PROCEEDINGS

of the

2021 Symposium on Information Theory and Signal Processing
in the Benelux

May 20-21, TU Eindhoven

Organizers: Ruud van Sloun and Boris Škorić

The symposium is organized under the the auspices of
the Werkgemeenschap voor Informatie- en Communicatietheorie (WIC)
and the IEEE Benelux Signal Processing Chapter

Supported by:
Gauss Foundation (sponsoring best student paper award)
IEEE Benelux Information Theory Chapter
IEEE Benelux Signal Processing Chapter
Werkgemeenschap voor Informatie- en Communicatietheorie (WIC)

A catalogue record is available from the Eindhoven University of Technology Library
ISBN 978-90-386-5318-1

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
36.	2015	Brussels, Belgium	ISBN 978-2-8052-0277-3
37.	2016	Louvain-la-Neuve, Belgium	ISBN 978-2-9601884-0-0
38.	2017	Delft, The Netherlands	ISBN 978-94-6186-811-4
39.	2018	Enschede, The Netherlands	ISBN 978-90-365-4570-9
40.	2019	Gent, Belgium	ISBN 9789491857034

Preface

This event is the 41st edition of a sequence of annual symposia that started in the 1980s, under the auspices of the Werkgemeenschap voor Informatie- en Communicatietheorie (WIC). Since 2011 the symposia are co-organized with the IEEE Benelux Signal Processing Chapter.
In 2020, the year of the Corona pandemic, the event was canceled. In the spring of 2021 it was still not possible to have gatherings on campus, so the symposium was held online.

We are very happy that our intended keynote speakers for 2020, Nir Shlezinger (School of Electrical and Computer Engineering, Ben-Gurion) and Slava Voloshynovskiy (Stochastic Information Processing, University of Geneva), were still keen to make an appearance. We had 35 contributions from researchers in the Benelux, which were presented in the form of a talk or a poster, and which are all documented in these proceedings, either as a full paper or as an (extended) abstract. The social part of the event was by necessity very different from the usual outings. Participants were mingling in a virtual room that allowed for easy ad-hoc formation of groups. The posters were also presented in this room.

We thank the keynote lecturers for accepting our invitation, all the authors for their contributions to the scientific program, all participants for their presence, and the Gauss Foundation for sponsoring the best student paper award.

Eindhoven, May 2021

Boris Škorić and Ruud van Sloun
(symposium organizers and proceedings editors)

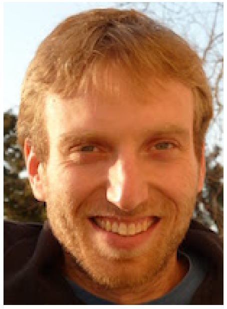
Table of Contents

Keynote 1

Model-based deep learning in signal processing and communications

Nir Shlezinger

Recent years have witnessed a dramatically growing interest in machine learning (ML) methods. These data-driven trainable structures have demonstrated an unprecedented empirical success in various applications, including computer vision and speech processing. The benefits of ML-driven techniques over traditional model-based approaches are twofold: First, ML methods are independent of the underlying stochastic model, and thus can operate efficiently in scenarios where this model is unknown, or its parameters cannot be accurately estimated; Second, when the underlying model is extremely complex, ML algorithms have demonstrated the ability to extract and disentangle the meaningful semantic information from the observed data. Nonetheless, not every problem can and should be solved using deep neural networks (DNNs). In fact, in scenarios for which model-based algorithms exist and are computationally feasible, these analytical methods are typically preferable over ML schemes due to their theoretical performance guarantees and possible proven optimality. Notable application areas where model-based schemes are typically preferable, and whose characteristics are fundamentally different from conventional deep learning applications, include signal processing and digital communications. In this talk, I will present methods for combining DNNs with traditional model-based algorithms. We will show how hybrid model-based/data-driven implementations arise from classical methods in signal processing, compressed sensing, control, and digital communications, and show how fundamental classic techniques can be implemented without knowledge of the underlying statistical model, while achieving improved robustness to uncertainty.



Nir Shlezinger is an assistant professor in the School of Electrical and Computer Engineering in Ben-Gurion University, Israel. He received his B.Sc., M.Sc., and Ph.D. degrees in 2011, 2013, and 2017, respectively, from Ben-Gurion University, Israel, all in electrical and computer engineering. From 2017 to 2019 he was a postdoctoral researcher in the Technion, and from 2019 to 2020 he was a postdoctoral researcher in Weizmann Institute of Science, where he was awarded the FGS prize for outstanding achievements in postdoctoral research. His research interests lie in the intersection of signal processing, machine learning, communications, and information theory.

Keynote 2

Information Bottleneck through variational glasses

Slava Voloshynovskiy

The Information Bottleneck (IB) principle has become an important element in information-theoretic analyses of deep models. Many state-of-the-art generative models of both Variational Autoencoder (VAE) and Generative Adversarial Networks (GAN) families use various bounds on mutual information terms to introduce certain regularisation constraints. Accordingly, the main difference between these models consists in added regularisation constraints and targeted objectives. However, it is not always obvious what the underlying assumptions behind these constraints are. In this talk, we will consider the IB framework for several applications covering supervised and semi-supervised classification, generative models based in VAE and GAN families and regression problems. We will will show how applying a variational decomposition to mutual information leads to a common structure and allows us to easily establish connections between these models and to analyze underlying assumptions. We will show the advantages of these models in semi-supervised classification and demonstrate some interesting connections to existing generative models such as VAE, β-VAE, AAE, InfoVAE and VAE/GAN. We show that many known methods can be considered as a product of variational decomposition of mutual information terms in the IB framework. Finally, we extend the same framework to anomaly detection problems and several classes of regression problems.

Slava Voloshynovskiy (IEEE SM'11) received a radio engineer degree from Lviv Polytechnic Institute, Lviv, Ukraine, in 1993 and a PhD degree in electrical engineering from the State University Lvivska Polytechnika, Lviv, Ukraine, in 1996. From 1998 to 1999, he was a visiting scholar with the University of Illinois at Urbana-Champaign. Since 1999 he has been with the University of Geneva, Switzerland, where he is currently a Professor with the Department of Computer Science and head of the Stochastic Information Processing group. His research interests are in information-theoretic aspects of stochastic image modeling, multimedia security and privacy, machine learning and physical object security.

Slava served as Associate Editor for IEEE TIFS (2013-2015). He was an elected member of the IEEE Information Forensics and Security Technical Committee (2011-2013) where he was an area chair in information-theoretic security and an associated member since 2015. He served as a guest editor to several special issues dedicated to information theory and security. He is a member of Eurasip BForSec SAT. He has served as a consultant to private industry and co-founded three companies. He leads several Swiss projects in Big Data acquisition, processing and analysis related to solar imaging and collaborates with CERN on machine learning in physics. S. Voloshynovskiy co-organized several interdisciplinary events between AI and physics. He was a recipient of the Swiss National Science Foundation Professorship Grant in 2003.

Contextual Pyramid Attention Network for Building Segmentation in Aerial Imagery

Clint Sebastian Raffaele Imbriaco Egor Bondarev Peter H.N de With
Eindhoven University of Technology,
Dept. of Electrical Engineering , SPS-VCA
Eindhoven, The Netherlands
{c.sebastian, r.imbriaco, e.bondarev, p.h.n.de.with}@tue.nl

Abstract

Building extraction from aerial images has several applications in problems such as urban planning, change detection, and disaster management. With the increasing availability of data, Convolutional Neural Networks (CNNs) for semantic segmentation of remote sensing imagery has improved significantly in recent years. However, convolutions operate in local neighborhoods and fail to capture non-local features that are essential in semantic understanding of aerial images. In this work, we propose to improve building segmentation of different sizes by capturing long-range dependencies using contextual pyramid attention (CPA). The pathways process the input at multiple scales efficiently and combine them in a weighted manner, similar to an ensemble model. The proposed method obtains state-of-the-art performance on the Inria Aerial Image Labelling Dataset with minimal computation costs. Our method improves 1.8 points over current state-of-the-art methods and 12.6 points higher than existing baselines on the Intersection over Union (IoU) metric without any post-processing.

1 Introduction

The developments in the systematic collection and organization of remote sensing imagery have resulted in several high-resolution aerial imagery datasets. Information from aerial imagery plays a key role in urban planning, disaster aversion, and change detection. Building detection is a crucial aspect for the aforementioned applications. Depending on the geographical region and conditions, building structures have different shapes and sizes. This challenge is particularly addressed by Maggiori *et al.* [1]. They created a dataset of labeled aerial imagery from different locations for this problem, such that a model trained from a variety of sources generalizes to the task of segmentation. Semantic segmentation in aerial imagery is challenging due to variable lighting conditions, shapes/sizes, and large intraclass variations. In this research, we address the problem of improving the building segmentation by utilizing attentive multi-scale pathways. Each of the paths exploits non-local neighborhoods that account for buildings of varying sizes. This allows our network to learn long-range dependencies at various scales with minimal computation costs. In addition to attentive multi-scale pathways, we also incorporate a channel-wise attention module to model interdependencies across channels. In summary, our contributions are as follows.

- We introduce a self-attention based contextual pyramid attention (CPA) module that accounts for various building sizes to segment buildings in aerial images. The proposed module outperforms current state-of-the-art methods by about 2% on the IoU metric and 12.6% over FCN baselines.

- Through experiments, we also show that our base model offers competitive performance to current state-of-the-art methods, while having much lower inference costs. We also provide ablation studies on the impact of our proposed module and other comparisons.

2 Related Work

Convolutional Neural Networks (CNNs) have improved the performance of semantic segmentation significantly in recent years. Encoder-Decoder architectures are a popular choice for segmentation. Using fully convolutional architectures for segmentation is proposed in [2]. Architectures such as FCN, U-Net, DenseNet, etc., are often applied to obtain higher quality segmentations [2, 3, 4]. Nowadays, most segmentation architecture designs are based on using a pre-trained backbone such as VGGNet, ResNet, and a complex decoder such as Pyramid Scene Parsing (PSP) or Atrous Spatial Pyramid Pooling (ASPP)[5, 6, 7, 8, 9, 10].

Due to the unique nature of remote sensing imagery, several custom architectures and loss functions have been developed for aerial image building segmentation [11, 12]. In [11], a multi-task learning approach is introduced, where a distance-transform loss function is applied in conjunction with the cross-entropy loss. Similar to our work, [13] proposes to combine local and global features to improve semantic segmentation of buildings in aerial imagery. However, the combination of two VGGNets makes the overall inference computationally expensive. Our proposed approach circumvents this by splitting only the last convolutional block to operate at multiple scales. In [14], a joint multi-stage multi-task approach is used, where the first stage trains a segmentation network, and the second stage trains for geo-localization using a multi-task loss function. Apart from these, post-processing techniques such as Conditional Random Fields (CRFs) and test-time augmentations are applied to improve segmentation performance. In [15], a recurrent network is applied in a fully convolutional network that exploits a decoder network fusing features from the encoder layer in a similar fashion to feature pyramid networks (FPN). An FPN utilizes a pyramidal hierarchy of features extracted from an encoder that is later combined with the decoder via lateral connections [16]. However, as FPNs utilize only a few of the earlier layers of the encoder for lateral connections, the features might not be rich as extracting from deeper layers.

A common practice to retain feature resolution is using dilated convolutions. Due to the larger feature resolution and the repeated object patterns in aerial imagery, capturing long-range relations is beneficial. However, a single-sized feature map may not be able to capture all objects of variable sizes. For example, spatial relationships across large buildings are easier to capture when feature maps have a low resolution. When buildings are smaller in size, larger feature resolution is suited to capture more fine-grained details. Using a decoder such as FPN is beneficial for high-resolution features, however, as they are not obtained from deeper layers, they lack rich semantics. The proposed approach in this research is suited for these properties that are typically exhibited in aerial imagery. We particularly address the problem of improving semantic segmentation of buildings at various sizes.

3 Method

The architecture is composed of an encoder and decoder, similar to other segmentation networks. An overview of the proposed method is shown in Fig. 1. Each component of our model is described in the following subsections.

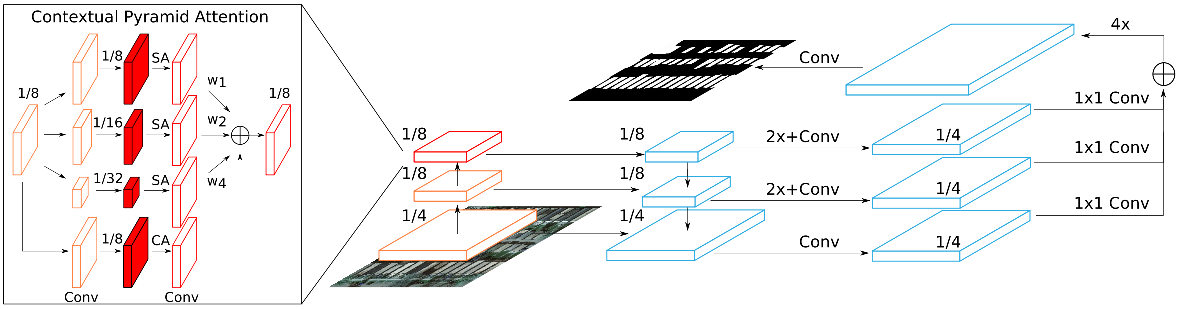


Figure 1: Overview of the proposed network. The network uses a ResNet backbone for the encoder and Feature Pyramid Network decoder. The proposed context pyramid attention (CPA) is applied after Block 4 of ResNet. CPA consists of contextual and channel-wise attention modules, where the contextual attention operates at multiple scales to produce a weighted output that is fed to the FPN decoder (CPA module at the left).

3.1 Architecture

Encoder: We consider three backbones for training, ResNet18, ResNet101 and Squeeze and Excitation ResNeXt101 (SEResNeXt101). ResNet18 serves as the light architecture that may be suited for fast inference, whereas SEResNeXt101 offers higher performance with additional computation cost. SEResNeXt101 incorporates an additional Squeeze and Excitation module along with aggregated residual connections.

Decoder: The decoder is a Feature Pyramid Network (FPN), similar to Semantic FPN in [17]. FPN combines features at different spatial resolutions through lateral connections to a top-down decoder. The lateral connections are established by combining the bottom-up outputs at each corresponding level of the top-down decoder. Each of the combined top-down decoder features is upsampled to $1/4^{th}$ of the input resolution and is combined via summation to produce rich features that are transformed into a segmentation mask.

3.2 Contextual Pyramid Attention block

The proposed Contextual Pyramid Attention block is composed of two parts, a contextual attention module that captures long-range spatial dependencies and a channel-wise attention module. Both proposed attention modules are based on self-attention [18, 19].

Contextual attention: Buildings in aerial imagery appear in various sizes, and structural features are often redundant in a given region (e.g. repeated houses structure in a neighborhood). This non-local information is not immediately processed by convolutions as they operate in a limited region defined by the kernel size. Self-attention, on the other hand, can bring in the understanding of long-range dependencies capturing the relation between repeated structures. It is a standard practice in semantic segmentation to use dilated convolutions in deeper layers, to increase the feature resolution. This is beneficial for smaller buildings and furthermore, self-attention blocks can capture fine-grained relations across the image. However, the contextual information of buildings of different sizes becomes challenging for two reasons. First, the dilated convolutions retain the feature resolution and self-attention may then retain spatial relations that are too fine-grained for larger buildings, which may not provide sufficient spatial context. Second, if the dilated convolutions are not applied to retain feature resolution, this may fail to capture relations between smaller buildings. Our

motivation using multi-scale pathways stems from these two reasons.
Method. Given a feature $\mathbf{F} \in \mathbb{R}^{C\times H\times W}$, we apply convolutions to generate key, query and value features $\mathbf{K}, \mathbf{Q}, \mathbf{V} \in \mathbb{R}^{C\times H\times W}$. The tensors $\mathbf{K}, \mathbf{Q}$ and $\mathbf{V}$ are reshaped into $N \times C$, where $N = H \times W$. The self-attention (SA) operation is defined as

$$\mathbf{A}_s = \mathrm{SA}(\mathbf{F}) = \gamma_s \cdot (\mathrm{Softmax}(\mathbf{K}\mathbf{Q}^T))\mathbf{V}, \tag{1}$$

where $\mathbf{A}_s \in \mathbb{R}^{C\times N}$ is then reshaped into $C \times H \times W$ and γ_s is a learned parameter. The output of the self-attention is the sum of $\mathbf{A}_s$ and $\mathbf{F}$. The resolution of $\mathbf{F}$ is retained by dilated convolutions ($1/8^{th}$ of the input image resolution). To capture contextual long-range dependencies, we operate the deep features $\mathbf{F}$ at various scales s. Contextual attention is

$$\mathbf{C}_s = \sum_{s\in(1,2,4)} w_s \cdot \mathrm{Conv}((\mathbf{F}_{1/s} + \mathrm{SA}(\mathrm{Conv}(\mathbf{F}_{1/s})))_s), \tag{2}$$

where Conv is the convolution operation, followed by ReLU activation and Batch Normalization [20, 21]. The features are downsampled by a factor of $1/s$, followed by Conv and SA operations. The resultant features are reshaped and upsampled to the original size, followed by another Conv operation. Each of these paths is weighted by a parameter w_s. The sum of the pathways results in the contextual attention output $\mathbf{C}_s$.

Channel-wise attention: Contextual attention employs both spatial axes to establish interdependencies to model spatial information. However, higher-level class or object information is prevalent across channels, and hence, we apply channel-wise attention to model relations across them. The last feature block of ResNet is large with 2,048 channels and is computationally expensive to perform self-attention based operations. Therefore, the features are compressed to 1/4 of the total number of channels by 1 × 1 convolutions to obtain the features $\mathbf{F}_{1/c}$, where c is the compression factor across channels (note the difference with scale factor s). The key, query and value are all the same in this case ($\mathbf{F}_{1/c}$). However, to apply self-attention, the features are reshaped into $C \times N$, where $N = H \times W$. The channel-wise attention is now

$$\mathbf{A}_{cw} = \gamma_c \cdot (\mathrm{Softmax}(\mathbf{F}_{1/c}\mathbf{F}^T_{1/c}))\mathbf{F}_{1/c}. \tag{3}$$

The final output is the sum of $\mathbf{F}_{1/c}$ and $\mathbf{A}_{cw}$. Note that the affinity matrices (inputs to the softmax) generated through both attention mechanisms have different shapes. The contextual attention affinity matrix has a size of $N \times N$, whereas the channel-wise affinity matrix has a size of $C \times C$. The final contextual pyramid attention block is the sum of both contextual and channel-wise attention.

4 Experiments

4.1 Dataset and Evaluation

Dataset: The Inria Aerial Image Labeling Dataset consists of 360 RGB ortho-rectified aerial images at a spatial resolution of 0.3 meters. Each of the ortho-rectified images has a resolution of 5000 × 5000 pixels, covering a region of 1500 square meters per image. The dataset comprises of 10 different cities, out of which 5 cities are available for training. The images are taken at different urban conditions from various cities in Europe and America. The groundtruth is available only for the training set, where information is available for two classes: buildings and non-buildings. For fair comparisons, we follow the same evaluation protocol for testing as in [11, 22, 23]. Images 1–5 of each location are used for validation and Images 6–36 for training.

Method		Austin	Chicago	Kitsap Co.	West Tyrol	Vienna	Overall
FCN + MLP [1]	IoU	61.20	61.30	51.50	57.95	72.13	64.67
	Acc.	94.20	90.43	98.92	96.66	91.87	94.42
SegNet MT-Loss [11]	IoU	76.76	67.06	**73.30**	66.91	76.68	73.00
	Acc.	93.21	**99.25**	97.84	91.71	96.61	95.73
MSMT-Stage-1 [22]	IoU	75.39	67.93	66.35	74.07	77.12	73.31
	Acc.	95.99	92.02	99.24	97.78	92.49	96.06
2-levels U-Net +	IoU	77.29	68.52	72.84	75.38	78.72	74.55
aug. [23]	Acc.	96.69	92.40	99.25	98.11	93.79	96.05
ICT-Net [24]	IoU	-	-	-	-	-	75.50
	Acc.	-	-	-	-	-	96.05
ResNet18-FPN	IoU	77.89	69.73	65.65	76.95	79.49	75.51
-CPA (ours)	Acc.	96.77	92.70	99.25	98.13	94.09	96.19
ResNet101-FPN	IoU	78.93	**71.57**	68.06	78.29	81.03	77.03
-CPA (ours)	Acc.	96.98	93.20	99.27	98.29	94.61	96.47
SEResNeXt101-FPN	IoU	**80.15**	69.54	70.36	**80.83**	**81.43**	**77.29**
-CPA (ours)	Acc.	**97.18**	92.78	**99.32**	**98.46**	**94.67**	**96.48**

Table 1: (Left) Performance comparison of CPA with other state-of-the-art methods.

Evaluation: For evaluation, we use both Intersection over Union (IoU) and accuracy. IoU is measured across the building class. IoU is the ratio of true positive pixels to the pixels that are labeled as positive in the ground-truth or predictions. We also utilize accuracy, which is the ratio of correctly classified pixels to the total. Each dataset is evaluated separately and jointly, for both of these metrics.

Implementation details: Each of the networks are initialized with weights pre-trained on ImageNet [25]. The models are trained using the Adam optimizer with a learning rate of 0.00001 [26]. The objective function is cross-entropy and is trained for 35 epochs. The input images have a size of 500 × 500 pixels and are randomly rotated by 90°or flipped horizontally and vertically while training. Images are also augmented with minor changes in brightness and contrast. The image augmentations are applied using Albumentations [27], and the models are implemented in PyTorch. Each of the contextual and self-attention weights are intialized to 1 and 0.05, respectively.

4.2 Comparison with state-of-the-art

With the proposed module, we observe an improvement of 3-5% IoU across all the datasets except Kitsap Co.. Our method with a simple backbone such as ResNet18 offers competitive performance to current state-of-the-art results. Using a deeper backbone such as ResNet101 or SE-ResNeXt101 further improves performance. Compared to the previous state-of-the-art ICT-Net that uses U-Net with Dense Blocks and SE Blocks [8], we observe a gain of 1.8 IoU points. The improvement is higher than previous increases in recent years. Our method does not rely on post-processing, or multi-task learning methods, however, applying this may further improve performance. A visual comparison of our results is in Fig. 3. We note that on Kitsap Co., which contains the fewest buildings, our performance decreases slightly. We observe very high accuracy (99.32%), whereas IoU is lower, indicating skewness to classifying background better. On visual inspection, we have noticed errors in the ground truth where the background is annotated as buildings. The incorrect ground truth, when combined with the sparsely covered building regions, penalizes the IoU metric, thereby lowering performance.

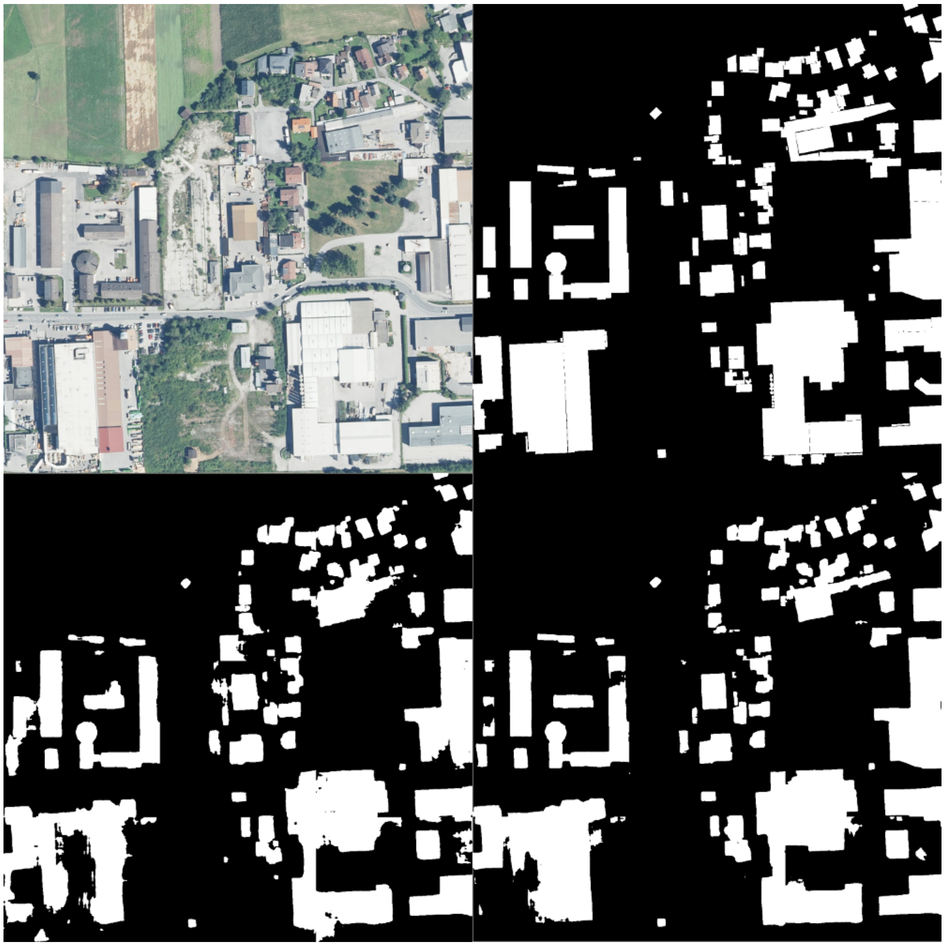


Figure 2: Comparison of building segmentation on Inria Aerial Image Labeling Dataset. RGB, GT and outputs (bottom row) ResNet101-FPN without (left) and with our module (right).

