.

Now, consider $C_{l\times m}$. We clearly have that $|C_{l\times m}| \leq 2^m(2^m - 1)\cdots(2^m - n + 1)$. Furthermore, if $m = n$, $|C_{(n+1)\times(n+1)}| \geq |C_{n\times n}| \cdot (2^{2n+1} - 2n2^n + n^2)$.
This implies that, asymptotically:

$$R(C_{2m\times 2m}) \approx 1, \quad \text{for } m \to \infty$$

The size of $D_{2m\times 2m}$ can be approximated by smaller building blocks such that the conditions are still satisfied [4]. There are exactly two building block of size 2×2.

Hence, $R(D_{2m\times 2m}) \geq \frac{1}{(2m)^2}\log_2(2^{m^2}) = \frac{1}{4}$, for $m \geq 1$.

Numerically, we have

m	$A_{2m\times 2m}$		$B_{2m\times 2m}$		$C_{2m\times 2m}$		$D_{2m\times 2m}$	
	Size	Rate	Size	Rate	Size	Rate	Size	Rate
1	16	1	2	0.25	10	0.83	2	0.25
2	2030	0.69	90	0.41	33864	0.94	76	0.39
3	3858082	0.61	?	?	?	?	5868	0.34

3 Erasure Channel

Suppose Q is a set of an alphabet and C is a code in Q^n.
Define $\hat{Q} = Q \cup \{-\}$, where the symbol "-" denotes a blank, that is an erasure.

Suppose $\mathbf{r}$ is the received word given that $\mathbf{c}$ is sent. We have $d(\mathbf{r},\mathbf{c})$ is the Hamming distance between $\mathbf{r}$ and $\mathbf{c}$. Since the errors are only blanks, $d(\mathbf{r},\mathbf{c})$ equal to the number of blanks in $\mathbf{r}$. Let $\mathbf{c}(\mathbf{r})$ be a closest codeword to $\mathbf{r}$, then $d(\mathbf{r},C) = d(\mathbf{r},\mathbf{c}(\mathbf{r}))$. Let p be the probability that a symbol is erased, and let $P_{ed,C}(p)$ denote the probability of correct erasure decoding. Then

$$P_{ed,C}(p) = \sum_{\mathbf{c}\in C} P(\mathbf{c}) \sum_{\substack{\mathbf{r}\in\hat{Q} \\ \mathbf{c}(\mathbf{r})=\mathbf{c}}} P(\mathbf{r}|\mathbf{c})$$

Suppose $\mathcal{E}_i(C) = \{\mathbf{r} \in \hat{Q}^n | d(\mathbf{r},C) = i\}$ and $E_i(C) = |\mathcal{E}_i(C)|$.
Define the homogenous erasure distance enumerator for code C by

$$E_C(X,Y) = \sum_{i=0}^{n} E_i(C) X^{n-i} Y^i$$

Proposition 3.1

$$P_{ed,C}(p) = \frac{1}{|C|} E_C(1-p,p)$$

Proposition 3.2 *Let $C \subseteq Q^m$ and $D \subseteq Q^n$. We have*

$$E_{C\times D}(X,Y) = E_C(X,Y) \cdot E_D(X,Y)$$

Corollary 3.3

$$P_{ed,C\times D}(p) = (P_{ed,C}(p)) \cdot (P_{ed,D}(p))$$

Corollary 3.4

$$P_{ed,C^n}(p) = (P_{ed,C}(p))^n$$

4 Binary Puzzle Solver

Binary puzzle can be seen as a SAT problem. Since each cell in the binary puzzle can only take the values '0' and '1', we can express the puzzle as an array of binary variables, where false corresponds to '0' and true to '1'. Next, we express each condition in terms of a logical expression.

Suppose we have an $2m \times 2m$ array in the variables x_{ij}. The array satisfies the first condition, that there are no three consecutive ones and also no three consecutive zeros in each row and each column, if and only if the expression below is true:

$$\left(\bigwedge_{i=1}^{2m}\left\{\bigwedge_{k=1}^{2m-2}\left(\left[\neg\left(\bigwedge_{j=k}^{k+2}x_{ij}\right)\right]\wedge\left[\neg\left(\bigwedge_{j=k}^{k+2}\neg x_{ij}\right)\right]\right)\right\}\right)\wedge$$

$$\left(\bigwedge_{j=1}^{2m}\left\{\bigwedge_{k=1}^{2m-2}\left(\left[\neg\left(\bigwedge_{i=k}^{k+2}x_{ij}\right)\right]\wedge\left[\neg\left(\bigwedge_{i=k}^{k+2}\neg x_{ij}\right)\right]\right)\right\}\right)$$

For satisfying the second condition on balancedness, the following expression must be true

$$\left(\bigwedge_{j=1}^{2m}\left[\bigwedge_{1\le i_1<\cdots<i_m\le 2m}\left(\bigvee_{k=1}^{m}x_{i_k,j}\right)\right]\right)\wedge\left(\bigwedge_{i=1}^{2m}\left[\bigwedge_{1\le j_1<\cdots<j_m\le 2m}\left(\bigvee_{k=1}^{m}x_{i,j_k}\right)\right]\right)\wedge$$

$$\left(\bigwedge_{j=1}^{2m}\left[\bigwedge_{1\le i_1<\cdots<i_m\le 2m}\left(\bigvee_{k=1}^{m}\neg x_{i_k,j}\right)\right]\right)\wedge\left(\bigwedge_{i=1}^{2m}\left[\bigwedge_{1\le j_1<\cdots<j_m\le 2m}\left(\bigvee_{k=1}^{m}\neg x_{i,j_k}\right)\right]\right).$$

Note that the complexity of this expression grows as $\binom{2m}{m}$ which is exponentially in m. An alternative polynomial expression can be obtained.

The satisfiability of the third condition, that every two rows and every two columns must be distinct, is equal to

$$\left(\bigwedge_{1\le j_1<j_2\le 2m}\left\{\bigwedge_{i=1}^{2m}\left[(x_{i,j_1}\wedge x_{i,j_2})\vee(\neg x_{i,j_1}\wedge\neg x_{i,j_2})\right]\right\}\right)\wedge$$

$$\left(\bigwedge_{1\le i_1<i_2\le 2m}\left\{\bigwedge_{j=1}^{2m}\left[(x_{i_1,j}\wedge x_{i_2,j})\vee(\neg x_{i_1,j}\wedge\neg x_{i_2,j})\right]\right\}\right).$$

It is shown in [3] that the binary puzzle is NP-complete.

References

[1] C. Atkins and J. Sayir, "Density evolution for SUDOKU codes on the erasure channel," *Proc. 8th Int. Symp. on Turbo Codes and Iterative Information Processing (ISTC),* 2014.

[2] R.A. Bailey, P.J. Cameron and R. Connelly, "Sudoku, gerechte designs, resolutions, affine space, spreads, reguli, and Hamming codes," *Amer. Math. Monthly*, 115(5):383–404, 2008.

[3] M. De Biasi, "Ninary puzzle is NP-complete,"

`http://nearly42.org`

[4] T. Etzion and K.G. Paterson, "Zero/positive capacities of two-dimensional runlength-constrained arrays," *IEEE Trans. Information Theory*, 51(9):3186–3199, Sept 2005.

[5] B. Felgenhauer and A.F. Jarvis, "Mathematics of Sudoku I, II" *Mathematical Spectrum* 39:15–22, 54–58, 2006.

[6] H.C. Ferreira, J.H. Weber, C.H. Heymann and K.A.S. Immink "Markers to construct dc free (d, k) constrained balanced block codes using Knuths inversion" *Electronics Letters*, 48(19):1209-1211, Sept 2012.

[7] A.M. Herzberg and M. Ram Murty, "Sudoku squares and chromatic polynomials," *Notices Amer. Math. Soc.* 54 (6):708–717, July 2007.

[8] H.D.L. Hollmann. *Modulation Codes*. Philips Electronics N.V., PhD thesis Techn. Univ. Eindhoven, 1996.

[9] K.A.S. Immink, P.H. Siegel, and J.K. Wolf, "Codes for digital recorders," *IEEE Transactions on Information Theory*, 44(6):2260–2299, Oct 1998.

[10] K.A.S. Immink and J.H. Weber, "Very efficient balanced codes," *IEEE Transactions on Selected Areas in Communications,* , 28(2):188–192, Feb 2010.

[11] D.E. Knuth, "Efficient balanced codes," *IEEE Transactions on Information Theory,* , 32(1):51–53, Jan 1986.

[12] J.H. van Lint. *Introduction to Coding Theory*. Springer-Verlag New York, Inc., Secaucus, NJ, USA, 1982.

[13] L.A. Phillips, S. Perkins, P.A. Roach and D.H. Smith, "Erasure correction capabilities of Sudoku and related combinatorial structures," *Proc. 4th Research Student Workshop, University of Glamorgan*, 65–69, 2009.

[14] `http://staffhome.ecm.uwa.edu.au/~00013890/sudokumin.php`

[15] C.E. Shannon, "A mathematical theory of communication," *Bell System Technical Journal*, 27(10):379–423, October 1948. Reprinted in [**?**].

[16] E. Soedarmadji and R. McEliece, "Iterative Decoding for Sudoku and Latin Square Codes," *Proceeings 45th Annual Allerton Conference on Communication, Control, and Computing,* 488–494, 2007.

[17] T. Yato and T. Seta, "Complexity and completeness of finding another solution and its application to puzzles," *Proc. National Meeting of the Information Processing Society of Japan (IPSJ),* 2002.

[18] `http://en.wikipedia.org/wiki/Mathematics_of_Sudoku`

Multiband distributed sensing based on compressed measurements via two AICs implementations

J. Verlant-Chenet F. Horlin
Université Libre de Bruxelles
OPERA department, Group WCG
jverlant@ulb.ac.be fhorlin@ulb.ac.be

Abstract

We develop a distributed multiband spectrum sensing detector for cognitive radios based on compressed measurements that does not rely on signal reconstruction. A fusion centre collects the measurements from different sensing nodes and then makes a sensing decision based on a simplified maximum likelihood criterion which is valid for both analog to information implemented in the paper (MWC and NUP) and does not require prior signal information. Simulation results for probability of erroneous detection and ROC curves show that the performance of the proposed detector is good. Plus, it has a low computational complexity.

1 Introduction

It is well known that static frequency allocation has lead to the scarcity of spectral resources. Assigning fixed frequency bands to ever-evolving new applications is both expensive as well as unfeasible. However, this necessitates very large sampling rates (proportional to the spectrum bandwidth) which can heavily stress analog-to-digital converters (ADCs) in terms of power consumption. Recently, the theory of compressed sampling (CS) [3] has received considerable attention among research community as a means to reduce the sampling-rate constraints on the design of CR systems. In the context of CRs, CS is based on the fact that given the sparsity of the signal in the frequency domain, sampling rates can be made significantly lower than the Nyquist rate without loosing much information. This may potentially facilitate simpler implementation of the ADCs and digital processors.

In the traditional CS framework, signal needs to be recovered from its compressed samples. A plethora of algorithms are available to provide reliable recovery of the sparse signal: matching pursuit (MP), orthogonal matching pursuit (OMP) [4], compressive sampling matching pursuit (CoSaMP) [5], basis pursuit (BP) [6], least absolute shrinkage and selection operator (LASSO) [7]. Note that most of these algorithms are quite complex and often consume a lot of computational resources. However, signal reconstruction may not be necessary in many signal processing applications as one may only be interested in solving an inference problem. Davenport et al. have demonstrated that it is possible to tackle the problem of detecting a known signal buried in noise, i.e., classification of the signals, directly in the compressive domain without first resorting to a complex signal reconstruction [8], [9]. Many other works focusing on solving a detection problem directly from the compressed samples are also available, e.g., [10], [11]. Continuing this direction of research, we focus on developing efficient detectors for CR systems, based on compressed measurements.

We have recently extended [9] to the optimal maximum likelihood (ML) detection of linearly modulated signals of unknown parameters occupying unknown frequency subchannels [12] by using the compressed measurements only. We basically focused on the

functionality of an individual CR. However, the use of distributed spectrum sensing algorithms is recommended to cope with the fading phenomenon present in all wireless communications systems. In this paper, we basically combine our proposed approach in [12] with the distributed signal processing, where instead of focusing on a single CR, multiple CRs generate compressed measurements which are then transmitted to the fusion centre (FC). This results in saving sensing resources at individual CRs as well as reduces the capacity requirements of the control channels. This is in contrast with the existing contributions in literature which aim at reconstructing the wideband signal spectrum at the FC by defining a sparsity model common to all sensing nodes [16], [14]. We propose the detection of a primary signal directly in the compressive domain which does not rely on signal reconstruction from its compressed measurements. The proposed multiband signal detection procedure is optimized according to the ML detection criterion for a distributed CR scenario. A closed-form expression of the detection metric is obtained by carefully approximating the likelihood function. We demonstrate that the CS analog-to-information converter (AIC) can advantageously be implemented by a distributed network of sensing nodes. In this paper we consider two different realizations of the AIC. The first is a modulated wideband converter (MWC) [15], where the received signal is first multiplied with a high-rate binary spread code and then sampled at a low sampling rate after a low-pass filtering stage. The second is a non-uniformly periodic (NUP) sampling, where the received signal is delayed then sub-sampled. Each sensing node implements a branch of MWC or the NUP. The sensing nodes then transfer the low-rate sample sequence of the received signal to the FC for multiband detection. The performance of the proposed detector is exhibited by means of numerical simulations for probability of erroneous detection and receiver operating characteristic curves.

2 System model

The primary network consists of multiple mobile terminals communicating to a base station (uplink transmission). The overall bandwidth is divided into M frequency bands that may be allocated to different terminals for their communication in a frequency division multiple access (FDMA) fashion. Each frequency band has a bandwidth $\frac{1}{T}$, where T is the symbol duration for each user. We assume there are $K \leq M$ active terminals in the network. The secondary network are mobile platforms that embed a MWC or NUP AIC and send all the information to the fusion centre, where a maximum likelihood based detection algorithm is applied. Since both AIC are meant to be implemented on USRP2 platforms, the following system model comprises imperfections that might occur in a real-time demonstration: carrier frequency offset (CFO), sampling clock offset (SCO), channel effect, etc.

2.1 Transmitter architecture

Fig. 1 describes the transmitter architecture for primary user k. Symbols I_k are convoluted with a halfroot Nyquist filter $g[n]$ after being upsampled by a factor M. The modulated symbols are then converted to an analog signal $s_k(t)$ at rate $\frac{1}{T}$. The result is

$$s_k(t) = \sum_{n=1}^{N} I_k[n]\ g(t - nT). \tag{1}$$

The modulated signal is then shifted to the allocated band of center frequency Δf_k, and then passed through a typical analog front-end. The baseband signal transmitted

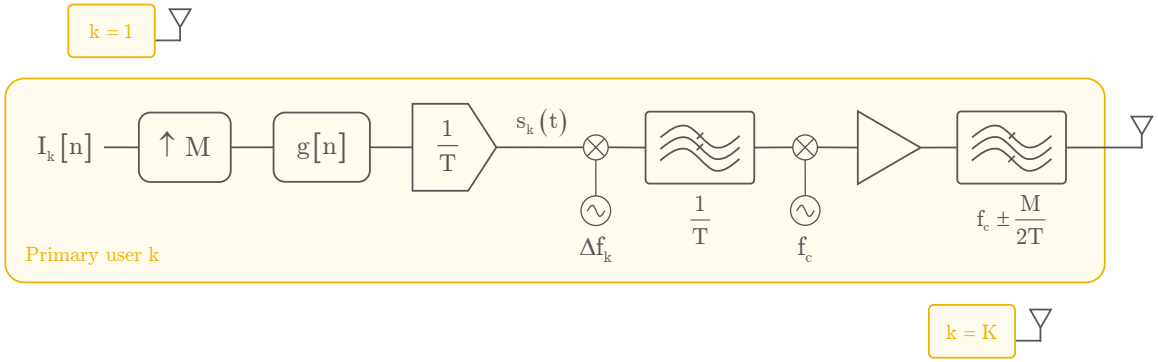


Figure 1: Primary user transmitter architecture

by the primary user k is expressed as $s_k(t)\ e^{j2\pi\Delta f_k t}$.

At any sensing node, the received signal is the sum of the signals for all contributing primary users, that is

$$s(t) = \sum_{k=1}^{K} s_k(t)\ e^{j2\pi\Delta f_k t}. \tag{2}$$

At the sensing node q, this signal is corrupted by a channel response $c_q(t)$ and additive white Gaussian noise $w_q(t)$. The received signal at node q is then

$$x_q(t) = c_q(t) * s(t) + w_q(t). \tag{3}$$

Fig. 2 represents the spectrum of this signal, assuming that $K = 3$ and that the channel response is flat in the band of interest.

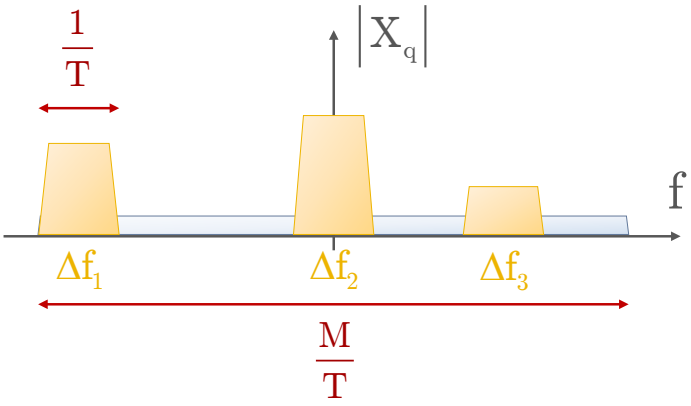


Figure 2: Spectrum of received signal $x_q(t)$ at node q

2.2 Receiver architecture

Fig. 3 describes the receiver architecture for secondary user q. Receiver analog front-end comprises a band-pass filter, a low noise amplifier, an automatic gain control, a

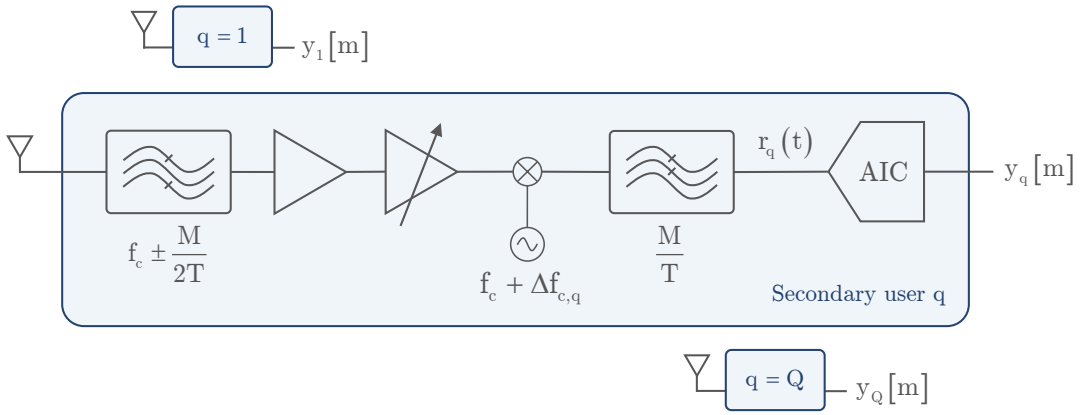


Figure 3: Secondary user receiver architecture

down-converter and a low-pass filter. The latter impulse response is denoted $h_{\text{LPF}}(t)$. Note that the down-converter center frequency might generally not match that of the transmitter, hence the CFO term $\Delta f_{c,q}$. The resulting received signal is

$$r_q(t) = x_q(t)\ e^{j2\pi\Delta f_{c,q}t} * h_{\text{LPF}}(t) = c_q(t) * s(t)\ e^{j2\pi\Delta f_{c,q}t} * h_{\text{LPF}}(t) + v_q(t), \tag{4}$$

where

$$v_q(t) = w_q(t)\ e^{j2\pi\Delta f_c t} * h_{\text{LPF}}(t) \tag{5}$$

is the noise restricted to the band of interest $\frac{M}{T}$. Since the sensing nodes focus their detection on all available sub-bands, we further only consider the power $\sigma^2_{v_q}$ of the noise in the overall bandwidth $\frac{M}{T}$. The power σ^2_s of the signal $s(t)$ depends on the number of active primary users K. Since we want to compare our algorithms performance for different values of K, we need to define the SNR so that it is independent of the number of that parameter. Thus, we further only consider the average signal power per sub-band $\left\langle \sigma^2_{s_k} \right\rangle = \frac{\sigma_s}{K}$. In our case, the SNR is defined as follows :

$$\text{SNR} = \frac{\left\langle \sigma^2_{s_k} \right\rangle}{\sigma^2_{v_q}}. \tag{6}$$

As shown in Fig. 3, the received signal is then passed through an AIC. In the following sections, we consider two different AIC implementations : NUP and MWC. We also show that the output signal $y_q[n]$ of both AIC can be expressed in the same model.