4.3 Ablation studies

To study the impact of attention, an ablation study is conducted without any attention, with self-attention and the proposed CPA. This is shown in Table 1 (top-right). The IoU improves by 0.87 and 1.54 points with the addition of self-attention and CPA over the ResNet-FPN. Furthermore, we conduct experiments to study the impact of different models on computation costs (Table 1 mid-right). For a tile of 5000 × 5000 pixels, ResNet18-FPN-CPA takes only 4.67 seconds on a GTX 1080Ti GPU, whereas ResNet101 and SE-ResNeXt101 take 12.16 and 14.07 seconds, respectively. In [23], a tile takes 160 seconds for processing, however, they use a K80 GPU, which is typically two times slower than our setup. Even so, compared to [23], our largest model (SE-ResNeXt101) is five times faster for inference, while offering higher performance.

5 Conclusions

We have presented a novel and effective method to improve building segmentation in aerial imagery. It consists of the contextual pyramid and channel-wise attention blocks to model long-range dependencies across spatial contexts to account for buildings of different sizes. The contextual pyramid attention combines contextual infor-

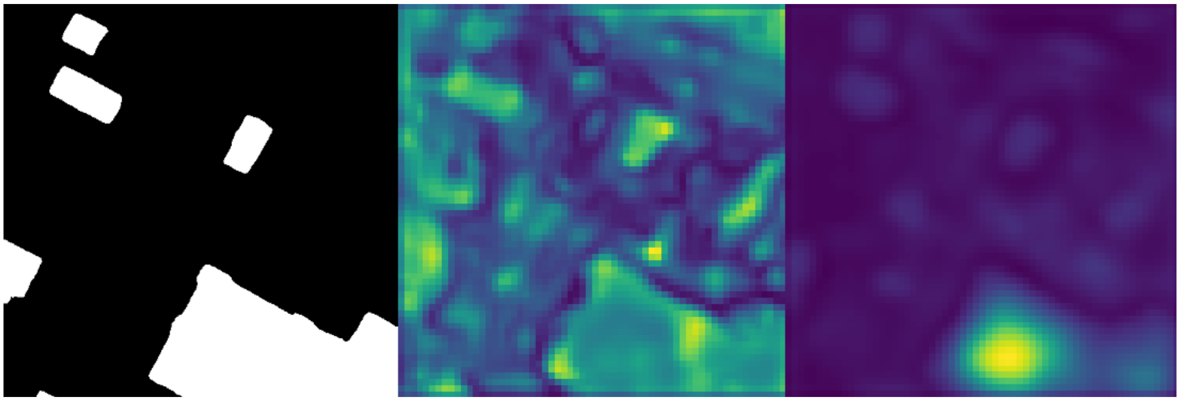


Figure 3: Prediction, high- and low-resolution attention map. The high-resolution learns semantics related to small buildings and fine-grained details of large buildings whereas low resolution learns semantics related to large buildings.

Model	Accuracy	IoU
ResNet18-FPN	95.9	73.97
-Self-attention	96.0	74.64
-CPA (ours)	96.2	75.51

Model	Time/tile	IoU
ResNet18	4.67 s	75.51
ResNet101	12.16 s	77.03
SE-ResNeXt101	14.07 s	77.29

Table 2: (Right) Comparison of our method with/without the CPA module. (Left) Inference time (in sec.) per tile of 5000 × 5000 pixels with different backbones with CPA module.

mation at multiple scales in an efficient manner with minimal computation overhead, whereas channel-wise attention captures interdependencies across channels. The proposed method achieves state-of-the-art performance on the Inria Aerial Image Labelling dataset by 1.8 and 12.6 IoU points over previous state of the art and baseline. Furthermore, we perform ablation experiments to study the impact of the CPA module and provide comparisons for computation costs. Our model offers high performance without any post-processing with low inference times.

References

[1] E. Maggiori, Y. Tarabalka, G. Charpiat, and P. Alliez, "Can semantic labeling methods generalize to any city? the inria aerial image labeling benchmark," in *2017 IEEE International Geoscience and Remote Sensing Symposium (IGARSS)*. IEEE, 2017, pp. 3226–3229.

[2] J. Long, E. Shelhamer, and T. Darrell, "Fully convolutional networks for semantic segmentation," in *Proceedings of the IEEE conference on computer vision and pattern recognition*, 2015, pp. 3431–3440.

[3] G. Huang, Z. Liu, L. van der Maaten, and K. Q. Weinberger, "Densely connected convolutional networks," in *The IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, July 2017.

[4] S. Jégou, M. Drozdzal, D. Vazquez, A. Romero, and Y. Bengio, "The one hundred layers tiramisu: Fully convolutional densenets for semantic segmentation," in *Proceedings of*

the IEEE Conference on Computer Vision and Pattern Recognition Workshops, 2017, pp. 11–19.

[5] K. Simonyan and A. Zisserman, "Very deep convolutional networks for large-scale image recognition," *arXiv preprint arXiv:1409.1556*, 2014.

[6] K. He, X. Zhang, S. Ren, and J. Sun, "Deep residual learning for image recognition," in *Proceedings of the IEEE conference on computer vision and pattern recognition*, 2016, pp. 770–778.

[7] S. Xie, R. Girshick, P. Dollár, Z. Tu, and K. He, "Aggregated residual transformations for deep neural networks," in *Proceedings of the IEEE conference on computer vision and pattern recognition*, 2017, pp. 1492–1500.

[8] J. Hu, L. Shen, and G. Sun, "Squeeze-and-excitation networks," in *Proceedings of the IEEE conference on computer vision and pattern recognition*, 2018, pp. 7132–7141.

[9] H. Zhao, J. Shi, X. Qi, X. Wang, and J. Jia, "Pyramid scene parsing network," in *Proceedings of the IEEE conference on computer vision and pattern recognition*, 2017, pp. 2881–2890.

[10] L.-C. Chen, Y. Zhu, G. Papandreou, F. Schroff, and H. Adam, "Encoder-decoder with atrous separable convolution for semantic image segmentation," in *Proceedings of the European Conference on Computer Vision (ECCV)*, 2018, pp. 801–818.

[11] B. Bischke, P. Helber, J. Folz, D. Borth, and A. Dengel, "Multi-task learning for segmentation of building footprints with deep neural networks," in *2019 IEEE International Conference on Image Processing (ICIP)*. IEEE, 2019, pp. 1480–1484.

[12] C. Sebastian, R. Imbriaco, E. Bondarev, and P. H. de With, "Adversarial loss for semantic segmentation of aerial imagery," *arXiv preprint arXiv:2001.04269*, 2020.

[13] A. Marcu and M. Leordeanu, "Dual local-global contextual pathways for recognition in aerial imagery," *arXiv preprint arXiv:1605.05462*, 2016.

[14] A. Marcu, D. Costea, E. Slusanschi, and M. Leordeanu, "A multi-stage multi-task neural network for aerial scene interpretation and geolocalization," *ArXiv*, vol. abs/1804.01322, 2018.

[15] L. Mou and X. X. Zhu, "Rifcn: Recurrent network in fully convolutional network for semantic segmentation of high resolution remote sensing images," *arXiv preprint arXiv:1805.02091*, 2018.

[16] T.-Y. Lin, P. Dollár, R. Girshick, K. He, B. Hariharan, and S. Belongie, "Feature pyramid networks for object detection," in *Proceedings of the IEEE conference on computer vision and pattern recognition*, 2017, pp. 2117–2125.

[17] A. Kirillov, R. Girshick, K. He, and P. Dollár, "Panoptic feature pyramid networks," in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, 2019, pp. 6399–6408.

[18] A. Vaswani, N. Shazeer, N. Parmar, J. Uszkoreit, L. Jones, A. N. Gomez, Ł. Kaiser, and I. Polosukhin, "Attention is all you need," in *Advances in neural information processing systems*, 2017, pp. 5998–6008.

[19] H. Nam, J.-W. Ha, and J. Kim, “Dual attention networks for multimodal reasoning and matching,” in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, 2017, pp. 299–307.

[20] V. Nair and G. E. Hinton, “Rectified linear units improve restricted boltzmann machines,” in *Proceedings of the 27th international conference on machine learning (ICML-10)*, 2010, pp. 807–814.

[21] S. Ioffe and C. Szegedy, “Batch normalization: Accelerating deep network training by reducing internal covariate shift,” *arXiv preprint arXiv:1502.03167*, 2015.

[22] A. Marcu, D. Costea, E. Slusanschi, and M. Leordeanu, “A multi-stage multi-task neural network for aerial scene interpretation and geolocalization,” *arXiv preprint arXiv:1804.01322*, 2018.

[23] A. Khalel and M. El-Saban, “Automatic pixelwise object labeling for aerial imagery using stacked u-nets,” *arXiv preprint arXiv:1803.04953*, 2018.

[24] B. Chatterjee and C. Poullis, “Semantic segmentation from remote sensor data and the exploitation of latent learning for classification of auxiliary tasks,” *arXiv preprint arXiv:1912.09216*, 2019.

[25] J. Deng, W. Dong, R. Socher, L.-J. Li, K. Li, and L. Fei-Fei, “Imagenet: A large-scale hierarchical image database,” in *2009 IEEE conference on computer vision and pattern recognition*. IEEE, 2009, pp. 248–255.

[26] D. P. Kingma and J. Ba, “Adam: A method for stochastic optimization,” *arXiv preprint arXiv:1412.6980*, 2014.

[27] A. Buslaev, A. Parinov, E. Khvedchenya, V. Iglovikov, and A. Kalinin, “Albumentations: fast and flexible image augmentations,” *arXiv preprint arXiv:1809.06839*, 2018.

Analyzing InfoGAN's potential for exploring demographic information in facial images

Zohra Rezgui
EEMCS-DMB,
University of Twente
z.rezgui@utwente.nl

Jasper Goseling
EEMCS-MOR,
University of Twente & CWI
j.goseling@utwente.nl

Raymond Veldhuis
EEMCS-DMB,
University of Twente
r.n.j.veldhuis@utwente.nl

Abstract

Recent face datasets tend to contain images encompassing a large variety of contexts. It may thus, be difficult to anticipate all the possible classification models that can be built from them. Investigating this aspect is interesting from a privacy point of view. We attempt this task by carrying out experiments on the CelebA dataset from the InfoGAN [1]. We show the factors of variations generated in the latent space using a prior code of 10 discrete uniform variables with 10 possible values. We report the behaviour of the auxiliary network, the generator network and the discriminator network separately during training. We show the vulnerability of InfoGAN to different prior codes in addition to those reported in the original paper. We discuss the instability challenges encountered that are caused both by the architecture proposed in the original paper and the impact of the number of prior variables and we propose a modified architecture for the discriminator and the generator based on DC-GAN [2] recommendations as a remedy for training instability. Finally, we argue qualitatively that some of the latent factors are not completely disentangled.

1 Introduction

With the recent development in deep neural networks, a lot of interest has been directed to the inference of soft biometric information from facial images, such as gender, age, ethnicity, profession or even health condition in both constrained and unconstrained settings[3],[4],[5],[6]. This raises consent-related privacy issues, especially in Europe after the GDPR regulations for data protection and privacy recently became operational. In order to study the scope of potential privacy preserving measures, it would be beneficial to explore the information that is captured in facial images. This calls for unsupervised methods of disentanglement as we are interested in a broad range of possible classification tasks that can be performed on a facial dataset. To do that, we study latent space disentanglement using InfoGAN [1]. InfoGAN is an unsupervised generative model that aims to discover meaningful representations of the data and eventually, to generate images by controlling these representations. In the context of protecting facial soft biometric attributes from being inferred with an attribute classifier, InfoGAN can be useful to discover possible classification tasks that can be performed on an image dataset with minor assumptions on what kind of attribute an attacker is interested in inferring. In the original paper, InfoGAN is applied using combinations of continuous and discrete prior codes, on several datasets including CelebA[7]. For the latter, only categorical prior variables are approximated by the latent representation. In this work, we evaluate the results of InfoGAN on the CelebA dataset for the task of exploring the potential classification tasks that can be performed on facial images.

2 Related work

Extensive work in learning meaningful disentangled representations has been done the last few years. Most of it focuses on a class of supervised methods to achieve disentanglement with regards to a wanted factor of variation, namely pose and illumination [8]. In order to disentangle identity from pose, [8] uses a multi-task discriminator that performs pose and identity classification as well as the original GAN's discriminator task; discriminating between real and fake samples. Some works have also addressed facial characteristics such as identity [9], age and gender [10]. Authors in [10] use two inputs of images of opposite facial attributes as a way to learn the discriminating features of the wanted attribute in the latent space and thus, implicitly use labels. In [9], the proposed method achieves disentanglement of identity by minimizing a loss measure between a pre-trained face recognition features of a source image and the features of its corresponding generated image while attribute encodings have to be learned. However, a supervised disentanglement is not relevant to our use case in this paper as it is more restrictive on the latent space than unsupervised methods and it therefore, would not allow us to explore more of the variety of the inherent factors of variation for a given dataset. On the other hand, a number of works perform disentanglement in an unsupervised way, mostly based on Variational Auto-Encoders (VAEs), such as β-VAE [11] and Factor-VAE [12]. For a variational autoencoder, the objective function is an addition of a reconstruction term and a regularization term in the form of a KL divergence between the latent distribution and a prior distribution. The second term is responsible for disentanglement in most of the disentanglement methods based on VAEs. The authors in [11] emphasize disentanglement by adding a magnitude factor β to the regularization term of the VAE's objective function. More recently, Factor-VAE [12] is derived from the Evidence Lower Bound decomposition known as ELBO surgery [13]. It shows the weakness of the β-VAE in penalizing the mutual information between the source distribution and the latent distribution which deteriorates reconstruction and overcomes it by only penalizing the total correlation of the latent distribution. We chose the InfoGAN approach for the option it offers in choosing to approximate either discrete or continuous latent distributions.

3 Preliminaries

3.1 Generative Adversarial Networks

Generative adversarial networks (GANs) were introduced in [14] and have been since, an active area of research as well as a revolutionary tool to learn complex distributions particularly for image datasets, GANs can produce realistic looking samples. They are generative models with a game-theoretic approach: Two networks, a generator and a discriminator, compete against each other in a min-max game. The generator produces fake samples similar to real samples from a given dataset and the discriminator evaluates them as real or fake. In theory, the game would converge to a Nash equilibrium. This framework is described by the following optimization problem with $V(D,G)$ being the GAN objective function:

$$\min_G \max_D V(D,G) = E_{X\sim P(data)}[\log D(X)] + E_{Z\sim p_z(Z)}[\log(1-D(G(Z)))]. \tag{1}$$

The real samples are represented by x while z corresponds to the random input of the generator. G stands for the function mapping the generator's noise input z to a fake image while D represents the function mapping an image input to its probability of being a real sample.

3.2 InfoGAN loss function

The authors of InfoGAN add a new term to the original GAN loss function in order to learn meaningful semantic features from the data. They add a coded part c to the noise input z of the generator that approximates meaningful latent features. As an example, digit type, rotation and width of the digits were recovered in the latent space for the MNIST dataset. To ensure that the generated output takes the coding into account, they propose maximizing the mutual information between the prior code and the generator output $I(c, G(Z,c))$; the new objective function is :

$$V^I(D,G) = V(D,G) - \lambda I(C, G(Z,C)). \tag{2}$$

Since mutual information is not tractable in this setting, they maximize a lower bound of it $L_1(G,Q)$ instead (3) where Q is the latent distribution approximating the prior P using the Variational Lower Bound method as following:.

$$L_1(G,Q) = E_{C\sim P(C),X'\sim G(Z,C)}[\log Q(C|X')] + H(C) \leq I(C, G(Z,C)). \tag{3}$$

The entropy of the latent codes $H(C)$ in (3) is treated as a constant, similarly as in the paper. Hence, maximizing the mutual information lower bound implies minimizing:

$$-E_{C\sim P(C),X'\sim G(Z,C)}[\log Q(C|X')], \tag{4}$$

which can be seen as the cross-entropy between the prior distribution P and the distribution approximated by the auxiliary network $Q(C|X')$. Since the CelebA experiment only requires categorical variables, the loss for the auxiliary network can be the categorical cross entropy. We show the different components of the InfoGAN's framework in Figure 1.

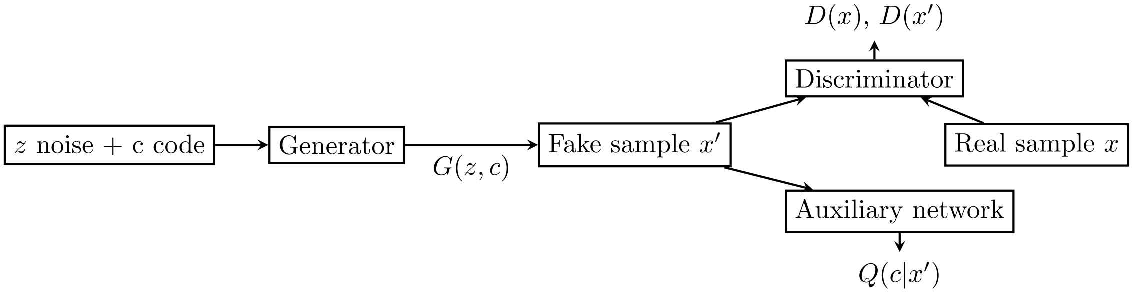


Figure 1: Framework of InfoGAN's components

3.3 Description of the CelebA dataset

CelebA [7] is a large dataset consisting of more than 200.000 images that was designed for multiple face attribute predictions. It contains 40 attribute annotations describing fine scale characteristics of the facial images, such as details about hair style and face shape as well as the usual demographic annotations such as gender and ethnicity. In [7], impressive results were reported on the ability to predict some of these attributes; this makes the dataset a good candidate for exploring and recovering the different factors of variation in an unsupervised manner. We use the entire training set without the labels to train InfoGAN.

4 Results

4.1 Original InfoGAN on CelebA

In the original paper, the authors provide different architectures for every experiment on each dataset. The architectures they used for the experiment on the CelebA dataset are reported in Figures 2 and 4. They consist of successive convolutional layers with fully connected layers at the beginning for the generator and towards the end of the network for the discriminator. The experiment failed using this architecture as shown in Figure 6. We suspected that the presence of the fully connected layers contribute to the instability of training and we decided to remove them following DC-GAN recommendations to stabilize GAN training [2].

4.2 Modified InfoGAN

We modified the architectures for the generator and the discriminator from the model based on the DC-GAN recommendations [2] as we were not able to reproduce the results from the original InfoGAN paper using their given architectures. Figures 3 and 5 show the architectures for the generator and the discriminator respectively.

Similarly as in the original paper, the prior control noise is composed of 10 discrete uniform variables, each with 10 possible values. This noise is then concatenated with the GAN's incompressible noise vector of size 128 sampled from a standard Gaussian distribution. This results in a 228-long noise vector. The λ term in (2) is fixed to 1 as mentioned in the paper for the case of a discrete code.

The loss for the discriminator is the binary cross entropy between the ground truth labels describing whether the images are sampled from the CelebA dataset or generated by the generator and the predictions of the discriminator for both real and fake images. The loss for the generator is composed of the binary cross entropy between the "real" labels and the predictions of the discriminator on the generated images, as well as the loss of the auxiliary network Q. The hyper-parameters are summarized in Table 1.

4.3 Discussion

Instability: We found the training of InfoGAN particularly unstable as following the instructions of the original paper for the CelebA experiment did not produce the expected outcome and resulted in a convergence failure as demonstrated in Figure 6. After changing the parameters as reported in [1], we were able to avoid the convergence failure for the experiment reported in the paper using prior code of 10 discrete uniform variables with 10 possible values as shown in Figure 7. On the other hand, the experiment failed for 10 discrete uniform variables with fewer possible values as demonstrated in Figures 8, 9 and 10. We also found that our generated images are of an inferior quality to the images reported in the original paper. One possible reason for this is that we do not crop the facial images strictly around the face; we use aligned facial images that include the hair.

Interpretation derived from generated images is unclear and prone to bias: We observe in Figure 11, that the most noticeably disentangled latent factor corresponds to the sixth row, where we have some confidence that it is consistently skin color varying considerably from one image to the next. In the next row, we notice the hair style and color changing in certain categories. The factor of variation illustrated in the last row seems to encode pose however, it looks entangled with other factors. We also note that it is quite difficult to objectively assign meaning to the latent features that are learned. We can notice dissimilarities between the different categories in Figure 11 but cannot guess the meaning of each category. We hold the same remark for the images reported in the original paper. We believe a possible way to assign meaning would be to establish a perceptual study by different parties where they would indicate the difference that they see between categories. This could still be biased with regards to the real variation factor described by the categories but it would reveal an agreement over a part of the information that can be the subject of a classification task.

Dependence on priors: The images obtained with InfoGAN describe the latent distribution that is forced to approximate a prior distribution chosen arbitrarily. This is quite restrictive as there could be a meaningful disentangled representation of which we do not know the prior distribution. Figures 8 to 9 also show that the choice of priors has a substantial impact on the convergence outcome of the model.

Table 1: Comparison of hyper-parameters for CelebA experiment

Hyper-parameters	**InfoGAN paper**	**Ours**
image size	(32, 32, 3)	(32, 32, 3)
batch size	not specified	64
number of epochs	not specified	100
learning rate G	0.001	0.0002
learning rate D	0.0002	0.0002
λ	1	1

5 Conclusion

In this paper, we evaluated InfoGAN [1] as a tool of investigating potential demographic attributes present in facial images that would lead to unanticipated classification tasks. We show the shortcomings of the original paper in terms of instability of training, dependence on the prior code, as well as the qualitative interpretation of disentanglement. We also propose modifications to the originally proposed architecture based on DC-GANs [2]. This improved convergence for the prior code that was used in [1]. However, the instability remained when we used a different prior code. Furthermore, we observed that even though some demographic attributes such as skin color, hair style/color and pose seem to be encoded as factors of variation, most of the factors do not seem to be fully disentangled and remain ambiguous. Therefore, we conclude that InfoGAN in its present form, is not suitable for

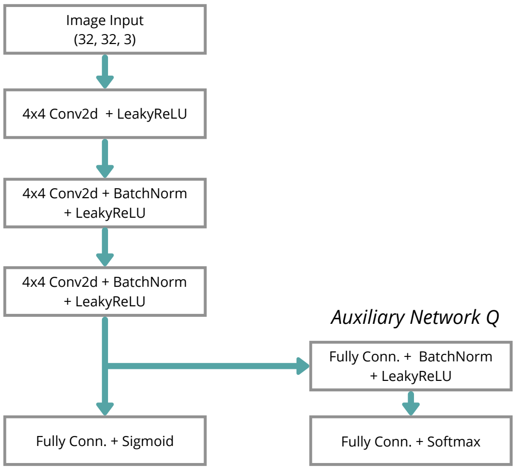


Figure 2: Original InfoGAN's discriminator architecture.

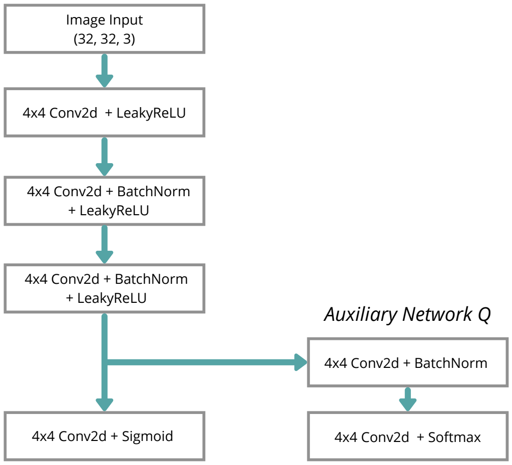


Figure 3: Proposed InfoGAN's discriminator architecture.

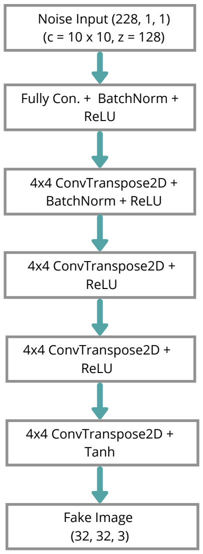


Figure 4: Original InfoGAN's generator architecture.

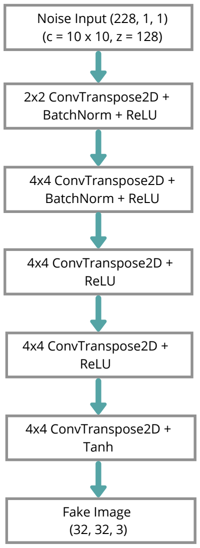


Figure 5: Proposed InfoGAN's generator architecture.

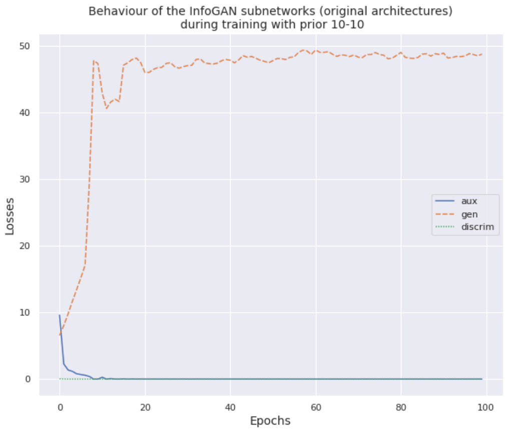


Figure 6: Average loss values per epoch for each sub-network using the original paper's architectures of the model during training when the injected prior codes are 10 discrete uniform variables with 10 possible values each.

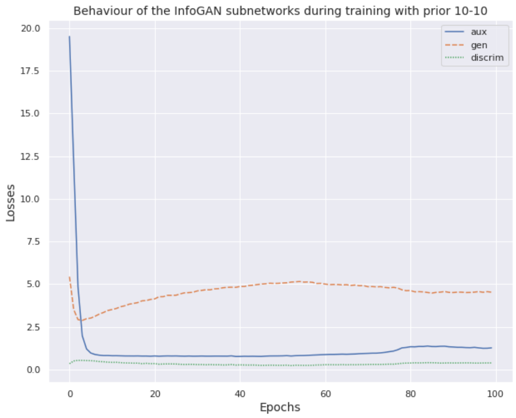


Figure 7: Average loss values per epoch for each sub-network of the model during training when the injected prior codes are 10 discrete uniform variables with 10 possible values each. The model stabilizes around epoch 80.

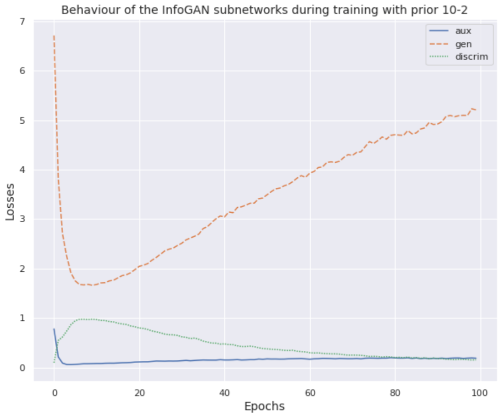


Figure 8: Average loss values per epoch for each sub-network of the model during training when the injected prior codes are 10 discrete uniform variables with 2 possible values each.

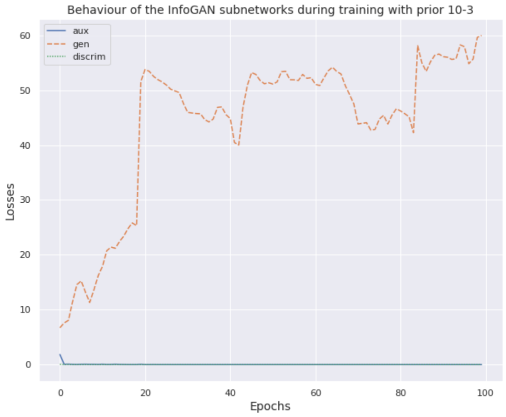


Figure 9: Average loss values per epoch for each sub-network of the model during training when the injected prior codes are 10 discrete uniform variables with 3 possible values each.

reliably exploring the possible classification tasks for demographic attributes. Part of our future work is to develop more suitable techniques.

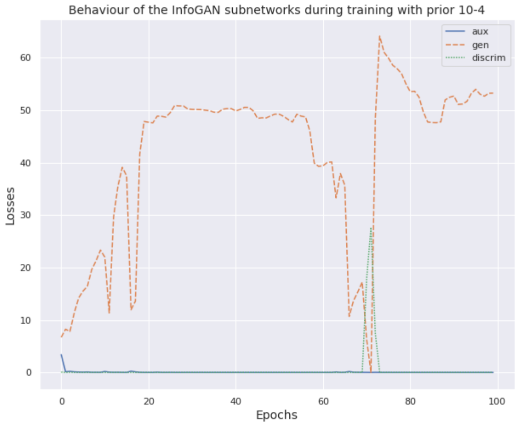


Figure 10: Average loss values per epoch for each sub-network of the model during training when the injected prior codes are 10 discrete uniform variables with 4 possible values each.

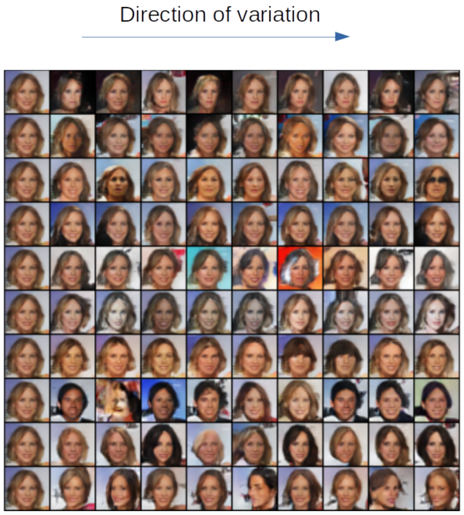


Figure 11: Generated images according to the prior variables. For each row, we explore the categories of one of the variables and fix all the other variables to their first categories.

6 Acknowledgement

This paper is supported by European Union's Horizon 2020 research and innovation programme under grant agreement No 860315, project PriMa (Privacy Matters).

References

[1] X. Chen, Y. Duan, R. Houthooft, J. Schulman, I. Sutskever, and P. Abbeel, "InfoGAN: interpretable representation learning by information maximizing generative adversarial nets," in *NIPS*, vol. 29, 2016.

[2] A. Radford, L. Metz, and S. Chintala, "Unsupervised representation learning with deep convolutional generative adversarial networks," *ICLR 2016*.

[3] G. Levi and T. Hassner, "Age and gender classification using convolutional neural networks," in *Proceedings of the IEEE conference on computer vision and pattern recognition workshops*, 2015, pp. 34–42.

[4] M. Afifi and A. Abdelhamed, "AFIF4: deep gender classification based on adaboost-based fusion of isolated facial features and foggy faces," *Journal of Visual Communication and Image Representation*, vol. 62, pp. 77–86, 2019.

[5] E. Eidinger, R. Enbar, and T. Hassner, "Age and gender estimation of unfiltered faces," *IEEE Transactions on Information Forensics and Security*, vol. 9, no. 12, pp. 2170–2179, 2014.

[6] J. I. Stoker, H. Garretsen, and L. J. Spreeuwers, "The facial appearance of CEOs: Faces signal selection but not performance," *PloS one*, vol. 11, no. 7, pp. 1–11, July 2016.

[7] Z. Liu, P. Luo, X. Wang, and X. Tang, "Deep learning face attributes in the wild," in *Proceedings of International Conference on Computer Vision (ICCV)*, December 2015.

[8] L. Tran, X. Yin, and X. Liu, "Disentangled representation learning GAN for pose-invariant face recognition," in *Proceedings of the IEEE conference on computer vision and pattern recognition*, 2017, pp. 1415–1424.

[9] Y. Nitzan, A. Bermano, Y. Li, and D. Cohen-Or, "Face identity disentanglement via latent space mapping," *ACM Transactions on Graphics (TOG)*, vol. 39, no. 6, pp. 1–14, 2020.

[10] T. Xiao, J. Hong, and J. Ma, "ELEGANT: Exchanging latent encodings with GAN for transferring multiple face attributes," in *Proceedings of the European Conference on Computer Vision (ECCV)*, September 2018, pp. 172–187.

[11] I. Higgins, L. Matthey, A. Pal, C. Burgess, X. Glorot, M. Botvinick, S. Mohamed, and A. Lerchner, "Beta-VAE: Learning basic visual concepts with a constrained variational framework," in *ICLR 2017*.

[12] H. Kim and A. Mnih, "Disentangling by factorising," in *International Conference on Machine Learning*. PMLR, 2018, pp. 2649–2658.

[13] M. D. Hoffman and M. J. Johnson, "ELBO surgery: yet another way to carve up the variational evidence lower bound," in *Workshop in Advances in Approximate Bayesian Inference, NIPS*, vol. 1, 2016, p. 2.

[14] I. J. Goodfellow, J. Pouget-Abadie, M. Mirza, B. Xu, D. Warde-Farley, S. Ozair, A. Courville, and Y. Bengio, "Generative adversarial networks," *Commun. ACM*, 2014.

Can Fake News Detection be Accountable? The Adversarial Examples Challenge

(Extended Abstract)

Jérémie Bogaert, Quentin Carbonelle,
Antonin Descampe, François-Xavier Standaert

UCLouvain, Louvain-la-Neuve, Belgium

Abstract

Automated fake news detection is an important challenge in view of the increasing ability of statistical language models to generate large amounts of (possibly fake) articles, so that recognizing them manually becomes unrealistic. Yet, the reliable deployment of such automated detection tools would require ensuring that they are accountable. Algorithmic accountability is known to be difficult to reach, especially when adversarial behaviors aim to make algorithms deviate from their expected mode of operation. In this paper, we illustrate with a case study that this challenge is further amplified in contexts where the labeling of the articles is prone to errors, which is the case of fake news detection.

1 Introduction

The proliferation of fake news on social media platforms has created an increasing need of solutions to detect them. While the generation of fake news and the necessary fact checking that it implies started as a mostly manual process (e.g., discussed in [17]), recent advances in natural language generation with machine learning algorithms have amplified the risk of so-called neural fake news [19]. The massive amount of articles that such tools can generate makes manual fact checking unrealistic and raises the question of their automated detection. As recently surveyed in [1, 20], solutions combining natural language processing and machine learning are attractive for this purpose, due to their ability to capture various features of newspaper articles. Besides, the sensitive nature of the fake news detection problem also requires that its deployment comes with guarantees of algorithmic accountability [3]. But as discussed in [2], such guarantees are hard to obtain in contexts where adversarial behaviors can target the robustness of machine learning classifiers, which is the case of fake news detection.

In this paper, we therefore study the robustness of fake news detection against adversarial examples [4, 8]. For this purpose, we first build a database of fake and reliable news. As already observed in the literature, this step is inherently challenging since there is no strict definition of what a fake news is [6]. We deal with this difficulty by combining a public dataset of fake news (`https://www.kaggle.com/mrisdal/fake-news`), which were scraped from blacklisted websites over a period of 30 days around the 2016 US election, and reliable news collected from the New York Times and the Guardian over approximately the same period. We additionally explore both data sets to show that they cover similar topics. While inevitably imperfect, this database is then used to show that standard machine learning classifiers can detect fake news with a good accuracy, but are also easy to fool with adversarial examples. Interestingly, it appears that the difficulty to define what a fake news is (possibly combined with label errors in the training sets) gives adversaries additional opportunities to generate fake news classified as reliable. So our results question again the possibility to rely on accountable algorithms for sensitive tasks that can be targeted by adversarial behaviors and suggest different research avenues related to fake news in general.

We note that the topic of fake news is widely multidisciplinary [7] and even the more technical topic of automated fake news detection is already covered by a broad literature (e.g., the already mentioned [1, 19, 20] but also [5, 11, 13, 12] to name a few). Our contribution is not to improve automated fake news detection algorithms but to illustrate the difficult interplay between such algorithms and the need of algorithmic accountability when adversarial behaviors are considered. As a natural starting point in this direction, we show that there exist (for now simple) examples of fake news detectors that are weak against such adversarial behaviors, raising the question of how to prevent them, with technical and non-technical means. In other words, we put forward an under-discussed risk for the reliable deployment of such systems.

The rest of the paper is structured as follow. We start by describing the database we used for our investigations and discussing its unavoidable limitations in Section 2. We follow by showing simple examples of supervised fake news detection tools that can detect fake news with reasonable accuracy in Section 3. We finally exhibit how to craft adversarial examples against these classifiers in Section 4. We conclude by analyzing the impact of these findings and tracks for further research in Section 5.

2 Building a fake news database

Research on fake news detection has often been limited by the quality of existing datasets and their specific application contexts [12]. As already mentioned, this is mostly due to the difficulty of precisely defining what a fake news is, making their labeling for supervised learning challenging [6]. Besides, our goal to investigate adversarial examples is typically calling for "not too short" texts, excluding popular databases such as [18]. To the best of our knowledge, the database that comes the closest to our needs is the fake news corpus (`https://github.com/several27/FakeNewsCorpus`). However, preliminary investigations suggested quite significant dissimilarities between the topics of reliable and fake articles (namely, football for fake articles and politics for reliable ones). This similarity makes it unsuitable for our purposes, since it implies risks to classify the articles based on their topic more than their fake/reliable nature. We therefore prepared our investigations by attempting to build a better database.

Concretely, we started from a public dataset of 3500 fake news from Kaggle (`https://www.kaggle.com/mrisdal/fake-news`), which were scraped from 207 blacklisted websites over a period of 30 days around the 2016 US election. Some preliminary processing was applied to remove unwanted features of the articles (e.g., meta-data that is not in the original articles and was added by the blacklisted websites). No website contributed to more than 1% of the database to ensure that imperfect scraping, preprocessing or labeling for some websites cannot significantly impact the overall results. We then tried to build a complementary database of reliable news, covering the same period of time. For this purpose, we used the API developed by the New York Times and The Guardian, and scraped papers about World news and US news for the intended period. We had to slightly increase the window of time (Oct. 16 to Dec. 4 for the reliable news vs. Oct. 26 to Nov. 25 for the fake news), in order to collect 3500 articles (1500 from the New York Times, 2000 from The Guardian).

We performed a preliminary exploration of the topics covered by this database using the Term Frequency–Inverse Document Frequency (TF-IDF) statistic. It is a standard tool for information retrieval or summarization, which indicates the relevance of the words in some documents [10]. We first launched it on the full database to identify the most relevant words overall. Next, we launched it on the fake news corpus, leading to the results of Figure 1 (where the X axis lists the 40 most relevant words of the full database). It confirms the coverage of the US elections (e.g., with Trump and Clinton in the first places), but also highlights the weight of some neutral words like ‘said’.

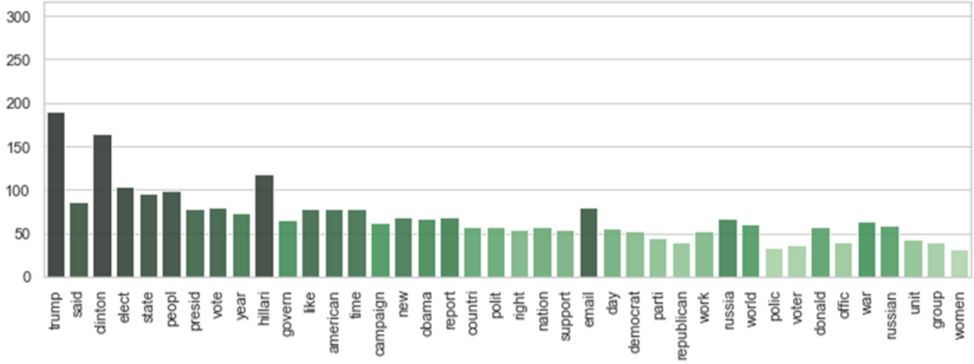


Figure 1: Evaluation of the fake news database’s topics with TF-IDF score.

We finally computed the same TF-IDF scores for the reliable news, which are given in Figure 2 and confirm the coverage of similar topics. They also highlight some specific features of the fake news corpus already: for example the more frequent use of the (misspelled) first name Hilari or the importance of the word e-mail (relating to an ongoing affair of Mrs. Clinton’s e-mail leaks during the 2016 elections).

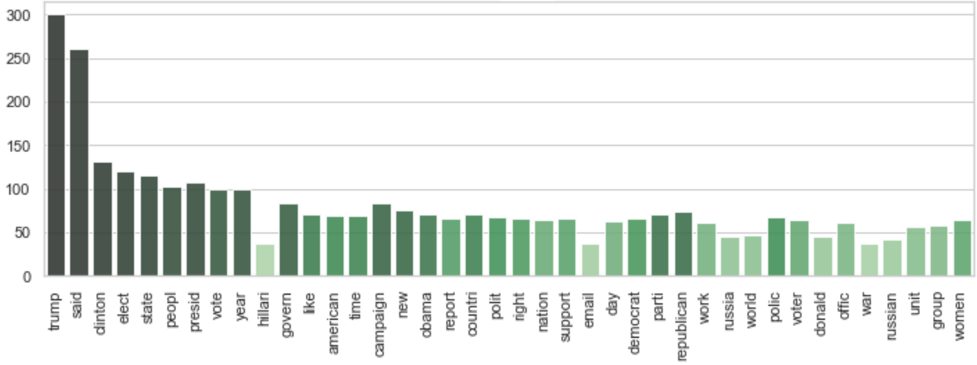


Figure 2: Evaluation of the reliable news database’s topics with TF-IDF score.

In order to confirm that the resulting (7000-paper) database did not contain obvious parasitic patterns, we additionally visualized the data by feeding the t-distributed Stochastic Neighbor Embedding (t-SNE) tool of [16] with the vectors output by the TF-IDF transform.* t-SNE projects each high-dimensional object towards a two-dimensional point in such a way that similar objects are modeled by nearby points and dissimilar objects are modeled by distant points with high probability. As illustrated in Figure 3, the distribution of the topics is similar for fake and reliable articles while, for example, the separation between World and US (reliable) news can be distinguished in the right part of the figure. It suggests that the topics do not create obvious (parasitic) ways to discriminate fake and reliable news. So while the evaluations in this section do admittedly not provide any formal guarantee that no such patterns exist, we assume in the following that this database is good enough for our purposes.

* Reduced to 500 dimensions thanks to truncated Singular Value Decomposition (SVD). We also tried larger number of dimensions but it did not change our main observations.

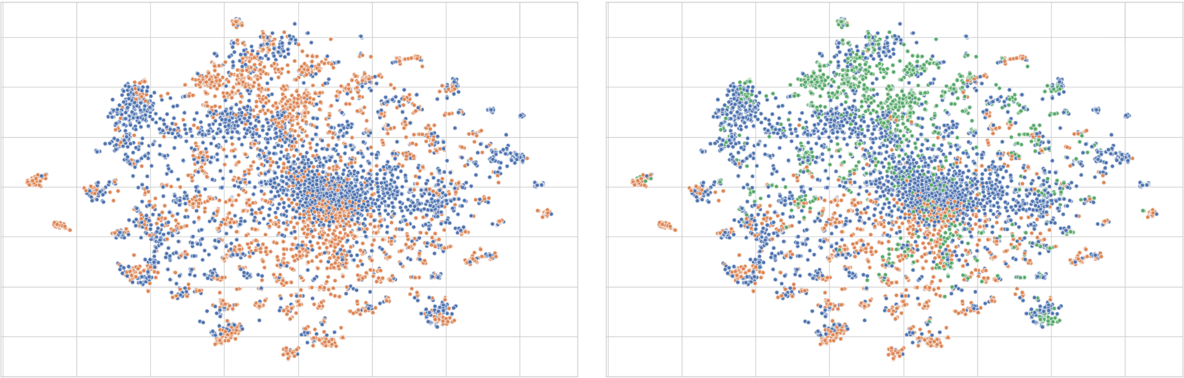


Figure 3: Visualization of the news' topics with t-SNE. Left: fake news (●) and reliable news (●). Right: fake news (●), reliable World news (●) and reliable US news (●).

3 Exemplary classifiers

The next step of our investigations was to build statistical classifiers for fake and reliable news, thanks to supervised machine learning. We considered various options for this purpose: logistic regression, naive Bayes, random forests and Long Short-Term Memory (LSTM). For place constraints, we focus our following descriptions on logistic regression and LSTM (the other classifiers did not significantly affect our main conclusions).

Concretely, all our classifiers were built starting with the same preprocessing: we removed punctuation, non-alphanumeric characters, multiple whites spaces, websites, stop words and short words). For the logistic regression, we then vectorized the articles as bag of words using TF-IDF while for the LSTM we used a Word2Vec model trained on our full dataset [9]. As Google's Word2Vec (`https://code.google.com/archive/p/word2vec/`), we fixed the embedding to 300-dimensional vectors. The rationale behind this choice is that the LSTM can take advantage of the words' order. The model hyperparameters were then tuned using a 5-fold cross-validation.

The accuracy of these classifiers is illustrated in Figure 4. It is significantly better than a random guess in both cases, and reaches > 90% for the logistic regression (we expect that the LSTM would reach this accuracy or even improve it with more training samples). Despite further optimizations are certainly possible, we assume these values to be sufficient for trying to decrease them with adversarial examples.

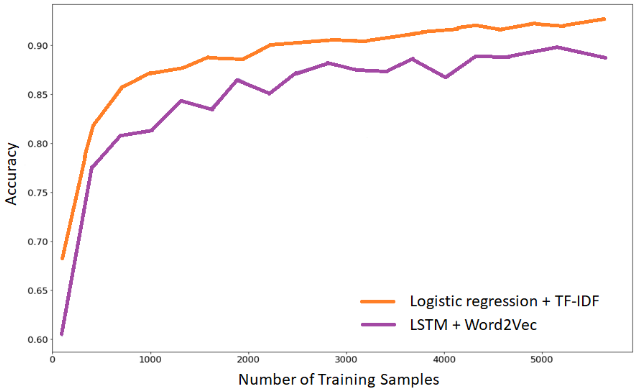


Figure 4: Learning curve of exemplary fake news detectors.

4 Crafting adversarial examples

An adversarial example is an input, unknown to the target machine learning algorithm, that makes it deviate from its public specifications and is purposely generated by an adversary having an interest in such a deviated behavior [4]. In the context of fake news detection, the goal of the adversary is to force the misclassification of a fake news as reliable, despite not changing its semantic content from the human perception viewpoint. The main question we tackle next is whether such adversarial examples can be crafted for the detectors of the previous section. In this context, the adversarial goal could vary from producing such adversarial examples at a low rate, possibly taking advantage of some manual processing, or to produce them at a high rate, automatically. As a first step to show the existence of a risk, we consider the easiest (low rate with manual intervention) context. We briefly discuss the relaxation of this context at the end of the section. The threat model could also vary from white box access to the models (i.e., knowing their parameters) to only black box access (i.e., only being able to observe input/output pairs). We discuss both options in the following.

4.1 Methodology

The high-level approach we used to craft adversarial examples at low rate is in 3 steps:

1. Identify Reliable Hot Words (RHW), which are the words having high probability to push the classification of any article towards the reliable class.

2. Identify the Target's Fake Hot Words (TFHW), which are the words having high probability to push the classification of a target article towards the fake class.

3. Human intervention to replace TFHW by RHW in a semantic-preserving manner.

The identification of the RHW and TFHW was performed using high-level ideas similar to [8, 2]. In the white box setting, we applied the Fast Gradient Sign Method (FGSM) which essentially boils down to maximizing the classifier's loss function, then using the gradient information to add or remove words independent of their position.[†] This gradient can be computed analytically for the logistic regression, and we used the numerical estimation provided by TensorFlow (`https://www.tensorflow.org/`) for the LSTM. In the black box setting, we used a more straightforward approach where we just removed words one by one and feed the classifier with the modified articles in order to identify words that impact the detection probability the most.

4.2 Experimental results

We will discuss the type of results we obtain with Example 1, which is initally detected as a fake news with 90% probability by the logistic regression, and 65% probability by the LSTM. As aforementioned, the first step in trying to fool the detection is to identify RHW and TFHW. RHW are identified on the reliable news corpus while TFHW are identified on the target article only. We illustrate this step with our black box approach applied to the beginning of the example (namely, the words *In what is being described as another 'bizarre'*), where the short words 'In', 'what', 'is', 'being' and 'as' do not impact the classification, while removing the words 'described' and 'bizarre' respectively decrease its probability of being detected as a fake news by 2.8% and 3.5%. By extending such an approach to the whole article (and the whole corpus of reliable news for RHW), we can then build lists of TFHW and RHW.

[†] This position could be exploited in the case of the LSTM, which we leave as a scope for further investigations (a direct exploitation would result in a quite computationally intensive strategy).

> In what is being described as another ‘bizarre’ attempt to sabotage her own campaign, Hillary Clinton has desecrated a series of beloved US symbols, including punching a bison, setting fire to the Stars & Stripes and spitting at Jerry Seinfield. The Presidential hopeful seems determined to make a series of unprovoked errors, not least of which was agreeing to Bill hosting a sleepover for a group of Girl Guides. Short of dressing the Statue of Liberty in a Burka, Mrs Clinton has lurched from one PR blunder to another. Commented one journalist: ‘The Presidential race is entering the final furlong and if Mrs Clinton was horse – and before you can say Benghazi – she’s gone from bookie’s favourite to an ingredient at the local glue factory’. Having already become the unwitting focus of various health scares and FBI investigations, Mrs Clinton’s campaign is as orderly as a Marx Brothers movie. Her lead in the polls has been cut as video emerges of her lighting a cigar with a rolled up Bill of Rights, then proceeding to take a dump on the White House lawn. Hillary’s erratic behaviour has seen her sing the Star-Spangled Banner in Korean, dress as Oprah Winfrey for Halloween and pebble-dash Mount Rushmore. Remarked a flummoxed advisor: ‘She keeps doing the unthinkable – like making Donald Trump electable’. Share this story...

Example 1: Sample of the fake news database.

The main high-level observations that can be extracted from these lists are:

1. That they contain both semantically tainted words (e.g., ‘Hilary’, ‘FBI’, ‘Donald’) and semantically neutral words (e.g., ‘said’, ‘commented’, ‘campaign’).
2. That there is a higher proportion of semantically tainted words for TFHW.
3. That the lists of (ordered) words identified as RHW or TFHW for the logistic regression and the LSTM, significantly overlap but are not identical.

We can then build adversarial examples, e.g., for the (simpler) LSTM:

> [...] ‘bizarre’ attempt to sabotage her own campaign, ~~Hillary~~ **Mrs** Clinton has desecrated
> [...] Mrs Clinton’s campaign is as ~~orderly~~ **neat** as a Marx Brothers movie [...]

By changing only two words (which do not affect the meaning of the article), it is now classified as a fake with only 45% probability (so as reliable with 55% probability). Interestingly, exactly those changes are not sufficient to misclassify the article with the logistic regression (which still classifies it as a fake, but with a probability reduced from 90% to 55%). Yet, another change of the second word does the job (suggesting that as usual with adversarial examples, they have a certain level of transferability):

> [...] ‘bizarre’ attempt to sabotage her own campaign, ~~Hillary~~ **Mrs** Clinton has desecrated
> [...] Remarked a flummoxed ~~advisor~~ **minister**: ‘She keeps doing the unthinkable [...]

Since based on manual interventions, we did not craft such examples for large number of articles. We repeated the process for 5 fake news and succeeded to force a misclassification by changing a maximum of 15 words. The LSTM classifier was usually fooled with slightly less words which we assume is due to its initially lower accuracy.