2.2.1 NUP AIC

Fig. 4 describes the AIC architecture for NUP. It simply waits a time $\tau_{\text{NUP},q}$ before sub-sampling at rate

$$T_s = \frac{c}{M}T + \Delta T_s \tag{7}$$

where c is the sub-sampling factor, and ΔT_s is the sampling clock offset (SCO) between the transmitter and the receiver (assuming all transmitters clocks are frequency synchronized). Not that to achieve a non-uniformly periodic sampling, $\tau_{\text{NUP},q}$ has to be strictly different for each sensing node.
The AIC output is given by

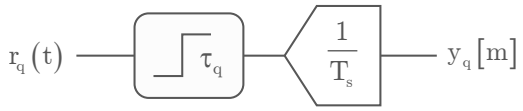


Figure 4: NUP AIC architecture

$$y_q[m] = r_q(t)\Big|_{t=mT_s-\tau_q} = \alpha_q \; s(mT_s - \tau_q) \; e^{j2\pi\Delta f_{c,q}(mT_s-\tau_q)+\varphi_q} + v_q(mT_s - \tau_q). \tag{8}$$

where α_q and φ_q represents the effect of the flat channel. In the expression of the delay

$$\tau_q = \tau_{p,q} + \tau_{s,q} + \tau_{\text{NUP},q}, \tag{9}$$

$\tau_{p,q}$ is the propagation delay, $\tau_{s,q}$ is the phase difference between transmitter and receiver sampling clocks (assuming all transmitters clocks are phase synchronized), and $\tau_{\text{NUP},q}$ is the artificial delay specific to the NUP AIC. Giving (2), the output (8) can be written in the general form

$$y_q[m] = \sum_{k=1}^{K} c_{k,q}[m] s_{k,q}[m] + n_q[m] \tag{10}$$

where

$$c_{k,q}[m] \triangleq \alpha_q \; e^{j2\pi(\Delta f_k+\Delta f_{c,q})(mT_s-\tau_q)+\varphi_q} \tag{11}$$

$$s_{k,q}[m] \triangleq s_k(mT_s - \tau_q) \tag{12}$$

$$n_q[m] \triangleq v_q(mT_s - \tau_q) \tag{13}$$

2.2.2 MWC AIC

Fig. 5 describes the AIC architecture for MWC. The received signal is first mixed with a chip sequence $p_q(t)$. The sequence is T-periodic and can thus be expanded as a Fourier series, i.e.,

$$p_q(t) = \sum_{m=-\infty}^{+\infty} c_q[m] \; e^{j2\pi m\frac{t}{T}} \tag{14}$$

where $c_q[m]$ are the Fourier coefficients of $p_q(t)$ expansion. The mixing of (14) and (4) implies the product

$$p_q(t)s(t) = \sum_{m=-\infty}^{+\infty} \sum_{k=1}^{K} c_q[m] \; e^{j2\pi m\frac{t}{T}} \; s_k(t) \; e^{j2\pi\Delta f_k t}. \tag{15}$$

By defining m_k as

$$\Delta f_k \triangleq \frac{m_k}{T}, \tag{16}$$

this product becomes

$$p_q(t)s(t) = \sum_{m=-\infty}^{+\infty} \sum_{k=1}^{K} c_q[m]s_k(t) \; e^{j2\pi(m_k+m)\frac{t}{T}}, \tag{17}$$

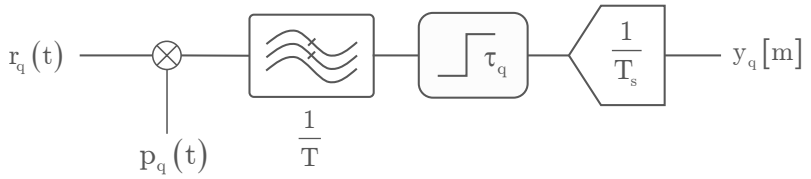


Figure 5: MWC AIC architecture

which means that each occupied sub-band m_k is now shifted in every other part of the spectrum and summed with other PU contribution. In particular, after the low-pass filter, we only keep the contributions in (18) for every $m = -m_k$. The product becomes

$$p_q(t)s(t) = \sum_{k=1}^{K} c_q[-m_k]s_k(t). \tag{18}$$

Thus, the output of the AIC is

$$y_q[m] = r_q(t)p_q(t) * h_T(t)\Big|_{t=mT_s-\tau_q} \tag{19}$$

where $h_T(t)$ is the impulse response of a low-pass filter of bandwidth $\frac{1}{T}$, and

$$T_s = T + \Delta T_s. \tag{20}$$

Similarly to (8), and with (18), we find

$$y_q[m] = \alpha_q \sum_{k=1}^{K} c_q[-m_k]s_k(mT_s - \tau_q)\ e^{j2\pi\Delta f_{c,q}(mT_s-\tau_q)+\varphi_q} + n_q(mT_s - \tau_q), \tag{21}$$

where $n_q(mT_s-\tau_q) = v_q(t)p_q(t)*h_T(t)\Big|_{t=mT_s-\tau_q}$ is still an AWGN. (21) can be expressed in the form of (10) if we define

$$c_{k,q}[m] \triangleq \alpha_q\ c_q[-m_k]\ e^{j2\pi\Delta f_{c,q}(mT_s-\tau_q)+\varphi_q} \tag{22}$$

$$s_{k,q}[m] \triangleq s_k(mT_s - \tau_q) \tag{23}$$

$$n_q[m] \triangleq n_q(mT_s - \tau_q) \tag{24}$$

2.2.3 Common model

If we assume we only keep W samples from the AIC output $y_q[m]$ of each sensing node q, (10) can be expressed in the following matrix form :

$$\mathbf{y}_q = \sum_{k=1}^{K} \mathbf{C}_{k,q}\ \mathbf{s}_{k,q} + \mathbf{n}_q \tag{25}$$

where

$$\mathbf{y}_q \triangleq [y_q[0], \ldots, y_q[W-1]]^T \tag{26}$$

$$\mathbf{s}_{k,q} \triangleq [s_{k,q}[0], \ldots, s_{k,q}[W-1]]^T \tag{27}$$

$$\mathbf{n}_q \triangleq [n_q[0], \ldots, n_q[W-1]]^T \tag{28}$$

$$\mathbf{C}_{k,q} \triangleq \begin{bmatrix} c_{k,q}[0] & & 0 \\ & \ddots & \\ 0 & & c_{k,q}[W-1] \end{bmatrix} \tag{29}$$

This model is valid for both AICs.

2.3 Information gathering at the FC

If we concatenate all AIC outputs from all Q sensing nodes at the FC, we obtain

$$\mathbf{y} = \sum_{k=1}^{K} \mathbf{C}_k \ \mathbf{s}_k + \mathbf{n} \tag{30}$$

where

$$\mathbf{y} \triangleq \left[\mathbf{y}_1^T \cdots \mathbf{y}_Q^T\right] \tag{31}$$

$$\mathbf{s}_k \triangleq \left[\mathbf{s}_{k,1}^T \cdots \mathbf{s}_{k,Q}^T\right] \tag{32}$$

$$\mathbf{n} \triangleq \left[\mathbf{n}_1^T \cdots \mathbf{n}_Q^T\right] \tag{33}$$

$$\mathbf{C}_k \triangleq \begin{bmatrix} \mathbf{C}_{k,1} & & 0 \\ & \ddots & \\ 0 & & \mathbf{C}_{k,Q} \end{bmatrix} \tag{34}$$

3 Distributed Maximum Likelihood detector

If we apply the same method and assume the same hypotheses as in [12] with FC information (30), it is possible to demonstrate that the Maximum Likelihood approximate becomes

$$\left\{\Delta \widehat{f}_k\right\} = \arg \max_{\{\Delta f_k\}} \prod_{k=1}^{K} e^{-\rho_k} \left\| \sum_{q=1}^{Q} \mathbf{C}_{k,q}^H \ \mathbf{y}_q \right\|^2, \tag{35}$$

in which

$$\rho_k = W \frac{\sigma_s^2}{2K\sigma_n^2} \sum_{q=1}^{Q} \sum_{m=0}^{W-1} \left|c_{k,q}[m]\right|^2, \tag{36}$$

where $\frac{\sigma_s^2}{K}$ is the mean power of the signal in one single band. Note that for the NUP AIC, the definition (11) leads to

$$\sum_{m=0}^{W-1} \left|c_{k,q}[m]\right|^2 = W \left|\alpha_q\right|^2 \tag{37}$$

which means that ρ_k is independent of k in that case. For that AIC, the ML criterion becomes

$$\left\{\Delta\widehat{f}_k\right\} = \arg\max_{\{\Delta f_k\}} \left\| \sum_{q=1}^{Q} \mathbf{C}_{k,q}^{H}\ \mathbf{y}_q \right\|^2 \tag{38}$$

4 Performance results

4.1 Simulation setup

The performance of our proposed detector is assessed numerically by computing probability of erroneous detection (PED) and receiver operating characteristic (ROC) curves. PED gives the average error rate. A detection is regarded as erroneous for even a single miss or false detection. For ROC curves, we also provide their theoretical limits. Further, we evaluate the performance of our proposed AIC MWC detector.

Assuming that the activity of each sub-band follows a Bernoulli distribution with probability p and the overall bandwidth is sliced into M uniform sub-bands, the number of enabled sub-bands K follows a binomial distribution, i.e., $B(M, p)$. It is then possible to establish the extreme ROC curves for a perfect detection, i.e., for the noiseless case. In this case, a false alarm (FA) occurs when $\hat{K} > K$ while there is no misdetection (MD), whereas a MD is observed when $\hat{K} < K$ while there is no FA. Thus, the probability of false alarm p_{FA} is given by the expectation of $\frac{\hat{K}-K}{\hat{K}}$ over K and the probability of misdetection p_{MD} is found by computing the expectation of $\frac{K-\hat{K}}{K}$ over K. Both can be analytically computed for each value of $\hat{K}$ ranging from 1 to M.

We consider an overall bandwidth to be sensed as $\frac{M}{T} = 6.25\,\text{MHz}$ which is sliced into $M = 31$ uniform sub-bands of bandwidth 201.6 kHz each. When the number of licensed users K is known and fixed, then $K = 6$. Otherwise, K is a parameter. When evaluating ROC curves, the amount of used sub-bands follows a Bernoulli distribution with probability $p = 0.20$ and thus, the average usage of overall bandwidth is 20%. The total number of sensing node is fixed to $Q = 16$ except when the impact of this parameter on the performance is studied. Each simulation curve is generated by averaging over 1000 to 5000 realizations.

4.2 PED vs SNR

We obtain PED results against different values of SNR. Figure 6 and 7 show the PED simulation results for our proposed MLA based detector and the PSD based detector, respectively, for varying values of sensing nodes. In general, the performance improves with an increase in the number of sensing nodes Q and with a decrease in the number of primary users K. Indeed, increasing the number of sensing nodes increases the average sampling rate and thus, more linear combinations of the enabled sub-bands are available which result in improved performance.

4.3 ROC curves

We generate the ROC curves for different values of $\hat{K}$, i.e., $\hat{K} \in [1; M]$. Figure 8 show the ROC curves for the proposed detector with MWC AIC for varying values of SNR. We also plot the curve for theoretical limit as a reference. Once again, the performance of both the detectors is comparable and also reaches the theoretical limit for very low values of SNR ($-9\,\text{dB}$). However, our proposed detector has an edge in terms of low computational complexity.

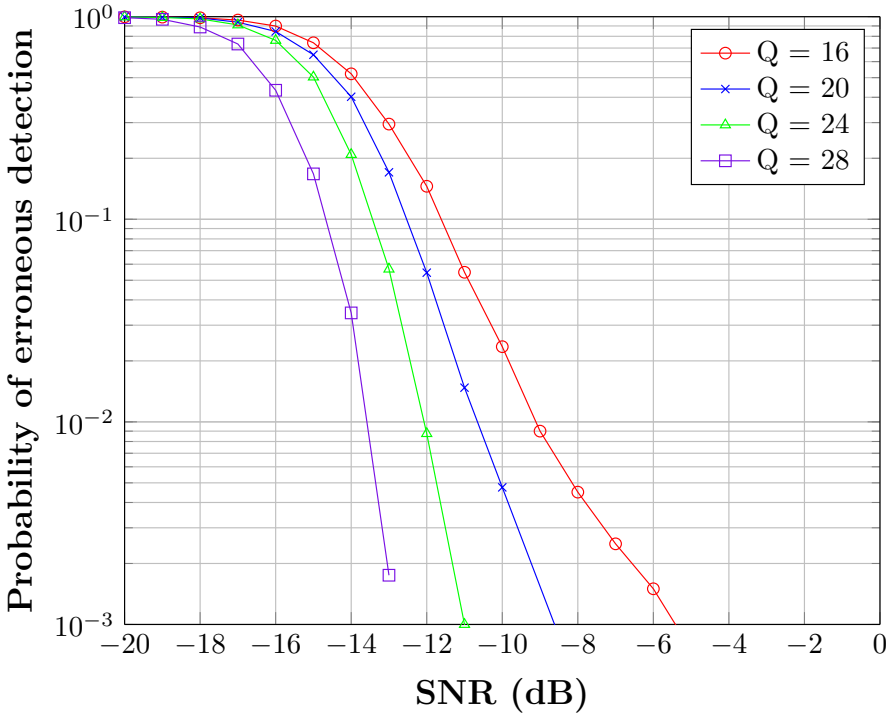


Figure 6: PED versus SNR for $K = 6$

5 Conclusion

In this paper, we have developed a distributed multiband spectrum sensing detector for cognitive radios which is based on compressed measurements and does not rely on signal reconstruction. The detector uses a simplified maximum likelihood metric which is valid for both MWC and NUP AIC and does not require prior signal information. Simulation results for probability of erroneous detection and ROC curves show that the performance of the proposed detector is good. Plus, it has a low computational complexity.

References

[1] A. Sahai AND D. Cabric, "Spectrum Sensing: Fundamental Limits and Practical Challenges", IEEE Workshop on Networking Technologies for Software Defined Radio Networks, 2006

[2] A. Sahai AND D. Cabric, "Spectrum Sensing: Fundamental Limits and Practical Challenges", IEEE Proceedings of DySPAN, 2005

[3] E. J. Candes AND M. B. Waykin, "An Introduction to Compressive Sampling", IEEE Signal Processing Magazine volume 25 number 2 pages 21-30, 2008

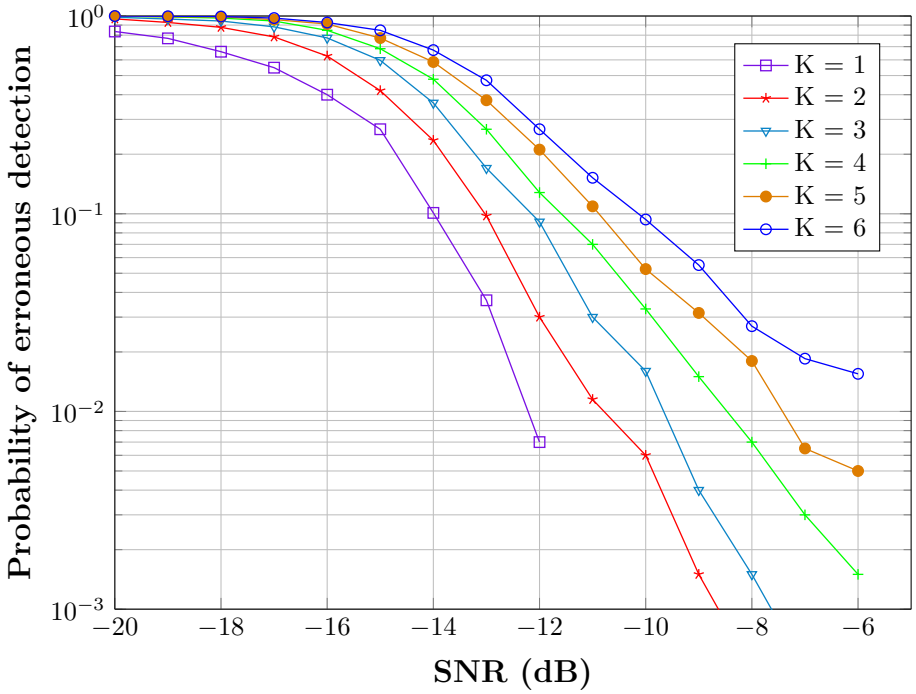


Figure 7: PED versus SNR for $Q = 16$

[4] J. A. Tropp AND A. C. Gilbert, "Signal Recovery from Partial Information via Orthogonal Matching Pursuit", IEEE Transactions on Information Theory volume 53 number 12 pages 4655 - 4666, 2007

[5] Deanna Needell and Joel A. Tropp, "Cosamp: iterative signal recovery from incomplete and inaccurate samples", Commun. ACM volume 53 number 12 pages 93 100, 2010

[6] Scott Shaobing Chen, David L. Donoho, Michael, and A. Saunders, "Atomic decomposition by basis pursuit", SIAM Journal on Scientific Computing volume 20 pages 33-61, 1998

[7] Robert Tibshirani, "Regression shrinkage and selection via the lasso", Journal of the Royal Statistical Society, Series B volume 58 pages 267-288, 1994

[8] M. A. Davenport AND M. Wakin AND R. Baraniuk, "Detection and Estimation with Compressive Measurements", Technical report, University of Rice, 2006

[9] M. A. Davenport AND P. T. Boufounos AND M. B. Wakin AND R. G. Baraniuk, "Signal Processing with Compressive Measurements", IEEE Journal on Selected Topics in Signal Processing volume 4 number 2 pages 445-460, 2010

[10] S. Gishkori and G. Leus, "Compressive sampling based energy detection of ultra-wideband pulse position modulation", IEEE Transactions on Signal Processing volume 61 number 15 pages 3866-3879, 2013

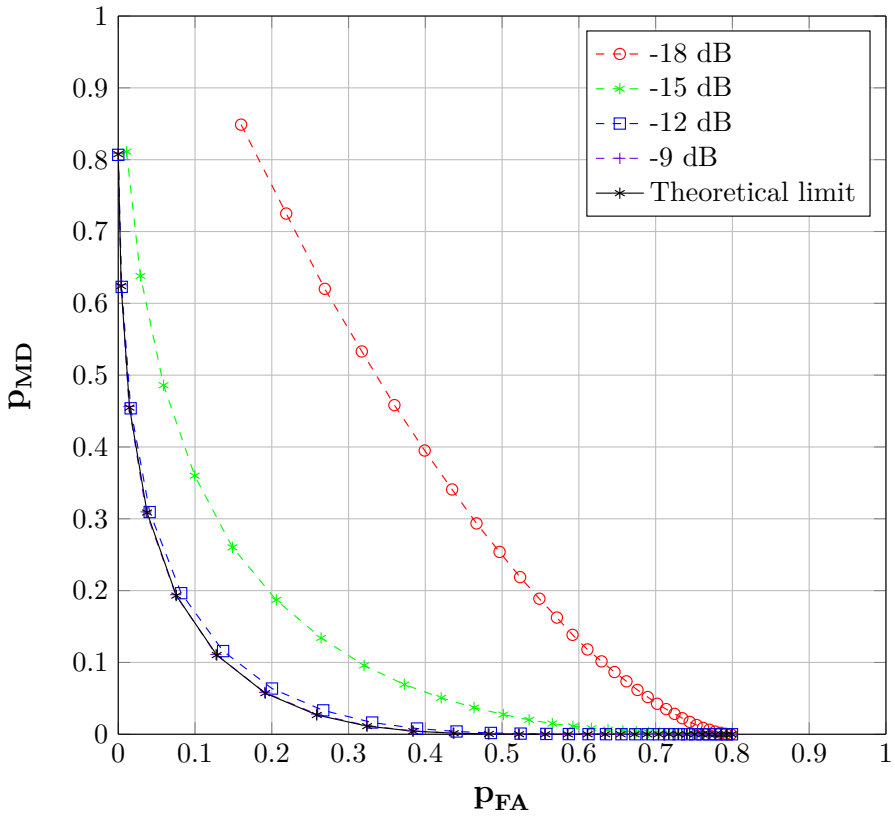


Figure 8: ROC curves for $Q = 16$

[11] S. Gishkori, V. Lottici, and G. Leus, "Compressive sampling-based multiple symbol differential detection for uwb communications", IEEE Transactions on Communications volume 13 number 7 pages 3778-3790, 2014

[12] J. Verlant-Chenet AND A. Bourdoux AND J.-M. Dricot AND P. De Doncker AND F. Horlin, "Multiband Maximum Likelihood Signal Detection based on Compressive Measurement", IEEE Proceedings of Globecom, 2012

[13] M. F. Duarte AND S. Sarvotham AND D. Baron AND M. B. Wakin AND R. G. Baraniuk, "Distributed Compressed Sensing of Jointly Sparse Signals", IEEE Proceedings of Asilomar, 2005

[14] Y. Wang AND A. Pandharipande AND Y. L. Polo, G. Leus, "Distributed Compressive Wide-Band Spectrum Sensing", IEEE Proceedings of ICASSP, 2010

[15] M. Mishali, A Elron, and Y.C. Elda, "Sub-nyquist processing with the modulated wideband converter", 2010 IEEE International Conference on Acoustics Speech and Signal (ICASSP), 2010

[16] DD Arianandа and Geert Leus, "A Study on Cooperative Compressive Wideband Power Spectrum Sensing", 33rd WIC Symposium on Information Theory in the Benelux, 2012

Understanding high-order correlations using a synergy-based decomposition of the total entropy

Fernando Rosas*, Vasilis Ntranos†, Christopher J. Ellison‡,
Marian Verhelst* and Sofie Pollin*

* Departement Elektrotechniek, KU Leuven
† Communication Sciences Institute, University of Southern California
‡ Center for Complexity and Collective Computation, University of Wisconsin-Madison
fernando.rosas@esat.kuleuven.be

Abstract

The interactions between three or more variables are frequently nontrivial, poorly understood, and yet, are paramount for future advances in fields such as multiuser information theory, neuroscience, and genetics. We introduce a novel framework that characterizes the ways in which random variables can share information, based on the notion of information synergy. The framework is then applied to several network information theory problems, providing a more intuitive understanding of their fundamental limits.

1 Introduction

Developing a framework for understanding the correlations that can exist between multiple signals is crucial for the design of efficient and distributed communication systems. For example, consider a network that measures the weather conditions (e.g. temperature, humidity, etc) in a specific region. Given the nature of the underlying processes being measured, one should expect that the sensors will generate strongly correlated data. A haphazard design will not account for these correlations and, undesirably, will process and transmit redundant information across the network.

Higher-order correlations are also of more general interest. In neuroscience, researchers desire to identify how various neurons affect an organism's overall behavior, asking to what extent the different neurons are providing redundant or synergistic signals [1]. In genetics, the interactions and roles of multiple genes with respect to phenotypic phenomena are studied, e.g. by comparing results from single and double knockout experiments [2].

In this work we propose a new framework for understanding complex correlations, which is novel in combining the notion of hierarchical decomposition as developed in [3], with the notion of information synergy as proposed in [4]. In contrast to [3], we focus on the dual total correlation instead of the total correlation, which is more directly related to the shared information within the system. In contrast to [4], we analize the joint entropy instead of the mutual information. Our framework provides new insight to various problems of Network Information Theory. Interestingly, many of the problems of Network Information Theory that have been solved are related to systems which present a simple structure in terms of synergies and redundancies, while most of the open problems possess a more complex mixture of them.

In the following, Section 2 introduces the notions of hierarchical decomposition of correlations and synergistic information, providing the necessary background for an unfamiliar reader. Then, Section 3 presents our decomposition for the joint entropy, focusing on the case of three variables and leaving its generalization for a future work. Section 4 applies this framework in settings of fundamental importance for Network Information Theory. Finally, Section 5 summarizes our main conclusions.

2 Preliminaries

One way of analyzing the interactions between the random variables $\mathbf{X} = (X_1, \ldots, X_N)$ is to study the matrix properties of $\mathcal{R}_{\mathbf{X}} = \mathbb{E}\{\mathbf{X}\mathbf{X}^t\}$. However, this only captures linear relationships and hence the picture provided by $\mathcal{R}_{\mathbf{X}}$ is incomplete. Another possibility is to study the matrix $\mathcal{I}_{\mathbf{X}} = [I(X_i; X_j)]_{i,j}$ of mutual informations. This matrix captures the existence of both linear and nonlinear dependencies, but its scope is restricted to pairwise relationships and thus, it misses all higher-order structure. To see how this can happen, consider two independent fair coins X_1 and X_2 and let $X_3 \coloneqq X_1 \oplus X_2$ be output of an `XOR` logic gate. The mutual information matrix $\mathcal{I}_{\mathbf{X}}$ has all off-diagonal elements equal to zero, making it indistinguishable from an alternative situation where X_3 is another independent fair coin.

For the case of $\mathcal{R}_{\mathbf{X}}$, the standard next step would be to consider higher order moment matrices such co-skewness and co-kurtosis. We seek their information-theoretic analogs, which complement the description provided by $\mathcal{I}_{\mathbf{X}}$. One method of doing this is by studying the information contained in marginal distributions of increasingly larger sizes; this approach is presented in Section 2.1. Other methods try to provide a direct representation of the information that is shared between the various random variables; they are discussed in Section 2.2.

2.1 Negentropy and total correlation

When the random variables that compose a system are independent, their joint distribution is given by the product of their marginal distributions. Hence, in this case the marginals contain all that is to be learned about the statistics of the entire system. However, arbitrary joint p.d.f.s can contain information that is not present in their marginals. To quantify this idea, let us consider N discrete random variables $\mathbf{X} = (X_1, \ldots, X_N)$ with joint p.d.f. $p_{\mathbf{X}}$, where each X_j takes values in a finite set with cardinality Ω_j. The maximal amount of information that could be stored in any such system is $H^{(1)} = \sum_j \log \Omega_j$, which corresponds to the entropy of the p.d.f. $p_{\mathbf{U}} \coloneqq \prod_j \overline{p}_{X_j}$, where $\overline{p}_{X_j}(x) = 1/\Omega_j$ is the uniform distribution for each random variable X_j. On the other hand, the joint entropy $H(\mathbf{X})$ with respect to the true distribution $p_{\mathbf{X}}$ measures the actual uncertainty that the system possesses. Therefore, the difference $\mathcal{N}(\mathbf{X}) \coloneqq H^{(1)} - H(\mathbf{X})$ corresponds to the decrease of the uncertainty about the system that occurs when one learns its p.d.f. – i.e. the information about the system that is contained in its statistics. This quantity is known as *negentropy* [5], and can be also computed as

$$\mathcal{N}(\mathbf{X}) = D\big(\textstyle\prod_j p_{X_j} \,\big|\big|\, p_{\mathbf{U}}\big) + D\big(p_{\mathbf{X}} \,\big|\big|\, \prod_j p_{X_j}\big) \ , \tag{1}$$

where p_{X_j} is the marginal of the variable X_j and $D(\cdot||\cdot)$ is the Kullback-Leibler divergence. In this way, (1) decomposes the negentropy into a term that corresponds to the information given by simple marginals and a term that corresponds to higher order marginals. The second term is known as the *Total Correlation* (TC) and has been suggested as an extension of the notion of mutual information for multiple variables.

An elegant framework for decomposing the TC can be found using the framework presented in [3]. Let us call k-marginals the distributions that are obtained by marginalizing the joint p.d.f. over $N - k$ variables. In the case where only the 1-marginals are known, the simplest guess for the joint distribution is $\tilde{p}^{(1)}_{\mathbf{X}} = \prod_j p_{X_j}$. One way of generalizing this for when the k-marginals are known is by using the *maximum entropy principle*, which suggests to choose the distribution that maximizes the joint entropy while satisfying the constrains given by the partial (k-marginal) knowledge. Let us denote by $\tilde{p}^{(k)}_{\mathbf{X}}$ the p.d.f. which achieves a maximum entropy while being consistent with the k-marginals, and let $H^{(k)} = H(\{\tilde{p}^{(k)}_{\mathbf{X}}\})$ denote its entropy. Then, it can be

showed the following generalized Pythagorean relationship for the total correlation:

$$\mathrm{TC} = \sum_{k=2}^{N} D(\tilde{p}^{(k)}||\tilde{p}^{(k-1)}) = \sum_{k=2}^{N} \left(H^{(k-1)} - H^{(k)}\right) \triangleq \sum_{k=2}^{N} \Delta H^{(k)} \ . \tag{2}$$

Above, $\Delta H^{(k)} \geq 0$ measures the information that is provided by the k marginals and not by the $k-1$ ones. In general, information located in terms with high values of k correspond to complex correlations between many variables, which cannot be reduced to a combination of simpler correlations between smaller groups.

2.2 Yeung's decomposition and synergistic information

Another approach to study the correlations between many random variables is to analyze the way in which they share information, which can be done by decomposing the joint entropy of the system. For the case of two variables, the joint entropy can be decomposed as $H(X_1, X_2) = I(X_1; X_2) + H(X_1|X_2) + H(X_2|X_1)$, suggesting that it can be divided into shared information, $I(X_1; X_2)$, and informations that are exclusively located in just one variable, $H(X_1|X_2)$ and $H(X_2|X_1)$. In systems with more than two variables, one can still compute the information that is exclusively located in one element as $H_{(1)} \coloneqq \sum_j H(X_j|\mathbf{X}_j^c)$, where $\mathbf{X}_j^c$ denote all the system variables except X_j. The difference between the joint entropy and the sum of informations contained in just one location defines the *Dual Total Correlation* (DTC),

$$\mathrm{DTC} = H(\mathbf{X}) - H_{(1)}, \tag{3}$$

which measures the portion of the joint entropy that is shared between two or more variables of the system. As in (2), it would be appealing to look for a decomposition of the DTC of the form $\mathrm{DTC} = \sum_{k=2}^{N} \Delta H_{(k)}$, where $\Delta H_{(k)} \geq 0$ would measure the information that is shared by k variables.

One possible decomposition for the DTC is provided by the *I-measure* [6]. For the case of three variables, this decomposition can be written as

$$\mathrm{DTC}_{N=3} = [I(X_1; X_2|X_3) + I(X_2; X_3|X_1) + I(X_3; X_1|X_2)] + I(X_1; X_2; X_3) \ . \tag{4}$$

The last term is known as the *co-information* [7] and can be calculated as $I(X_1; X_2; X_3) = I(X_1; X_2) - I(X_1; X_2|X_3)$, being other candidate for extending the mutual information to multiple variables. Although it is tempting to associate the term in square brackets of (4) with $\Delta H_{(2)}$ and the co-information with $\Delta H_{(3)}$, this would not be very intuitive since the co-information can be negative. Conventionally, we think of the conditional mutual information as the information contained in X_1 and X_2 that is not contained in X_3, but this quantity should be strictly less than the total information shared by X_1 and X_2. The counterintuitive fact that sometimes $I(X_1; X_2) \leq I(X_1; X_2|X_3)$ suggests that the conditional mutual information can capture information that extends beyond X_1 and X_2, incorporating higher-order effects with X_3.

An extended treatment of the conditional mutual information and it relationship with the mutual information can be found in [4]. For presenting those ideas, let us consider two random variables X_1 and X_2 which are used to predict X_3. The total predictability, i.e., the information X_1 and X_2 provide about X_3, is given by $I(X_1, X_2; X_3) = I(X_1; X_3) + I(X_2; X_3|X_1)$. Is natural to think that the information provided by X_1, $I(X_1; X_3)$, can be unique or redundant with respect of the information provided by X_2. On the other hand, $I(X_2; X_3|X_1)$ must contain the unique contribution of X_2. However, the fact that $I(X_2; X_3|X_1)$ can be larger than $I(X_2; X_3)$ (while the latter contains both the unique and redundant contributions of X_2) suggests that

there can be an additional predictability that is accounted only by the conditional mutual information. This predictability, which is not contained in any single predictor but is only revealed by both X_1 and X_2, is called *synergistic mutual information*. As an example of this, consider again the case in which X_1 and X_2 are independent random bits and $X_3 = X_1 \oplus X_2$. Then, it can be seen that $I(X_1; X_3) = I(X_2; X_3) = 0$ but $I(X_1, X_2; X_3) = 1$. Hence, neither X_1 nor X_2 individually provide information about X_3, although together they fully determine it.

Further discussions about the notion of information synergy can be found in [8–11].

3 A non-negative joint entropy decomposition

In this section we present our non-negative decomposition of the joint entropy, which is based on the notion of information synergy. It is important to note that there is an ongoing debate about the best way of characterizing and computing the synergy in arbitrary systems, as the commonly used axioms are not enough for specifying a unique formula [9]. Nevertheless, our approach in this work is to explore how far one can reach based only on the axioms. In this way, our results are going to be consistent to any choice of formula that is consistent with the axioms.

In the following, Section 3.1 presents the axioms of Information Synergy that are used in this work. Then, Section 3.2 will first present the decomposition for an arbitrary system of three variables. Sections 3.2.1 and 3.2.2 specify the decomposition for the important cases of Markov chains and pairwise independent predictors, which provide the basis for the applications explored in Section 4.

3.1 Information synergy axioms

We proceed to determine a number of desired properties that a decomposition of the mutual information should possess. Note that we initially privilege X_3, but our decomposition will end up being symmetric in each random variables.

Definition A decomposition of the mutual information is provided by the functions $I_\cap(X_1X_2; X_3)$, $I_\mathrm{S}(X_1X_2; X_3)$ and $I_\mathrm{un}(X_1; X_3|X_2)$ which satisfy the following axioms:

(1) $I(X_1; X_3) = I_\cap(X_1X_2; X_3) + I_\mathrm{un}(X_1; X_3|X_2)$.

(2) $I(X_1; X_3|X_2) = I_\mathrm{un}(X_1; X_3|X_2) + I_\mathrm{S}(X_1X_2; X_3)$.

(3) *Weak symmetry*: $I_\cap(X_1X_2; X_3) = I_\cap(X_2X_1; X_3)$, $I_\mathrm{S}(X_1X_2; X_3) = I_\mathrm{S}(X_2X_1; X_3)$ and $I_\mathrm{un}(X_1; X_3|X_2) = I_\mathrm{un}(X_3; X_1|X_2)$.

(4) Non-negativity: $I_\cap(X_1X_2; X_3) \geq 0$, $I_\mathrm{S}(X_1X_2; X_3) \geq 0$, and $I_\mathrm{un}(X_1; X_3|X_2) \geq 0$.

Intuitively, $I_\cap(X_1X_2; X_3)$ measures the redundancy of X_1 and X_2 for predicting X_3, $I_\mathrm{un}(X_1; X_3|X_2)$ quantifies the unique information that is provided by X_1 (and not X_2) about X_3, and $I_\mathrm{S}(X_1X_2; X_3)$ is the synergistic mutual information between X_1 and X_2 about X_3. Note that the weak symmetry of the unique information is not strictly necessary for proving our results, but is adopted here because it allows for a more intuitive development of our ideas.

Using the symmetry of the mutual information and Axiom (1), we can show that

$$I_\cap(X_1X_2; X_3) + I_\mathrm{un}(X_1; X_3|X_2) = I(X_3; X_1) = I_\cap(X_3X_2; X_1) + I_\mathrm{un}(X_3; X_1|X_2) \quad (5)$$

Then, by using the weak symmetry of the unique information, it follows that the redundancy also satisfies *strong symmetry*, i.e. $I_\cap(X_1X_2; X_3) = I_\cap(X_3X_2; X_1)$. In a similar way, using the symmetry of the conditional entropy one can show that

$$I_\mathrm{un}(X_1; X_3|X_2) + I_\mathrm{S}(X_1X_2; X_3) = I(X_3; X_1|X_2) = I_\mathrm{un}(X_3; X_1|X_2) + I_\mathrm{S}(X_3X_2; X_1). \quad (6)$$

Using again the weak symmetry of the unique information, one can prove the strong symmetry of the synergy. In order to reflect the strong symmetry of these functions, we will henceforth denote the redundancy and synergy as $I_\cap(X_1; X_2; X_3)$ and $I_\mathrm{S}(X_1; X_2; X_3)$, respectively.

3.2 Decomposition for three variables

Inspired by the non-negative decomposition of the TC, our approach is to build a non-negative decomposition of the joint entropy which is based on a non-negative decomposition of the DTC. For the case of three variables, we let

$$H_{(1)} = H(X_1|X_2, X_3) + H(X_2|X_1, X_3) + H(X_3|X_1, X_2) \tag{7}$$

$$\Delta H_{(2)} = I_\mathrm{un}(X_1; X_2|X_3) + I_\mathrm{un}(X_2; X_3|X_1) + I_\mathrm{un}(X_3; X_1|X_2) \tag{8}$$

$$\Delta H_{(3)} = I_\cap(X_1; X_2; X_3) + 2I_\mathrm{S}(X_1; X_2; X_3) \tag{9}$$

and define the decomposition of the joint entropy as:

$$H(X_1, X_2, X_3) = H_{(1)} + \Delta H_{(2)} + \Delta H_{(3)}. \tag{10}$$

Comparing (10) with (3) yields $\mathrm{DTC} = \Delta H_{(2)} + \Delta H_{(3)}$. Each $\Delta H_{(k)}$ term is non-negative because of Axiom (4), and hence (10) yields a non-negative decomposition of the joint entropy, where each of the corresponding terms captures the information that is shared by one, two or three variables.

In the following, we will analyze two scenarios for which explicit formulas for (8) and (9) can be found.

3.2.1 Markov chains

Let us consider the case in which $X_1 - X_2 - X_3$ form a Markov chain. Because of the conditional independence of X_1 and X_3 with respect to X_2 one has that $I(X_1; X_3|X_2) = 0$. Therefore, by using Axiom (2), it is clear that $I_\mathrm{un}(X_1; X_3|X_2) = 0$, which is consistent with the fact that X_1 and X_3 should not share information that is not also present in X_2. Moreover, using this and Axiom (1), one can find that the redundant information of the Markov chain is $I_\cap(X_1; X_2; X_3) = I(X_1; X_3)$. Using this and Axiom (1), one can show that $I_\mathrm{un}(X_1; X_2|X_3) = I(X_1; X_2) - I(X_1; X_3)$ and $I_\mathrm{un}(X_2; X_3|X_1) = I(X_2; X_3) - I(X_1; X_3)$. Therefore, the information that is shared by pairs of variables in a Markov chain can be found to be

$$\Delta H_{(2)} = I(X_1; X_2) + I(X_2; X_3) - 2I(X_1; X_3) \ . \tag{11}$$

Using again $I(X_1; X_3|X_2) = 0$ and Axiom (2), it is direct to see that $I_\mathrm{S}(X_1; X_2; X_3) = 0$. Therefore, in this case

$$\Delta H_{(3)} = I_\cap(X_1; X_2; X_3) = I(X_1; X_3). \tag{12}$$

3.2.2 Pairwise independent predictors (PIP)

Let us assume that X_1 and X_2 are pairwise independent, and therefore $I(X_1; X_2) = 0$. Then, using Axiom (1), it is direct to see that $I_\mathrm{un}(X_1; X_2|X_3) = I_\cap(X_1; X_2; X_3) = 0$, which in turn allows to show that $I_\mathrm{un}(X_1; X_3|X_2) = I(X_1; X_3)$ and $I_\mathrm{un}(X_2; X_3|X_1) = I(X_2; X_3)$. Therefore, in this case

$$\Delta H_{(2)} = I(X_1; X_3) + I(X_2; X_3) \ , \tag{13}$$

which shows that the positive mutual information terms correspond to information that is shared only by two variables. Using these results and Axiom (2), one can also compute the synergy directly as $I_\mathrm{S}(X_1; X_2; X_3) = I(X_1; X_3|X_2) - I(X_1; X_3) = I(X_1; X_2|X_3)$. Therefore, in this case we have

$$\Delta H_{(3)} = 2I(X_1; X_2|X_3), \tag{14}$$

which measures the correlations between X_1 and X_2 that are introduced by X_3.

4 Applications to Network Information Theory

In this section we will apply the framework presented in Section 3 to develop new intuitions over three fundamental scenarios in Network Information Theory [12]. In the following, Section 4.1 uses the general framework to analyze the Slepian-Wolf coding for three sources, which is a fundamental result in the literature of distributed source compression. Then, Section 4.2 applies the results for PIP to the multiple access channel (MAC), which is one of the fundamental settings in multiuser information theory. Finally, Section 4.3 applies the results for Markov chains to the wiretap channel, which constitutes one of the main models of information-theoretic secrecy.

4.1 Slepian-Wolf coding

The Slepian-Wolf coding gives lower bounds for the data rates that are required to transfer the information contained in various data sources. Let us denote as R_k the data rate of the k-th source and define $\Delta R_k = R_k - H(X_k|\mathbf{X}_k^{\mathrm{c}})$ as the extra data rate that each source has above what is needed for their own exclusive information (c.f. Section 2.2). Then, in the case of two sources X_1 and X_2, the well-known Slepian-Wolf bounds can be re-written as $\tilde{R}_1 \geq 0$, $\tilde{R}_2 \geq 0$, and $\tilde{R}_1 + \tilde{R}_2 \geq I(X_1; X_2)$. The last inequality states that $I(X_1; X_2)$ corresponds to shared information that can be transmitted by any of the two sources.

Let us consider now the case of three sources, and denote $R_S = I_{\mathrm{S}}(X_1; X_2; X_3)$. The Slepian-Wolf bounds provide seven inequalities, which can be re-written as

$$\tilde{R}_i \geq 0, \ i \in \{1, 2, 3\} \tag{15}$$

$$\tilde{R}_i + \tilde{R}_j \geq I_{\mathrm{ex}}(X_i; X_j|X_k) + R_S, \ i, j \in \{1, 2, 3\}, i < j \tag{16}$$

$$\tilde{R}_1 + \tilde{R}_2 + \tilde{R}_3 \geq \Delta H_{(2)} + \Delta H_{(3)} \tag{17}$$

Above, (17) states that all shared information (i.e. the DTC) needs to be accounted by the extra rate of the sources, and (16) that every pair needs to to take care of their unique information and the synergy. Note that, because of (9), the redundancy can be included in only one of the rates while the synergy has to be included in at least two.

4.2 Multiple Access Channel

Let us consider a multiple access channel (MAC), where two pairwise independent transmitters send X_1 and X_2 and a receiver gets X_3 as shown in Fig. 1, forming a PIP system (c.f. Section 3.2.2). It is well-known that, for a given distribution $(X_1, X_2) \sim p(x_1)p(x_2)$, the achievable rates R_1 and R_2 satisfy the capacity constrains $R_1 \leq I(X_1; X_3|X_2)$, $R_2 \leq I(X_2; X_3|X_1)$ and $R_1 + R_2 \leq I(X_1, X_2; X_3)$.

As the transmitted random variables are pairwise independent, one can apply the results of Section 3.2.2. Hence, there is no redundancy and $I_{\mathrm{S}} = I(X_1; X_3|X_2) - I(X_1; X_3)$. Let us introduce a shorthand notation for the remaining three terms : $C_1 = I_{\mathrm{un}}(X_1; X_3|X_2) = I(X_1; X_3)$, $C_2 = I_{\mathrm{un}}(X_2; X_3|X_1) = I(X_2; X_3)$ and $C_S = I_{\mathrm{S}}(X_1; X_2; X_3)$. Then, one can re-write the bounds for the transmission rates as

$$R_1 \leq C_1 + C_S, \quad R_2 \leq C_2 + C_S \quad \text{and} \quad R_1 + R_2 \leq C_1 + C_2 + C_S. \tag{18}$$

From this, it is clear that while each transmitter has an unique portion of the channel with capacity C_1 or C_2, their interaction creates *synergistically* an additional capacity that is given by $C_S = I_{\mathrm{S}}(X_1; X_2; X_3)$.

There exists an interesting relationship between (18) and the bounds provided by Slepian-Wolf coding for two sources A and B. In effect, $H(A|B)$ and $H(B|A)$ correspond to exclusive information contents that needs to be transmitted by each source,

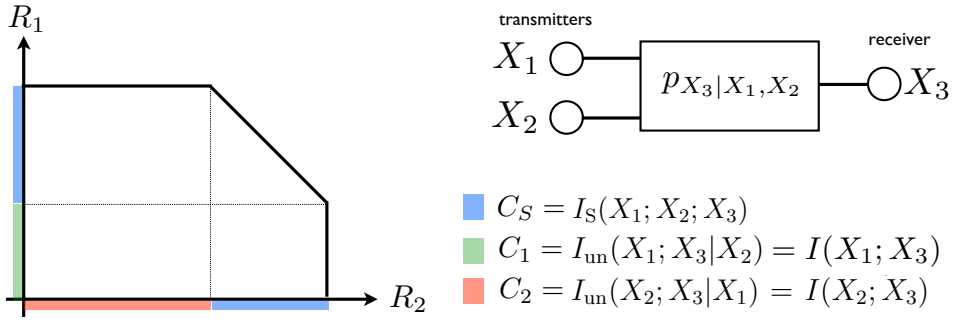


Figure 1: Multiple Access Channel

while C_1 and C_2 are the capacities of the unique portions of the channel that cannot be shared. Also, the mutual information $I(A; B)$ is the information that can be transmitted by either of the variables, while the synergetic capacity C_S corresponds to the part of the channel that can be shared between the users.

4.3 Degraded Wiretap Channel

Consider a communication system with an eavesdropper (shown in Fig. 2), where the transmitter sends X_1, the intended receiver gets X_2 and the eavesdropper receives X_3. For simplicity of the exposition, let us consider the case of a degraded channel where $X_1 - X_2 - X_3$ form a Markov chain. Using the results of Section 3.2.1, one can see that in this case there is no synergy but only redundancy and unique information between X_1 or X_3 with X_2.

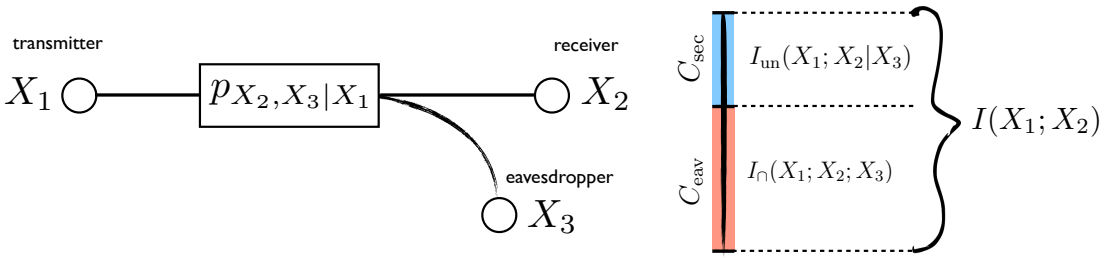


Figure 2: Wiretap Channel

In this scenario, it is known that for a given input distribution p_{X_1} the rate of secure communication that can be achieved is upper bounded by

$$C_{\text{sec}} = I(X_1; X_2) - I(X_1; X_3) = I_{\text{un}}(X_1; X_2|X_3), \tag{19}$$

which is precisely the unique information between X_1 and X_2. Also, as intuition would suggest, the eavesdropping capacity is equal to the redundancy and is given by

$$C_{\text{eav}} = I(X_1; X_2) - C_{\text{sect}} = I(X_1; X_3) = I_\cap(X_1; X_2; X_3). \tag{20}$$

5 Conclusions

We proposed a framework for understanding how multiple random variables can share information, based on a novel decomposition of the joint entropy. We showed how the axioms, on which our framework is based, allow us to find concrete expressions for all the terms of the decomposition for Markov chains and for the case where two variables

are pairwise independent. These results allow for an intuitive understanding of the optimal information-theoretic strategies for several fundamental scenarios in Network Information Theory.

The key insight that this framework provides is that while there is only one way in which information can be shared between two random variables, it can be shared in two different ways between three: redundantly or synergistically. This important distinction has shed new light in the understanding of high-order correlations, whose consequences have only begun to be explored.

Acknowledgments

We want to thank David Krakauer and Jessica Flack for providing the inspiration for this research. We also thank Bryan Daniels, Vigil Griffith and Martin Ugarte for helpful discussions. This work was partially supported by a grant to the Santa Fe Institute for the study of complexity and by the U.S. Army Research Laboratory and the U.S. Army Research Office under contract number W911NF-13-1-0340. FR would also like to acknowledge the support of the F+ fellowship from KU Leuven and the SBO project SINS, funded by the Agency for Innovation by Science and Technology IWT, Belgium.