Overall, and despite drawing general conclusions based on such small-scale examples is hard, one important observation is that in contrast with the case study of [2] where adversarial examples had to be mostly based on neutral words (to avoid being easy to spot by the human perception), fooling a fake news detection algorithm can be done with more tainted words (e.g., by changing ‘Hillary’ into ‘Mrs’ in our example). In other words, the fact that the fake or reliable nature of an article does not have a definition as clear as the topics used in the case study [2] makes the adversary’s task easier.

4.3 Towards automation

Our handmade experiments naturally raise the question whether adversarial examples can be automated. As a first step in this direction (i.e., to evaluate the sensitivity of our fake news detectors to semantically-neutral modifications), we evaluated a greedy strategy where we just substituted words by synonyms (sometimes causing syntax problems). One example obtained for the logistic regression is given below:

> [...] symbols, including punching a ~~bison~~ **buffalo** [...] [...] a group of ~~Girls~~ **Woman** guides [...] various health scares and FBI ~~investigations~~ **inquiry**, Mrs Clinton's campaign [...]

Out of 100 test articles, we could misclassify 22% (resp., 32%) with the LSTM (resp., logistic regression) classifier (presumably because our greedy strategy is better suited to models that do not exploit the words' order). sing advanced statistical language models should lead to more adversarial opportunities, which we leave as an open problem.

5 Conclusions and open problems

Advances in digital media are responsible for journalists to loose the monopoly on information production and dissemination, and raise new legitimacy concerns (e.g., regarding whether journalism can still offer quality and reliable news in the digital era) [15]. Algorithmic accountability is one of the emerging (and difficult to reach) goals aiming to mitigate such concerns [3]. In this work, we confirm that ensuring algorithmic accountability with robustness against adversarial behaviors is especially challenging in contexts such as fake news detection where the definition of optimization criteria is inherently fuzzy due to the lack of well defined classes. Our results are preliminary in many respects and therefore suggest various directions for further investigations. First, the difficult collection of a fake news / reliable news database would be improved by designing a tool able to automatically generate large amounts of fake news from reliable ones (e.g., by biasing them in a given direction). It would allow a better analysis of the syntactic or semantic patterns that fake news detection can exploit. Second, and as already mentioned, the generation of adversarial examples would benefit from automation in order to enable larger-scale experiments. Third, the evaluation of countermeasures (i.e., fake news detection tools able to cope with adversarial examples) is an important long-term goal as well. Early discussions in [4, 2] suggest that a purely technical solution may not be possible. The perspective of such negative conclusions therefore questions the need of complementary approaches, e.g., relying on the trust in the journalists who write stories more than on content, as suggested in [14].

References

[1] Nadia K. Conroy, Victoria L. Rubin, and Yimin Chen. Automatic deception detection: Methods for finding fake news. Proceedings of the Association for Information Science and Technology, 52(1):1–4, 2015.

[2] Antonin Descampe, Clément Massart, Simon Poelman, François-Xavier Standaert, and Olivier Standaert. Automated news recommendation in front of adversarial examples and the technical limits of transparency in algorithmic accountability. AI & Society, 2021.

[3] Nicholas Diakopoulos. Algorithmic accountability. Digital Journalism, 3(3):398–415, 2015.

[4] Ian J. Goodfellow, Patrick D. McDaniel, and Nicolas Papernot. Making machine learning robust against adversarial inputs. Commun. ACM, 61(7):56–66, 2018.

[5] Naeemul Hassan, Fatma Arslan, Chengkai Li, and Mark Tremayne. Toward automated fact-checking: Detecting check-worthy factual claims by claimbuster. In KDD, pages 1803–1812. ACM, 2017.

[6] Edson C. Tandoc Jr., Zheng Wei Lim, and Richard Ling. Defining “fake news”. Digital Journalism, 6(2):137–153, 2018.

[7] David M. J. Lazer, Matthew A. Baum, Yochai Benkler, Adam J. Berinsky, Kelly M. Greenhill, Filippo Menczer, Miriam J. Metzger, Brendan Nyhan, Gordon Pennycook, David Rothschild, Michael Schudson, Steven A. Sloman, Cass R. Sunstein, Emily A. Thorson, Duncan J. Watts, and Jonathan L. Zittrain. The science of fake news. Science, 359(6380):1094–1096, 2018.

[8] Bin Liang, Hongcheng Li, Miaoqiang Su, Pan Bian, Xirong Li, and Wenchang Shi. Deep text classification can be fooled. In IJCAI, pages 4208–4215. ijcai.org, 2018.

[9] Tomás Mikolov, Kai Chen, Greg Corrado, and Jeffrey Dean. Efficient estimation of word representations in vector space. In ICLR (Workshop Poster), 2013.

[10] Juan Ramos. Using tf-idf to determine word relevance in document queries. In Proceedings of the First Instructional Conference on Machine Learning, pages 29–48, 2003.

[11] Natali Ruchansky, Sungyong Seo, and Yan Liu. CSI: A hybrid deep model for fake news detection. In CIKM, pages 797–806. ACM, 2017.

[12] Karishma Sharma, Feng Qian, He Jiang, Natali Ruchansky, Ming Zhang, and Yan Liu. Combating fake news: A survey on identification and mitigation techniques. ACM Trans. Intell. Syst. Technol., 10(3):21:1–21:42, 2019.

[13] Kai Shu, Amy Sliva, Suhang Wang, Jiliang Tang, and Huan Liu. Fake news detection on social media: A data mining perspective. SIGKDD Explor., 19(1):22–36, 2017.

[14] David Sterrett, Dan Malato, Jennifer Benz, Liz Kantor, Trevor Tompson, Tom Rosenstiel, Jeff Sonderman, and Kevin Loker. Who shared it?: Deciding what news to trust on social media. Digital Journalism, 7(6):783–801, 2019.

[15] Jingrong Tong. Journalistic legitimacy revisited. Digital Journalism, 6(2):256–273, 2018.

[16] Laurens van der Maaten and Geoffrey Hinton. Visualizing data using t-sne. Journal of Machine Learning Research, 9(86):2579–2605, 2008.

[17] Chris J. Vargo, Lei Guo, and Michelle A. Amazeen. The agenda-setting power of fake news: A big data analysis of the online media landscape from 2014 to 2016. New Media Soc., 20(5):2028–2049, 2018.

[18] William Yang Wang. "Liar, liar pants on fire": A new benchmark dataset for fake news detection. In ACL (2), pages 422–426. Association for Computational Linguistics, 2017.

[19] Rowan Zellers, Ari Holtzman, Hannah Rashkin, Yonatan Bisk, Ali Farhadi, Franziska Roesner, and Yejin Choi. Defending against neural fake news. In NeurIPS, pages 9051–9062, 2019.

[20] Xinyi Zhou and Reza Zafarani. A survey of fake news: Fundamental theories, detection methods, and opportunities. ACM Comput. Surv., 53(5):109:1–109:40, 2020.

Learning robust image representations with autoencoders

Shunxin Wang Raymond Veldhuis Nicola Strisciuglio
University of Twente
Dept. EEMCS, Group Data Management and Biometrics
Enschede, The Netherlands
s.wang-2@utwente.nl r.n.j.veldhuis@utwente.nl n.strisciuglio@utwente.nl

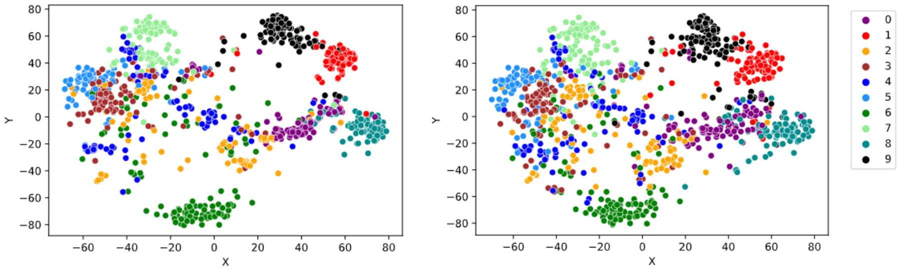
Abstract

Recently, the robustness of deep learning models has raised concerns. The performance of a trained model usually degrades when its inputs have data distributions different from the distributions of the training data. In the field of computer vision, data distribution shift occurs when training images contain different corruptions such as noise, blur, among others, or the images are captured under different conditions than those used for testing (e.g. when models are deployed in real-world scenarios).

We consider a classifier that reaches an accuracy of around 87% on CIFAR-10 test set; it can generate partially discriminative latent representations as shown in Fig. 1a. Some examples of the class 4 (deer) are mixed with class 7 (horse) while the rest of class 4 are located in two different clusters. Importantly, a discriminative latent space is not equal to an invariant latent space, meaning that adding small perturbations to an input can largely change its latent representation (Fig. 1b). It further reduces the robustness of the classifier and thus leads to wrong classification results. Based on this observation, we hypothesize that a well regularized latent space which is, to some extent, invariant to dataset shift or corruptions in the input, can benefit the robustness of a model. Hence, we are investigating the contributions of autoencoders to the learning of invariant latent space representations of corrupted images.

To learn a regularized latent space, we combined autoencoders with classifiers, aiming at learning image representations in a semi-supervised way. Multiple state-of-the-art architectures such as ResNet [1], DenseNet [3] and U-Net [5] are utilized and compared. Two autoencoders applying ResNet as encoder are used, which are called ResNet-Sym and ResNet-Skip. The ResNet-Sym has a partially-symmetric architecture with its encoder, while the ResNet-Skip applies skipping-connection from encoder to decoder, and its decoder has similar structure to the decoder of U-Net. In the experiments, we train the models on images without corruptions (CIFAR-10 [4]) and test on images with different corruptions (CIFAR-10-C [2]) as well as the clean CIFAR-10 test images.

We observe that a well-designed architecture is helpful in learning discriminative latent representations, such as ResNet, DenseNet. However, the robustness towards corruptions is not guaranteed. Different architectures perform differently to corruptions, but in general, the influence of corruptions such as Gaussian noise, impulse noise, speckle noise and shot noise is significant even if the noise level is low. Attractively, classifiers combined with autoencoders are slightly more robust to noise of high severity.



(a) Without corruption

(b) With low-severity Gaussian noise

Figure 1: Distributions of latent representations by a Regular encoder (0 to 9 indicate the class labels)

References

[1] Kaiming He, Xiangyu Zhang, Shaoqing Ren, and Jian Sun, “Deep Residual Learning for Image Recognition”, arXiv: 1512.03358 [cs.CV], 2015.

[2] Dan Hendrycks and Thomas Dietterich, “Benchmarking Neural Network Robustness to Common Corruptions and Perturbations”, arXiv: 1903.12261 [cs.LG], 2019.

[3] Gao Huang, Zhuang Liu, Laurens van der Maaten, and Kilian Q. Weinberger, “Densely Connected Convolutional Networks”, arXiv:1608.06993 [cs.CV], 2018.

[4] Alex Krizhevsky, “Learning Multiple layers of features from tiny images”, 2009.

[5] Olaf Ronneberger, Philipp Fischer and Thomas Brox, “U-Net: Convolutional Networks for Biomedical Image Segmentation”, arXiv:1505.04597,2015.

Wasserstein Generative Adversarial Privacy

Kars Mulder Jasper Goseling
Stochastic Operations Research
University of Twente
mail@karsmulder.nl j.goseling@utwente.nl

Abstract

We consider the problem of sharing data without revealing sensitive information, quantified through mutual information. We do so in a generative adversial setting, i.e., training a GAN. In particular we consider training under a Wasserstein distance. Our main contribution is to bound the resulting mutual information in terms of the Wasserstein distance.

1 Introduction

We consider the setting in which a user has data $(X, Y) \in \mathbb{R}^k \times \{0, 1\}$ from which they want to share X without revealing information about the sensitive information Y. The challenge is that Y is in general correlated with X. Therefore, the user can only share a privatised version Z of X. The goal is to obtain Z as the solution of

$$\text{minimise } I(Z;Y) \text{ subject to } \mathrm{E}[d(Z,X)] \leq \ell,$$

where d is a distance metric between Z and X. Summarizing: we minimise privacy leakage w.r.t. a minimum utility constraint on the data.

Following [4, 8, 3] we consider a generative adversarial (GAN) [2] approach to obtain Z. In [4, 8] it was proposed to train two networks simultaneously: a generator $Z = G(X, Y)$ and an adversary $\hat{Y} = D(Z)$, where a cross-entropy loss for the adversary minimises $I(Z;Y)$. However, a known problem with GANs, including the above proposed solution by [4, 8], is that their training process suffers from vanishing gradients and mode collapse. One of the solutions that has been proposed in the literature is to train under a Wasserstein distance [1]. Therefore, we propose that instead of asking the adversary to estimate Y, we ask it to estimate the 1-Wasserstein distance between the two distributions $\mathrm{P}_{Z|Y=0}$ and $\mathrm{P}_{Z|Y=1}$.

Our main contribution in this paper, which is based on [7], is to bound the resulting $I(Z;Y)$ in terms of the 1-Wasserstein distance that is obtained after training. We obtain this bound not only for $I(Z;Y)$, but for general f-informations.

The remainder of this paper is organized as follows. In Section 2 we present background information. In Section 3 we formulate our problem. In Section 4 we present our results. A sketch of the proofs of some of these results is given in Section 5. Finally, in Section 6 we provide a discussion of our results. Proofs are omitted due to space constraints and can be found in [7].

2 Preliminaries

2.1 Mathematical model

We denote the public information that we want to share with the random variable $X \in \mathcal{X}$ where $\mathcal{X}$ is some metric space, and the private information that we don't want to share with $Y \in \mathcal{Y}$. We also define a random variable $N \in \mathcal{N}$ representing some random noise, for example N could be a standard Gaussian in $\mathbb{R}^k$.

We denote the privatiser with a function $G_\omega : \mathcal{X} \times \mathcal{Y} \times \mathcal{N} \to \mathcal{X}$ and the adversary with a function $D_\psi : \mathcal{X} \to \mathcal{Y}$.

The privatiser is a function that takes as input the public and private information X, Y, and gives a possibly random output. The noise N allows the privatiser to produce random output, which is important to prevent G_ω from being a deterministic and possibly reversible mapping. We use the random variable Z to denote the output of the privatiser: $Z = G_\omega(X, Y, N)$.

Sometimes we let the privatiser G_ω be decomposed into two parts: an artificial neural network, and some algorithmic postprocessing of the output of the neural network. In such a case we use the random variable Z' to denote the direct output of the neural network, and Z to denote the output with all postprocessing steps applied.

The discriminator, also called adversary, is supposed to take as input the output of the generator, and output its best guess of what Y was. We denote this guess with $\hat{Y} = D_\psi(Z) = D_\psi(G_\omega(X, Y, N))$. We further assume that there is a loss function $\ell : \mathcal{Y} \times \mathcal{Y} \to \mathbb{R}$, such that $\ell(Y, \hat{Y})$ determines how good the guess $\hat{Y}$ was. The goal of the adversary is to minimise the loss $\ell(Y, \hat{Y})$.

2.2 Loss functions

In machine learning classification tasks, a popular loss function is the cross-entropy loss. The cross-entropy loss is applicable on classification tasks with a finite amount of classes.

Assume that there are n classes and each entry corresponds to one of those classes. The classifier (the adversary in our case) should take as input a privatised datapoint Z and guess to which class Y it belongs. The cross entropy loss expects the output of the classifier to be a set of n neurons, each neuron i representing the adversary's estimate of the probability $\mathrm{P}(Y = i)$.

Assume the correct class is labelled using a one-hot vector $(q_i)_{i \in \{0,\ldots,n\}}$ such that $q_i = 1$ if the input belongs to class i and $q_i = 0$ otherwise, and the classifier activates the output neurons with activation $a_i \in [0, 1]$ for $i = 1, \ldots, n$, then the cross-entropy loss can be computed as:

$$\ell(a, q) = -\sum_{i=1}^{n} q_i \ln a_i.$$

One good property of the cross-entropy loss is that when the cross-entropy loss is used for the adversary, the privatiser's goal of maximising the adversary's loss is equivalent to minimising the amount of mutual information between Y and Z.

2.3 Alternative distance measures between distributions

2.3.1 f-information

There is a more general form of mutual information, called f-information. Let $f : (0, \infty) \to \mathbb{R}$ be a convex function such that $f(1) = 0$. The f-divergence between two probability distributions P and Q, where P is absolutely continuous with respect to Q, is defined as

$$D_f(\mathrm{P} \mid\mid \mathrm{Q}) = \int f\left(\frac{\mathrm{dP}}{\mathrm{dQ}}\right) \mathrm{dQ},$$

where $\frac{\mathrm{dP}}{\mathrm{dQ}}$ represents the Radon-Nikodym derivative of P with respect to Q. When the function $f(t) = t \ln t$ is used, the f-divergence is equivalent to the Kullback-Leibler divergence [5, 6].

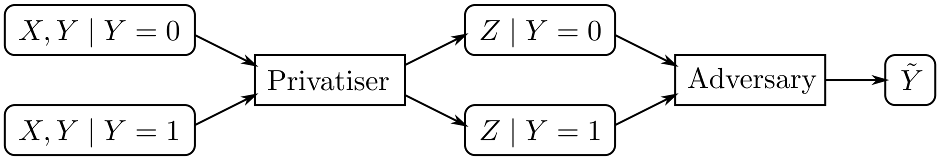


Figure 1: The distribution of the output of the decomposed into two separate distributions, separated on basis of what the underlying private variable was.

The f-information between two not necessarily independent probability distributions P and Q is defined as the f-divergence between their joint probability distribution and the product of their marginal probability distributions. In the special case where the function $f(t) = t \ln t$ is used, the f-information is equal to the mutual information.

2.3.2 Wasserstein distance

The 1-Wasserstein distance is a metric between probability distributions. It is defined using the transport-theoretic notion of an optimal transport plan. Specifically, assume we have two random variables A and B on some shared metric space X with probability distributions respectively P_A and P_B, then the 1-Wasserstein distance $d_W(\mathrm{P}_A, \mathrm{P}_B)$ is defined as

$$d_W(\mathrm{P}_A, \mathrm{P}_B) = \inf_{(X,Y)\in\Pi(A,B)} \mathrm{E}\left[||X - Y||\right],$$

where $\Pi(A, B)$ is the set of all jointly distributed random variables whose marginal distributions are equal to those of A and B. Intuitively, this can be thought of as a transport problem where the probability mass of P_A needs to be transported to the probability mass of P_B, and the cost of transporting mass is the amount of mass to be transported times the distance it must be transported over. The 1-Wasserstein distance between two random variables is the cost of the optimal transport plan.

If the distributions P_A and P_B are compactly supported, the Kantorovich-Rubenstein duality [9] gives another equivalent way to compute the 1-Wasserstein distance between them:

$$d_W(\mathrm{P}_A, \mathrm{P}_B) = \sup_{\substack{f:X\to\mathbb{R} \\ f\text{1-Lipschitz continuous}}} \mathrm{E}_{x\sim\mathrm{P}_A}[f(x)] - \mathrm{E}_{x\sim\mathrm{P}_B}[f(x)].$$

A function $f : \mathbb{R}^n \to \mathbb{R}^m$ is said to be L-Lipschitz continuous if for all $x, y \in \mathbb{R}^n$ holds: $||f(x) - f(y)|| \leq L \cdot ||x - y||$.

2.4 Wasserstein distance as a loss function

When the private information Y is binary, the adversary's task reduces to trying to differentiate between two different distributions $Z \mid Y = 0$ and $Z \mid Y = 1$, see Figure 1. In this case we can try to train the adversary to output a single scalar $\hat{Y}$ which should have a low value when $Y = 0$ and a high value when $Y = 1$:

$$\ell(Y, \hat{Y}) = \begin{cases} \hat{Y}, & \text{if } Y = 0, \\ -\hat{Y}, & \text{if } Y = 1. \end{cases}$$

If this loss function is used, the distribution of the privatiser's output has bounded support, the adversary is constrained to be a 1-Lipschitz-continuous function, and the adversary's performance is theoretically optimal, then the Kantorovich-Rubenstein duality [9] says that the average loss of the adversary $\mathrm{E}\left[\ell(Y, D_\psi(G_\phi(X,Y)))\right]$ is equal to the negative 1-Wasserstein distance between the two distributions $Z \mid Y = 0$ and $Z \mid Y = 1$. The generator's task of maximising the adversary's loss then becomes equivalent to minimising the 1-Wasserstein distance between those two distributions. Training GANs with this kind of loss function forms the basis for a variant of GANs known as "Wasserstein GANs" [1].

3 Problem Statement

Wasserstein GANs are favourable in terms of stability during training. However, there is no direct connection between the 1-Wasserstein distance between $Z \mid Y = 0$ and $Z \mid Y = 1$, and the amount of private information that is leaked $I(Z;Y)$. As such, we do not know whether the privatiser's goal of fooling the adversary actually reduces the amount of private information leaked.

Indeed, it is possible to create pairs of random variables that have arbitrarily small 1-Wasserstein distance but still leak information. A simple example of such variables are the following point-mass distributions. Let $\alpha \neq 0$ arbitrarily and define:

$$\begin{aligned} &Y \sim \text{Bernoulli}(0.5), \\ &Z = \begin{cases} 0 & \text{if } Y = 0, \\ \alpha & \text{if } Y = 1. \end{cases} \end{aligned}$$

The 1-Wasserstein distance between these two distributions equals $|\alpha|$, but by observing the value of Z it is possible to determine the value of Y with 100% certainty.

This is possible because the 1-Wasserstein distance takes into account how far the supports of two distributions are separated. This is an important part of what makes Wasserstein GANs easier to train, but a real adversary doesn't care about how close the supports of $Z \mid Y = 0$ and $Z \mid Y = 1$ lie to each other: if an adversary observes a value of Z which lies in the support of $Z \mid Y = 0$ but not in the support of $Z \mid Y = 1$, then it follows that $Y = 0$.

To address the problem of having arbitrarily close but nonoverlapping supports, it is possible to add random noise to the output of the neural network. E.g. if Z' is the output of the privatiser's neural network layers and N is a Gaussian random variable, we define $Z = Z' + N$ and then consider Z to be the output of the privatiser. This will cause all probability mass of Z' to be smeared out over an area with a certain width determined by the distribution of the noise.

The big question is then whether adding noise to the output of the privatiser's network is sufficient to guarantee that a low 1-Wasserstein distance between $Z \mid Y = 0$ and $Z \mid Y = 1$ implies low mutual information between Z and Y. The answer to this question turns out to be "yes, subject to certain conditions."

If Z' is some random variable and N is independent Gaussian noise, then the random variable $Z = Z' + N$ is continuously distributed with a Lipschitz continuous probability density function regardless of the distribution of Z'. As such, we have analysed the properties of continuous random variables with Lipschitz-continuous probability density functions. In Section 6 we discuss this assumption in more detail.

4 Results

As the privatiser/adversary neural networks are trained, the distribution of their outputs change. We model this training process by considering the output of a privatiser/adversary after n training steps as a separate random variable for each n.

If we let the random variable A_n represent distribution of $Z \mid Y = 1$ after n training iterations, and B_n represent the distribution of $Z \mid Y = 0$ after n iterations, and let the private information Y be Bernouilli($\frac{1}{2}$) distributed, then $Z_n = YA_n + (1 - Y)B_n$ represents the distribution of the output of the privatiser after n iterations.

In the remainder we assume that A_n and B_n are continuously distributed, and that their probability density functions a_n, b_n are L-Lipschitz continuous for some $L \in \mathbb{R}$. One way to satisfy this requirement is by adding random noise to the output of the privatiser, e.g. $Z = Z' + N$ where N may be a Gaussian random variable or a random variable with some other appropriate distribution.

Our first result states that if A_n converges to B_n under the 1-Wasserstein metric, then its probability density function a_n converges uniformly to b_n as $n \to \infty$.

Theorem 1. $\lim_{n\to\infty} d_W(A_n, B_n) = 0 \implies \lim_{n\to\infty} ||a_n - b_n||_\infty = 0$.

Our next result relates to the f-information between A_n and B_n, where f is a convex function $f : (0, \infty) \to \mathbb{R}$ such that $f(1) = 0$, and where our theorems also require that $\lim_{t\to 0} f(t) < \infty$.

The next result requires a slightly stronger condition on A_n and B_n. In particular, it requires the distributions of the A_n (or by symmetry, B_n) to be tight, i.e. for all $\epsilon > 0$ there must be a set $K_\epsilon \subset \mathbb{R}^k$ with $\lambda^k(K_\epsilon) < \infty$ such that for all $n \in \mathbb{N}$ we have $\mathrm{P}(A_n \in K_\epsilon) > 1 - \epsilon$.

Theorem 2. *If the random variables* $(A_n)_{n\in\mathbb{N}}$ *are tight and* $\lim_{n\to\infty} ||a_n - b_n||_\infty = 0$ *then* $\lim_{n\to\infty} I^f(Z_n; Y) = 0$.

Provided that the requirements of the theorems hold, Theorem 1 and 2 together state that if the 1-Wasserstein distance between A_n and B_n were to converge to zero as the privatiser is trained, then the f-information between the output of the privatiser $YA_n + (1 - Y)B_n$ and the private information Y converges to zero as well.

Finally, imposing slightly stronger conditions, we bound $I^f(Z_n; Y)$ in terms of $d_W(A_n, B_n)$.

Theorem 3. *If* A_n *and* B_n *are supported in a compact set* $K \subset \mathbb{R}^k$ *and*

$$d_W(A_n, B_n) \leq \frac{1}{8L^2\lambda^k(K)},$$

then

$$I^f(Z_n; Y) \leq c_1 \cdot \sqrt{d_W(A_n, B_n)}\big(c_2 - \ln d_W(A_n, B_n)\big),$$

where c_1 *and* c_2 *are constants whose values depend only on* f, $\lambda^k(K)$ *and* L.