References

[1] L. Martignon, G. Deco, K. Laskey, M. Diamond, W. Freiwald, and E. Vaadia, "Neural coding: higher-order temporal patterns in the neurostatistics of cell assemblies," *Neural Computation*, vol. 12, no. 11, pp. 2621–2653, 2000.

[2] D. Deutscher, I. Meilijson, S. Schuster, and E. Ruppin, "Can single knockouts accurately single out gene functions?" *BMC Systems Biology*, vol. 2, no. 1, p. 50, 2008.

[3] S.-I. Amari, "Information geometry on hierarchy of probability distributions," *Information Theory, IEEE Transactions on*, vol. 47, no. 5, pp. 1701–1711, Jul 2001.

[4] V. Griffith and C. Koch, "Quantifying synergistic mutual information," in *Guided Self-Organization: Inception*, ser. Emergence, Complexity and Computation, M. Prokopenko, Ed. Springer Berlin Heidelberg, 2014, vol. 9, pp. 159–190.

[5] L. Brillouin, "The negentropy principle of information," *Journal of Applied Physics*, vol. 24, no. 9, pp. 1152–1163, 1953.

[6] R. W. Yeung, "A new outlook on shannon's information measures," *Information Theory, IEEE Transactions on*, vol. 37, no. 3, pp. 466–474, 1991.

[7] A. J. Bell, "The co-information lattice," in *Proceedings of the Fifth International Workshop on Independent Component Analysis and Blind Signal Separation: ICA*, vol. 2003. Citeseer, 2003.

[8] P. L. Williams and R. D. Beer, "Nonnegative decomposition of multivariate information," *arXiv preprint arXiv:1004.2515*, 2010.

[9] V. Griffith, E. K. Chong, R. G. James, C. J. Ellison, and J. P. Crutchfield, "Intersection information based on common randomness," *Entropy*, vol. 16, no. 4, pp. 1985–2000, 2014.

[10] M. Harder, C. Salge, and D. Polani, "Bivariate measure of redundant information," *Physical Review E*, vol. 87, no. 1, p. 012130, 2013.

[11] N. Bertschinger, J. Rauh, E. Olbrich, J. Jost, and N. Ay, "Quantifying unique information," *Entropy*, vol. 16, no. 4, pp. 2161–2183, 2014.

[12] A. El Gamal and Y.-H. Kim, *Network information theory*. Cambridge University Press, 2011.

Phase Synchronisation for FBMC/OQAM Fiber-Optic Transmissions

Mathieu Navaux François Rottenberg Jérôme Louveaux
ICTEAM Université Catholique de Louvain
mathieu.navaux@student.uclouvain.be

Abstract

In this paper, we investigate two phase synchronisation schemes for FBMC/OQAM (filter-bank based multicarriers) in fiber-optic transmissions. We start by investigating pilot-based phase synchronisation, using the so-called auxiliary pilot method [2] consisting in using the pilot's neighbour symbol to compensate for the imaginary intersymbol interference. We derive the maximum likelihood estimator based on the observation of the pilots and analyse its performance. In the second part, taking advantage of the particular shape of the received constellation, a semi-blind channel estimation scheme presented in [3] is adapted to the phase synchronisation issue. It uses the spatial-sign covariance matrix to identify the main variance axis and infer the rotation of the constellation. Both methods are compared and their performance is analyzed by simulations.

1 Introduction

With the rise of more and more powerful signal processing units, it becomes possible to use more complex modulation formats for fiber-optic transmissions. Foreseen as a candidate for 5G, the feasibility of using filter bank based multicarriers (FBMC) for fiber-optic communication is demonstrated in [1]. FBMC, as other multicarriers (MC) scheme, allows to divide a wide bandwidth into several narrowband parallel subchannels. Providing scalability and flexibility when configuring the communication link [2].

In [1], the problems of fiber chromatic dispersion, polarisation mode dispersion and phase noise are addressed. The chromatic dispersion is a-priori known and can be equalised using the overlap and save algorithm [1]. The phase noise results from the phase difference between the transmitter and receiver lasers, which imprints itself on the signal. As a consequence, the received signal shows a phase drift in accordance with the laser linewidths [1]. The study performed on phase noise in [1] focuses mainly on adapting single carrier fiber-optic equalisation techniques to FBMC/OQAM and exhibits limited performance for high number of subcarriers. However, many other approaches like preamble based, training sequences based or blind could be studied. In this paper, the first approach is based on a pilot and an auxiliary pilot used to compensate for the interference from adjacent symbols. An adaptation of the scatter-pilot-based approach presented in [2] is presented. The work in [2] focuses on channel estimation and is here adapted to the issue of phase noise estimation. Then a semi-blind approach taken from [3] is studied. This technique uses the statistical properties of the signal through the spatial sign covariance matrix and identify the rotation introduced by the phase noise. For all these methods, it is assumed that the phase noise is varying sufficiently slowly so that the phase is approximately constant on the duration $T/2$ of an FBMC symbol.

2 System model

Figure 1 shows the general transmultiplexer configuration of the filter-bank multicarrier system (FBMC). The analysis filter banks are located in the transmitter (AFB) and the synthesis filter banks (SFB) in the receiver. In FBMC/OQAM, the complex modulated symbols of duration T are decomposed into their real and imaginary parts. These are transmitted alternatively for a duration of $T/2$ and are denoted by $d_{k,n}$ on figure 1. After multiplication

by $\theta_{k,n} = j^{k+n}$, the purely real $d_{k,n}$ are now alternatively purely real and imaginary according to the OQAM pattern. Then, the samples are upsampled and filtered by the filters $g_k[m]$ obtained thanks to the frequency shift of a prototype filter $p[m]$. The time index n and m are used respectively for the low $2/T$ and high $1/(MT)$ sampling frequency.

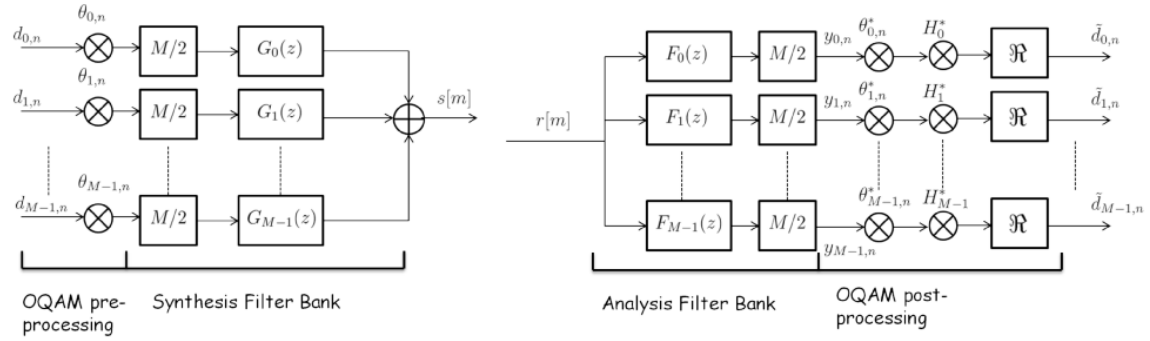


Figure 1: (a) Synthesis and (b) analysis filter banks for complex FBMC transmultiplexer (TMUX)

The output of the SFB can be expressed as:

$$s[m] = \sum_{k=0}^{M-1} \sum_{n=-\infty}^{+\infty} d_{k,n}\theta_{k,n} g_k \left[m - n\frac{M}{2} \right] \tag{1}$$

where

$$g_k[m] = p\,[m] \exp\left[j\frac{2\pi}{M} k \left(m - \frac{L_p - 1}{2} \right) \right] \tag{2}$$

for $k = 0, 1, \ldots, M-1$ and $m = 0, 1, \ldots, L_p - 1$. The subchannel spacing is given by $\Delta f = 1/T$. The received signal in presence of phase noise and additive gaussian noise is

$$r[m] = e^{j\phi[m]}(s[m] * h[m]) + w_m \tag{3}$$

where w_m is a zero mean circularly symmetric complex white Gaussian noise, $e^{j\phi[m]}$ is the phase noise and $h[m]$ is the channel impulse response.

At the receiver, the signal is filtered with $f_k[m] = g_k^*[L_p - 1 - n]$ then downsampled by a factor $M/2$. Assuming a flat channel (hypothesis of perfectly compensated chromatic dispersion) and a slowly varying $\phi[m]$ with regards to the length $L_p = KM - 1$ of $f_k[m]$, the AFB outputs are given by:

$$y_{k,n} = e^{j\phi_n} \sum_{n'=-K}^{K} \sum_{k'=k-1}^{k+1} d_{k',n'}\theta_{k',n'} t_{k-k',n-n'} + w_{k,n} \tag{4}$$

$$= e^{j\phi_n}\theta_{k,n}\left(d_{k,n} + j u_{k,n}\right) + w_{k,n} \tag{5}$$

where $w_{k,n}$ denotes the noise w_m filtered and downsampled at subcarrier k and where $t_{k-k',n-n'}$ is given in Table 1 and denotes the transmultiplexer response of the filter bank. The chosen prototype filter is named C4 [4]. It is designed to minimize the total out of band interference. Its length L_p depends on the size of the filter bank and the overlapping factor K such that $L_p = MK - 1$. The overlapping factor K is set to 4 in this paper.

	n=-4	n=-3	n=-2	n=-1	n=0	n=1	n=2	n=3	n=4
k=-2	0	-0.0006	-0.0001	0	0	0	-0.0001	-0.0006	0
k=-1	0.0049	j0.0422	-0.125	-j0.2065	0.2403	j0.2065	-0.125	-j0.0422	0.0049
k=0	0	0.0658	0.0002	-0.5637	1	-0.5637	0.0002	0.0658	0
k=1	0.0049	-j0.0422	-0.125	j0.2065	0.2403	-j0.2065	-0.125	j0.0422	0.0049
k=2	0	-0.0006	-0.0001	0	0	0	-0.0001	-0.0006	0

Table 1: Interference weight $t_{k,n}$

Table 1 demonstrates the interest of the OQAM modulation. The interference $ju_{k,n}$ on the useful symbol is purely imaginary except for a small negligible term contribution from $t_{k,n\pm2}$ due to the non-perfect reconstruction filter design.

After multiplication by $\theta^*_{k,n}$ and equalisation, only the real part of the signal is kept and the OQAM symbols are recovered.

$$\tilde{y}_{k,n} = \theta^*_{k,n} y_{k,n} \tag{6}$$

$$= e^{j\phi_n} (d_{k,n} + ju_{k,n}) + \tilde{w}_{k,n} \tag{7}$$

and

$$\tilde{d}_{k,n} = \Re\{H^*_n \tilde{y}_{k,n}\} \tag{8}$$

The phase noise $\phi[m]$ can be modeled as

$$\phi(t) = \int_{-\infty}^{t} \delta_\omega(\tau) d\tau \tag{9}$$

where $\delta_\omega(\tau)$ is a zero mean gaussian process with variance $\sigma^2_\phi = 2\pi\Delta_\nu$, where Δ_ν is the 3 dB linewidth of the laser's output and it's value is traditionally around the MHz.

3 Pilot based estimation

Pilots can be used to estimate the phase noise. However, their derivation is not as straightforward as for OFDM. Indeed, in FBMC, the orthogonality only holds for the real components. The auxiliary pilots methods taken from [2] is considered here. It cancels the intrinsic interference at the pilot positions. Then the maximum likelihood estimator for the phase is derived on the basis of the information available at the pilot positions.

Considering a sufficiently frequency selective prototype filter, only adjacent subchannels overlap and (5) holds. Without phase noise, (7) simply becomes:

$$\tilde{y}_{k,n} = (d_{k,n} + ju_{k,n}) + \tilde{w}_{k,n} \tag{10}$$

The expression of the intrinsic interference is given by

$$u_{k_0,n_0} = \sum_{(k,n)\in\Omega_{k_0,n_0}} d_{k,n} \hat{t}_{k,n} \tag{11}$$

where $\hat{t}_{k,n} = \Im\left[\theta^*_{k,n} t_{k,n}\right]$ and Ω_{k_0,n_0} is the set of subcarriers and time indices contributing to the interference at (k_0, n_0). Assuming a pilot is positioned in (k_p, n_p), an auxiliary pilot

positioned in k_a, n_a is computed in order to cancel the intrinsic interference at the pilot position, i.e. such that $u_{k_p,n_p} = 0$:

$$d_{k_a,n_a} = -\frac{1}{\hat{t}_{k_p-k_a,n_p-n_a}} \sum_{\substack{(k,n)\in\Omega_{k_0,n_0} \\ (k,n)\neq(k_p,n_p) \\ (k,n)\neq(k_a,n_a)}} d_{k,n}\hat{t}_{k_p-k,n_p-n}. \tag{12}$$

The auxiliary pilot position is usually chosen as $n_a = n_p+1$ in order to maximize $\hat{t}_{k_p-k_a,n_p-n_a}$.

Thanks to this method, pilots without intrinsic interference are available at the receiver side. From (7), the maximum log-likelihood estimator for the phase is derived at each FBMC symbols of duration $T/2$:

$$\hat{\phi}_n = \underset{\phi_n}{\operatorname{argmax}} \left\{ \sum_{i\in\Theta_n} \ln\left(f\left(\tilde{y}_{i,n}|\phi_n\right)\right) \right\} \tag{13}$$

where Θ_n is the set of subcarriers occupied by a pilot and $f\left(\tilde{y}_{i,n}|\phi_n\right)$ follows a complex Gaussian distribution with mean $\mu = \left(\begin{smallmatrix}\cos\phi_n \\ \sin\phi_n\end{smallmatrix}\right)$ and the same variance as w if the total energy of the filter is 1. This leads to

$$\tan\hat{\phi}_n = \frac{\sum_{i\in\Theta_n} \Im\{\tilde{y}_{i,n}\}}{\sum_{i\in\Theta_n} \Re\{\tilde{y}_{i,n}\}}. \tag{14}$$

Following this method, a phase noise estimate is obtained for each FBMC symbol n that carries pilots on some of its subcarriers. Interpolation is then performed for the other FBMC symbols.

4 Semi-blind channel estimation

Another technique that does not need pilots is presented in [3]. It is based on the different statistical properties of the intrinsic interference and the channel. A theoretical presentation is first made, then the method is adapted for phase noise mitigation.

Starting from (7) and introducing the complex-valued effective transmitted signal $S_{k,n} = d_{k,n} + ju_{k,n}$, the equation is rewritten in a matrix form in the real domain:

$$\underbrace{\begin{bmatrix}\tilde{y}^R_{k,n} \\ \tilde{y}^I_{k,n}\end{bmatrix}}_{\tilde{\mathbf{y}}_{\mathbf{k,n}}} = \begin{bmatrix}h^R_{k,n} & -h^I_{k,n} \\ h^I_{k,n} & h^R_{k,n}\end{bmatrix}\begin{bmatrix}d_{k,n} \\ u_{k,n}\end{bmatrix} + \begin{bmatrix}\eta^R_{k,n} \\ \eta^I_{k,n}\end{bmatrix} \tag{15}$$

where $h_{k,n}$ is the channel impulse response and the exponents R and I denote respectively the real and imaginary part of the corresponding variable.

In absence of phase noise, $d_{k,n}$ takes discrete values on the real axis and $u_{k,n}$ is scattered continuously along the imaginary axis. Under the assumption of slowly varying phase noise, the phase noise only creates a small rotation of this situation. This means that identifying the direction of the highest variance in the real-imaginary plane is equivalent to identifying the phase noise contribution (up to the sign ambiguity). So, the spatial sign covariance matrix can be used to identify the main rotation axis in the constellation. The spatial sign covariance matrix is given by

$$C_{k,n} = \mathbb{E}\{\tilde{\mathbf{y}}'_{\mathbf{k,n}}\tilde{\mathbf{y}}'^{\mathbf{T}}_{\mathbf{k,n}}\} \tag{16}$$

where

$$\tilde{\mathbf{y}}'_{\mathbf{k,n}} = \begin{cases} \frac{\tilde{\mathbf{y}}_{\mathbf{k,n}}}{||\tilde{\mathbf{y}}_{\mathbf{k,n}}||}, & \text{if } ||\tilde{\mathbf{y}}_{\mathbf{k,n}}|| \neq 0 \\ 0, & \text{if } ||\tilde{\mathbf{y}}_{\mathbf{k,n}}|| = 0. \end{cases} \tag{17}$$

The dominant eigenvector of $C_{k,n}$ is the rotation of the signal up to a sign ambiguity,

$$C_{k,n} = V_{k,n}\Sigma_{k,n}V_{k,n}^T \tag{18}$$

where $V_{k,n} = [\mathbf{v^1_{k,n}}, \mathbf{v^2_{k,n}}]$ is an orthogonal matrix and $\Sigma_{k,n}$ =diag(σ_1, σ_2) is a diagonal matrix with $\sigma_1 > \sigma_2$. As previously explained, we can now consider:

$$\begin{bmatrix} h^R_{k,n} \\ h^I_{k,n} \end{bmatrix} = \alpha_n \mathbf{v^1_{k,n}} \qquad \alpha_n \in \{-1, +1\}. \tag{19}$$

The phase noise ϕ_n can be identified from (19) as:

$$e^{j\phi_n} = h^R_{k,n} + jh^I_{k,n} \tag{20}$$

However, in practice, the spatial-sign covariance matrix (SSCM) is not available. Under the assumption of constant phase noise over one FBMC symbol, the SSCM is approximated as

$$\hat{C}_n = \frac{1}{M}\sum_{k=0}^{M-1} \tilde{\mathbf{y}}'_{\mathbf{k,n}}\tilde{\mathbf{y}}'^{\mathbf{T}}_{\mathbf{k,n}}. \tag{21}$$

Eigenvalue decomposition is then performed on $\hat{C}_n$ and an estimate of the phase noise for each FBMC symbol is obtained as presented above. To cope with the sign ambiguity α_n, a first training FBMC symbol must be sent and a tracking must be performed.

5 Simulation results

Simulations are performed on a fixed 30GHz bandwith divided between M subcarriers and with the following parameters:

Type of modulation on $d_{k,n}$		BPSK
Frame size in samples	MP	2^{16}
Sampling frequency	$1/(MT)$	30GHz
Laser's linewidth	Δ_ν	0.5MHz
Prototype filter's overlapping factor	K	4

5.1 Benchmark

Two simulations are performed as benchmark. The first one assumes perfectly corrected phase noise. The second one is meant to check the assumption of a slowly varying phase noise and averages the true value of the phase noise over a duration of $T/2$. The obtained value is then used for phase noise correction over each FBMC symbols of length $T/2$. Note that the sampling frequency is constant, so that increasing the number of subcarriers results in longer FBMC symbols. The assumption of constant phase noise may become less accurate with higer number of subcarriers.

Figure 2 shows the BER as a function of the SNR for both benchmark simulations. In the case of the average correction of the (true) phase noise, a penalty appears at high SNR when the number of subcarriers increases. This is due to the fact that the average is less representative when the size of the time window increases.

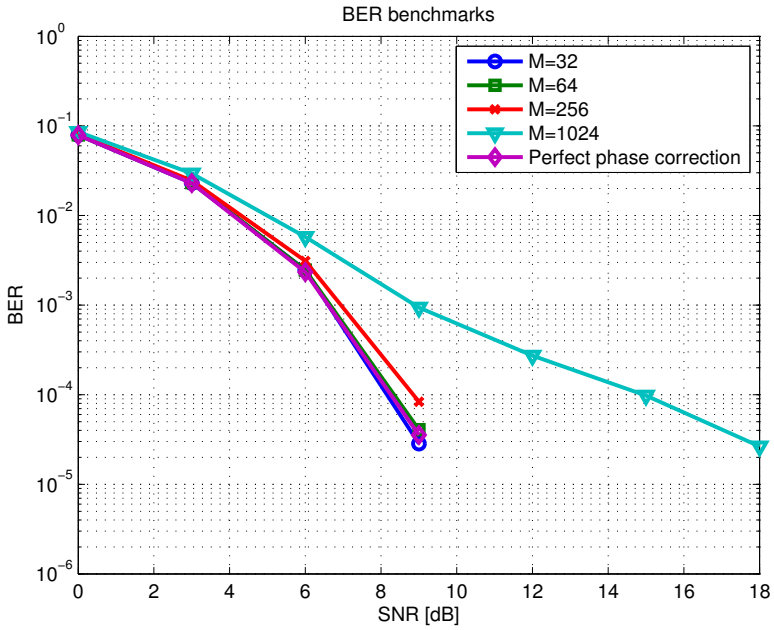


Figure 2: Benchmark's BER

5.2 Pilot based estimation

Using the estimator presented in section 3, it is possible to estimate the phase noise for each FBMC symbol under the assumption of a constant phase rotation for all subcarriers. The pilots are disposed every 5 subcarriers and every $8T$. The FBMC symbols without any pilot are equalised with the same phase as the closest FBMC symbol with pilot (similar results are obtained with linear interpolation between pilots).

The plain line on figure 3 presents the results after equalisation thanks to the pilot method. Good performances are observed at low SNR but at high SNR the floor is high. This effect is particularly strong for a high number of subcarriers. The dashed line on figure 3 shows the BER computed only on FBMC symbols including pilots. The performances are really close to ideal. This shows that most of the error comes from the interpolation between symbols containing pilots, as the phase changes too rapidly to be considered constant.
This leads us to the conclusion that the pilots should be as close as possible in time. However, due to the auxiliary pilots method and the size of the intrinsic interference, the interval between pilots can not be arbitrary small (computing the d_{k_a,n_a} with the contribution of an other auxiliary pilot is really complex). A solution could be to consider specific pilot patterns with changing subcarrier positions keeping in mind the trade-off between the quality of the phase estimate and the time interval between the various estimations. This is left for future work.

5.3 Semi-blind phase estimation

This method presents the advantage of being semi-blind. In theory, it only requires a few known symbols at the beginning of the transmission to resolve the sign ambiguity. In this paper, the issue of sign tracking is not considered and the sign ambiguity is assumed to be perfectly resolved.

Figure 4 shows the results after equalisation thanks to the SSCM method. In order to decrease the impact of the number M of subcarriers on the estimation accuracy, $\hat{C}_n$ is computed

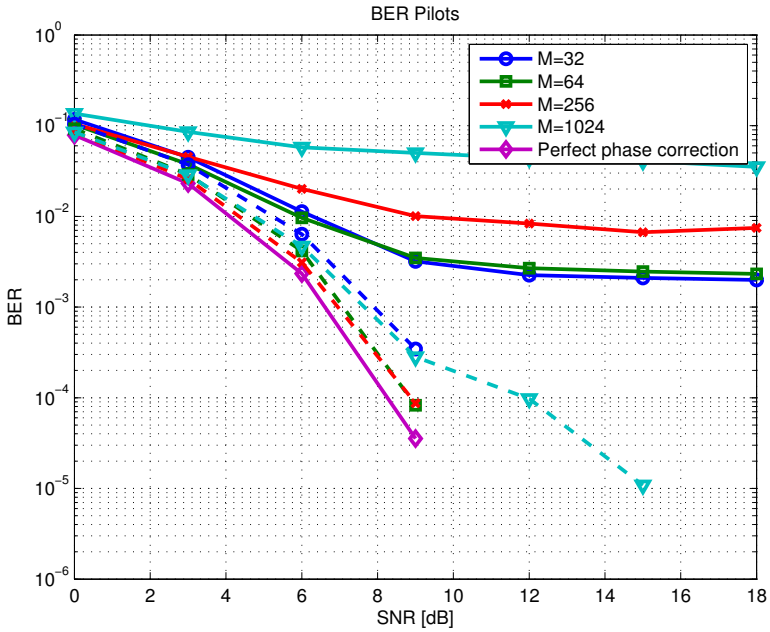


Figure 3: BER with correction thanks to the pilots. Plain line: metric evaluate on every sample. Dashed line: metric for FBMC symbols with pilots

on a fixed number of samples chosen as 1024 for these simulations. The performances are really close to the benchmark with average phase noise correction, showing that the estimation accuracy is very good.

6 Conclusion

Two phase noise equalisation methods for FBMC/OQAM fiber optic transmissions have been studied. The first one rely on the auxiliary pilot method. The performance are good for FBMC symbols with pilots but poor for FBMC symbols without pilots. Closer pilots in time should solve the problem but the intrinsic interference imposes constrains on the relative position of the auxiliary pilots. Varying pilot subcarrier patterns seem a good improvement path.

The second method is semi-blind and only requires a preamble to solve the sign ambiguity. The problem of the sign tracking was not addressed in this paper but the performances considering a known sign provide very good performances if the number of samples used for the spatial-sign covariance matrix estimation is sufficient.

References

[1] J.Fickers, "Modulation formats and digital signal processing for fiber-optic communications with coherent detection", Thesis ULB, 2014.

[2] Stitz, Tobias Hidalgo and Ihalainen, Tero and Viholainen, Ari and Renfors, Markku,"Pilot-Based Synchronization and Equalization in Filter Bank Multicarrier Communications", EURASIP Journal on Advances in Signal Processing, 741429, December & 2009

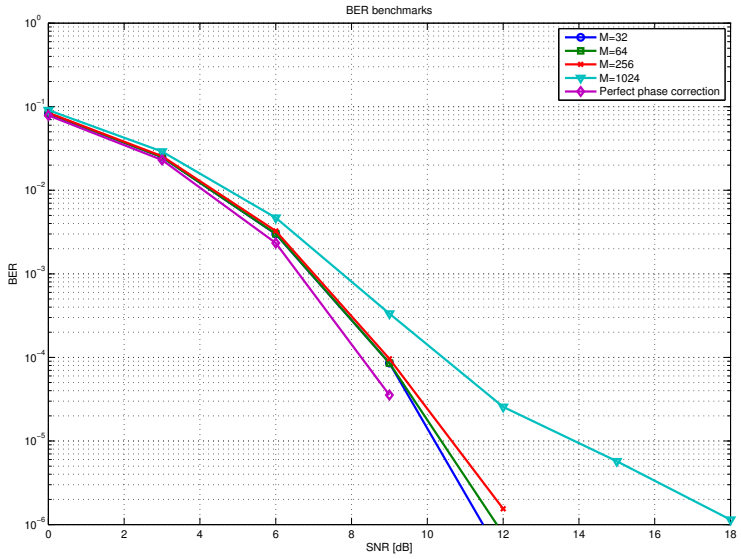


Figure 4: BER for the semi-blind method. $\hat{C}_n$ computed on 1024 samples

[3] Weikun, Hou and Benoit Champagne, "Semiblind Channel Estimation for OFDM/OQAM Systems", IEEE Signal Processing Letters, Vol.22, No.4, April & 2015.

[4] Ari Viholainen and Tero Ihalainen and Tobias Hidalgo Stitz and Markku Renfors and Maurice Bellanger, "Prototype Filter Design for Filter Bank Based Multicarrier Transmission", 17th European Signal Processing Conference, August & 2009.

[5] Jérôme Louveaux, Leonardo Baltar, Dirk Waldhauser, Markku Renfors, Mario Tanda, Carlos Bader and Eleftherios Kofidis, "Equalization and demodulation in the receiver (single antenna)", PHYDYAS D3.1, 2008

[6] Gianni Di Domenico, Stephane Schilt, and Pierre Thomann, "Simple approach to the relation between laser frequency noise and laser line shape", Optical Society of America APPLIED OPTICS, Vol. 49, No. 25, September & 2010

[7] P. Agrawal, "Fiber-Optic Communication Systems", Wiley, 3rd edition, 2002

Spamming the Code Offset Method

Niels de Vreede Boris Škorić
Eindhoven University of Technology
Department of Mathematics and Computer Science
5600 MB Eindhoven, The Netherlands
n.d.vreede@tue.nl b.skoric@tue.nl

Abstract

This is an extended abstract of the work published in [5].
We propose an extension of the Code Offset Method, the 'mother of all Secure Sketches', in which we hide the error correction data in a large list of random decoy values. Secure Sketches are an important ingredient for building privacy-preserving biometric databases. Our scheme, the "Spammed Code Offset Method" (SCOM), improves the level of privacy at the cost of extra storage or computational requirements.

1 Introduction

1.1 Helper Data Schemes

Helper Data Schemes (HDSs) are a security primitive that allows for reliable extraction of secret information from noisy data, e.g. biometric data or data from a physical unclonable function (PUF). They make use of a special form of redundancy information, 'helper data', to correct measurement noise. HDSs can be used to construct e.g. privacy-preserving biometric databases.
The functionality of a generic HDS is shown in Fig. 1. There is an enrollment phase and a reconstruction phase. The enrollment procedure Enroll takes as input a measurement value X and optionally a random value R. The output is helper data W and secret data S. The reconstruction procedure Rec takes the helper data W and a fresh sample X', which is a noisy version of X, and produces $\hat{S}$, which is an estimate of S. If the noise between X and X' is not too large, then $\hat{S} = S$. Furthermore, W should not reveal too much information about the secret, ideally none at all. Secrecy of S is preserved even if W is stored publicly. It is always assumed that attackers have access to W.
Two special types of HDS with additional properties are the fuzzy extractor and secure sketch. A fuzzy extractor requires that the secret is uniformly distributed. For a secure sketch, the secret is identical to the measured value, $S = X$, and no uniformity is required.

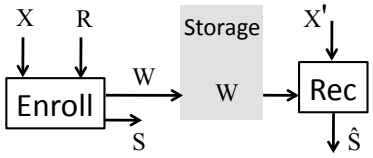


Figure 1: *A generic helper data scheme.*

1.2 The Syndrome-Only Code Offset Method

One of the first introduced helper data schemes is the Code Offset Method (COM)[4, 2]. The COM employs a linear error correcting code to compensate for measurement noise. Below we describe the *Syndrome-Only* COM: a modified version of the COM that is more suitable for our purposes. It additionally requires the existence of an efficient syndrome decoder. The Enroll procedure consists of nothing more than computing a syndrome (Syn),

$$W = \mathsf{Syn}\, X.$$

This construction is a secure sketch, i.e., the secret is the measurement value itself. The X is reconstructed from W and a sample X' as follows:

$$\hat{X} = X' \oplus \mathsf{SDec}(W \oplus \mathsf{Syn}\, X').$$

The linearity of the code ensures that, if X' is sufficiently close to X, then $\hat{X}$ will be equal to X.
In a biometric database, the values stored for each enrolled person would be W and a hash of X. As long as X given W has sufficient entropy, it is infeasible to guess X from the enrolled data.

2 Adding Fake Helper Data

Consider an attacker who tries to guess X given the helper data W. Consider a low-entropy source X, such that the attacker's task is difficult but feasible. We propose to increase the attacker's workload by hiding the real helper in a list of fake entries. If we store m helper data items, only one of which is real, the attacker's average workload increases by a factor of about $m/2$. (For very large values of m, the attacker is even forced to ignore the helper data altogether.) We refer to this technique as *spamming*. The technique can be applied in any HDS, as long as there exists an efficient way to select the true helper data given X'. For the (syndrome only) COM, this is achieved by employing a *Low Density Parity Check* (LDPC) code.
The idea of adding chaff data to hide information is not new [3, 1], but, whereas previous work considered adding chaff points directly to the stored feature vectors, we are the first to apply chaffing in the helper data domain. Our data hiding technique allows us to make more effective use of the source entropy, but it comes at the cost of increased storage requirement or computational workload. An advantage of adding spam in the helper data domain (instead of e.g. X-space) is that it allows for a very precise security analysis.

2.1 The Enrollment and Reconstruction Algorithms

We modify the enrollment procedure such that W is replaced by a list $\mathbf{\Omega}$ of length m. The list $\mathbf{\Omega}$ consists of $m-1$ fake items and W hidden at random secret position Z. One way to do this [5] is to generate the fake items in $\mathbf{\Omega}$ according to the prob. distribution of $\mathsf{Syn}\, X$ and then store the full list. However, this blows up the storage requirements by a factor m. We present an alternative in which $\mathbf{\Omega}$ is generated 'on the fly' from a seed S. We call this the 'generative' Spammed Code Offset Method. The scheme needs a one-way function f and a fast Pseudo Random Number Generator (PRNG) γ that generates uniform bit strings of same length as our code's syndrome. By $\gamma^i(S)$ we denote the i-th string derived from seed S.
Enrollment

1. Measure X.

2. Compute $W = \mathsf{Syn}\, X$.
3. Uniformly draw index $Z \in \{1, \ldots, m\}$.
4. Uniformly draw seed S.
5. Compute mask $B = W \oplus \gamma^Z(S)$.
6. Compute $G = f(S \| B \| X)$.
7. Store public data $P = (S, B, G)$.

We can think of the list $\{B \oplus \gamma^i(S)\}_{i \in \{1,\ldots,m\}}$ as the list $\mathbf{\Omega}$ of fake helper data which contains the real W at position Z.
For clarity, we present a simplified version of the reconstruction algorithm; see [5] for more details. The reconstruction algorithm inspects the Hamming distance d_H between the syndrome of the measured value X' and the candidate helper data items and only carries out the expensive decoding step if the Hamming distance is below a threshold θ.
Reconstruction

1. Read $P' = (S', B', G')$.
2. Measure X'.
3. Compute $M = B' \oplus \mathsf{Syn}\, X'$
4. For $i = 1$ to m:
   (a) If $d_H(M, \gamma^i(S')) \geq \theta$, then next i.
   (b) Compute $\hat{X} = X' \oplus \mathsf{SDec}(M \oplus \gamma^i(S'))$.
   (c) If $G' = f(S' \| B' \| \hat{X})$ then return $\hat{X}$.
5. If the loop is exhausted, then return failure.

Because we use a LDPC code, a small Hamming distance between X' and X implies a small Hamming distance between $\mathsf{Syn}\, X'$ and $\mathsf{Syn}\, X$. For example, a column weight 3 LDPC code ensures that every bit flip between X' and X causes at most three bit flips between $\mathsf{Syn}\, X'$ and $\mathsf{Syn}\, X$.

3 Security Analysis

We express the security properties of our scheme in terms of Shannon entropy H and mutual information I. We start with a general theorem that holds for any method of inserting fake helper data.

Theorem 1 *Let $\mathbf{\Omega}$ be the list of fake helper data in which the real helper data W are inserted at a random position Z. Then the entropy improvement compared to the plain COM is given by*

$$\begin{aligned} \mathsf{H}(X|\mathbf{\Omega}) - \mathsf{H}(X|W) &= \mathsf{H}(W|\mathbf{\Omega}) \\ &= \underbrace{\mathsf{H}(Z)}_{\substack{\text{entropy} \\ \text{gain}}} - \underbrace{\mathsf{H}(Z|W\mathbf{\Omega})}_{\substack{\text{collision} \\ \text{penalty}}} - \underbrace{I(Z;\mathbf{\Omega})}_{\substack{\text{distribution} \\ \text{mismatch} \\ \text{penalty}}}. \end{aligned} \tag{1}$$

In the first term of (1), we recognize the entropy gained from hiding the real helper data at a random position in the list. There are also two clearly interpretable penalty terms in (1). The 'collision penalty' $\mathsf{H}(Z|W\mathbf{\Omega})$ increases with m. It becomes non-negligible when $\mathbf{\Omega}$ contains so many entries that it becomes likely that there exist entries with the same value; then even knowing W and $\mathbf{\Omega}$ does not fix Z.
The 'distribution mismatch penalty' occurs when the fake entries in $\mathbf{\Omega}$ do not look statistically the same as W; then some information about Z can be obtained already from inspecting $\mathbf{\Omega}$.
Next, we provide two lower bounds on the entropy. These bounds follow from (1). Theorem 2 is relevant for the case in which the fake helper data is distributed identically to the real helper data; Theorem 3 is relevant for the generative SCOM.

Theorem 2 *If the distribution of the fake helper data is identical to the distribution of the real helper data and the index Z is drawn uniformly, then*

$$\mathsf{H}(X|\mathbf{\Omega}) - \mathsf{H}(X|W) \geq \log m - \frac{m-1}{\ln 2} \sum_w \left(\Pr[W = w]\right)^2 . \tag{2}$$

If the fake entries are drawn from the same distribution as W, then the distribution mismatch penalty vanishes. Furthermore, if W is not uniform, then this affects the probability of encountering a collision. This is reflected in the $\sum_w$ term of (2). The summation runs over all possible helper data values. As long as W is not too wildly non-uniform and m is not too large, the $\sum_w$ term is negligible w.r.t. $\log m$.

Theorem 3 *Let $W \in \mathcal{W}$. Let U denote a random variable uniform on $\mathcal{W}$. If the fake helper data and the index Z are drawn uniformly, then*

$$\mathsf{H}(X|\mathbf{\Omega}) - \mathsf{H}(X|W) \geq \log m - \frac{m-1}{|\mathcal{W}| \ln 2} - (1 - \frac{1}{m})[D(W\|U) + D(U\|W)], \tag{3}$$

where D is the Kullback-Leibler divergence.

Here the collision penalty has a simple from since it pertains to collisions of uniform variables. In both Theorem 2 and Theorem 3 we see that for $m \ll |\mathcal{W}|$ the improvement in the entropy of X given the public information is approximately $\log m$, as one would intuitively expect.

References

[1] Claude Barral. *Biometrics & Security: Combining Fingerprints, Smart Cards and Cryptography*. PhD thesis, École Polytechnique Fédérale de Lausanne, Switzerland, 2010.

[2] Yevgeniy Dodis, Rafail Ostrovsky, Leonid Reyzin, and Adam Smith. Fuzzy extractors: How to generate strong keys from biometrics and other noisy data. *SIAM J. Comput.*, 38(1):97–139, 2008.

[3] Ari Juels and Madhu Sudan. A fuzzy vault scheme. In *Proc. IEEE ISIT*, pages 408–410, 2002.

[4] Ari Juels and Martin Wattenberg. A fuzzy commitment scheme. In *Proc. ACM CCS*, pages 28–36, 1999.

[5] Boris Škorić and Niels de Vreede. The Spammed Code Offset Method. *IEEE Transactions on Information Forensics and Security*, 9(5):875–884, 2014.

Pilots allocation for sparse channel estimation in multicarrier systems

François Rottenberg*, Kévin Degraux*, Laurent Jacques*, François Horlin** and Jérôme Louveaux*
*ICTEAM, Université catholique de Louvain, Belgium
{francois.rottenberg,kevin.degraux,laurent.jacques,jerome.louveaux}@uclouvain.be
**OPERA, Université libre de Bruxelles, Belgium
fhorlin@ulb.ac.be

Abstract

Wireless channels experience multipath fading which can be modeled by a sparse discrete multi-tap impulse response. Estimating this channel is of crucial importance to allow the receiver to properly recover the transmitted signal. This paper investigates the issue of allocating the pilots for sparse channel estimation applied to multicarrier systems. When the number of pilots is larger than or equal to the channel maximal length, this issue is well-known and the optimal allocation is equispaced. However for long channels, this would require a very large number of pilots decreasing the throughput of the system. Therefore, compressed sensing (CS) techniques are considered to estimate the sparse channel from a limited number of pilots. In that case, the problem of placing the pilots remains an open issue. This paper proposes a two-step hybrid allocation of the pilots that takes the maximal channel length into account to restrict the frequency candidates. The performance of this allocation is demonstrated through simulations and comparisons with other classical allocations.

1 Introduction

In wireless telecommunication systems, multiple reflections induce distortion on the transmitted signal. The receiver has to estimate the channel impulse response to properly compensate this effect commonly known as multipath fading. The wireless channel is often characterized by a sparse discrete multi-tap impulse response of maximal delay L. In a multicarrier system with M subcarriers, the receiver probes the channel frequency response thanks to L_p known pilot symbols transmitted at well-chosen subcarriers. Using this partial information ($L_p \ll M$) the receiver has to estimate or interpolate the entire channel frequency response. In this paper, we investigate the optimal choice for the pilot subcarrier positions as a function of L_p and leveraging sparsity of the channel impulse response.

In practice, the maximal channel length L (number of time samples before the last tap), is much smaller than the number of subcarriers M, *i.e.*, $L \ll M$. For $L_p \geq L$, traditional approaches based on the least squares (LS) criterion have shown that the optimal pilot subcarrier positions are equispaced [1]. However, for very long channels, the required number of pilots L_p can be prohibitive, wasting subcarriers that are therefore unavailable for data transmission.

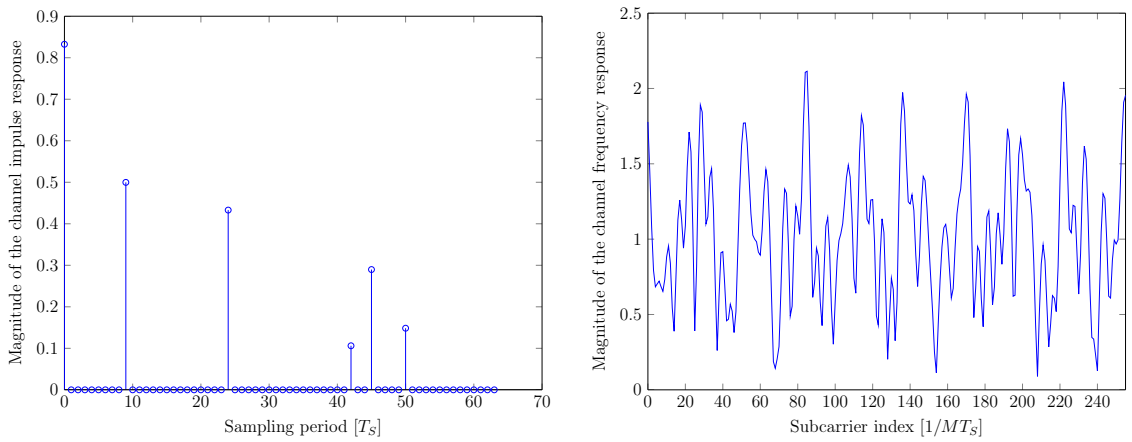
Compressed sensing (CS) theory allows to robustly estimate a sparse signal from a limited number of incoherent linear measurements [2]. In particular, a signal of length M with only K nonzero elements is perfectly recovered with high probability

from about $K\log(M)$ randomly selected Fourier samples [3]. Some works have applied CS recovery methods to channel estimation showing that significant gains can be obtained [4] compared with classical methods. The problem of optimally placing the pilots has been investigated only very recently. Some works proposed sub-optimal research algorithms to select the pilot locations based on different criteria, i.e. the average mean squared error (MSE) using known channel models in [5], the coherence of the measurement matrix in [6] or its mutual and modified mutual coherence in [7]. However, none of those approaches effectively take the maximal channel length L into account to restrict the pilot possible positions.

Instead of randomly selecting the pilots among the M subcarriers, this paper proposes to restrict the frequency candidates to a subset of L equispaced subcarriers and randomly select L_p positions among these L candidates. This hybrid allocation strategy between equispaced and purely random gives a clear gain in performance and reduces the required number of pilots with respect to L.

The rest of this paper is structured as follows. Section 2 describes the system under study by introducing the linear forward model. It also presents three options for pilots allocation and gives the main idea behind CS principles and the recovery algorithm. Section 3 compares the different pilots allocations through numerical simulations and shows that the proposed hybrid allocation allows one to significantly reduce the number of pilots.

2 System model



(a) Channel impulse response

(b) Channel frequency response

Figure 1: Sparse channel model, $M = 256$, $K = 6$, $L = M/4 = 64$

The channel impulse response (CIR) $h[l]$ $(0 \geq l \geq L-1)$ is modeled as a discrete channel as depicted in Figure 1a for $K = 6$ non zero taps, $M = 256$ subcarriers, maximal channel length $L = 64$ and sampling period T_S. This is a typical length for very long channels, e.g. ITU Vehicular B channel model. Moreover an exponentially decaying power delay profile (PDP) is assumed for the channel. Figure 1b depicts the channel frequency response (CFR), obtained by performing a discrete Fourier trans-

form (DFT) on $h[l]$ evaluated at M equispaced frequency points.

The system model for channel estimation in a multicarrier system based on transmitted pilots can be formulated as follows

$$\boldsymbol{H}_{L_p} = \underbrace{\boldsymbol{S}\boldsymbol{F}_{M\times M}\boldsymbol{\Sigma}}_{\boldsymbol{\Phi}}\boldsymbol{h} + \boldsymbol{n} \tag{1}$$

where $\boldsymbol{H}_{L_p}$ is the $L_p \times 1$ stacked vector of observations at pilot subcarriers, $\boldsymbol{S}$ is a $L_p \times M$ matrix which select the L_p rows of the full $M \times M$ DFT matrix $\boldsymbol{F}_{M\times M}$ corresponding to the pilot subcarrier indexes, $\boldsymbol{\Sigma}$ is a $M \times L$ matrix which selects the L first columns of $\boldsymbol{F}_{M\times M}$, $\boldsymbol{h}$ is a $L \times 1$ vector corresponding to the stacked CIR $h[0], ..., h[L-1]$ and $\boldsymbol{n}$ is an additive white Gaussian noise of covariance $\boldsymbol{C}_n = \mathbb{E}(\boldsymbol{n}\boldsymbol{n}^H) = \frac{2N_0}{E_{\text{Pilot}}}\boldsymbol{I}_{L_p}$. A constraint on the total energy E_T used by the pilots is assumed such that $E_{\text{Pilot}} = \frac{E_T}{L_p} = \frac{ME_S}{L_p}$. The $L_p \times L$ matrix $\boldsymbol{\Phi}$ is commonly known as the measurement matrix.