4.1 On the tightness requirements

Even if the random variables A_n, B_n satisfy all the requirements for Theorem 1, it is still possible for them to leak private information, which is why Theorem 2 has an additional requirement: that either $(A_n)_{n\in\mathbb{N}}$ or $(B_n)_{n\in\mathbb{N}}$ have tight distributions.

Take a look at the following probability density functions, which have been plotted in Figure 2:

$$\begin{aligned} a_n(x) &= \mathbf{1}_{[0,2^{n-1}]}(x) \cdot 2^{-n+1} \cdot \left(\sin(2^n \pi x - \tfrac{1}{2}\pi) + 1\right), \\ b_n(x) &= \mathbf{1}_{[0,2^{n-1}]}(x) \cdot 2^{-n+1}. \end{aligned}$$

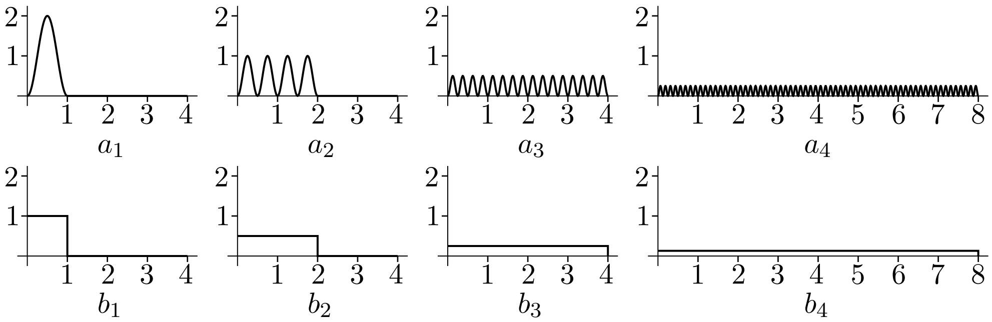


Figure 2: Two series probability density functions which uniformly converge to each other in the supremum-norm, but do nevertheless leak a constant amount of information.

The derivative of a_n is

$$a_n'(x) = \mathbf{1}_{[0,2^{n-1}]}(x) \cdot 2\pi \cos(2^n \pi x - \tfrac{1}{2}\pi),$$

which is bounded between $[-2\pi, 2\pi]$ for all values of n. Hence the series a_n is uniformly Lipschitz continuous and satisfies the requirements of Theorem 1. The series b_n is strictly speaking not Lipschitz continuous at the border of its support, but that could be fixed by adding a small slope at the border, which we haven't done for simplicity. Indeed, a_n and b_n do converge uniformly to each other as $\lim_{n\to\infty} ||a_n - b_n||_\infty = 0$.

Nevertheless, a combination $Z_n = YA_n + (1-Y)B_n$ leaks a constant information about Y: if we observe that Z_n takes a value that is unlikely to be taken by A_n, e.g. $Z_n \in \{x \mid x \in [0, 2^{n-1}] \mid a_n(x)/b_n(x) < \frac{1}{10}\}$), then we know that this value of Z_n was probably sampled from B_n and therefore the private information Y is most likely 0. The chance that Z_n observes one such value is independent of n. This counterexample shows that we need some constraint on the support of the distributions A_n and B_n.

5 Sketch of proofs

Sketch of proof of Theorem 1. If $||a_n - b_n||_\infty > \gamma$, then there is some $x \in \mathbb{R}^k$ such that $||a_n(x) - b_n(x)|| > \gamma$. Because a_n and b_n are L-Lipschitz-continuous, we would have $a_n(y) - b_n(y) > 0$ for all $y \in B(x; \gamma/2L)$. This mass must be directly or indirectly moved from this surplus area to a deficit area, so the probability mass in $B(x; \gamma/4L)$ would have to be moved over a distance of at least $\gamma/4L$, which means that the 1-Wasserstein distance between A_n and B_n is bounded from below by $\frac{1}{4L}\gamma \cdot \lambda^k(B(x; \gamma/4L))$.

Thus, if the $d_W(A_n, B_n) \to 0$, then it must follow that $\frac{1}{4L}\gamma \cdot \lambda^k(B(x; \gamma/4L)) \to 0$, hence $\gamma \to 0$ and $||a_n - b_n||_\infty \to 0$. □

Sketch of proof of Theorem 2. Because the probability distributions of A_n are tight, the probability mass of them cannot be spread over a too large area. As result, for every $\epsilon > 0$ we can find a δ such that for all $n \in \mathbb{N}$ we have $\mathrm{P}(a_n(A_n) > \delta) > 1 - \epsilon$.

If we know that $a_n(A_n)$ is most likely over a given threshold δ, then we can choose a N sufficiently large such that for all $n > N$ the distance $||a_n - b_n||_\infty$ is much smaller than either δ or $\lambda^k(\{a_n > \delta\})$, from which follows that B_n has to be very similar to A_n.

We use the similarity between A_n and B_n similarity to find subsets $S_n \subset \mathbb{R}^k$ with all of the following nice properties:

- The probability that A_n and B_n lie in S_n is arbitrarily close to 1;

- The probability density functions a_n and b_n have lower bounds greater than zero on S_n;
- The distance $||a_n - b_n||_\infty$ is arbitrarily small compared to the last-mentioned lower bounds.

From this follows that $a_n(x)/b_n(x)$ arbitrarily close to 1 for all $x \in S_n$. When finally computing the f-information between $YA_n + (1-Y)B_n$ and Y by computing an integral over $x \in \mathbb{R}^k$, one of the two following things can happen:

- If $x \in S_n$, then $a_n(x)/b_n(x)$ and similar formula's are arbitrarily close to 1 and the contributed divergence is arbitrarily small;
- If $x \notin S_n$, then the argument of $f(\dots)$ is bounded between $[0, 2]$, on which f is bounded, and thus the amount of divergence contributed is bounded by a constant times $P(A_n \notin S_n \vee B_n \notin S_n)$, which can be made arbitrarily small. □

Sketch of proof of Theorem 3. This proof happens in two steps: first, we bound the L_1 distance between a_n and b_n in terms of the 1-Wasserstein distance, and then we bound the leaked information in terms of the L_1 distance.

To get the bound on the L_1 distance, we first compute a bound on the L_2 distance. Due to the L-Lipschitz continuity of a_n, b_n, it follows that if $|a_n(x) - b_n(x)| = \delta$ for some $x \in \mathbb{R}^k$, we also have $|a_n(x') - b_n(x')| > 0$ for all $x' \in B(x; \delta/2L)$. Since in this setting, there is an optimal transport plan where probability mass is only transported from surplus areas to deficit areas, the probability density $|a_n(x) - b_n(x)|$ needs to be transported over a distance of at least $|a_n(x) - b_n(x)|/2L$ and thus contributes a density of at least $|a_n(x) - b_n(x)|^2/2L$ to the 1-Wasserstein distance.

We then use the property that a_n, b_n have compact support and Jensen's inequality to transform the bound on the L_2 distance into a bound on the L_1 distance.

The f-information $I^f(Z_n; Y)$ is defined as $\int f\big(\frac{\mathrm{dP}}{\mathrm{dQ}}\big)\mathrm{dQ}$, where P is the joint distribution between $(Y, YA_n + (1-Y)B_n)$ and Q is the product of their marginal distributions. Since the ratio $\frac{\mathrm{dP}}{\mathrm{dQ}}$ is bounded in $[0, 2]$ and f is a convex function such that $f(1) = 0$, $f(2) < \infty$ and $\lim_{x\to 0} f(x) < \infty$, we can bound the function f as $f(x) < C \cdot |x - 1|$ for $x \in (0, 2]$ and some constant C. We then use this to bound the f-information with $\mathrm{E}_{v,x\sim\mathrm{Q}}\big[f\big(\frac{\mathrm{dP}}{\mathrm{dQ}}(v, x)\big)\big] \leq C \cdot \mathrm{E}_{x\sim\mathrm{Q}}\big[\big|\frac{\mathrm{dP}}{\mathrm{dQ}}(v, x) - 1\big|\big]$

The bulk of the rest of the proof then involves showing a bound upon $\mathrm{E}_{v,x\sim\mathrm{Q}}\big[\big|\frac{\mathrm{dP}}{\mathrm{dQ}}(v, x) - 1\big|\big]$ in terms of $||a_n - b_n||_1$. □

6 Discussion

The core problem with the relation between the 1-Wasserstein distance and the leaked information is that the 1-Wasserstein distance between distributions can be decreased by decreasing the distance over which the mass can be transported, whereas the leaked information doesn't care about that distance. An arbitrarily close distance can lead to an arbitrarily small 1-Wasserstein distance without impacting the amount of leaked information.

To avoid the issue of having arbitrarily small distances, we tried requiring a_n and b_n to be Lipschitz continuous. The Lipschitz constraint on the densities ensures that if a_n and b_n differ by a certain amount, this difference has to be transported over a certain distance and therefore have a nonnegligible impact on the 1-Wasserstein distance.

In the provided counterexample we use a series of sinuses $(a_n)_{n\in\mathbb{N}}$ of which the amplitude decreases simultaneously with its period. L-Lipschitz continuity only ensures that a difference in probability density of magnitude $|a_n(x) - b_n(x)|$ has to be transported over a distance $|a_n(x) - b_n(x)|/L$. Therefore, as the magnitude of $a_n(x) - b_n(x)$ decreases, so does the minimum distance over which the mass needs to be transported.

By creating series of functions where the amplitude of $a_n - b_n$ and period of $a_n - b_n$ decrease simultaneously, we can let the 1-Wasserstein distance $d_W(A_n, B_n)$ to decrease quadratically while various other measures such as $||a_n - b_n||_1$ and $||a_n - b_n||_\infty$ decrease only linearly.

Our proposed method to enforce Lipschitz continuity of a_n and b_n is to add appropriately-shaped random noise to the output of the privatiser. However, adding random noise to another random variable guarantees more properties than just Lipschitz continuity.

For example, we suspect that adding Gaussian noise would prevent the period from decreasing linearly with the amplitude, thereby making the worst-case scenario's we've thought of impossible. This could lead to the rate of convergence becoming approximately $O(x)$ instead of $O(-\sqrt{x}\ln x)$, making Theorem 3 a lot more applicable in practice.

References

[1] Martin Arjovsky, Soumith Chintala, and Léon Bottou. "Wasserstein generative adversarial networks". In: *International Conference on Machine Learning.* 2017, pp. 214–223.

[2] Ian Goodfellow et al. "Generative adversarial nets". In: *Advances in neural information processing systems.* 2014, pp. 2672–2680.

[3] Jihun Hamm. "Minimax filter: learning to preserve privacy from inference attacks". In: *The Journal of Machine Learning Research* 18.1 (2017), pp. 4704–4734.

[4] Chong Huang et al. "Context-aware generative adversarial privacy". In: *Entropy* 19.12 (2017), p. 656.

[5] Solomon Kullback and Richard A Leibler. "On information and sufficiency". In: *The annals of mathematical statistics* 22.1 (1951), pp. 79–86.

[6] Friedrich Liese and Igor Vajda. "On divergences and informations in statistics and information theory". In: *IEEE Transactions on Information Theory* 52.10 (2006), pp. 4394–4412.

[7] Kars Mulder. "Wasserstein Generative Adversarial Privacy Networks". MA thesis. Department of Applied Mathematics, University of Twente, July 2019. URL: `http://essay.utwente.nl/79054/`.

[8] Ardhendu Tripathy, Ye Wang, and Prakash Ishwar. "Privacy-preserving adversarial networks". In: *Allerton.* 2019, pp. 495–505.

[9] Cédric Villani. *Optimal transport: old and new.* Vol. 338. Springer Science & Business Media, 2008.

Finger Vein Verification with a Convolutional Auto-encoder

Tugce Arican Raymond Veldhuis Luuk Spreeuwers
University of Twente
EEMCS, DMB
Enschede, Netherlands
t.arican@utwente.nl r.n.j.veldhuis@utwente.nl l.j.spreeuwers@utwente.nl

Abstract

Unsupervised learning methods can learn generalized features without label information, which can be advantageous for small datasets like finger veins. This work explores the possibility of learning finger vein representations with an unsupervised learning method called Convolutional Auto-encoder(CAE), and proposed a modification to the loss function to learn better representations of the vein patterns. The results indicate that the CAE is powerful in reconstructing global structures, yet failed to reconstruct vein patterns. The CAE with a Log-likelihood ratio classifier achieved 6.74% EER on UTFVP dataset. Though the performance of the system is far from the state-of-the-art, the findings imply that the global finger structures contribute to the identity, and are powerful enough to be used as an additional information source for finger vein recognition.

1 Introduction

Finger vein recognition is a biometric authentication method relying on comparing finger vein patterns beneath the skin. First step in finger vein recognition is to extract the vein representations. Classical methods extract the finger vein features based on the vein appearance such as vein profile, vein directions, bifurcation points of veins, or texture possessed by the finger vein patterns. These methods are powerful on extracting vein patterns[1, 2], but they rely on a single type of feature, and the image quality highly affects the quality of the extracted veins.

Rather than directly comparing vein patterns, supervised training-based methods, like convolutional Neural Networks(CNNs), can generalise finger vein representations. Due to the small sizes of finger vein databases, generally transfer learning is preferred for finger vein recognition[3, 4]. However, the models used in transfer learning are very deep and trained on large object recognition datasets. Hence, many layers and features might not be suitable for finger veins, which increases redundancy.

In addition to the supervised methods, unsupervised methods learn input representations without requiring label information. Convolutional Auto-encoder(CAE) is a type of unsupervised learning method which takes an image as an input and compresses it to a lower dimensional representation, called latent space, then aims to reconstruct the input image from the latent space. During this process, CAE learns the most informative representation of the input image in order to reconstruct it again as good as possible. In literature, the approaches using this learned representation as a feature vector achieved promising results on face recognition[5, 6].

This work investigates whether an unsupervised learning method can reconstruct finger vein images and learn finger vein representations. For this purpose, a CAE model is constructed, and its loss function is modified to help learning better vein representations. Then, a Log-likelihood classifier is trained on top of the latent space of the CAE.

This paper is organised as follows. In Section 2, previous studies on autoencoders are briefly reviewed. Sections 3 and 4 present the methodology and the dataset used in this work, respectively. Section 5 presents the results of this work on a publicly available figner vein dataset. Section 6 discusses the findings of this work and presents a feature work. Finally, Section 7 concludes this paper.

2 Related Work

Learning powerful and generalised representations is crucial for biometric recognition. Recent works show that unsupervised learning methods can learn good representations by easing the demand for large datasets. Yang et al.[5] implemented a convolutional denoising autoencoder for facial feature extraction. The authors achieved a higher recognition performance with XGBoost classifier compared to other methods like PCA, Gabor, and a CNN. Bhaswara[6] explored the potential of autoencoders as a feature extractor on a face recognition system. In this work, several types of autoencoders were explored including the regular ones and the generative models. The results showed that autoencoders are capable to extract powerful features. Moreover, the generative model achieved a better generalisation over the latent variables then the regular one. Silva et.al.[8] trained a CAE on lung images in order to learn representations of lung nodules. Later, the encoder part of the network is transferred and used to determine the risk of malignancy. The results indicated that such an unsupervised learning step could help learn better image representations with small datasets also encountered in the medical imaging field.

CAEs can learn powerful representations, however, they fail to reconstruct fine details in the input image. Ghodrati et.al.[9] evaluated MR image reconstruction with several network structures and loss functions such as Mean Square Error(MSE), Mean Absolute Error(MAE), structural dissimilarity(Dssim), and perceptual loss. The results showed that pixel-wise loss functions such as MSE and MAE produced blurry outputs by removing a part of image texture information. Ichimura[11] utilised a CAE for feature extraction from images. The author claimed that the pixel based loss functions such as MSE leads a loss on high spatial frequency details such as edges in the reconstructed image, which are important for object detection. In this work, a spatial frequency loss(SFL) function was proposed to mitigate the problem faced with MSE loss. A convolution layer with a Laplacian filter bank was used to select high frequency components of the input and the reconstructed images. Then, SFL was computed by taking the mean square error of these filtered images. The results indicates that the proposed loss function clearly reduced the blur on the reconstructed image. SFL provides a basis for the proposed loss function in this paper.

3 Methodology

3.1 Autoencoders and Feature extraction

Convolutional autoencoders(Figure 1) are a type of neural networks which consist of two parts, namely an encoder and a decoder. The encoder part consists of a number of convolution layers which are responsible for dimensionality reduction. The decoder is generally a reversed version of the encoder which aims to reconstruct the input from the latent space representation as good as possible. The goodness of the reconstruction is measured by a loss function by comparing the reconstructed image against the input of the network. It is expected that the CAE eventually learns a representation of the input data which can be further used as a feature vector.

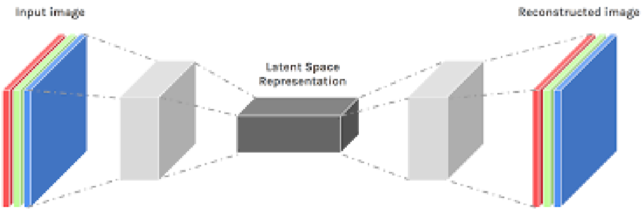


Figure 1: Convolutional auto-encoder

3.2 Loss Function

Loss function measures the goodness of a reconstruction of CAE by comparing the reconstructed image against the input of the network. Mean Square Error(MSE)(Equation 1) and Mean Absolute Error(MAE)(Equation 2) are the most commonly used loss functions in CAE training.

$$MSE(x, y) = \frac{1}{n}\sum(x - y)^2 \tag{1}$$

$$MAE(x, y) = \frac{1}{n}\sum|x - y| \tag{2}$$

Both MAE and MSE are pixel-wise loss functions, which means that the difference between the images are computed at the pixel level. MAE loss treats all the errors equally, while MSE loss tend to prefer larger errors over the smaller ones, which generally leads loss of small details. On finger vein images, it is observed that MSE loss causes the loss of all fine vein details while keeping the global structures such as joints and illumination patterns. Although MAE loss is able to put more attention to finer veins than MSE, this is not powerful enough to properly reconstruct the finger veins. Encoding the finger veins properly is crucial since the vein patterns carry identity information, and a good reconstruction can be considered as an indicator of a good encoding.

Loss function of the CAE is modified such a way to exploit benefits of MSE and MAE and avoid from their drawbacks. The proposed loss function combines the error on finger vein images with the error on the vein patterns(Equation 3). The vein patterns are extracted by applying a gabor filter bank on the input and the reconstructed finger vein images. α is a hyper parameter which determines the contribution of finger vein patterns to the loss calculation, and it is set to 5 after exhaustive experiments.

$$\begin{aligned} L &= MSE(x, y) + \alpha MAE(G(x), G(y)) \\ G &= GaborFilter \end{aligned} \tag{3}$$

3.3 Log-likelihood Ratio Classifier

Let $x \in R^M$ and $y \in R^N$ be two biometric feature vectors, where M and N are the dimensions of these vectors and $M \geq N$. These feature vectors should support the hypothesis H_s where the feature vectors belong to the same subject, or H_d where the features are coming from different subjects. Then, the likelihood ratio can be defined as,

$$l(x, y) = \frac{p(\binom{x}{y}|H_s)}{p(\binom{x}{y}|H_d)} \tag{4}$$

The vectors x and y are feature vectors of random individual which are specified by its feature mean. Hence, these feature vectors can be expressed as,

$$x = \mu_\omega + w_\omega \ , \ y = \mu^{'}_\theta + w^{'}_\theta \tag{5}$$

where μ_ω and $\mu^{'}_\theta$ are the subject specific mean modeling within subject variations, and w_ω and $w^{'}_\theta$ are statistically independent and have zero-mean within subject variations. ω and θ represent the identity of x and y. The co-variance and cross-variance of x and y are

$$\begin{gathered}\Sigma_{xx} = E\{xx^T\} \in R^{MxM} \ , \ \Sigma_{yy} = E\{yy^T\} \in R^{NxN} \\ \Sigma_{xy} = E\{xy^T\} \in R^{MxN} \ , \ \Sigma_{yx} = \Sigma_{xy}\end{gathered} \tag{6}$$

During training of the classifier, these co-variance and cross-variance matrices are estimated. The estimated cross-variance $\hat{\Sigma}_{xy} = \frac{1}{K}\sum_{i=1}^{K} \hat{\mu}_i \hat{\mu}_i^{'T}$, where K is the number of individuals, and $\hat{\mu}_i$ and $\hat{\mu}^{'}_i$ are the estimated sample means of subject i. Assume the probability density of the feature vector pairs for μ_ω, $\mu^{'}_\theta$, w_ω, and $w^{'}_\theta$ with unknown ω and θ have normal and zero-mean. The similarity score of x and y

$$s(x,y) = (x^T y^T)(\begin{pmatrix}\Sigma_{xx} & 0 \\ 0 & \Sigma_{yy}\end{pmatrix}^{-1} - \begin{pmatrix}\Sigma_{xx} & \Sigma xy \\ \Sigma_{yx} & \Sigma_{yy}\end{pmatrix}^{-1})\begin{pmatrix}x \\ y\end{pmatrix} \tag{7}$$

Equation 7 can be further simplified by reducing the dimensionality and applying whitening transforms to x and y. The reduced equation

$$s(x_w, y_w) = (x^T{}_w y^T{}_w)(\begin{pmatrix}I & 0 \\ 0 & I\end{pmatrix}^{-1} - \begin{pmatrix}I & \Sigma^w{}_{xy} \\ \Sigma^w{}_{yx} & I\end{pmatrix}^{-1}\begin{pmatrix}x_w \\ y_w\end{pmatrix} \tag{8}$$

where I is the identity matrix, $x_w = W_H x \in R^{M_w}$ and $y_w = W_L y \in R^{N_w}$. M_w and N_w are the new dimensions of x_w and y_w where $M_w < M$, $N_w < N$, and $M_w \geq N_w$. The identity matrices from Equation 8 are obtained from $\Sigma^w{}_{xx} = E\{x_w x^T{}_w\} = I$, $\Sigma^w{}_{yy} = E\{y_w y^T{}_w\} = I$, and $\Sigma^w{}_{xy} = W_H \Sigma_{xy} W_L$.

$\Sigma^w{}_{xy} = W_H \Sigma_{xy} W_L$ can later be decomposed into $\Sigma^w{}_{xy} = UDV^T$ by using singular value decomposition(SVD), where $U \in R^{M_w x M_w}$ and orthonormal, $V \in R^{N_w x N_w}$, and orthonormal, and $D \in R^{M_w x N_w}$. The first N_w rows of D forms a diagonal matrix containing singular values $v_i = 1, ..., N$. Also, the last $M_w - N_w$ rows of D are all 0, and the rank of D is $D = min(N_w, K-1)$ where K is the number of individuals. The feature vectors x an y can be transformed using SVD components.

$$x_c = (U_{*,1:D})^T x_w \in R^D \ , \ y_c = (V_{*,1:D})^T y_w \in R^D \tag{9}$$

Now, Equation 8 can be written as follows,

$$s(x_w, y_w) = (x^T{}_w y^T{}_w)(\begin{pmatrix}I & 0 \\ 0 & I\end{pmatrix}^{-1} - \begin{pmatrix}I & D \\ D & I\end{pmatrix}^{-1}\begin{pmatrix}x_w \\ y_w,\end{pmatrix} \tag{10}$$

and then can be simplified into,

$$s(x_c, y_c) = -\sum_{i=1}^{D} \frac{v_i}{1-v_i}(x_{c,i} - y_{c,i})^2 + \sum_{i=1}^{D} \frac{v_i}{1+v_i}(x_{c,i} + y_{c,i})^2 \tag{11}$$

Then, the log-likelihood ratio classifier is defiend as

$$log(l(x_c, y_c)) = -\frac{1}{2}\sum_{i=1}^{D} log(1 - v^2{}_i) + \frac{1}{4}s(x_c, y_c) \tag{12}$$

3.4 Architecture and Training

The CAE architecture is inspired by [7], and is adapted to the finger vein images. The model involves only convolution layers, and dimensionality reduction is performed only with the convolution layers. Table1 shows the encoder architecture. The output of the last layer of the encoder serves as the latent space of the CAE.

Layer	Kernel size / Stride	Output shape
conv(nb=16, BN, ReLU)	3x3 / 2	64x128x16
conv(nb=32, BN, ReLU)	3x3 / 2	32x64x32
conv(nb=64, BN, ReLU)	3x3 / 2	16x32x64
conv(nb=128, BN, ReLU)	3x3 / 2	8x16x128
conv(nb=256, BN, ReLU)	3x3 / 2	4x8x256
conv(nb=512, BN, ReLU)	3x3 / 2	2x4x512
conv(nb=1024, BN, ReLU)	3x3 / 2	1x2x1024

Table 1: Encoder Architecture. BN: Batch Normalisation

The decoder involves transposed convolution layers to perform up-sampling in the reconstruction stage, and a convolution layer is added to generate one channel output image. Detailed architecture of the decoder network is given in Table2.

Layer	Kernel size / Stride	Output shape
convT(nb=512, BN, ReLU)	3x3 / 2	2x4x512
convT(nb=256, BN, RELU)	3x3 / 2	4x8x256
convT(nb=128, BN, ReLU)	3x3 / 2	8x16x128
convT(nb=64, BN, ReLU)	3x3 / 2	16x32x64
convT(nb=32, BN, ReLU)	3x3 / 2	32x64x32
convT(nb=16, BN, ReLU)	3x3 / 2	64x128x16
convT(nb=16, BN, ReLU)	3x3 / 2	128x256x16
convT(nb=1, BN, ReLU)	1x1 / 1	128x256x1

Table 2: Decoder Architecture. BN: Batch Normalisation

The CAE is trained with both MSE loss and the modified loss for 50 epochs. Adam optimiser is chosen for training with a learning rate of 10^{-4}. Batch size is set to 64.