If the number of transmitted pilots is larger than or equal to the maximal channel length, $L_p \geq L$, a classical approach in the literature [1] is to use a least square (LS) criterion to estimate $\boldsymbol{h}$. The estimator of the CIR is then given by

$$\begin{aligned}\hat{\boldsymbol{h}}_{\text{LS}} &= \text{argmin}_{\tilde{\boldsymbol{h}}} \|\boldsymbol{H}_{L_p} - \boldsymbol{\Phi}\tilde{\boldsymbol{h}}\|^2 \\ &= (\boldsymbol{\Phi}^H\boldsymbol{\Phi})^{-1}\boldsymbol{\Phi}^H\boldsymbol{H}_{L_p} \\ &= \boldsymbol{h} + (\boldsymbol{\Phi}^H\boldsymbol{\Phi})^{-1}\boldsymbol{\Phi}^H\boldsymbol{n}.\end{aligned} \tag{2}$$

This estimator converges asymptotically towards the true CIR $\boldsymbol{h}$ at high SNR. The work in [1] has shown that to minimize the noise amplification due to the inverse in last expression, one should place the pilot subcarriers in an equispaced way. This could actually be predicted by Shannon sampling theory. Since $h[l]$ has a length limited to L samples, there will not be any aliasing if its DFT is sampled uniformly at maximal sampling period $1/L$ which is obtained as soon as $L_p \geq L$.

However, if $L_p < L$, the least squares problem is underdetermined. Inspired by sparse approximation theory, most CS reconstruction techniques leverage the sparsity of $\boldsymbol{h}$ to regularize the least squares problem, *i.e.*,

$$\hat{\boldsymbol{h}} = \text{argmin}_{\tilde{\boldsymbol{h}}} \|\boldsymbol{H}_{L_p} - \boldsymbol{\Phi}\tilde{\boldsymbol{h}}\|^2 \quad \text{s.t.} \quad \|\tilde{\boldsymbol{h}}\|_0 \leq K, \tag{3}$$

where $\|.\|_0$ is the number of non-zero coefficients in a vector. The estimator $\hat{\boldsymbol{h}}$ of (3) is NP-hard to compute due to the combinatorial number of possible supports for $\tilde{\boldsymbol{h}}$. However, there exist several ways to find a good approximation to $\hat{\boldsymbol{h}}$, *e.g.*, by ℓ_1-norm relaxation [2]. In this work, the Iterative Hard Thresholding (IHT) algorithm is for estimating $\hat{\boldsymbol{h}}$. This procedure is simple, fast and it provides guarantees of good reconstruction quality [8].

In this particular setup, the optimal pilot positions are not known. The problem addressed in this paper is the choice of the pilot subcarrier positions in order to recover $\boldsymbol{h}$ based on the channel observations $\boldsymbol{H}_{L_p}$.

Equispaced allocation One could think at first that placing the subcarriers in an equispaced manner over the band is still a good idea. Nevertheless, regarding Shannon

sampling theory, this allocation would result in strong aliasing since the frequency band is not sampled at a sufficient rate. Sparse approximation theory also supports this fact. Indeed, it is known that one can find a unique sparse representation (here $\boldsymbol{h}$), of a signal (here $\boldsymbol{H}_{L_p}$) in a redundant dictionary $\boldsymbol{\Phi}$ only when the coherence of that dictionary is small [9]. The coherence μ of $\boldsymbol{\Phi}$ is defined as the maximum normalized absolute correlation between two columns,

$$\mu = \max_{0 \leq m < n \leq L-1} \frac{\boldsymbol{\phi}_m^H \boldsymbol{\phi}_n}{\|\boldsymbol{\phi}_m\|_2 \|\boldsymbol{\phi}_n\|_2}, \tag{4}$$

that is, in this case,

$$\begin{aligned} \mu &= \max_{0 \leq m < n \leq L-1} \frac{1}{L_p} \sum_{l=0}^{L_p-1} e^{-j\frac{2\pi}{M} k_l (n-m)} \\ &= \max_{1 \leq c \leq L-1} \frac{1}{L_p} \sum_{l=0}^{L_p-1} \omega_M^{k_l c}, \end{aligned} \tag{5}$$

where $\omega_M^k = e^{-j\frac{2\pi}{M}k}$ and $c = n-m$. Let's assume that L_p divides M and the subcarriers are placed equispaced, that is, $k_l = l\frac{M}{L_p}$. While the condition $L_p \geq L$ gives a zero coherence due to the orthogonality property of the root of unity, the underdetermined case $L_p < L$ allows $c = L_p$, that is,

$$\mu = \max_{1 \leq c \leq L-1} \frac{1}{L_p} \sum_{l=0}^{L_p-1} e^{-j\frac{2\pi}{M} l \frac{M}{L_p} c} = \frac{1}{L_p} \sum_{l=0}^{L_p-1} \omega_{L_p}^{lL_p} = 1. \tag{6}$$

This shows that if $L_p < L$, placing the subcarrier uniformly is the worst choice in the sense of the minimal coherence criterion.

Fully random allocation Following CS theory, when $L = M$, *i.e.*, the time range of $\boldsymbol{h}$ is equal to the number of subcarriers, a good choice for the subcarrier assignment scheme (SAS) is simply to choose fully randomly L_p subcarriers and one could expect much better performance than with equispaced placement. In particular, it has been shown that in this case, the condition

$$L_p \geq CK \log(M) \tag{7}$$

with C a reasonably small fixed constant, guarantees perfect recovery in the noiseless case and robust recovery in the presence of noise [3]. This SAS will be further referred to as the fully random SAS.

The limitation of the fully random SAS is that it does not take into account the maximal channel length L that is smaller than M in practice. Fully randomly selecting the subcarriers would wastefully allow to have a time range up to M taps in time domain (TD). A second way to interpret this limitation is that the CFR should not be sampled too fast. Indeed, a small distance between pilot subcarriers would give useless information on the taps which are far from the origin and known to be zero. Yet another way to see this is by acknowledging the fact that two frequency samples that are next to each other are strongly correlated and their mutual information is very low.

Proposed hybrid allocation Considering the limitation of the fully random SAS, a constraint can be added to restrict the distance between pilot subcarriers to be bigger than $\frac{M}{L}$. Instead of fully randomly selecting among the M subcarriers this paper proposes to perform a two-step hybrid SAS:

- The frequency candidates are first restricted to a subset of L equispaced subcarriers.
- The L_p positions are randomly selected among these L candidates.

This allocation is referred to as "hybrid" between equispaced and purely random since it randomly selects L_p subcarriers among the frequency candidates of the $L_p = L$ equispaced SAS which corresponds to an optimal Shannon sampling. Doing so, we try to ensure minimum correlation between pilots.

Furthermore, if L is assumed[1] to divide M, this hybrid SAS leads to a simplified system model as

$$\begin{aligned} \boldsymbol{H}_{L_p} &= \boldsymbol{S}\boldsymbol{F}_{M\times M}\boldsymbol{\Sigma}\boldsymbol{h} + \boldsymbol{n} \\ &= \boldsymbol{S}'\boldsymbol{F}_{L\times L}\boldsymbol{h} + \boldsymbol{n}. \end{aligned} \tag{8}$$

where $\boldsymbol{S}'$ is the new selection matrix with indices belonging to $[0, L-1]$. The previous DFT simplification comes from the fact that the L first columns and the L equispaced rows of $\boldsymbol{F}_{M\times M}$ are selected.

This simplification has two advantages. On the one hand, the complexity of the reconstruction algorithm is reduced since the DFT size decreases by a factor $\frac{M}{L_p}$. On the other hand, this allows to use the results of [3] but with L instead of M, namely,

$$L_p \geq CK\log(L), \tag{9}$$

which implies a reduction in the required number of pilots to guarantee perfect reconstruction in the noiseless case and robust recovery in the presence of noise.

3 Simulation results

In the simulations, $M = 256$ subcarriers are assumed, the maximal length of the channel is set to $L = M/4 = 64$ and $K = 6$ taps are non negligible. The first tap is placed in 0 and the five remaining tap delays follow a uniform distribution in $\{1, \cdots, L-1\}$. A uniform PDP of the channel is considered such that each non zero tap has a zero mean and a variance exponentially decaying with the delay with a 20dB attenuation of the last tap with respect to the first. 100 channel realizations and 100 SAS realized for each of the hybrid and fully random SAS while only one realization for the equispaced SAS (since it is deterministic). The metric used to evaluate each method is the normalized mean squared error (NMSE) defined as

$$\text{NMSE} = \mathbb{E}\left\{\frac{\|\boldsymbol{h} - \hat{\boldsymbol{h}}\|_2^2}{\|\boldsymbol{h}\|_2^2}\right\}, \tag{10}$$

which is averaged over all channel and SAS realizations. The method used to reconstruct the channel based on the observations is Iterative Hard Thresholding (IHT) [8].

[1] If it is not true, one can increase L up to the point where it becomes true.

We fix the number of iterations to 50 and the gradient step size parameter of the line search is computed at each iteration to minimize the LS criterion. This method assumes the channel maximal length and the number of non zero taps are known by the receiver. However, one could as well use another CS technique which does not know exactly the sparsity a priori, *e.g.* basis pursuit denoising [2].

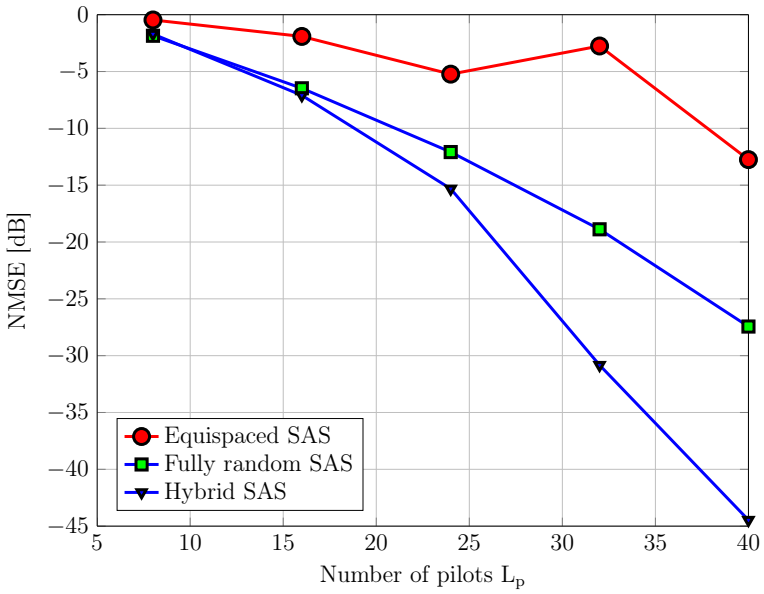


Figure 2: NMSE as a function of the number of pilots for different SAS and $E_S/N_0 = 30$dB.

Figure 2 shows the evolution of the NMSE as a function of the number of pilots for $E_S/N_0 = 30$dB. As expected, the equispaced SAS performs very badly[2]. Moreover, the hybrid SAS clearly outperforms the fully random SAS. There is a 12dB gap and a 17dB gap between the two methods for $L_p = 32$ and $L_p = 40$ respectively.

The results in Figure 2 are based on the NMSE which gives no information on the distribution of the normalized squared error (NSE) of reconstruction. Figure 3 depicts the cumulative density function (CDF) of the NSE for the different SAS and $L_p = 32$. As explained before, the reconstruction fails for almost all realizations using the equispaced SAS. We also see that the probability of good reconstruction for the hybrid SAS is higher than for the random SAS, e.g. about 99% of the NSE realizations are below -35dB for the hybrid allocation compared to about 70% of the NSE realization using the fully random SAS.

Figure 4 shows the NMSE as a function of the E_S/N_0 ratio for different SAS. As explained, if $L_p < L$, the equispaced SAS performs very badly. For $L_p = L$, a LS estimator can be computed for which the equispaced SAS is optimal. Moreover, the LS estimator thresholded to the 6 more significant taps is also shown[3] and is the best performance obtained. The fully random SAS performs well at low SNR while at high

[2]Note that the equispaced is only averaged over one SAS realization since it is deterministic versus 100 for the two other SAS.

[3]Since the IHT method is assumed to know the number of non zero taps, it is also more fair to threshold the result of the LS estimator.

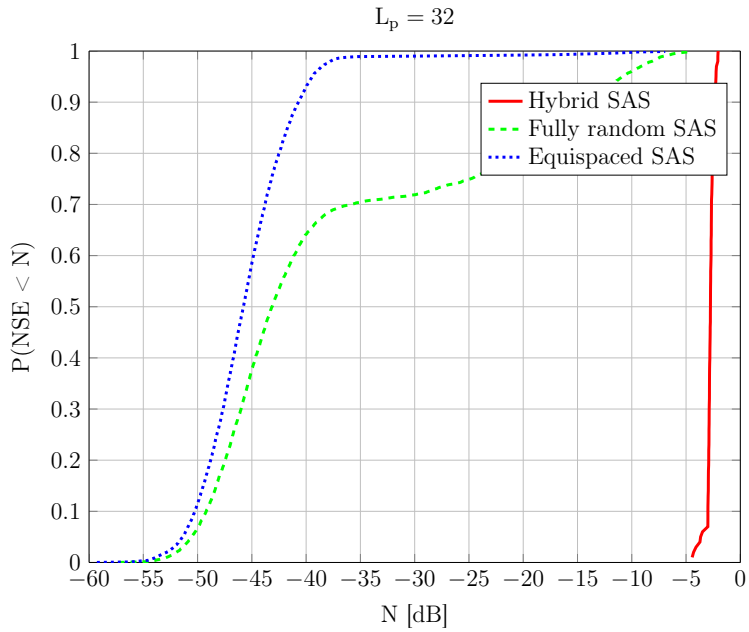


Figure 3: Cumulative density function (CDF) of the NSE for different SAS and $E_S/N_0 =$ 30dB, $L_p = 32$.

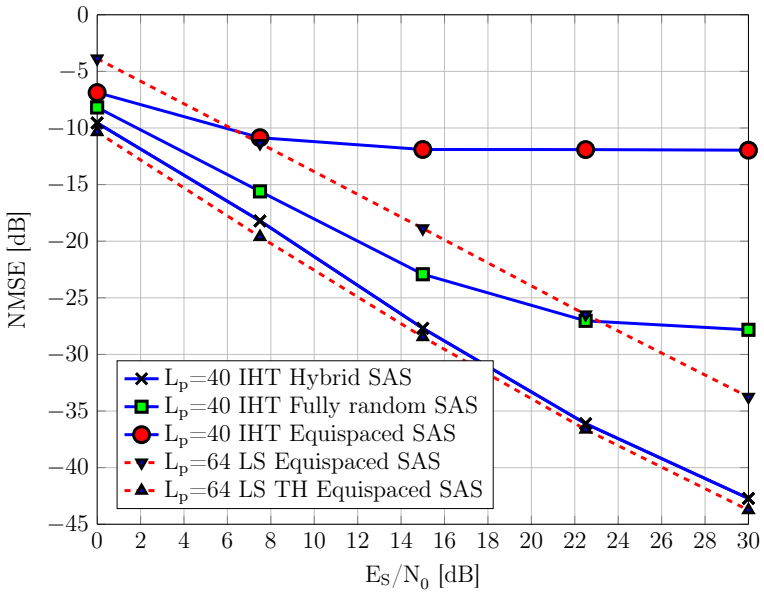


Figure 4: NMSE as a function of E_S/N_0 ratio ($L = 64$). The presented SAS are tested with $L_p = 40 < L$. A classical LS estimator is computed with $L_p = 64 = L$, for which the equispaced SAS is optimal. The LS estimator thresholded to its K highest coefficients is also shown for fair comparison with an ideal case.

SNR, the reconstruction error imposes a NMSE floor. Then, the two LS methods perform better at the cost of more pilots. However, the hybrid technique performs almost as well as the LS thresholded estimator still at high SNR and with only $L = 40$ pilots.

4 Conclusion

This paper investigated the issue of allocating the pilots for sparse channel estimation. In situations where the number of pilots is larger than or equal to the channel maximal length, this issue is well-known and the optimal allocation is equispaced. However, if the number of pilots is strictly smaller than the channel length, the problem remains open. This paper showed that still placing the subcarriers in an equispaced way is the worst choice in the sense of the minimal coherence criterion. Rather than selecting the pilots at random among the subcarriers, this paper proposed to use a two-step hybrid allocation. The first step restricts the frequency candidates to a subset of equispaced subcarriers and the second step randomly selects positions among these candidates. This allocation allows to significantly reduce the complexity of the reconstruction by decreasing the DFT matrix size. The performance of the method was demonstrated through simulations and compared with fully random allocation approach and other classical approaches based on the LS criterion. The hybrid allocation clearly outperforms the fully random allocation while reaching almost the same performance as a LS thresholded estimator requiring much more pilots.

Acknowledgment

The research reported herein was partly funded by F.N.R.S and F.R.I.A.

References

[1] R. Negi and J. Cioffi, "Pilot tone selection for channel estimation in a mobile OFDM system," *IEEE Transactions on Consumer Electronics*, vol. 44, no. 3, pp. 1122–1128, 1998.

[2] E. J. Candes, J. K. Romberg, and T. Tao, "Stable signal recovery from incomplete and inaccurate measurements," *Communications on pure and applied mathematics*, vol. 59, no. 8, pp. 1207–1223, 2006.

[3] E. J. Candes and Y. Plan, "A probabilistic and ripless theory of compressed sensing," *IEEE Transactions on Information Theory*, vol. 57, no. 11, pp. 7235–7254, 2011.

[4] S. F. Cotter and B. D. Rao, "Sparse channel estimation via matching pursuit with application to equalization," *IEEE Transactions on Communications*, vol. 50, no. 3, pp. 374–377, 2002.

[5] C. Qi and L. Wu, "Optimized pilot placement for sparse channel estimation in OFDM systems," *IEEE Signal Processing Letters*, vol. 18, no. 12, pp. 749–752, 2011.

[6] C. Qi and L. Wu, "A study of deterministic pilot allocation for sparse channel estimation in ofdm systems," *IEEE Communications Letters*, vol. 16, no. 5, pp. 742–744, 2012.

[7] X. He, R. Song, and W.-P. Zhu, "Optimal pilot pattern design for compressed sensing-based sparse channel estimation in ofdm systems," *Circuits, Systems, and Signal Processing*, vol. 31, no. 4, pp. 1379–1395, 2012.

[8] T. Blumensath and M. E. Davies, "Iterative hard thresholding for compressed sensing," *Applied and Computational Harmonic Analysis*, vol. 27, no. 3, pp. 265–274, 2009.

[9] J. A. Tropp, "Greed is good: Algorithmic results for sparse approximation," *IEEE Transactions on Information Theory*, vol. 50, no. 10, pp. 2231–2242, 2004.

Towards Closing the Gap between Theory and Practice in Open Data Publishing

R-J. Sips[1] Z. Erkin[2] A. Manta[2] B. Havers[1] R.L. Lagendijk[2]
[1]Center for Advanced Studies, IBM Nederland B.V., Amsterdam
[2]Cyber Security Group, Dep. of Intelligent Systems, Delft University of Technology
{robert-jan.sips,bram.havers}@nl.ibm.com
{z.erkin,r.l.lagendijk}@tudelft.nl, manta.s.andrei@gmail.com

Abstract

In recent years, the Open Data movement has gained momentum, mainly driven by the idea of Open Data as fuel for innovation. Unfortunately, the publication of large sets of Open Data may also lead to unforeseen breaches to the privacy of individuals. In order to investigate this, we conducted interviews with the policy-makers responsible for Open Data publication in the Netherlands, during which we observed that little is known about the possible privacy issues surrounding the publication of Open Data by the organizations responsible for its publication. In addition, we observed that little attention is given to the preservation of *utility* in anonymized Open Data sets. Following these central observations, we present an updated data publishing process, supported by an automated decision support system (ADSS) to help data publishers to make informed decisions on the choice of anonymization algorithm. We also provide a reference implementation to illustrate the use of ADSS that involves a number of commonly used anonymization algorithms.

1 Introduction

The online publication of (governmental) data which can be used and republished without restrictions, so called the Open Data movement, has gained momentum in recent years [8]. The main idea in Open Data movement is to increase governmental transparency by making public information more easily accessible. Many countries, including the Netherlands, have set targets on the adaptation to this movement [21]. An example of Open Data in Europe is the publication of patient data with medical details in the UK that aims to improve public health [2].

Open Data movement was initiated as it is believed that there are a number of advantages. Huijboom and van den Broek [8] summarized them as follows: (1) increasing democratic control and public participation, (2) fostering service and product innovation and (3) strengthening law enforcement. On the other hand, the movement has also a disadvantage. Although the Open Data movement is considered as a tool for transparency and economic growth, the open publication of governmental datasets, which often contain a wide range of governmental data, may lead to unforeseen privacy and security breaches due to unintended publication of sensitive data and unpredicted combinations with other data [7], which in turn may lead to legal and economic consequences.

A good example of such unforeseen uses of data for malicious purposes is the so-called "Makkie Klauwe"* app ("easy stealing"). The application directs burglars in Amsterdam to houses which are easy to break into and are expected to generate a nice profit, by combining public data such as area value, reported problems and how

*http://www.bramfritz.nl/makkieklauwe/

much the municipality can spend on area improvement and repairs. For example it may suggest a house in an expensive neighbourhood where the street-light is broken (easy to break into undetected).

The existing research in data anonymization is focused on data privacy. However, as demonstrated by the example above, we observe that the risks of data publication range beyond privacy breaches alone. To better understand whether these risks are understood and sufficiently mitigated in practice, we conducted interviews with policymakers responsible for the publication of Open Data in the Netherlands from the following institutions: Rijkswaterstaat (responsible for the roads, water and infrastructure in The Netherlands), Kadaster (responsible for parcel information), Amsterdam Economic Board (responsible for strategies for the economic development of the Amsterdam region) and Statistics Netherlands - CBS (supplier of most statistical information in The Netherlands). These interviews revealed that there are neither clear procedures to mitigate the risks related to Open Data publication, nor sufficient understanding on the impact of the risks within most of the institutions.

In this paper, we address the data privacy considerations in the Open Data publication process. We lay a foundation to an enhanced data publishing process, supported by an automated decision support system. This system assists policy makers (as well as data publishers) in assessing the risk of publishing data sets as Open Data, without assuming deep knowledge of privacy attack mechanisms. Therefore, we describe a system which provides insight on the potential risks associated with data leakage and which provides advice on how to anonymize data. By means of this enhanced data publishing process, we aim to build a bridge between theory and practice in open data publishing.

The main contributions to existing research in this paper are:

Improved Data publishing process We present an improved data publishing process which serves as the foundation towards automated data publishing.

Automated decision support system (ADSS) We present a reference implementation of an automated decision support system, which assists a data publisher in understanding the privacy risks. Unlike other frameworks, the proposed system evaluates and visualizes the performance of a number of state-of-the-art algorithms on a given data set, enabling the data publisher to make an informed choice based on the privacy and utility metrics presented.

Testing As a *proof of principle* we provide results of a comparative study of the algorithms included in the ADSS, to better understand the privacy-utility trade-off.

We believe that the improved Open Data publication process presented in this paper combined with the ADSS will help authorities to better mitigate the privacy risks associated with the publication of their datasets, while preserving the utility of the published data. Moreover, we provide a discussion on the issues we observe in practice, issues that may guide the research community to address the needs of the data publishing authorities.

The remainder of the paper is organized as follows. We present related work in the field in Section 2. A brief description of the used privacy models is presented in Section 3. In Section 4 we provide an overview of the data publishing process. Following this, we give a high level overview of the ADSS in Section 5. We provide some insights on the experiments conducted on the ADSS using a public data set in Section 6. Finally, we draw conclusions in Section 7.