3.5 Evaluation Metrics

Reconstruction performance of the CAE is visually evaluated by examining the input image and its reconstruction. Good reconstruction is an indicator of a good learning of finger vein representations. Therefore, it is important for the CAE to produce visually convincing reconstructions.

To evaluate the LLR classifier Equal Error Rate(EER) and Area Under Curve(AUC) metrics are used. EER indicates that the proportion of the false acceptances is equal to the proportion of the false rejections. The lower the EER, the higher the performance of the verification system. AUC implies how much the model is able to distinguish between classes. The higher the AUC, the better the classifier distinguishes between genuine and imposter pairs on a verification system.

4 Dataset

In this work, a finger vein dataset provided by University of Twente is used[12]. This dataset involves 1440 finger vein images of 60 subjects. 3 fingers, index, middle, and ring, of both hands are captured for each subject. The input size of the images are 380 x 672. First, the finger region is detected and extracted, then the extracted finger region is resized to 128 x 256 pixels in order to reduce the computational complexity. Figure 2 shows some samples from the dataset.

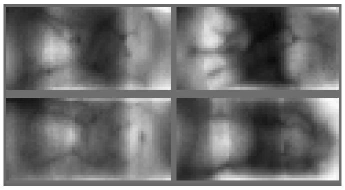


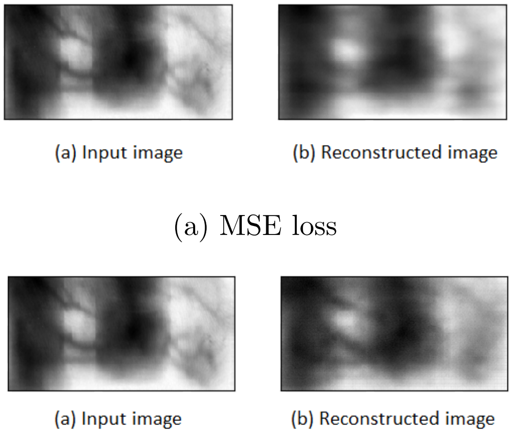
Figure 2: Input image samples

The first 20 subjects of the dataset are used in CAE training. This partition, later, split into training and validation sets by the fraction of 2/3 and 1/3, respectively. Basic augmentation techniques such as horizontal flip, image cropping, and brightness change, are applied on the training partition. The last 40 subjects are reserved for training and evaluation of the LLR classifier. The First 30 subjects of this partition are used in the training of the classifier, while the remaining 10 are kept for the evaluation.

5 Results

5.1 Reconstruction

[9, 11] claimed that MSE loss function causes a blur and loss of fine details in the input image. Figure 3a supports this argument by showing that the veins are mainly lost in the reconstruction, while the global structures like joints and illumination patterns are preserved well.



(a) MSE loss

(b) Modified loss

Figure 3: Reconstructions with (a) MSE loss (b) modified loss(alpha=5)

The modification on the loss function helps to recover some blur and achieves accurate reconstructions in terms of global structures and prominent veins(Figure 3b). Although the reconstructed image shows few finer vein structures, the clarity of these patterns falls short of what is expected from the modification.

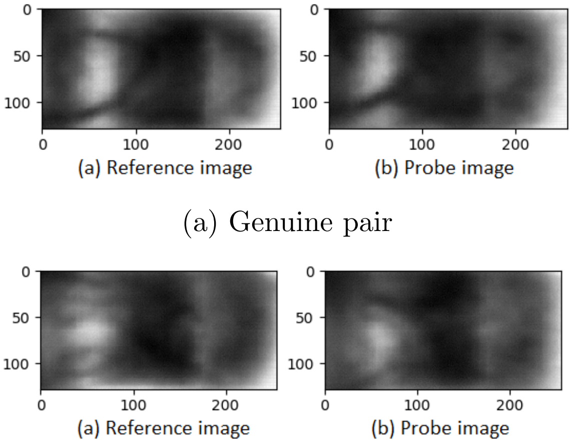
5.2 Verification

An LLR classifier is trained on top of the latent space of the CAE for a finger verification task. Table 3 compares the performances of the CAE trained with the aforementioned loss functions. MSE loss achieves 6.74% EER, while for the modified loss, the EER of the system is measured as 7.15%. This results indicate that the improvement in the reconstruction performance is not reflected on the verification stage.

Loss function	EER(%)	AUC
MSE loss	6.74	0.968
Modified loss	7.14	0.974

Table 3: Performance comparison of MSE loss and the modified loss

Falsely rejected(Figure 4a) and falsely accepted(Figure 4b) pairs show that the finger pairs are accepted or rejected based on the similarities on the global finger structures. Based on this finding, the classifier is likely to learn the global structures of the fingers, and defines the similarity based on the similar global structures appearing on the image pairs rather than the vein patterns.



(a) Genuine pair

(b) Imposter pair

Figure 4: (a) Falsely rejected pair, (b) Falsely accepted pair

On the other hand, the reconstruction and the verification results yield an interesting fact. The reconstructed images and the LLR classifier indicate that the latent space mostly involves the global structures rather than rich finger vein patterns. Yet, the verification performance measured as 6.39%. This performance is considered as poor for a biometric recognition system, still, it can imply that the finger itself highly contributes the identity information.

6 Discussion and Future Work

This paper aims to investigate whether an unsupervised learning method can learn finger vein representations good enough to perform biometric recognition. For this purpose, a Convolutional Auto-encoder and an improvement to the loss function are proposed.

The literature claims that MSE loss causes blurry reconstructions and loss of details. The reconstructed finger vein images in Figure 3 prove this claim. Vein patterns are mostly lost in the reconstruction, while global structures such as finger joints and illumination patterns are preserved well. This can be attributed to the fact that the finger veins are small structures and they are represented as dark regions in an image. Their contribution in the loss function is smaller compared to the larger regions such as joints. MSE loss prefers higher errors over the lower ones. As a result, finger veins are likely to be ignored by MSE, which causes a poor learning of the vein patterns.

The loss function is modified in order to recover the lost vein details. It yields slightly more accurate reconstruction of global structures and prominent veins. Contrary to expectations, fine vein structures are still lost in the reconstruction. Fine veins cover even smaller area in an image, hence, loss component on the vein patterns is not enough to put enough attention to the fine veins and properly learn them.

Even though the modified loss function achieves more accurate reconstructions compared to MSE, their verification performances are almost the same. It is observed that, the LLR classifier thinks that a finger pair is similar when the global structures are similar. The LLR classifier learns the most important components of the input features during training. The findings indicate that, the classifier prefers global structures over the vein patterns, and thinks that they are more important for verification.

Though the classifier performance is by far as not good as the state-of-the-art, the findings imply that the global structures of fingers also contribute to the identity. Finger joins and illumination patterns are pretty much similar on genuine pairs, and tend to differ on imposter pairs. The identity information carried by those global structures is not as strong as the vein patterns, yet, the classifier performance indicates they are powerful enough to be used as an additional information for finger vein recognition.

Although the proposed approach has not achieved the expected verification performance, the findings have greatly contributed to out understanding of how Convolutional Auto-encoders work. CAEs are powerful on learning global structures of fingers, however, they failed to capture the details like finger vein patterns. Since they are tiny, dark structures, the veins can not contribute to the loss function as much as the global structures. As a future work, small patches extracted from the finger region can be used as an input to the CAE instead of the whole finger. By doing this, both the area covered by the veins is increased, and the complexity of the input image is reduced. In this way, the CAE can learn the vein structures in a better way.

7 Conclusion

This paper proposes an unsupervised learning approach for learning finger vein representations. A convolutional autoencoder(CAE) model is constructed for unsupervised learning. The proposed approach achieves 7.15% with the modified loss function, while MSE loss achieves 6.74% EER on UTFVP dataset. Unlike the modified loss function improves the reconstruction performance, this improvement is not reflected on the verification performance.

Though the performance is by far from state-of-the-art, the experiments made an invaluable contribution to our knowledge about the Convolutional Auto-encoders and finger vein recognition. The CAE learns and successfully reconstructs global finger structures such as joints and illumination patterns. Moreover, the verification experiments imply that these global structures also contribute to the identity. In the light of our findings, we will be able to further improve this unsupervised approach to the state-of-the art.

References

[1] Miura, Naoto, Akio Nagasaka, and Takafumi Miyatake. ”Feature extraction of finger-vein patterns based on repeated line tracking and its application to personal identification.” Machine vision and applications 15.4 (2004): 194-203.

[2] Miura, Naoto, Akio Nagasaka, and Takafumi Miyatake. ”Extraction of finger-vein patterns using maximum curvature points in image profiles.” IEICE TRANSACTIONS on Information and Systems 90.8 (2007): 1185-1194.

[3] Tang, Su, et al. ”Finger vein verification using a Siamese CNN.” IET Biometrics 8.5 (2019): 306-315.

[4] Song, Jong Min, Wan Kim, and Kang Ryoung Park. ”Finger-vein recognition based on deep DenseNet using composite image.” IEEE Access 7 (2019): 66845-66863.

[5] Yang, Jun, et al. ”Facial Expression Recognition Based on Convolutional Denoising Autoencoder and XGBoost.” 2019 IEEE 8th Joint International Information Technology and Artificial Intelligence Conference (ITAIC). IEEE, 2019.

[6] Bhaswara, Irfan Dwiki. Exploration of autoencoder as feature extractor for face recognition system. MS thesis. University of Twente, 2020.

[7] Pinho, Eduardo, and Carlos Costa. ”Feature Learning with Adversarial Networks for Concept Detection in Medical Images: UA. PT Bioinformatics at ImageCLEF 2018.” CLEF (Working Notes). 2018.

[8] Silva, Francisco, et al. ”Pre-Training Autoencoder for Lung Nodule Malignancy Assessment Using CT Images.” Applied Sciences 10.21 (2020): 7837.

[9] Ghodrati, Vahid, et al. ”MR image reconstruction using deep learning: evaluation of network structure and loss functions.” Quantitative imaging in medicine and surgery 9.9 (2019): 1516.

[10] Pardo Ginés, Andrés. ”Study of brain magnetic resonance images reconstruction through convolutional autoencoders.”

[11] Ichimura, Naoyuki. ”Spatial frequency loss for learning convolutional autoencoders.” arXiv preprint arXiv:1806.02336 (2018).

[12] Ton, Bram T., and Raymond NJ Veldhuis. ”A high quality finger vascular pattern dataset collected using a custom designed capturing device.” 2013 International conference on biometrics (ICB). IEEE, 2013.

Unsupervised decoding of matched versus mismatched EEG responses to speech

Nicolas Heintz[*,†] Tom Francart[†] Alexander Bertrand[*]
KU Leuven
[*]Dept. of Electrical Engineering (ESAT), STADIUS
Kasteelpark Arenberg 10, B-3001 Leuven, and
[†]Dept. of Neurosciences, ExpORL
Herestraat 49 bus 721, B-3000 Leuven
nicolas.heintz@kuleuven.be

Abstract

In recent years, there has been a rapid development of new algorithms to model how the human brain responds to attended speech based on electroencephalography (EEG) signals. Such models are not only interesting to study the brain, but can also be used to create objective markers of speech intelligibility, which is in high demand in audiology, or to create novel brain-computer interfaces such as neuro-steered hearing aids [1, 2]. We designed an unsupervised algorithm that adaptively trains such models, and is ideally suited for online, wearable devices.

The exact interaction between the attended speech and the recorded EEG signals is different for each person. Subject-specific models are therefore required for optimal performance, yet they require long and impractical training sessions for each subject. Subject-independent models avoid such individualised training sessions, yet perform significantly worse [2]. Moreover, if these models are used in real-life applications such as neuro-steered hearing aids, they should ideally also adapt to changes in, e.g., the neural response to speech, the electrode characteristics, the listening environment, etc. This is possible by updating the model on the fly, using the newly recorded incoming EEG and speech signals [2]. Crucially, the model should only be updated when a person actually hears and attends to the speech, to avoid updating the model with irrelevant, noisy signals.

These models thus require a classifier that can detect when the person is actively listening to that speech, and when this is not the case. We model this case by detecting whether a given EEG segment is a response to a given speech segment, i.e., whether both are 'matched' or not [1]. However, such a classifier also needs to model the interaction between EEG signals and attended speech, and should thus also be updated adaptively, creating a chicken-and-egg problem. Inspired by [2], we designed a new, unsupervised algorithm that first correlates the recorded EEG to the speech with a canonical correlation analysis (CCA)-based model, then uses the correlations to perform the match-vs-mismatch classification and finally only updates the CCA-based model when this is the case. This iterative model quickly converges to an equilibrium operating point in which it only performs about 1% worse than the corresponding supervised model that has access to the match-vs-mismatch labels.

References

[1] A. de Cheveigné, M. Slaney, S. A. Fuglsang, and J. Hjortkjaer, "Auditory Stimulus-response Modeling with a Match-Mismatch Task," *Journal of Neural Engineering*, 2021.

[2] S. Geirnaert, T. Francart, and A. Bertrand, "Unsupervised Self-Adaptive Auditory Attention Decoding," *IEEE Journal of Biomedical and Health Informatics*, pp. 1–1, 2021.

Compressed Ultrasound Reconstruction Through Jointly Learned Analog-to-Digital Conversion and Adaptive Beamforming

Ben Luijten[1] Nir Shlezinger[2] Ariel Amar[3]
Yonina C. Eldar[3] Massimo Mischi[1] Ruud J. G. van Sloun[1,4]

[1] Eindhoven University of Technology, Eindhoven, The Netherlands
Dept. of Electrical Engineering (BM/d)
[2] Ben-Gurion University of the Negev, Beer-Sheva, Israel
Dept. of Electrical Engineering
[3] Faculty of Math and CS, Weizmann Institute of Science, Rehovot, Israel
Faculty of Math and Computer Science
w.m.b.luijten@tue.nl

Abstract

Digital ultrasound reconstruction relies on analog-to-digital conversion (ADC), mapping the continuous-time channel signals into discrete-time, and quantizing the continuous-amplitude signals in a finite number of quantization levels (or equivalently, bits). High bitrates minimize quantization distortion and maximize SNR, but also increase cost, power consumption, and the required bandwidth between probe and device. Recently, Shlezinger et al. [1] showed one can notably reduce the bitrate of analog signal acquisitions, while hardly degrading the performance, by taking into account the downstream reconstruction task. This can be achieved by jointly learning the quantization levels along with the subsequent the digital processing in and end-to-end fashion. In particular, the acquisition mapping is approximated as a differential layer operation, allowing its training along with the overall processing in the form of a deep neural network. For ultrasound imaging, the latter task can be facilitated through ABLE [2], an adaptive deep learning based beamformer. In this work, we propose joint learning of these quantization levels with adaptive beamforming, in order to significantly reduce data rates, while maintaining a high image quality. Once trained, these networks provide fast inference, and can facilitate portable and wireless ultrasound devices.

We combine the deep task-based acquisition as in [1] with the beamforming method proposed in [2]. As such, our processing chain optimizes the following within a predefined bit-budget: Linear combining of sensor channels, quantization levels of the ADC, and adaptive beamforming (i.e. the prediction of content adaptive apodization). For training, 1000 in-vivo images were recorded using the Verasonics L11-4v linear probe. As training targets, we employ Minimum Variance beamformed images. Furthermore, a separate dataset was acquired for testing purposes. We train our framework on different bit-budgets, yielding up to 16 times reduction in bandwidth requirements. As a reference, we compare against the original "uncompressed" 16-bit quantized recordings.

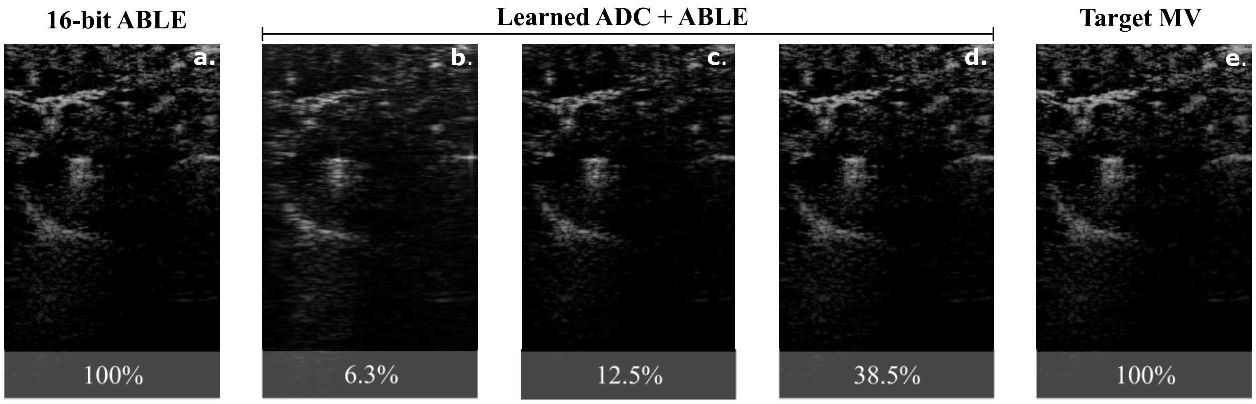


Figure 1: Results for a) Fully sampled 16-bit ABLE, b-d) deep task based ADC + ABLE at compression rates of 16x, 8x, and 2,6x, respectively, e) mimimum variance beamformed target image.

In figure 1, we show image reconstructions for the proposed method at bitrates 16, 8 and 2.6 times lower than the original recording. We show that we can reduce the bitrate by a factor of 8 without significantly compromising image quality, and maintaining high contrast reconstructions up to 16 times compression.

References

[1] N. Shlezinger et al., “Learning Task-Based Analog-to-Digital Conversion for MIMO Receivers,” ICASSP 2020, month & year.

[2] B. Luijten et al., ”Adaptive Ultrasound Beamforming Using Deep Learning,” IEEE TMI, 2020

Identification through Finger Bone Structure Biometrics

Muriel van der Spek Luuk Spreeuwers
University of Twente
Faculty of EEMCS, Group DMB
7500AE Enschede, The Netherlands
m.j.vanderspek@student.utwente.nl l.j.spreeuwers@utwente.nl

Abstract

Recent research on vascular biometrics shows that the bone structure is also contributing in the identification process. However, it is not yet documented whether bone structure can be used for identification by itself. In the NIR images of the fingerveins, the veins are dark shadow-like textures. The joints in these images are bright, where the bone itself appears darker. This research proposes a method to analyse intensity differences using two sliding windows. Brighter pixels around the joints, and darker pixels around the bones give that the difference of the sum of the pixels inside the windows give a robust representation of the rough shape of the bones. By dividing the two windows into three parts, creating six in total, the shape of the joint is taken into account more accurately. A small verification experiment resulted in an EER of 0.115 when all 6 windows have a width of 2 pixels and a gap of 10 pixels.

1 Introduction

Research in the field of biometric identification is rapidly growing due to the increasing interest in machine learning and development in neural network applications, and especially the methods using vascular biometrics is interesting. Finger vein patterns are unique and contain identity information, resulting in robust and secure identification applications [1].

The veins can be captured with near-infrared (NIR) light. An example of an image is shown in figure 1, where the green lines represent the extracted veins, and the two brighter areas are caused by two joints. The shadow-like patterns of the veins near the surface of the skin are caused by haemoglobin proteins in the blood that absorb the light [1]. Some of the green lines in figure 1 are resulting from the joints instead of the veins, indicating that these patterns also contribute to vascular biometric identification. In literature on fingervein recognition, the joints are mostly ignored, or even regarded as a nuisance. Interestingly, biometric identification on X-ray images using the bones is successfully done in [2, 3], which shows that bones do contain enough unique information for identification. Nevertheless, the bones and joints are much less visible on NIR images compared to X-ray.

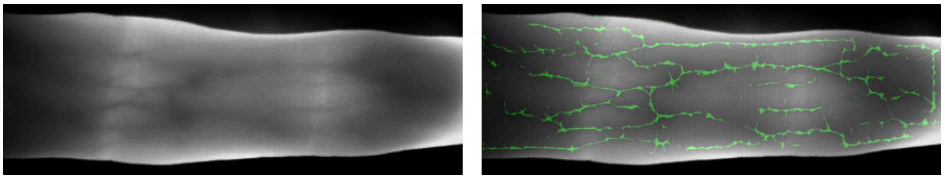


Figure 1: Example of NIR image without (left) and with veins extracted (right) [4].

It is yet unknown to which extent finger jonts and finger bone structures in NIR images can be used for identification. This paper will research whether finger bone information can be used for biometric identification with minimal contribution of the veins.

In the remainder of this paper, first relevant literature is presented in section 2. After that, a method to perform experiments is given in section 3. Then, some experiments will be discussed in section 4. The results with corresponding discussion are respectively presented in sections 5 and 6, together with some recommendations for future work. Finally, a conclusion is given in section 7.

2 Related work

2.1 Finger vein acquisition

The finger veins can be captured using a NIR capturing device, like the one available at the Univeristy of Twente. In 2012, the UTFVP dataset was recorded with this sensor [4]. For vascular biometrics, automatic non-uniform illumination settings are beneficial for sharp vein edges, but result in blurry joint shapes. Multiple ways for vein extraction are compared in [4] and an overview of vein extraction and using vascular biometrics is discussed in [1].

2.2 Pre-processing for finger vein recognition

Related work concerning pre-processing steps for vascular biometrics that can be applied for bone structure biometrics are discussed in this section, based on chapter 4.4.2 of [1]. The steps include boundary extraction, finger alignment and object enhancement.

The research in [5] presents an effective custom mask to extract the boundaries of fingers, based on the big contrast between the finger and background. The custom upper mask presented in this paper is a 4-by-20 matrix, with "1" on the first two rows, and "-1" on the second two rows. The lower mask is has the the entries with "-1" and "1" flipped.

A line in the middle of the finger, derived from the least squared error (LSE) method, determines the transformation needed for finger alignment using its ramp [6, 7]. Some form of column normalisation can be applied to cancel the finger width differences, but this removes finger geometry information.

Furthermore, vein and joint enhancement can be obtained using histogram equalisation, that cancels out illumination variations [4, 1].

2.3 Joint extraction

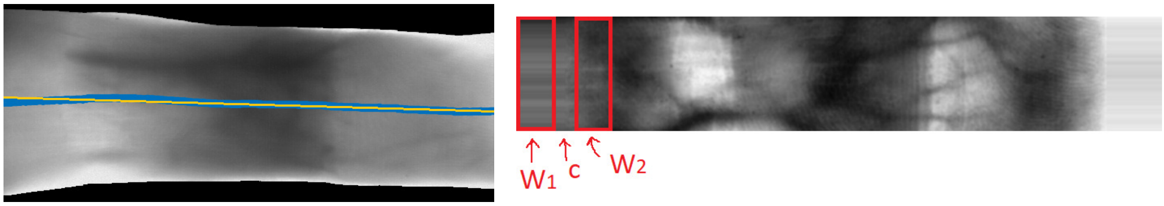
There are multiple possible ways for joint extraction. Two methods, will be discussed in this section. The most important aspect for this research is that the veins should contribute minimally to the identification. A first attempt at joint extraction is presented in [8], using the UTFVP dataset. Methods for both level sets and watershed are investigated, but the segmentation appears to be a difficult problem due to illumination variation, the images being blurry and the overlapping veins. Since most segmentation techniques are not designed to detect vague areas, this approach was not very successful.

The same study also presents a simple but promising optimal path approach, based on an applicable cost function. The Dijkstra implementation is relatively fast and makes a tree for multiple minimal cost paths in the image [9, 10]. This approach is expected to be less disturbed by to the veins and therefore suitable for joint segmentation.

Another approach, that uses the whole finger instead of only the joint segment, is based on the joint localisation algorithm presented in [11]. This research uses two sliding windows to analyse the intensity of the image, with the height of the finger and a predefined width. The local intensity is represented by the sum of the encapsulated pixels in a window. By sliding two windows over the whole image, and calculating the difference of the local intensities, a measure for the rough bone structure is obtained. Note that by taking the difference, a compensation for the intensity difference between images is made, making the algorithm more robust. Peaks result at the border of the joint, and low values occur at areas with similar summed intensity. Expected is that the veins will not influence this approach much, since they exist over the whole horizontal axis of the finger.

2.4 Conclusion

Comparing the discussed feature extraction methods give that the ones with joint segmentation will give rise to complications due to the illumination settings of the available NIR images, since it is difficult for the most segmentation techniques to detect vague areas if there are also veins that are much more visible. Nevertheless, the method using the intensity



(a) Blue gives the middle of the finger and yellow the LSE line.

(b) Image with the areas of W_1, c and W_2 selected. The first column of the image is at the centre of c.

Figure 2: Visualisation of alignment (left) and initial placement of the sliding windows (right)

differences caused by the bone structure is promising and will be carried out in this research. The detailed method and several experiments will be discussed in the next section.

3 Method

This section will briefly discuss the pre-processing steps, followed by the feature extraction method and a measure for performance evaluation and matching.

3.1 Pre-processing

Pre-processing includes extracting the boundaries and finger alignment. Boundary extraction and alignment can be performed the same as for vascular biometrics, as described in chapter 4.4.2 of [1]. Boundary extraction can effectively be done by using the custom masks from [5]. Horizontal finger alignment can be obtained using the ramp of the LSE line in the middle of the upper and lower boundary, see figure 2a.

3.2 Joint extraction

The chosen joint extraction method is based on two sliding windows based on the research in [11], from which the encapsulated pixel sum represents the local intensity. Taking the difference between the windows is a robust measure to represent the rough bone structure. The initial location of the two sliding windows is visualised in figure 2b, where the finger starts in the middle between the two windows. The sizes of the windows are defined using the following formulas. Here W_{1j} and W_{2j} are the windows, I the image, j the current row of the image, h the height of the image, w_{im} the width of the image, w the width of the window and c the gap between the two windows. The gap is defined by the number of columns between the end of the first window and beginning of the second one. Both windows have the same size.

$$\begin{aligned} W_{1j}(x,y) &= I[1:h,\ j:j+w],\ j \in [1, w_{im}-w-\frac{c}{2}] \\ W_{2j}(x,y) &= I[1:h,\ j+w+c:j+2w+c],\ j \in [1, w_{im}+\frac{c}{2}] \end{aligned} \tag{1}$$

The sums of the enclosed pixels in the sliding windows are referred to as G_{Wj1} and G_{Wj2}.

$$\begin{aligned} G_{W1j} &= \sum_{x=1}^{h}\sum_{y=j}^{j+w} W_{1j}(x,y), j \in [1, w_{im}-w-\frac{c}{2}] \\ G_{W2j} &= \sum_{x=1}^{h}\sum_{y=j+w+c}^{j+2w+c} W_{2j}(x,y), j \in [1, w_{im}+\frac{c}{2}] \end{aligned} \tag{2}$$

The difference $J(j)$ between the two results is a robust way to estimate the rough bone structure. Since the first row of the image is exactly between the two windows, the difference of the two intensity arrays will start at the first row of the image. $J(j)$ is at a maximum at the joint position.

$$J(j) = G_{W1j} - G_{W2j} \tag{3}$$

3.3 Matching and performance evaluation

The obtained intensity difference graphs can be matched using normalised cross-correlation in order to obtain comparison scores s. The genuine scores are obtained by comparing images of the same finger, and imposter scores by comparing two different fingers. These scores s are compared with a threshold th. If $s \geq th$ than the decision is positive, and if $s < th$ the result is a negative decision, thus the systems concludes the two fingers are different.

Cross-correlation is a method that uses convolution to slide one graph over another, searching for a maximum similarity overlay. The maximum overlay is the score of the cross-correlation and is a measure for how similar the two graphs are.

It is desired to minimise the equal error rate (EER), which is the threshold th for which the false match rate (FMR) and false non-match rate (FNMR) are the same, see the following equations. These values are calculated using the number of false positives (FP), false negatives (FN), true positives (TP) and true negatives (TN) at a certain threshold th. These four values respectively represent the occurrences of correct and incorrect classification. Note that the FN and TP represent the genuine, and the TN and FP the imposter scores.

$$FMR(th) = \frac{FP}{FP + TN} \qquad \text{and} \qquad FNMR(th) = 1 - \frac{TP}{TP + FN} \tag{4}$$

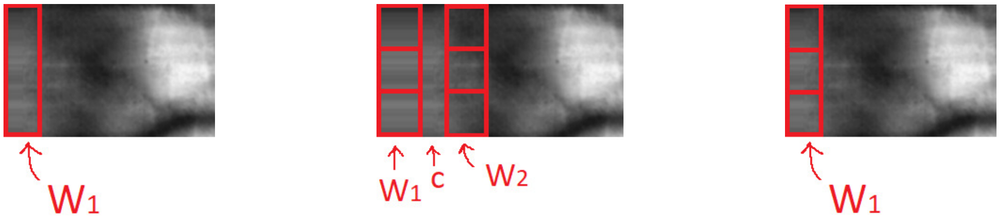
The EER can be calculated using the following equation, by searching for the threshold for which the absolute difference between the FMR and FNMR is minimal, so when these two are equal.

$$EER = \arg\min \left(|FMR(th) - FNMR(th)| \right) \tag{5}$$

4 Experiments

The research in [11] uses only one window with the height of the image. However, since the phalangeal joint is often not perfectly vertical, it is interesting to divide the height in several windows, to obtain a more accurate representation of the shape of the joint. However, the smaller these windows, the more influence the veins will have. The following experiments will be carried out, where the influence on the EER will be evaluated.

1. Varying w using the intensity difference between two sliding windows obtained from equations 1, 2 and 3.
2. Varying w using only one window, where the intensity is obtained with only G_{W1j} in equation 2.
3. Varying w, dividing the two windows of experiment 1 into three parts each. The difference between the top, middle and bottom areas are treated separately and result in three separate intensity difference plots.
4. Varying w, dividing the window of experiment 2 into three equal parts. Intensity is calculated per area by adjusting the window height h of G_{W1j} in equation 2.
5. Varying distance between windows, starting from $c = -w + 2$ with arbitrary w to let the second window start two columns after the first one causing them to overlap for the first few measurements.



(a) Experiment 2.

(b) Experiment 3.

(c) Experiment 4.

Figure 3: Visualisation of experiments 2, 3 and 4. Experiment 1 can be seen in figure 2b.

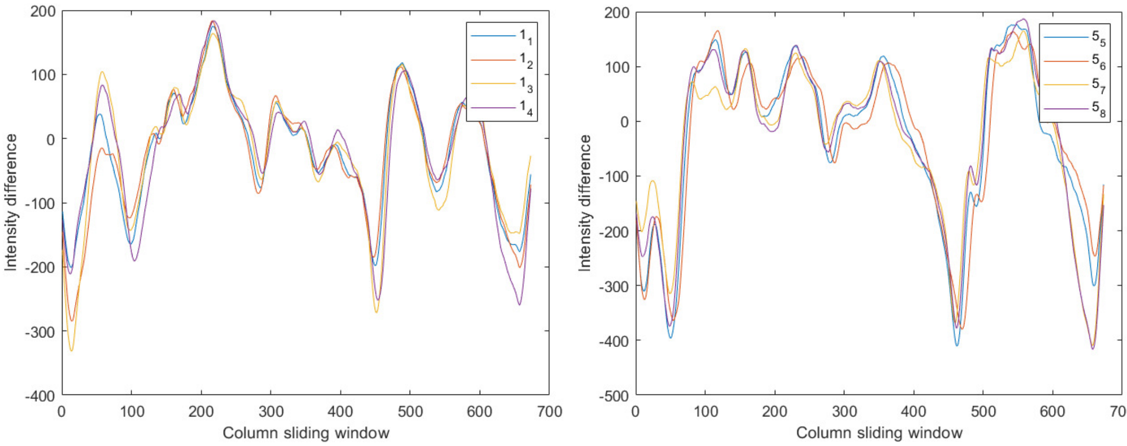
Figure 3 and figure 2b visualise the initial placements of the windows for experiments 1 to 4.

The area between the windows c is advised to be 3 to 6 pixels in [11]. Since the width of the images in the used UTFVP dataset is twice the size of the images in [11], c is chosen to be 10 for the experiments where the distance between the windows is not varying, experiment 1 and 3. For experiment 1 to 4, w will be ranging from 2 to 150. The last experiment will use $w = 30$ with a varying distance from -w+2 to 150.

5 Results

5.1 Intensity difference plot

An intensity difference plot of two randomly chosen subjects for experiment 1 is shown in figure 4, with subject 1 on the left and subject 5 on the right, with each four images. Both the window size w and gap c are 10 pixels. The plot shows a clear resemblance between the images of the same subject. All peaks per subject have approximately the same amplitude and are located on the same columns. The plots of the two subjects on the other hand differ considerably.



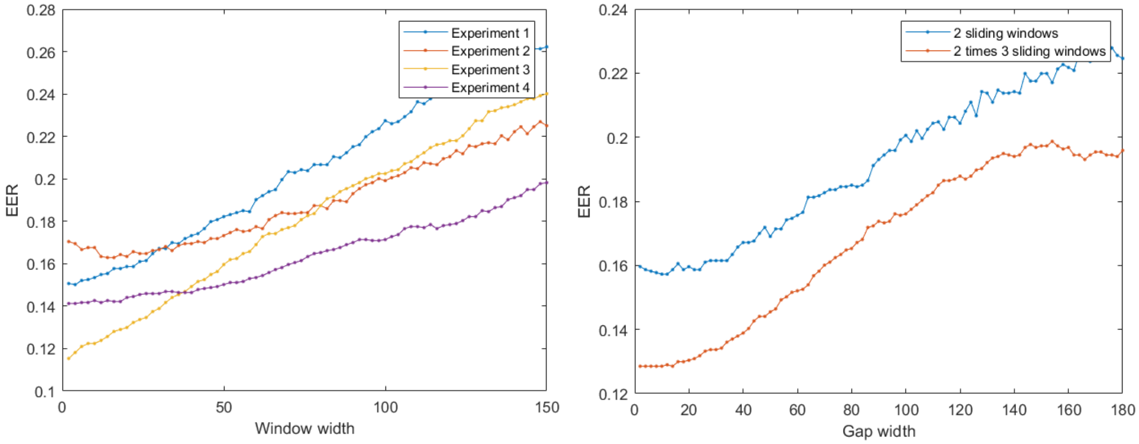
(a) Subject 1.

(b) Subject 5.

Figure 4: Intensity difference plots for subject 1 (left) and 5 (right), with both 4 images.

5.2 Influence of the window width

The results of experiment 1, 2, 3 and 4 are visualised in figure 5a, where the varying window size is plotted against the EER. The EER is calculated using the obtained genuine and imposter scores at a particular window size, and calculating the FMR and FNMR for a range of threshold values. All experiments result in a roughly linear behaviour with the EER, increasing the window size increases the EER. However, experiment 2 and 4 are more flat and perform worse with smaller windows, but outperform the other two experiments after a width of around 50 pixels. The lowest EER of 0.115 is obtained using experiment 3 at a window size of 2 pixels.



(a) EER in multiple situations for varying window size w for experiment 1 to 4.

(b) EER for varying gap size using the window-placement in figure 2b and 3b.

Figure 5: Results of experiment 1-4 (left) and experiment 5 (right)

5.3 Influence of the gap

The results of experiment 5 are shown in figure 5b. The two initial window placements are the ones shown in figure 2b and 3b. For this experiment $w = 30$, which gives a minimal EER of 0.139 for experiment 3 and 0.167 for experiment 1 with a gap of 10 columns. Note that the windows overlap for the first 13 measurements.

The middle part behaves roughly linear with the EER, and the first and last parts are more constant, especially the orange graph with three windows. With gaps of > 150 pixels, the EER even seems to decrease. This shows that a small gap is optimal for the EER, where the EER remains constant regardless of the gap size. The EER increases after a gap around 30 pixels, and for the situation with three windows the gap at > 150 pixels does not seem to influence the EER negatively. However, this graph does not show what happens after a gap of 180 pixels.

6 Discussion

6.1 Intensity plots

Inspecting figure 4 shows possible weaknesses of the algorithm. The seventh image of the fifth subject (red line for 5_7) in figure 4b appears slightly different from the other three lines, especially around column 100. This is probably caused by local illumination differences of the finger, see figure 6. This is arguably an artefact from the non-uniform automatic illumination settings of the NIR capturing device. These kinds of artefacts make the identification less accurate.

Furthermore, looking at the last 100 columns for both subjects shows that the lines corresponding to the same subject become less similar. Due to some fingers being shorter or smaller at the tip, light is able to travel along the finger causing overexposure, also seen in figure 2b. Apparently, this overexposure causes these images to have similar intensities at the end of the plot, and therefore reducing the identification accuracy.

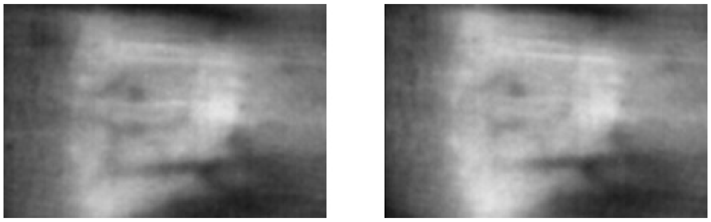


Figure 6: Difference of image 5_7 and 5_8 around the peak around column 100. The right joint is slightly brighter than the left one.

An improvement of the discussed problems could be to make a custom dataset with uniform illumination settings, and ignoring the top part of the finger. However, the latter is not optimal since it also removes information.

6.2 Influence of the window width

Figure 5a implies that all tested windows result in a system that can be used for identification, however smaller windows result in better applications. Using smaller windows, both bone structure details and vein information is included. With larger windows, the details of the bone structure and joint shape are lost, leaving only an expression for the rough bone structure. The results imply that the detailed information contains more unique information than the rough bone structure. This could be a result from the vein contribution.

Another interesting result is that the EER obtained from higher window sizes for experiment 2 and 4 outperforms the other two experiments, respectively $w > 30$ and $w > 47$. This is possibly caused by the addition to take the difference between the sliding windows. Bigger windows result in a rough representation of the bone structure, and taking the difference between the windows will make the results less deviate between subjects, resulting in similar intensity difference plots.

6.3 Influence of the distance between windows

The result in figure 5b shows that a small distance between the windows improves the results until a certain distance for this chosen window width, since it has a lower EER for both cases compared to the same window width in experiment 1 and 3. The EERs 0.139 and 0.167 for experiment 1 and 3 are respectively reduced to 0.128 at a distance of 10 pixels and 0.157 at a distance of 12 pixels. Thus, the optimal EER is obtained when the windows slightly overlap, which removes the intersection and the outer parts remain, resulting in two smaller windows. For the experiment with three windows, at the distance of 10 pixels, effectively the original window size w of 30 pixels is reduced to 10 pixels with a gap c of 20 pixels.

Reducing the distance between smaller windows, tested until $w = 8$, does also reduce the EER further. However, decreasing the gap c at optimal window size $w = 2$ according to the results of experiment 1 and 3 does not improve the EER any further.

7 Conclusion

Besides the shadow-like patterns of the veins, the near-infrared (NIR) images used for vascular identification also contain bright areas for the joints, and darker areas for bone tissue. This research proposes a successful method to evaluate whether the rough bone structure in human fingers can be used for identification systems, using the NIR images used for vascular biometrics. The method is based on two sliding windows with a gap, where the sum of the encapsulated pixels roughly represents the local intensity. The difference between the windows results in a robust intensity plot that accounts for the intensity differences between images. Four experiments were carried out that related the obtained EER with the window width.

The first is carried out using the two sliding windows. The second experiment is similar to the first one, but the two sliding windows are divided into three equal parts, resulting in three intensity difference plots. The last two experiments are used to evaluate the influence of the difference of the windows and contain only one window, thus one window in total or one divided into three parts. An additional experiment covers the relation between the gap between the windows and the EER.

The experiment where the two sliding windows are each divided into three parts proved to be very effective and managed to get an EER of 0.115 at a window size of 2 and a gap between the windows of 10 pixels by doing a small verification experiment. Decreasing the gap width also improves the result further for a larger window (30 columns). Interestingly, reducing the gap for the optimal window width of 2 with a gap of 10 columns does not improve the EER further. These results give that finger bone structure does contain useful identity information.

References

[1] Andreas Uhl, Christoph Busch, Sébastien Marcel, and Raymond Veldhuis. *"Handbook of vascular biometrics"*. Springer Nature, 2020.

[2] Yeihya Kabbara, Amine Nait-Ali, Ahmad Shahin, and Mohamad Khalil. "sc". In *2015 International Conference on Image Processing Theory, Tools and Applications (IPTA)*, pages 423–428. IEEE, 2015.

[3] Yeihya Kabbara, Ahmad Shahin, Amine Nait-Ali, and Mohamad Khalil. "An automatic algorithm for human identification using hand X-ray images". In *2013 2nd International Conference on Advances in Biomedical Engineering*, pages 167–170. IEEE, 2013.

[4] B Ton. "Vascular pattern of the finger: Biometric of the future?" *University of Twente*, 3, 2012.

[5] Eui Chul Lee, Hyeon Chang Lee, and Kang Ryoung Park. "Finger vein recognition using minutia-based alignment and local binary pattern-based feature extraction". *International Journal of Imaging Systems and Technology*, 19(3):179–186, 2009.

[6] Ramachandra Raghavendra, Kiran B Raja, Jayachander Surbiryala, and Christoph Busch. "A low-cost multimodal biometric sensor to capture finger vein and fingerprint". In *IEEE International joint conference on biometrics*, pages 1–7. IEEE, 2014.

[7] Beining Huang, Yanggang Dai, Rongfeng Li, Darun Tang, and Wenxin Li. "Finger-vein authentication based on wide line detector and pattern normalization". In *2010 20th international conference on pattern recognition*, pages 1269–1272. IEEE, 2010.

[8] ECM Plas. "A study of the shape of the phalangeal joint in finger vein images". B.S. thesis, University of Twente, 2020.

[9] Eric N Mortensen and William A Barrett. "Interactive segmentation with intelligent scissors". *Graphical models and image processing*, 60(5):349–384, 1998.

[10] Edsger W Dijkstra et al. "A note on two problems in connexion with graphs". *Numerische mathematik*, 1(1):269–271, 1959.

[11] Shirong Qiu, Yaqin Liu, Yujia Zhou, Jing Huang, and Yixiao Nie. "Finger-vein recognition based on dual-sliding window localization and pseudo-elliptical transformer". *Expert Systems with Applications*, 64:618–632, 2016.

Energy Dispersion Index for Finite-Length Probabilistic Shaping by 4D Signaling

Kaiquan Wu Gabriele Liga Yunus Can Gültekin Alex Alvarado
Eindhoven University of Technology
Department of Electrical Engineering, Signal Processing Systems Group
Eindhoven 5600 MB
k.wu@tue.nl g.liga@tue.nl y.c.g.gultekin@tue.nl a.alvarado@tue.nl

1 Abstract

Probabilistic shaping (PS) is a key enabler to realize near capacity-achieving transmission for the additive white Gaussian noise (AWGN) channel. In the context of fiber optical communications, the shaping gains provided by AWGN-optimal PS are undermined by the nonlinear interference (NLI), as fiber Kerr nonlinearities are enhanced by the Maxwell–Boltzmann distribution of amplitudes. To combat this NLI penalty, a straightforward approach is simply using short shaping blocklengths [1], which induces less NLI thanks to highly-correlated transmitted symbols. The effects of shaping blocklength on the NLI were investigated in [2]. A precise metric that quantifies such effects was missing, until recently we proposed energy dispersion index (EDI) [3]. The numerical results of coherent optical communications in [3] showed a strong association between EDI and effective signal-to-noise ratio (SNR). However, only 2D symbols using single polarization were considered in [3] for the sake of simplicity. By contrast, dual-polarization offers four carriers and, hence, naturally leads to 4D symbols. Besides shaping blocklength, mapping strategy also plays an important role in the temporal structure of 4D symbols and subsequently affects the NLI [4].

This work is a follow-up study of [3]. The performance of EDI in predicting effective SNR of constant-composition distribution matcher (CCDM) shaped 4D symbols is evaluated at different distances, as shown in Fig. 1. Generation of 4D symbols is realized at various shaping blocklengths and with different symbol mapping strategies. We first generalize the EDI definition to 4D symbols. Then we analyze the EDI of 4D symbol sequences and also provide the corresponding analytical expressions for the EDI. The numerical results show that EDI performs well, in particular for the single-span links.

References

[1] T. Fehenberger, H. Griesser, and J.-P. Elbers, "Mitigating fiber nonlinearities by short-length probabilistic shaping," in *Proc. Opt. Fiber Commun. Conf.* San Diego, CA, USA, Mar. 2020, Paper Th1I.2.

[2] T. Fehenberger, D. S. Millar, T. Koike-Akino, K. Kojima, K. Parsons, and H. Griesser, "Analysis of nonlinear fiber interactions for finite-length constant-composition sequences," *J. Lightw. Technol.*, vol. 38, no. 2, pp. 457–465, Jan. 2020.

[3] K. Wu, G. Liga, A. Sheikh, F. M. J. Willems, and A. Alvarado, "Temporal energy analysis of symbol sequences for fiber nonlinear interference modelling via energy dispersion index," Feb. 2021. [Online] Available: https://arxiv.org/abs/2102.12411v1.

[4] P. Skvortcov, I. D. Phillips, W. Forysiak, T. Koike-Akino, K. Kojima, K. Parsons, and D. Millar, "Huffman-coded sphere shaping for extended-reach single-span links," *IEEE J. Sel. Top. Quantum Electron.*, Jan. 2021, (early access).

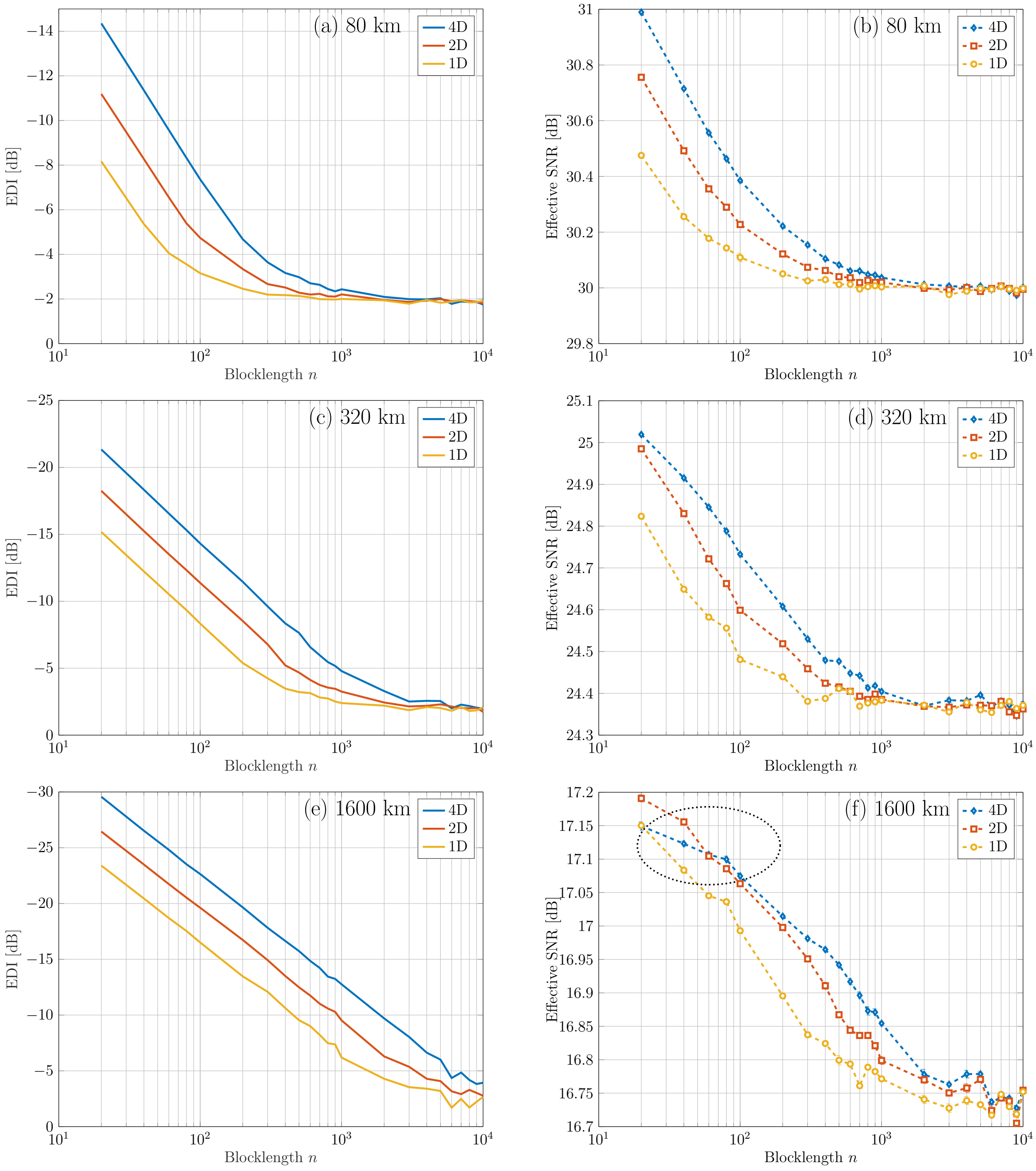


Figure 1: EDI (left column) and effective SNR (right column) vs. blocklength n after CCDM-shaped 64QAM WDM fiber transmission at various distances. The simulation setup is the same as in [3], except that dual-polarization is used here. The EDI is shown in dB and inverted for convenience of comparison. 1D, 2D and 4D (see [4, Fig. 3]) symbol mapping strategies are compared. Window lengths of 30, 150 and 1000 are used for the EDI calculation at distances of 80 km, 320 km and 1600 km, respectively. The dashed circle in (f) indicates the discrepancy between EDI and effective SNR for 4D symbol mapping, which could be due to the coupling effect between the two polarization components.

Progressing Insights in Signal Processing for Optical Wireless Communications

Jean-Paul M. G. Linnartz[1,2] and Xiong Deng[1]
[1]Eindhoven University of Technology (TU/e)
Department of Electrical Engineering, Signal Processing Systems (SPS) group
Flux Building, 5600 MB, Eindhoven, The Netherlands
[2]Signify, 5656 AE, Eindhoven, The Netherlands

Abstract

Significant progress has been made in optical wireless communication (OWC) for indoor applications. Scientific literature in this field shows an appetite for experimental results that contribute to a speed race. Compared to the tradition in radio communication, the (re-) building of a new theoretical basis is less emphasized. A number of practices, properties and optimizations have often adopted from radio communication, while these may not always be fully appropriate for OWC. We argue that it can be helpful to revisit some foundations and to verify and refine the models to better reflect the specificities of OWC, particularly using light emitting diodes (LEDs). This paper reviews some popular statements and comments on these, using recent insights.