2 Related Work

To the best of our knowledge, little research has been conducted on the privacy risks associated with Open Data. However, there does exist a large body of literature on anonymization of data and on the quantification of anonymity and utility, for an overview we refer readers to the paper by Fung *et al.* [6]. In the remainder of this section, we describe frameworks that use anonymization techniques and quantify the associated privacy-utility trade-off.

Duncan et al. [4] present the idea of the trade-off between privacy and utility and formally define the Risk-Utility maps as a way to visualize this trade-off. One limitation of their work is that applying the concept in practice requires expert level statistical and mathematical knowledge.

Sramka *et al.* [18] took another approach to the problem and defined two types of utility: bad utility and good utility. The former represents the usefulness to an attacker while the latter to a legitimate user. In this context, Sramka models privacy in terms of bad utility. The less bad utility a data set contains, the less useful it is to an attacker and thus, the more privacy preserving it is. The authors propose a framework to anonymize and then analyze the data from a data-mining perspective. The framework uses data-mining algorithms to compute the privacy and utility metrics. This idea complements our reference implementation since our metrics provide micro-level information, while the data- mining metrics provide macro-level information.

Lin and Kifer [16] propose another framework to extract semantic privacy guarantees from the anonymized data. In other words, the authors seek an answer for the question "what does privacy guarantee Y protect?". The authors rely on the change in beliefs of Bayesian attackers (attackers who use Bayesian inference to breach privacy) to build their proof. To achieve this, the privacy definition is restated in the language of set theory and then a geometric object called the row cone is extracted. This object encapsulates all the ways in which an attacker's prior beliefs can become posterior beliefs after seeing the data.

The framework of Beck and Marhöfer [3] is built on top of the UTD [17] anonymization toolkit and uses sdcMicro [20] module to measure risk. Unlike our system, the authors only test how a classifier behaves on anonymized user profiles.

The only framework we know of as being used in practice on large scale is μ-argus [9, 10]. The framework uses k-anonymity [19] and suppression to anonymize the data. sdcMicro [20] has been developed by the same community that developed μ-argus, so privacy is measured in a similar way.

In adddition to the above frameworks, a few tools are available to anonymize data. UTD [17] is a simple tool that provides the means to anonymize the data; sdcMicro [20] is an R language module which can also anonymize data and compute several metrics.

3 Preliminaries

One of the core contributions in this paper is an automated decision support system as part of Open Data publishing process. For our reference implementation of this system, we have chosen four privacy models: k-anonymity [19], t-closeness [14], (n,t)-closeness [15] and (n,t)-closeness together with k-anonymity. These models are widely used in literature as baseline or reference and some of them are applied in practice. Our framework relies on two implementations for k-anonymity: one is based on the Incognito [12] algorithm and the other is based on the Mondrian [13] algorithm. The other models have been implemented by relying on either Incognito or Mondrian as follows. t-closeness extends the Incognito implementation of k-anonymity, while (n,t)-closeness and (n,t)-closeness with k-anonymity both use the Mondrian algorithm. In the remainder of this section we provide brief descriptions of the aforementioned privacy models for the sake of completeness.

An important concept is that of a *Quasi-IDentifier* or QID. It represents a set of attributes of the current data set used for linking with external information in order to uniquely identify individuals. These include attributes, which at first glance may seem harmless, such as the combination of postcode, gender and age, that uniquely identifies people.

k-anonymity This model works as follows. Let *qid* represent a QID value combination for a record in a data set. k-anonymity only requires that every single *qid* value appears at least k times in the data set. This means that the QID values of the records in the data set are generalised in such a way that grouping by QID values generates bins called equivalence classes (EC) of size at least k. The effect of k-anonymity is that an attacker can link an individual to a record with a maximum probability of $1/k$. k can take any positive integer value.

t-closeness The privacy model t-closeness also uses generalisation of QID values to achieve its privacy requirement. But instead of requiring a minimum group size, it requires a maximum distance between two distributions. A data set is said to achieve t-closeness if for every EC, the distribution of the sensitive values in the EC is within t of the distribution of sensitive values in the whole data set. t can take any real values between 0 and 1. The reasoning behind it is to limit the information gain from an individual EC, compared to the information already gained from the whole data set.

(n,t)-closeness (n,t)-closeness builds on top of t-closeness. The main advantage is that it distorts the data less. It requires the distribution of sensitive values for every EC to be within t of a population of size at least n. This population must be sufficiently large of size at least n, and needs to be a natural superset of its respective EC. n can be a positive integer of at most the size of the full data set.

(n,t)-closeness with k-anonymity For this privacy model, Mondrian has been used as the base algorithm. The requirement for the cut step, explained above, has to respect both the k-anonymity and the (n,t)-closeness privacy requirements: the EC has to be at least of size k and within t distance of a natural superset of size at least n.

4 Data Publishing Process

In this section, we present a 6-step publishing guideline that considers the legal, ethical and technical aspects of the process. This is based on the guideline presented in [10, Ch. 3.2] and adapted to be applicable to Open Data and to account for the decision support system presented in Section 5.

Assess need for confidentiality protection. The first step involves analyzing the data set to be published for the presence of information requiring protection. The decision on what needs to be protected and what does not is based on legislation, common sense and experience. In the case of the Netherlands, legislation includes the Personal Data Protection Act and domain specific acts for, e.g., medical data.

Identifying data characteristics and data usage. This step involves gaining a better understanding of the characteristics of the data and how these data can and may be used by different parties. Characteristics include the type of data (e.g. textual, numerical), distribution of values, etc. The usage of the data is difficult to determine. In the context of Open Data, there is no intended way to use the data. One can only attempt to estimate the possible usage, but we may expect (and hope for) users to bring new insights by using and combining the data in novel and unforeseen ways.

Disclosure risk. Disclosure relates to re-identifying a sensitive (protected) piece of information in the published data. There are several types of disclosure including identity disclosure, attribute disclosure, inferential disclosure and table linkage. [10, Ch. 3]. Depending on the obtained data, different disclosure risks exist. Based on the gathered information from other sources, the data publisher tries to develop scenarios which demonstrate how the data can be misused. A possibly useful framework for this is described in [5].

Configuration of the automated decision support tool. In this step, the user configures the system by selecting the privacy and utility metrics to be included in the analysis of the data (and related parameters).
Selecting the algorithm to be used for publishing. The user runs the automated software to analyze the data set. As a result, he receives information on all algorithms and their performances given their Risk Utility maps [4]. This is a plot which shows an algorithm's performance, measured by the privacy and utility metrics, when executed using different parameter values.

If needed, the publisher can execute post-anonymization data processing. Examples include suppressing certain values or changing the format of the data(e.g. if the date should follow a specific standard) before it is published. Finally, the data is written out to the configured location.

Data audit and documentation. This last step is required in order to create valid expectations on behalf of the future data users. The data publisher should choose which pieces of information can be released to the public. Two important pieces of information are the results of the utility metrics and the methods used to protect the data. The former gives insight on how usable the data set can be towards certain tasks, The later should be made public for reasons of transparency. The data can be checked by an external party for compliance with the regulations. The documentation needs to explain the legal or administrative reasons behind the data anonymization process. Furthermore, information about the anonymization process can help users understand what has been changed and what the impact could be on their data usage. This is important because it is possible, for example, to calibrate data mining algorithms to account for modifications made by applying a privacy model such as k-anonymity.

5 Automated Decision Support System

The purpose of our automated decision support tool (ADSS) is to assist data publishers in anonymization of the data, without assuming deep knowledge of anonymization algorithms and utility metrics. It anonymizes a data set in many different ways and then presents the data publisher with the results of different metrics computed for each anonymization. Using this information, the data publisher can now make a better choice as to which anonymization to use. Our focuss is on the non-expert users because the government does not have the expert manpower to assess all the data sets that it needs to publish, within a reasonable amount of time.

The performance of each anonymization algorithm strongly depends on the data set. Currently, there is no single best anonymization algorithm. One needs to assess the guarantees offered by each technique. In the context of Open Data, utility deserves special attention. If one releases a data set with low utility, it might not be worth publishing at all.

Our proposed ADSS consists of 6 modules, depicted in figure 1. The *Configure* and *Control* modules are responsible for configuring the algorithms used and controlling the overall program flow respectively. The *Data operations* module reads and writes

data from and to persistent storage. The *Anonymize*, *Measure* and *Visualize* will be discussed below.

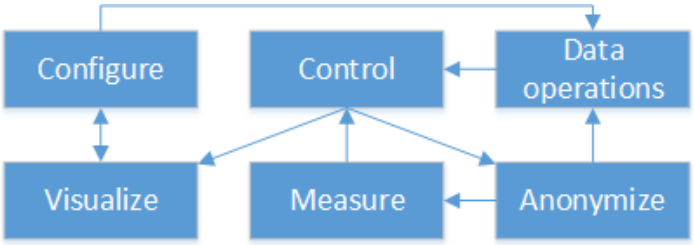


Figure 1: Modules and dataflow for the Automated Decision Support System

ANONYMIZE The anonymize module applies a (configurable) number of anonymization algorithms to a given dataset. Intermediate (sanitized) results are persisted for measurement and visualization. The anonymize module is designed to be easily extendible with anonymization algorithms.

MEASURE Within the measure module, *privacy* and *utility* metrics are applied to the anonymized datasets resulting by the anonymize module. Outcomes are normalized to allow visual comparison in the visualize module.

VISUALIZE Visualization of the results is done by plotting the outcomes of the various metrics and algorithms into a single image. In order to make the metric values meaningful for any non-expert, we defined a factor as the value of the utility metric applied to the anonymized data set divided by the value of the same utility metric applied to the original data set (ODS). This factor is a positive value: A value of less than 1 implies an improvement over the ODS, while a value greater than 1 implies a decrease in utility. A value smaller than 1 is possible because the anonymized data set removes some of noise contained by the ODS. Similarly, we use normalized results of the privacy metrics. The outcomes of the various metrics are visualized in a colored scatter-plot, privacy on the X-axis and utility on the Y-axis. A lower-left quadrant indicates the algorithms and configurations which are optimal, and boundaries can be visualized for the required levels of privacy and utility.

6 Reference Implementation and Experimental Results

In this section, we describe our reference implementation for the ADSS and provide a high level analysis of an experiment that compares the most commonly used anonymization techniques. The metrics used in this experiment give some insights in the success of the used algorithms, in terms of privacy-utility trade-off.

Our reference implementation has been built by extending and modifying the University of Dallas Texas anonymization toolbox [17]. The toolbox provides a number of anonymization algorithm implementations and has capabilities to read, write and transform data. However, the tool has been changed significantly to enable running multiple anonymizations, automatically generating algorithm configurations, and computing and visualising the results. Moreover, we've added Mondrian (n,t) and (k,n,t) to the suite.

For the experiments, the Adult data set[†] from the UC Irvine machine learning

[†]`http://archive.ics.uci.edu/ml/datasets/Adult`

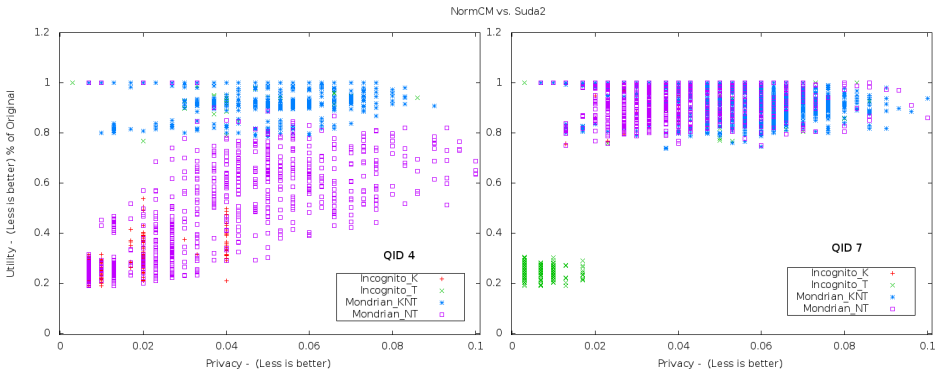


Figure 2: QID_4 and QID_7 comparison for NormCM.

repository has been used. It consists of data collected from the US census. Records with missing values have been removed, resulting in a data set consisting of 30162 records in total.

To simulate the effect of combining the Adult data set with external sources, we experimented with QID sizes of 4 and 7, which are represented as QID_4 and QID_7, respectively. The two sets of QIDs are {Age, Occupation, Race, Gender} and {Age, Education, Marital status, Occupation, Race, Gender, Native country}. For brevity reasons, the attribute details of this data set have been ommited.

The experiments were conducted on several machines with Intel(R) Xeon(R) CPU 2 GHz with at least 32 GB of RAM. The reference implementation runs on Java and makes use of an internal SQLite library for data storage and processing.

The metrics used is our experiments are the *Classification Metric (CM)* [11], the *Discernibility metric (DM)* [1] and the *Normalized average equivalence class size (NormAvgECSize)* [13]. Furthermore, since *CM* and *DM* are values that depend on the number of records, their normalized forms were used in our experiments.

6.1 Privacy-Utility Metrics

In our experiments we analyse the algorithms on an individual and collective level. However, for space constraints, we present only the latter in this paper, a comparison of the four algorithms based on the utility and privacy metrics mentioned above.

NormCM In Figure 2 we observe that for QID_4, Incognito_K and Mondrian_NT offer the best utility and second best privacy. Incognito_T, as expected, offers the best privacy but at a high utility cost. In this scenario, the worst case value for utility is on par with that of the ODS.

For a QID of size 7, it becomes clear that Incognito_T is the best choice. Because Mondrian tries to slice the QID space as uniformly as possible, it does not provide optimal aggregation of values and incurs a higher classification penalty for its two implementations. k-anonymity is limited by the value of k to the minimum bin size. This makes it possible for mixed values to be grouped together and incur a higher penalty. Incognito_T manages to achieve a grouping of values into smaller bins and yet preserve privacy.

NormDM In both cases in Figure 3, where the QID set size was 4 and 7, all algorithm anonymizations have, for a given parameter value, a global re-identification rate of 0.7% for approximately the same utility value. The only relevant fact is that the Mondrian based algorithms managed to find anonymizations with a better utility level than when the QID set size was equal to 4. KNT managed a factor of 5 w.r.t to the

ODS while NT a factor of 1.3 to 2. The reason why NT outperforms KNT is that the former is not limited by a minimum EC size of k.

NormAvgECSize We observe in Figure 4 that Incognito_T offers the best anonymization possible for QID_4. In QID_7, Incognito_T is on par with Mondrian_KNT. Incognito_T requires a taxonomy tree for every attribute in order to work. Having a better result than Mondrian_KNT means that the user defined taxonomy tree for QID_4 is better than the Mondrian self generated partitioning. In QID_7 we see that Mondrian_KNT is able to find a similar partitioning to that of Incognito_T.

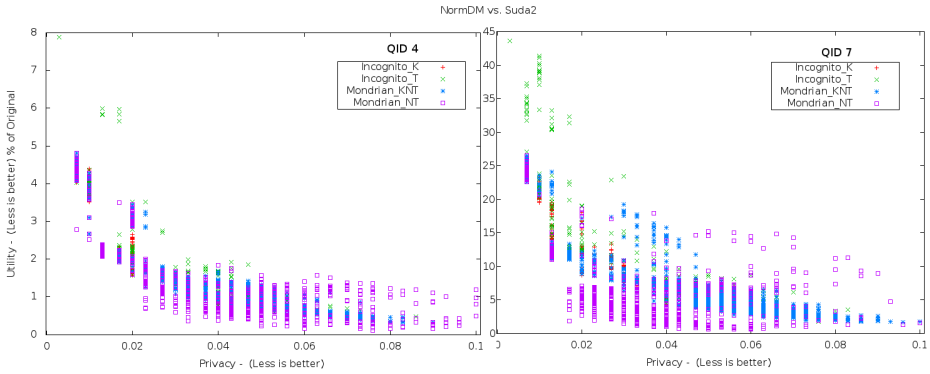


Figure 3: QID_4 and QID_7 comparison for NormDM.

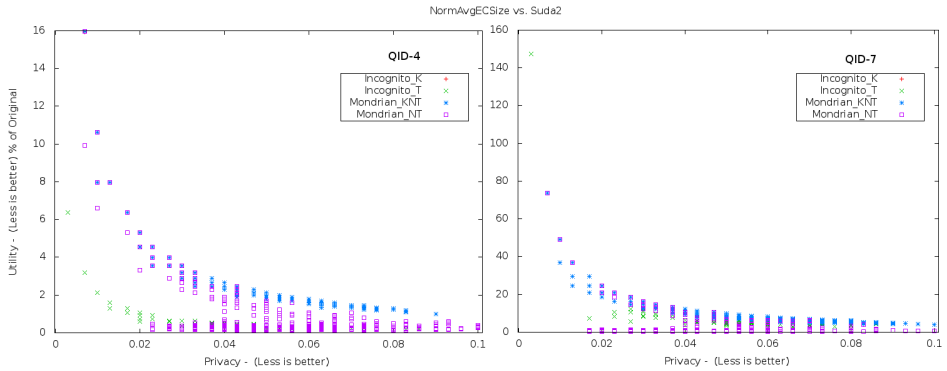


Figure 4: QID_4 and QID_7 comparison for NormAvgECSize

7 Conclusion

In this paper, we address the privacy considerations associated with the publication of Open Data in practice. We provide an enhanced guideline for the publication procedure and present a reference implementation of an automated decision support system that compares the performance of a number of widely used anonymization algorithms given certain metrics for a particular data set. Our goal is to enable decision makers to effectively compare the performance of different anonymization algorithms using privacy-utility metrics, without understanding these metrics and algorithms in detail, since in the end, there is no best algorithm. The data publishing procedure and the automated decision support system we propose make a step towards closing the gap between theory and practice in Open Data publishing. However, additional aspects

to improve the framework remain. Firstly, the proposed ADSS currently handles relational data only. But many other types of data are considered for publication, including, and not limited to transactional, locational, social and graphical data. Secondly, other algorithms which are deemed suitable for those new kinds of data should be integrated to the framework. Thirdly, the effects of combining information from other sources such as social networks, should be investigated in terms of privacy and utility. Nevertheless, we believe that our enhanced data publishing procedure and the ADSS provide an important step towards building a bridge between the theory and practice in open data publishing.

References

[1] Roberto J Bayardo and Rakesh Agrawal. Data privacy through optimal k-anonymization. In *Data Engineering, 2005. ICDE 2005. Proceedings. 21st International Conference on*, pages 217–228. IEEE, 2005.

[2] BBC. Everyone 'to be research patient', says David Cameron. `http://www.bbc.co.uk/news/uk-16026827`, 5 December 2011. Online.

[3] Martin Beck and Michael Marhofer. Privacy-preserving data mining demonstrator. In *Intelligence in Next Generation Networks (ICIN), 2012 16th International Conference on*, pages 210–216. IEEE, 2012.

[4] George T. Duncan, Sallie A. Keller-mcnulty, and S. Lynne Stokes. Disclosure risk vs. data utility: The r-u confidentiality map. Technical report, Chance, 2001.

[5] Mark Elliot and Angela Dale. Scenarios of attack: the data intruders perspective on statistical disclosure risk. *Netherlands Official Statistics*, 14(Spring):6–10, 1999.

[6] Benjamin Fung, Ke Wang, Rui Chen, and Philip S Yu. Privacy-preserving data publishing: A survey of recent developments. *ACM Computing Surveys (CSUR)*, 42(4):14, 2010.

[7] Srivatsava Ranjit Ganta, Shiva Prasad Kasiviswanathan, and Adam Smith. Composition attacks and auxiliary information in data privacy. In *Proceedings of the 14th ACM SIGKDD international conference on Knowledge discovery and data mining*, pages 265–273. ACM, 2008.

[8] Noor Huijboom and Tijs Van den Broek. Open data: an international comparison of strategies. *European journal of ePractice*, 12(1):1–13, 2011.

[9] A Hundepool et al. Mu-argus 4.2 users maual. *Statistics Netherlands*, 2008.

[10] Anco Hundepool, Josep Domingo-Ferrer, Luisa Franconi, Sarah Giessing, Eric Schulte Nordholt, Keith Spicer, and Peter-Paul de Wolf. *Statistical Disclosure Control*. Wiley, 2012.

[11] Vijay S Iyengar. Transforming data to satisfy privacy constraints. In *Proceedings of the eighth ACM SIGKDD international conference on Knowledge discovery and data mining*, pages 279–288. ACM, 2002.

[12] Kristen LeFevre, David J DeWitt, and Raghu Ramakrishnan. Incognito: Efficient full-domain k-anonymity. In *Proceedings of the 2005 ACM SIGMOD International Conference on Management of Data*, pages 49–60. ACM, 2005.

[13] Kristen LeFevre, David J. DeWitt, and Raghu Ramakrishnan. Mondrian multidimensional k-anonymity. In Ling Liu, Andreas Reuter, Kyu-Young Whang, and Jianjun Zhang, editors, *International Conference on Data Engineering*, page 25. IEEE Computer Society, 2006.

[14] Ninghui Li, Tiancheng Li, and Suresh Venkatasubramanian. t-closeness: Privacy beyond k-anonymity and l-diversity. In Rada Chirkova, Asuman Dogac, M. Tamer zsu, and Timos K. Sellis, editors, *International Conference on Data Engineering*, pages 106–115. IEEE, 2007.

[15] Ninghui Li, Tiancheng Li, and Suresh Venkatasubramanian. Closeness: A new privacy measure for data publishing. *IEEE Trans. Knowl. Data Eng.*, 22(7):943–956, 2010.

[16] Bing-Rong Lin and Daniel Kifer. A framework for extracting semantic guarantees from privacy. *CoRR*, abs/1208.5443, 2012.

[17] University of Texas at Dallas. Anonymization toolbox. `http://cs.utdallas.edu/dspl/cgi-bin/toolbox/index.php`, October 2013.

[18] Michal Sramka, Reihaneh Safavi-Naini, Jörg Denzinger, and Mina Askari. A practice-oriented framework for measuring privacy and utility in data sanitization systems. In *Proceedings of the 2010 EDBT/ICDT Workshops*, EDBT '10, pages 27:1–27:10, New York, NY, USA, 2010. ACM.

[19] L. Sweeney. k-anonymity: a model for protecting privacy. *International Journal on Uncertainty, Fuzziness and Knowledge-based Systems*, 10(5):557–570, 2002.

[20] Matthias Templ. Statistical disclosure control for microdata using the r-package sdcmicro. *Transactions on Data Privacy*, 1(2):67–85, 2008.

[21] Thijs van den Broek, Noor Huijboom, Arjanna van der Plas, Bas Kotterink, and Wout Hofman. Open overheid, January 2011. Retrieved May 2013.

Analysis of Direct Signal Recovery Scheme for DVB-T Based Passive Radars

Osama Mahfoudia Xavier Neyt
Royal Military Academy
Dept.CISS
Avenue de la renaissance 30, 1000 Bruxelles
osama.mahfoudia@rma.ac.be xavier.neyt@rma.ac.be

Abstract

In this work, a directed signal reconstruction scheme for Digital Video Broadcast Terrestrial-based passive radars is assessed. The direct signal reconstruction provides a noiseless and multipath-free estimate of the reference signal which improves the static clutter rejection (SCR) efficiency. The direct signal recovery is performed by demodulating and remodulating the base-band received signal. The recovery process induces errors leading to a mismatch between the estimated and the true copies of the reference signal. The impact of this mismatch on the SCR is studied and an expression is derived to evaluate the degradation of the SCR efficiency.

1 Introduction

Passive radars perform target detection using signals from non-cooperative sources of illumination in the environment. Target detection in passive radars requires reference and surveillance signals. On the principle, the reference signal is obtained by an antenna directed towards the transmitter and the surveillance signal is received by an antenna directed to the area of interest. In addition to the target echo, the surveillance signal contains direct path signal and multipath echoes which decreases the detection performances. To cope with this issue, undesirable echoes removal is performed using an adequate filter, this operation is named the static clutter rejection (SCR) [1].

The SCR process requires a noiseless multipath-free template of the reference signal to achieve the total undesirable echoes removal. However, the received reference signal is affected by reception noise and multipath fading which decreases the SCR efficiency. DVB-T based passive radars benefit of the reference signal reconstruction possibility; it is performed by demodulating and remodulating the received signal which increases the SCR efficiency. The process of the reference signal reconstruction and the encountered issues are detailed in the next sections.

This paper is organized as follows, section 2 presents the system model and details the demodulation/remodulation task. Section 3 treats the SCR operation and proposes an expression for SCR efficiency degradation. In the section 4, the simulation scheme is presented and simulation results are given to validate the derived expression. Section 5 concludes the paper.

2 DVB-T direct signal recovery

2.1 System model

Considering the DVB-T based passive radar presented in figure 1, we denote the received reference signal by $x_{ref}(n)$ and the surveillance signal by $x_s(n)$ [2]. The received reference signal $x_{ref}(n)$ is given by

$$x_{ref}(n) = \alpha_0 x(n-\tau_0) + \sum_{i=1}^{K-1} \alpha_i x(n-\tau_i) + \xi_r(n), \quad (1)$$

where, $x(n)$ is the transmitted signal after undergoing the effects of a frequency-selective channel H, α_0 is the complex gain of the direct path signal, the coefficient α_i represents the complex gain of the i^{th} static scatterer, τ_i is the delay corresponding to the i^{th} range-cell, K is the number of range-cells and $\xi_r(n)$ is the additive white Gaussian noise (AWGN) for the reference channel.

The surveillance signal $x_s(n)$ includes target returns in the form of delayed, attenuated and Doppler-shifted versions of the transmitted signal. In addition, it contains static clutter, noise and possible direct path signals. The surveillance signal model is

$$x_s(n) = \sum_{i=1}^{N} \beta_i x(n-\tau_i) + \sum_{l=1}^{M} \gamma_l x(n-\tau_l) exp(j\omega_l n) + \xi_s(n), \quad (2)$$

with β_i represents the scattering coefficient at the i^{th} static scatterer, N is the number of the considered range-cells, γ_l is the reflection coefficient for the l^{th} moving target, M is the moving targets number, ω_l the shift caused by the Doppler effect for the l^{th} moving target and $\xi_s(n)$ is the AWGN for the surveillance channel. If we note $z(n)$ the sum of the noise and the target echoes signal in the surveillance signal, we may write

$$x_s(n) = \sum_{i=1}^{N} \beta_i x(n-\tau_i) + z(n) \quad with \quad z(n) = \sum_{l=1}^{M} \gamma_l x(n-\tau_l) exp(j\omega_l n) + \xi_s(n). \quad (3)$$

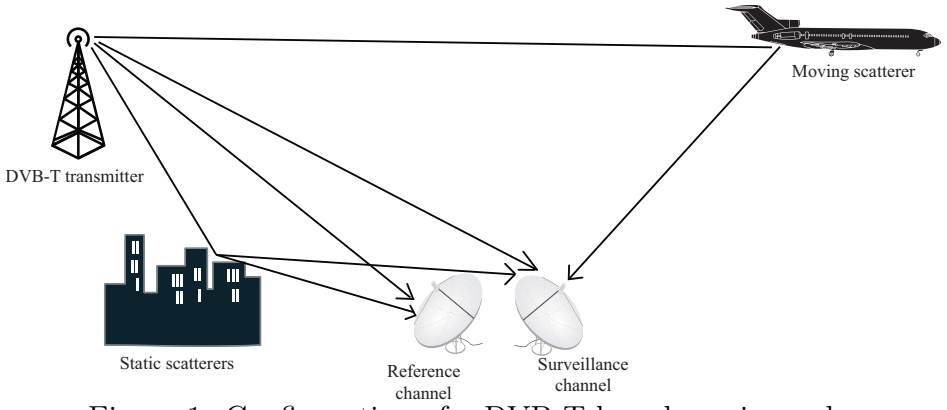


Figure 1: Configuration of a DVB-T based passive radar.

2.2 Synchronization

The demodulation of the base-band received reference signal is preceded by the transmitter-receiver synchronization. The synchronization is achieved by the estimation of the following parameters: the coarse time delay, the fractional frequency offset (FFO), the integer time delay and the integer frequency offset (IFO).

- The coarse time synchronization aligns the FFT window with the received DVB-T symbols by estimating the beginning of the DVB-T symbol.
- The integer time synchronization estimates the order of each DVB-T symbol.

- The frequency synchronization compensates the transmitter-receiver frequency offset, it consists of two steps: FFO and IFO compensations.

Coarse time and FFO estimation exploits the cyclic nature of DVB-T symbols; a cyclic prefix (called also the guard interval) is inserted at the beginning of each DVB-T symbol. The cyclic prefix is formed by the last N_g samples of the DVB-T symbol, with N_g the length of the guard interval. In the present work, an autocorrelation-based method is applied for coarse time and FFO estimation [3].

The subcarrier pilots (continuous pilots and the scattered pilots) are one type of the transmitted subcarriers, they are modulated by a known Pseudo-Random Binary Sequence (PRBS) with a boosted amplitudes compared to other subcarriers ($\pm 4/3$). In addition to the use for synchronization, subcarriers pilots are used for channel estimation and equalization [7].

After coarse time correction and FFO compensation, a pilot-aided method is applied for IFO and integer time estimations [4]. IFO compensation is required to align each subcarrier with the corresponding FFT bin and integer time synchronization estimates the scattered pilots pattern for the first symbol. The received reference signal synchronization is achieved by the frequency offset compensation (FFO and IFO) of the time synchronized signal (after considering the coarse time delay).

2.3 Demodulation

The demodulation of the synchronized signal is performed by removing the cyclic prefix from each DVB-T symbol and applying an FFT on the useful samples. The result for each DVB-T symbol is a constellation of coded symbols (64-QAM in our case). The received constellation is affected by propagation channel effect, noise and synchronization imperfections. The k^{th} coded symbol($k_{max} = 1705$ for the 2k-mode and $k_{max} = 6817$ for the 8k-mode) from the l^{th} DVB-T symbol is given by

$$X_{ref}(l,k) = H(l,k)X_t(l,k) + W(l,k), \tag{4}$$

where $H(l,k)$ is the channel weight, $X_t(l,k)$ is the exact transmitted QAM symbol and $W(l,k)$ includes the AWGN and the multipath.

The transmitted coded symbols for subcarrier pilots are known, which allows the channel response estimation over pilot subcarriers bins. Then, the resulting estimate is interpolated to obtain the channel response for the remaining subcarriers. In this work, the least-squares (LS) estimator is used for channel estimation [6]. The LS estimator ignores the effect of the noise $W(l,k)$ and gives the channel estimate for subcarrier pilots by

$$\hat{\boldsymbol{H}}_{\boldsymbol{p}}(l) = \boldsymbol{X}_{\boldsymbol{t,p}}^{-\mathbf{1}}(l)\boldsymbol{X}_{\boldsymbol{ref,p}}(l), \tag{5}$$

with $\boldsymbol{X}_{\boldsymbol{t,p}}$ is a matrix with the known transmitted pilot amplitudes on its diagonal ($\pm 4/3$) and $\boldsymbol{X}_{\boldsymbol{ref,p}}$ represents the array of the received symbols $X_{ref}(l,k)$ at the pilot subcarriers, i.e., $k \in P$ with P indicates the pilot subcarriers positions.

The channel response for the l^{th} DVB-T symbol, $\hat{\boldsymbol{H}}(l)$, is obtained by the interpolation of the pilots response $\hat{\boldsymbol{H}}_{\boldsymbol{p}}(l)$. After the channel estimation, the equalization of the received symbols is performed by

$$X_{ref,eq}(l,k) = X_{ref}(l,k)/\hat{H}(l,k). \tag{6}$$

The transmitted symbols, $\hat{X}_t(l,k)$, are estimated by approximating the equalized symbols, $X_{ref,eq}(l,k)$, to the nearest QAM symbol. The LS estimator is characterized by its simplicity and its sensitivity to noise. One can reduce the noise effect by averaging the channel response for pilot subcarriers $\boldsymbol{X}_{\boldsymbol{ref,p}}(l)$ over L DVB-T symbols.

2.4 Remodulation

It has been proven that remodulating the recovered QAM symbols without the reintroduction of the channel effects and the frequency offsets creates a mismatch between the reconstructed and the true reference signals [2]. Therefore, The channel effect is reintroduced as follows

$$\hat{X}(l,k) = \hat{H}(l,k)\hat{X}_t(l,k). \tag{7}$$

The channel estimate $\hat{H}$ is affected by errors caused by zero-forcing (LS estimator), interpolation of the subcarrier pilots channel response and synchronization imperfections. If we note the estimation error e, the channel estimate is

$$\hat{H}(l,k) = H(l,k) + e(l,k). \tag{8}$$

We define the normalized estimation error variance $\sigma^2_{\Delta H}$ as

$$\sigma^2_{\Delta H} = \sigma^2_e/\sigma^2_H = SNR_H^{-1}, \tag{9}$$

where σ^2_H is the channel variance, σ^2_e represents the channel estimation error variance and SNR_H is the signal-to-noise ratio for the channel estimate. We use (8) in (7), we get

$$\hat{X}(l,k) = H(l,k)\hat{X}_t(l,k) + V(l,k) \quad with \quad V(l,k) = e(l,k)\hat{X}_t(l,k). \tag{10}$$

The SNR for the symbols $\hat{X}$ is

$$SNR_{\hat{X}} = (\sigma^2_H\sigma^2_X)/(\sigma^2_e\sigma^2_X) = \sigma^2_H/\sigma^2_e = (\sigma^2_{\Delta H})^{-1}. \tag{11}$$

The noiseless multipath-free estimate of the reference signal, $\hat{x}(n)$, is obtained by applying an IFFT on the symbols $\hat{X}(l,k)$ [7]. The remodulation result is given by

$$\hat{x}(n) = x(n) + v(n), \tag{12}$$

with $x(n)$ is the true multipath-free estimate of the reference signal and $v(n)$ represents the estimation error.

If we consider $v(n)$ (with variance σ^2_v) as a noise uncorrelated with $x(n)$ (with variance σ^2_x), the SNR for the estimated signal $\hat{x}$ is determined by

$$SNR_{\hat{x}} = \sigma^2_x/\sigma^2_v. \tag{13}$$

Since $\hat{x}$ is the time-domain version of $\hat{X}$, we may write $SNR_{\hat{x}} = SNR_{\hat{X}}$. Using the SNR equality with (11) leads to

$$\sigma^2_{\Delta H} = SNR_{\hat{x}}^{-1}. \tag{14}$$

If we consider a channel response averaging along L DVB-T symbols, the channel estimate is

$$\hat{H}_{av}(k) = \frac{1}{L}\sum_{l=1}^{L}(H(l,k) + e(l,k)), \tag{15}$$

the averaging process reduces the estimation error variance by a factor of L, we may write the normalized estimation error in (9) as

$$\sigma^2_{\Delta\hat{H}_{av}} = \sigma^2_{\Delta H}/L. \tag{16}$$

Thus, equation (14) becomes

$$SNR_{\hat{x}} = L\,(\sigma^2_{\Delta H})^{-1}, \tag{17}$$

where, $SNR_{\hat{x}}$ is the SNR of the estimated reference signal, $\sigma^2_{\Delta H}$ is normalized estimation error variance and L is the channel averaging length.

3 Static clutter rejection

The static clutter rejection removes zero-Doppler echoes form the received signal, which allows the detection of targets with weak echoes. The SCR efficiency is evaluated with the residual power P_r, it is the power of the post-SCR signal [5]. For a perfect SCR, P_r represents the power of targets echoes and surveillance channel noise. Poor SCR leads to a P_r with residual static clutter. The SCR is performed using an adequate filter which requires a reference signal to operate. One of the SCR efficiency degradation factors is a noisy reference signal. The recovery of the reference signal provides a noiseless multipath free reference signal increasing the SCR efficiency. In practice, even the recovered signal is affected by channel estimation errors among other factors. In this section, a theoretical approach is applied to retrieve an expression relating the channel estimation errors to the post-SCR signal power. We consider a finite impulse response (FIR) Wiener filter [8] for the SCR, the filter weights, $\boldsymbol{w}$, are defined by

$$\boldsymbol{w} = \boldsymbol{R}_{\hat{\boldsymbol{x}},\hat{\boldsymbol{x}}}^{-1} \boldsymbol{r}_{\hat{\boldsymbol{x}},\boldsymbol{x_s}}, \tag{18}$$

with $\boldsymbol{R}_{\hat{\boldsymbol{x}},\hat{\boldsymbol{x}}}$ is the autocorrelation matrix of $\hat{x}$ and $\boldsymbol{r}_{\hat{\boldsymbol{x}},\boldsymbol{x_s}}$ is the cross-correlation of $\hat{x}$ and x_s. The values of the previous quantities can be approximated as follows

$$\begin{cases} \boldsymbol{R}_{\hat{\boldsymbol{x}},\hat{\boldsymbol{x}}} = diag(\sigma_x^2 + \sigma_v^2) \\ \boldsymbol{r}_{\hat{\boldsymbol{x}},\boldsymbol{x_s}}(i) = \beta_i \sigma_x^2 \end{cases} \tag{19}$$

Hence, using (19) in (18) yields to relate Wiener filter weights to the exact multipath coefficients;

$$w_i = \beta_i / (1 + SNR_{\hat{x}}^{-1}). \tag{20}$$

The SCR output signal is denoted by $y(n)$, it is given by subtracting the Wiener filter output $\hat{x}_s(n)$ from the surveillance signal $x_s(n)$;

$$y(n) = x_s(n) - \hat{x}_s(n) \quad with \quad \hat{x}_s(n) = \sum_{i=1}^{N} w_i \hat{x}(n - \tau_i). \tag{21}$$

After replacing $x_s(n)$ and $\hat{x}_s(n)$ by their values, we get

$$y(n) = \sum_{i=1}^{N} \beta_i x(n - \tau_i) + z(n) - \sum_{i=1}^{N} w_i \hat{x}(n - \tau_i), \tag{22}$$

where $z(n)$ includes targets echoes and surveillance channel noise; it is the residual signal after a perfect SCR (equation 3). It follows that

$$y(n) = \sum_{i=1}^{N} (\beta_i - w_i) x(n - \tau_i) + z(n) - \sum_{i=1}^{N} w_i v(n - \tau_i). \tag{23}$$

The difference $(\beta_i - w_i)$ can be defined from (20) as $\beta_i - w_i = w_i SNR_{\hat{x}}^{-1}$, this yields to

$$y(n) = SNR_{\hat{x}}^{-1} \sum_{i=1}^{N} w_i x(n - \tau_i) + z(n) - \sum_{i=1}^{N} w_i v(n - \tau_i). \tag{24}$$

The post-SCR signal power can be approximated by

$$P_y = SNR_{\hat{x}}^{-2} \sigma_x^2 \sum_{i=1}^{N} \mid w_i \mid^2 + P_z + \sigma_v^2 \sum_{i=1}^{N} \mid w_i \mid^2 . \tag{25}$$

Therefore,

$$P_y = (SNR_{\hat{x}}^{-2}\sigma_x^2 + \sigma_v^2)\sum_{i=1}^{N} \mid w_i \mid^2 + P_z, \tag{26}$$

we get

$$P_y = \sigma_x^2(SNR_{\hat{x}}^{-2} + SNR_{\hat{x}}^{-1})\sum_{i=1}^{N} \mid w_i \mid^2 + P_z. \tag{27}$$

We denote the static clutter power by P_{sc}

$$P_{sc} = \sigma_x^2 \sum_{i=1}^{N} \mid \beta_i \mid^2 . \tag{28}$$

To represent P_y as a function of P_{sc} and $SNR_{\hat{x}}$, the value of w_i in (27) is replaced by (20):

$$P_y = (SNR_{\hat{x}}^{-2} + SNR_{\hat{x}}^{-1})(1 + SNR_{\hat{x}}^{-1})^{-2}\sigma_x^2 \sum_{i=1}^{N} \mid \beta_i \mid^2 + P_z. \tag{29}$$

The post-SCR signal power is summarized by writing

$$P_y = P_z + P_{sc}/(1 + SNR_{\hat{x}}). \tag{30}$$

Finally, we replace (17) in (30)

$$P_y = P_z + P_{sc}/(1 + L(\sigma_{\Delta H}^2)^{-1}). \tag{31}$$

Thus, equation (31) gives an estimate of the residual static clutter power. It proves the impact of the channel estimation error on the SCR performances; the SCR efficiency decreases significantly for high channel estimation error.

4 Simulation

4.1 Simulation scheme

Figure 2 illustrates the simulation scheme. The reference signal is formed by a strong line of sight signal, multipath components and additive white Gaussian noise (AWGN). The surveillance signal comprises multipath returns, moving targets returns and AWGN. In the reference signal reconstruction stage, an estimate of the propagation channel is used: $\hat{H} = H + e$ with H denotes the exact channel and e represents the estimation error. The SCR stage is performed using a FIR Wiener filter.

To investigate the impact of the channel estimation error e on the SCR performances, the residual power for different values of e is calculated. The channel estimation error is modeled by a zero-mean complex Gaussian noise with variance σ_e^2. The channel H is considered time-unvarying during the observation time.

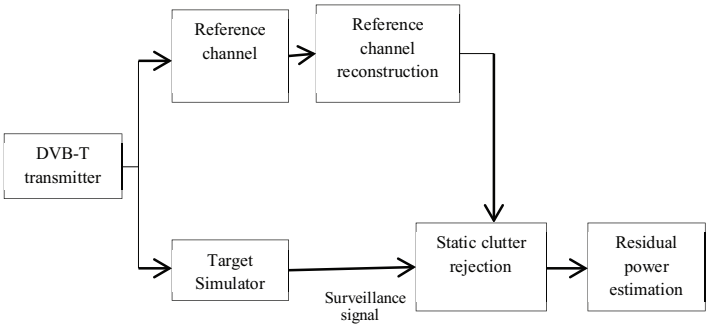


Figure 2: Simulation scheme.

4.2 Simulation results

Figure 3 is a comparison of the simulation results and the results from the model 31 for the case $L = 1$ (no channel averaging). We notice that the model fits perfectly the simulation. The results show the sensitivity of the SCR performances for channel estimation error. For large channel estimation errors (≥ 10 dB) the SCR effect vanishes.

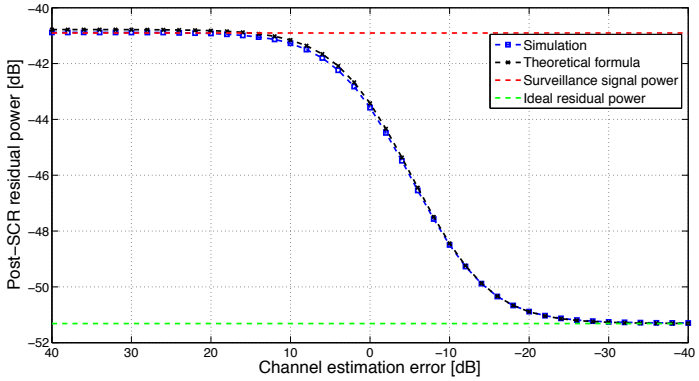


Figure 3: Validation of the theoretical formula.

To reduce the impact of the channel estimation errors, we perform an averaging of the subcarrier pilots response for L DVB-T symbols. Figure 4 shows the impact of channel response averaging ($L = 100$) on the SCR efficiency; a considerable improvement can be noticed.

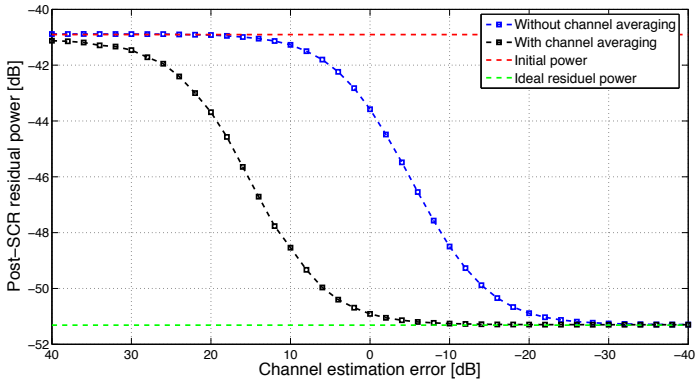


Figure 4: Averaging channel response impact on SCR efficiency.

5 Conclusion

In this paper, a reference signal recovery method is analyzed. The methods applied on the synchronization, demodulation and remodulation are tested on real-world data proving their efficiency. A theoretical analysis led to the expression for the post-SCR signal power. Simulation results proved that the channel estimation errors effect can be reduced by an averaging of the pilots channel response.

References

[1] D. Poullin, "Passive detection using digital broadcasters (DAB, DVB) with COFDM modulation," Radar, Sonar and Navigation, IEE Proceedings, vol 152 .number 3 pages 143-152, June 2005.

[2] J.E. Palmer, H.A. Harms, S.J. Searle and L.M. Davis, "DVB-T Passive Radar Signal Processing," Signal Processing, IEEE Transactions,vol. 61, pages 2116-2126, April 2013.

[3] S.H. Chen et al., "Mode detection, synchronization, and channel estimation for DVB-T OFDM receiver," Global Telecommunications Conference,GLOBECOM '03. IEEE, vol. 5, pages 2416-2420, December 2003.

[4] Peng Liu, Bing-bing Li, Zhao-yang Lu and Feng-kui Gong, "A new frequency synchronization scheme for OFDM," Consumer Electronics, IEEE Transactions on, Vol. 50, pages 823-828, August 2004.

[5] M.K. Baczyk and M. Malanowski, "Decoding and reconstruction of reference DVB-T signal in passive radar systems," 11^{th} International Radar Symposium(IRS), pages 1-4, June 2010.

[6] A.M. Khan V. Jeotiet and M.A. Zakariya, "Improved pilot-based LS and MMSE channel estimation using DFT for DVB-T OFDM systems," Wireless Technology and Applications, 2013 IEEE Symposium on, pages 120-124, September 2013.

[7] European Telecommunications Standards Institute (ETSI), "Digital Video Broadcasting (DVB); Framing Structure, Channel Coding and Modulation for Digital Terrestrial Television," June 2004.

[8] S. Haykin, "Adaptive Filter Theory - Third Edition," Prentice Hall Inc., pages 194-235, 1996.