1 Introduction

Optical wireless communication (OWC) borrows many principles that were originally developed for radio communication. However, the translation of these concepts to OWC requires attention for differences in channel properties, or in transmitter and receiver behaviors. There have been a number of models for the light emitting diode (LED) channel that zoom in on different aspects of the intensity-modulated with direct detection (IM/DD) of positive real-valued signals over a non-linear low–pass channel. Recently, also the photonic principles of photon generation have received more attention as a renewed basis for a more realistic channel model that better reflects modern LEDs [1–4].

This paper comments on a number of properties attributed to the use of orthogonal frequency division multiplexing (OFDM) for OWC over LED channels. It reviews these based on detailed results and insights gathered in recent publications of Signify and the Signal Processing Systems (SPS) group at Eindhoven University of Technology (TU/e). References to these papers are given for more details and further substantiation of the insights. The focus is on throughput that can be reached by OFDM or PAM modulation, on how these can be modelled using properties of the LED.

2 LED bandwidth limitations

An LED, particularly a high–power one has a large junction capacitance. This, in combination with the dynamic resistance of the LED, leads to a first–order low-pass behavior [5]. One of the main reasons to use OFDM over an LED channel is to optimize the modulation (bit/s/Hz) per frequency bin, far beyond this turn-over frequency. Pre-emphasis is said to extend the LED bandwidth by a frequency–response correction. However, that may come with a severe power penalty.

If the objective is to reach higher bit rates regardless of power consumption, the combination of pulse-amplitude modulation (PAM) and pre–emphasis can work [4]. However, in practical cases, the signal to noise penalty may not outweigh the increased bandwidth. If the objective is, for a given constraint of the variance of the input current to reach the highest throughput, pre–emphasis is not recommended [6]. An elaborate comparison of pre-emphasis versus optimized (waterfilling) usage of the LED bandwidth in [6] concludes that pre-emphasize can be reasonable only if the cut-off frequency is made adaptive to carefully match with the link budget. Pre-emphasis over a fixed bandwidth is very intolerant to a degrading signal-to-noise ratio (SNR) over the link and gives a much more dramatic cut-off if the receivers move out of range where a degrading SNR can be repaired by adapting the constellation. Also if the non-negativity of a light intensity and the mean current (or, equivalently, the optical power) are constrained, the ability to adaptively load the various frequency bins, as allowed by OFDM, seems attractive, instead of rigorously trying to invert the channel response [6].

3 Throughput, bandwidth and power

In first–order approximation, electrical *current* is converted into optical *power* (photon flows). However, the statement that electrical power, typically calculated as the current–squared, is also the square of the optical power needs to be used with care. At the receiving photodiode, photons are converted to currents. Evidently, after the trans-impedance amplifier (TIA), power relates to current–squared ($P_T \approx \mathrm{E}[i^2]$). However, the electrical power consumed in an emitting diode is not proportional to current-squared, but is governed by the Shockley equation [7]. The expectation (E) value is, using commonly used values n, k, T, I_0,

$$P_T = \frac{nkT}{q} \mathrm{E} \left\{ I \ln \left(\frac{I - I_0}{I_0} \right) \right\}. \tag{1}$$

This expression can be used provided that the signal is within the 3 dB bandwidth of the LED, to avoid extra reactive power losses due to the junction capacitance. This P_T can be approximated by a fixed junction voltage V_j multiplied by a varying current, thus on average $P_T \approx V_j \mathrm{E}[i]$. For thin modulation where the approximation by a fixed voltage is most accurate, this power is proportional to the *average* current [8, 9], not to the current–squared. Nonetheless, we see a majority of papers model electrical power as $\mathrm{E}[I^2]$.

If we challenge $P \approx \mathrm{E}[i^2]$ also the concept of energy per bit of $E_b = R\mathrm{E}[\int i^2 dt]$ must be abandoned, either because the non-linear LED does not act as a resister or because the power consumption actually is another function than a squaring i^2. Some intuitions and interpretations of signal dimensions in a Euclidean space loose accuracy. As we showed by specific examples in [8], the received energy per bit and emitted energy per bit no longer have a linear relation. Applying half the power over double the time interval does not retain equal received energy per bit.

3.1 Throughput equations using bandwidth and SNR

The Shannon expression, derived for a power-limited linear time-invariant (LTI) channel with additive white Gaussian noise (AWGN) is often (mis-) used as a "capacity" for an OWC system. In fact, the use of an expression like "bandwidth B_w times the 2-logarithm of one plus the signal–to–noise ratio", thus

$$R = \alpha B_w \log_2 \left(1 + \frac{SNR}{\Gamma} \right) \tag{2}$$

is challenged in some papers and reconfirmed in other papers. We inserted α as a dimensionality penalty and Γ as a power penalty, to allow broader use. This expression is in fact an "achieved throughput". It can be seen as a "conditional capacity" subject to the limitation of using a particular subclass of modulation methods, e.g. DC-biased optical (DCO-) OFDM. It can be derived just by inverting the bit error rate (BER) Q–function [6]. In fact, one can specify a symbol rate r_s and a constellation M that carries m bits ($M = 2^m$) per dimension. Or, equivalently, M^2-quadrature amplitude modulation (QAM) per subcarrier. These specify the bite rate R and B_w. Demanding adequate BER sets the required SNR. This derivation of (2) [6] yields the same "BW times the log of one plus SNR" expressions, where Γ then is a function of the required uncoded BER. This suggests that it is acceptable to use Shannon-like expressions to derive DCO-OFDM throughput. However, the word "capacity" would incorrectly suggest that it is also an upper limit to what can be achieved.

3.2 The LED channel is "power" limited

A common objection from the Information Theory community against the use of the above expression is that it does not reflect the non-negativity of light intensities. It seems inappropriate to use part of the power budget as a DC bias and assume that the system is efficient. In fact, capacity–achieving distributions are known for non–negative frequency–flat channels, and these differ from the shifted Gaussian of DCO-OFDM. Nonetheless, [10,11] showed that optimized OFDM can performs well over a non-negative channel. Moreover, known non-negative capacity–achieving distributions do not address the low–pass nature of the LED. So, these are only usable to build a communication system that works within the 3-dB bandwidth of the channel. Here, DCO-OFDM can exploit a much larger bandwidth, e.g. quantified by [6]. For the majority of LED based systems, the low-pass effect seems a more restricting throughput limitation than the power lost in using non-optimum distributions. The more rewarding approach is to push beyond the 3 dB bandwidth. Optimized power loading over this gentle (first–order) roll-off, say according to waterfilling, relies on a multi-frequency modulation method, such as variants of OFDM. By virtue of the central limit theorem, this unavoidably leads to a Gaussian signal. In other words, a choice to addresses frequency–selectivity via OFDM may restrict the freedom to optimize signal distribution. For a Gaussian, the variance, thus the power of the modulation is constrained [6]. Thus, the use of a variance (power) constraint is appropriate for LED channels.

4 Orthogonal Frequency Division Multiplexing

OFDM is said to be have a superior spectrum efficiency. OFDM can densely pack subcarriers and thereby it is more spectrum efficient than FDM with guard bands. However, OFDM is not necessarily more efficient than PAM. The casual use of the claim "OFDM is used because it is spectrum efficient" may have to be read with caution: The number of orthogonal signal dimensions available in a certain bandwidth B_w for a duration T_f equals $2B_wT_f$. A rotation of the signal space can fundamentally not increase the number of dimensions. Equivalently, seeing the fast Fourier transform (FFT) used in OFDM as an invertible unitary matrix rotation of dimensions, OFDM is equally spectrum efficient as PAM. Used in optical intensity modulation, thus in baseband settings it converts $N/2$ complex values QAM signals into N real–valued signals, thereby maintaining exactly the same number of signal dimensions. If incoming signals arrive at a rate of r_s real–valued symbols per second, basedband PAM requires a Nyquist bandwidth of at least $r_s/2$. Alternatively, complex QAM symbols can be constructed at a rate of $r_s/2$. $N/2$ such samples form an OFDM frame, that results in N real samples that need to be clocked at rate r_s. According to Nyquist, this requires a bandwidth of $r_s/2$ to carry OFDM frames of N samples at a rate of r_s/N frames per second. In other words, spectrum efficiency is preserved, and OFDM cannot break through the Nyquist limit of number of orthogonal dimension that are available per second in a given bandwidth [12].

4.1 Optical Intensity Modulated OFDM

Baseband optical signals are real-valued and cannot be imaginary. Therefore, usually the imaginary part of the output of the inverse fast Fourier transform (IFFT) of an optical OFDM system is forced to zero by applying an Hermetian symmetry at the IFFT input. This reduces the number of complex valued independent QAM symbols by one half. Yet, this is not a loss of spectrum efficiency, as argued above.

In fact, radio signals are also real. It has nonetheless been proposed to discount the Shannon expression, not only with an appropriate Γ but also by $\alpha = 1/2$. The I and Q modulation in carrier–based OFDM, with complex QAM symbols may trigger some confusion here. While a real signal of bandwidth B_w can be conveyed over a baseband channel of width B_w, amplitude modulation (AM) on a carrier doubles the bandwidth to $2B_w$, but leaves room for quadrature modulation of a second (quote unquote "imaginary") signal. Carrier–based RF carries double the rate in double bandwidth. Thus, $\alpha = 1$ both for OWC and RF, if we consider the bandwidth in an end-to-end OWC channel from LED input to photodiode output. This agrees with the fact that Shannon developed his theory for strictly real–valued channels.

Possibly, the debate of the factor one half ($\alpha = 1/2$) originates from system evaluations that use a laser with optical spectrum $S_{source}(f) = \delta(f - c/\lambda)$. One compares coherent modulation of the laser-based optical carrier both in inphase and in quadrature phase, versus laser–based intensity modulation, thus amplitude modulation (AM) with double side bands:

$$S_{optical}(f) = A_c S_{source}(f) + \sum_{\pm} \frac{A_m}{2} S_{mod}\left(\frac{c}{\lambda} \pm f\right) \tag{3}$$

where ± is shorthand for the addition of an upper and lower side band. It occupies in the optical domain twice the bandwidth of the modulation ($BW_{optical} = 2BW_{modulation}$), and the use of $\alpha = \frac{1}{2}$ could be justified if one inserts $B_w = BW_{optical}$. However for LEDs, and certainly for VLC, the

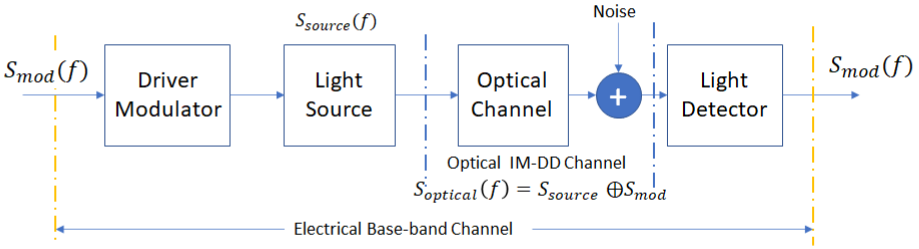


Figure 1: In VLC and LED-based OWC, the electrical channel acts as a baseband channel. For perfectly narrow sources (e.g. ideal lasers), the bandwidth in the optical domain is twice the bandwidth of the modulation.

optical emission spectrum $S_{optical}(f) = S_{source} \otimes S_{mod} \approx S_{source}(f)$ is anyhow very many orders of magnitude larger than the modulation, and the $\alpha = \frac{1}{2}$ becomes meaningless.

In a comparison between (equalized) baseband PAM (or on-off keying (OOK)) and OFDM over the electrical end-to-end channel, there is no justification for the $\frac{1}{2}$ penalty. In fact, one would also not apply a factor $\frac{1}{2}$ correction for a twisted pair telephone, which also is a base-band channel. If $\alpha = 1/2$ is used for laser-based IM-DD viewed in terms of the optical spectrum occupancy, the double-side bands occur both for OFDM and PAM, although in literature it is sometimes only applied for OFDM, and said to be an artefact of the Hermetian symmetry. In either case, the optical channel could accommodate an second orthogonal data stream on the quadrature phase of the laser.

4.2 Hermetian Symmetry

In contrast to common belief, there is no need to explicitly impose an Hermitian symmetry at the inverse fast Fourier tramsform (iFFT) input, to transmit a real-valued OFDM over a baseband OWC channel. Zero-padding of the higher subcarriers suffices, while the resulting imaginary output is simply discarded [13],

$$z_k = DC + \sum_{n=1}^{N/2-1} X_n e^{\frac{j2\pi nk}{N}} + \overline{X_n} e^{\frac{j2\pi(N-n)k}{N}}. \tag{4}$$

Taking the real part of the first term already fully describes the transmission data. The second term is created by the Hermetian-symmetric input to cancel the imaginary output. The $\exp(\pm j)$ rotation of the second term is the complex conjugate of that of the first term. However, as the imaginary outcome can be truncated anyhow, zero-padding (thus eliminating the second term) works equally well without imposing Hermetian symmetry. Alternatively, for reasons of processing efficiency, one may multiplex (add) an Hermetian symmetric input and an Hermetian anti-symmetric input into the FFT and use the respectively resulting real and purely imaginary outputs sequentially in separate frames. This can lead to less compute-intense solutions [13].

5 Key Performance Indicators

In RF communications, it is an appropriate common practice to benchmark different system solutions for the same SNR. However, for communication over an LED, the use of SNR is ambiguous. The choice of the communication bandwidth is not dictated by regulatory constraints in spectrum usage, but it is an engineering trade–off. The signal structure is usually chosen to go beyond the 3 dB bandwidth of the LED, subject to some strategy. Waterfilling theory shows it is often preferred to go more than an order of magnitude beyond the 3 dB bandwidth.

It may be more useful to benchmark systems for their achieved performance within a given budget $Q = P_T H_0^2 / N_0$. Here H_0 is the path loss, N_0 the noise spectral density, P_T is transmit power, so Q has the unit sec^{-1} (rate). To create a dimensionless parameter and to achieve generic capacity results in closed form in [6], we used $\gamma = P_T H_0^2 / (N_0 f_{LED})$, which is normalised to the bandwidth f_{LED} of the LED rather than to the signal bandwidth. An illustrative example of the effectiveness of the above normalisation is the capacity of a first–order low pass channel [5, 6]. For the given power budget, waterfilling arguments reveal that the maximum (or optimum) frequency $f_{\max}$ in the modulation bandwidth which has nonzero power level adheres to

$$f_{\max} = \sqrt[3]{\frac{3\gamma}{2}} f_{LED}. \tag{5}$$

A unique relation between the normalized maximum throughput R_w (for waterfilling) and the power budget γ was derived as [5]

$$\frac{R_w}{f_{LED}} = 2\sqrt[3]{\frac{3\gamma}{2}} - 2\tan^{-1}\sqrt[3]{\frac{3\gamma}{2}}. \tag{6}$$

The normalization strategy used in (5) and (6) makes the expressions generic, hence independent of any specific LED channel property other than f_{LED}. In the limit for emitters with large bandwidth relative to the available power needed to use that bandwidth effectively, thus with $\gamma \to 0$, R_w becomes a constant times Q, regardless of f_{LED}. Power (or equivalently photon flux) is what then matters.

In communication over optical fiber, the *bit rate multiplied by the distance* is limited by a fundamental mechanism, namely that the dispersion caused by the fiber increase with distance in a particular manner. Although in wireless indoor communication this effect is absent, we see that also OWC systems are sometimes bench-marked for reaching the highest bitrate-distance (m Gbit/s) performance. The use of optics to narrow down the optical bandwidth is included as a tool to enhance the bit rate-distance, thereby obscuring whether progress is made in terms of modulation processing and modulator technology. We know that increasing the current density in LEDs, e.g. by going to micro-LEDs, increases the bandwidth, but decreases the power efficiency [4]. In a "bit rate distance product race", power efficiency can be compensated by focusing optics, which reduces this effort to one of increasing current density and proper thermal handling. The winning strategy will be one of increasing LED bandwidth, jeopardizing coverage.

We argue that a more appropriate performance indicator would be the product: *bit rate times covered area*, thus to consider the area in which the signal is picked up without re-orientating the transmitter and receiver. In fact, if we progress more towards laser–based OWC in which bandwidth is not constrained, we may take the limit of $BW \to \infty$ of

$$B_w \log_2\left(1 + \frac{P_T H_0^2}{\Gamma N_0 B_W}\right) \xrightarrow{B_W \to \infty} \frac{P_T H_0^2}{\Gamma N_0} = c_\gamma H_0^2 Q \tag{7}$$

with some proportionality constant c_γ. Thus, rate is proportional to Q, i.e., power over the noise-spectral density. In other words, an OWC system is preferably bench–marked against photons per bit. The aspect of coverage then automatically comes into the equation, via the path loss H_0. The latter is preferably expressed as the ratio of the received number of photons over the emitted number of photons integrated over all directions. Thus, H_0 can be expressed as the detector size divided by the coverage area.

6 Non-negativity and frequency selectivity

Asymmetrically clipped optical OFDM (ACO-OFDM) and Flip-OFDM have been designed to eliminate the need for a DC bias. These schemes create redundancy by ensuring that every OFDM sample occurs twice, once with positive and once, in a second set of time samples, with negative polarity [13]. This operation reduces the throughput by 50%, but it allows removal of all negative samples by clipping these to zero. The claim is that this obviates the DC bias.

A derivation repeated in several papers confirms that subcarriers remain orthogonal despite the clipping. However, a non-selective channel was modeled. Secondly, the receiver considered is in fact a correlation with a reference copy of the non-clipped signal, which is executed as an implicit correlation with FFT tones. We showed in [13] that this operation is blind to 50% of the data symbol energy and thereby looses 3 dB of SNR. The other 50% falls outside the signal space seen by the detector and is hard to disentangle. In other words, orthogonality of the signal waveform is not preserved, but an orthogonal fraction of one half of the power (- 3 dB) can be detected by a simple FFT receiver.

Only half of the symbol energy remains located on the subcarrier frequencies [13]. This portion of the signal is subject to one subcarrier frequency channel response. The other 50% is smeared across the band and is affected by many other channel transfer values. Yet, these are discarded both for the test subcarrier and for other subcarriers by an FFT detector. Crosstalk falls at even subcarriers in ACO-OFDM and exactly in between subcarriers for Flip-OFDM and wastes as much as 50% of the signal power. This partly defeats the gain of omitting the DC-bias [7].

An analysis in [13] was inspired by curiosity of the authors that it was not obvious to them that orthogonality is maintained if randomly delayed copies of the clipped signal are added in. In fact, calculations show that this orthogonality can be shown to hold for a (3 dB sub-optimum) FFT receiver. However, in another variant, namely PAM-OFDM [14], the copied part is not only flipped but also reversed in time. After folding back during detection, samples of the delayed copies randomly appear pre and post cursor. This changes the frequency response of the channel and affects orthogonality.

Several simulations revealed that Flip-OFDM and ACO-OFDM have an almost similar performance, even though these are said to be generated by different signal processing recipes. In fact, we showed in [13] that both systems share an identical mathematical concept and can both be generated by calculating a first OFDM block and copy this into a second block that is flipped in polarity. So their equivalence (except for a frequency up-shift by half a subcarrier spacing) can also be proven mathematically. An intuitive way to see this is to write the ACO signal generation as one using an FFT of size $2N$. ACO-OFDM maps the n-th complex–valued QAM symbol to (odd) subcarrier $2n+1$ of the $2N$-IFFT. For the $2N$ IFFT, ACO-OFDM creates

$$z_k = \sum_{n=1}^{N/2-1} X_n e^{\frac{j2\pi(2n+1)k}{2N}} + \overline{X_n} e^{\frac{j2\pi(2N-2n-1)k}{2N}}, \tag{8}$$

for $k = 0, 1, ..., 2N-1$. ACO-OFDM, then transmits z_k^+, thus clips away any negative signal parts, with $z_k^+ = \max(0, z_k)$. ACO-OFDM is known to satisfy the symmetry property (before clipping) $z_{N+k} = -z_k$. In fact, we can re-write ACO-OFDM in as

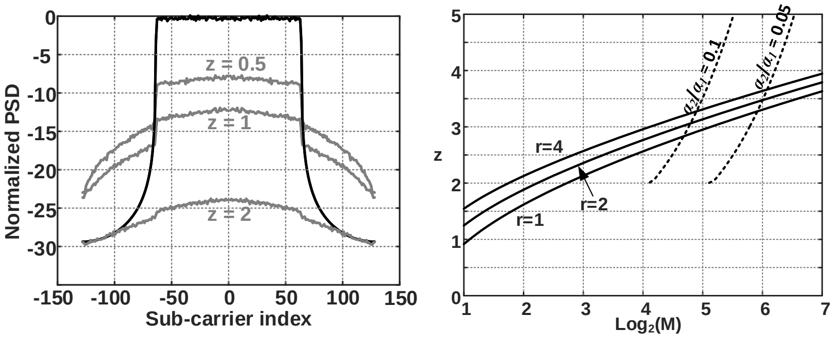
$$z_k = \mathrm{Re}\left[2e^{\frac{j\pi k}{N}} \mathrm{FFT_N}(X_n)\right],$$

for $k = 0$, 1,..., $N-1$, and zero-padded inputs. For $k = N, N+1, ..., 2N-1$, the FFT output samples $k-N$ are used, while the rotation $e^{\frac{j\pi k}{N}} = -e^{\frac{j\pi(k-N)}{N}}$ proves that the second block is a copied - flipped version of the first one. Interestingly, ACO-OFDM can be created by doing an N-sized FFT, a phase ramp-up, followed by a truncation of the imaginary part and a copy-flip-clip operation. This can be used to implement ACO-OFDM more efficiently, but also proves that Flip-OFDM and ACO-OFDM only differ in a frequency uplift of half a subcarrier thus perform equal.

7 Clipping and Distortion

In some papers, the LED is modelled as a peak–limited channel. When early LEDs were weak, inefficient and expensive, they needed to be used close to their maximum light output while further increase would be damaging. Modern LEDs are operated near their maximum external quantum efficiency (EQE) which is far below their limits [4, 15]. That makes a peak–limitation a less realistic model.

One may not only challenge the existence of high–light limit as a real restriction to throughput, the low limit near the LED turn-on current also deserves some debate. Light intensities cannot be negative and the LED is a diode. Nonetheless, one cannot draw the conclusion that negative currents are clipped. In fact, short peaks of negative currents can help to deplete the LED junction charge faster, thereby accelerating the response. This is known as carrier sweep out [1]. A more realistic clipping model would be to constrain the hole and electron concentrations in the junction to be non-negative. In other words, the LED first applies a low–pass filter and that filtered signal, with lower variance, may experience clipping of carrier concentrations. For frequencies substantially above the 3 dB bandwidth of the LED, that occurs only if the driver current makes large excursions into negative current.



(a)

(b)

Figure 2: (a): PSD of an OFDM signal (black) and clipping noise (gray), for bias-to-rms ratio of $z =$ 0.5, 1 and 2. LED low-pass response not included. [12]. (b): Ratio z between bias and rms modulation versus number of bits $b = \log_2 M$ per sub-carrier in one dimension. Noise-free ($r = 1$) and leaving a 3 and 6 dB power margin ($r = 2$ and $r = 4$, respectively) to ensure that distortion stays below the noise. Solid line: clipping limit. Dashed line: invertible second–order distortion limit. $r = 1$. [12]

Of course, current clipping may be caused by the driver [8,9,12]. Clipping noise is non-white. One clipped OFDM sample has an error that is a delta function in time domain. Its Fourier transform is a flat spectrum. However, a low-pass filtered OFDM signal has strong correlation, such that clips most likely come in bursts, which has relatively strong low frequency components, thus is non-white [12].

7.1 Inverting and mitigating LED non-linearities

LED clipping may be hard to repair. However, particularly at good SNR, it is most attractive to use a relatively large bias to avoid clipping to about 4σ: thus with a headroom that is more than 10 to 12 dB. At this point, invertible [8,16] second-order distortion starts to dominate over clipping [12]. This distortion is mainly-second order [4]. This LED distortion severely limits the ability to reach LED throughputs largely above 1 Gbit/s even if the SNR is very good, but can be removed by reasonably simple non-linear equalizers. 50 % rate improvements have been reported in [12]. The photonic mechanisms inside the LED can be translated into a model that is practical for simulation and to design a distortion–mitigation algorithm [1–4].

8 Conclusions

OWC is a field that overlaps both with wireless radio frequency communication and with optical fiber communication. Fiber links are very well–conditioned, so bit rate can be dominant objective. In contrast to this, wireless propagation is much less predictable and requires systems to be highly robust under a wide range of SNRs and propagation anomalies. We see that in radio communication, this has led to a wide range of study areas to improve various aspects of the system. Moreover, in radio communications, channel models, despite their variability, are well established, such that simulations and purely theoretical studies are deemed trustworthy. On these aspects, OWC seems to lack a parallel theory–developing track that addresses differences with radio. This has motivated research at Signify and at the SPS group of TU/e to better understand the performance of wireless optical links. We concluded that several commonly used statements deserve further attention.

Acknowledgments

This research was supported by EU H2020 Project under Grant ELIOT 825651 and acknowledges participation in EU COST action NEWFOCUS CA19111.

References

[1] J. P. M. G. Linnartz, X. Deng, A. Alexeev, and S. Mardanikorani, "Wireless communication over an led channel," *IEEE Comm. Mag.*, vol. 58, no. 12, pp. 77–82, 2020.

[2] S. Mardani, A. Alexeev, and J.-P. Linnartz, "Modeling and compensating dynamic nonlinearities in led photon-emission rates to enhance owc," in *Light-Emitting Dev., Mat., and App.*, vol. 10940. Intern. Soc. for Opt. and Phot., 2019, p. 109400U.

[3] X. Deng *et al.*, "Mitigating led nonlinearity to enhance visible light communications," *IEEE Trans. Comm.*, vol. 66, no. 11, pp. 5593–5607, 2018.

[4] A. Alexeev *et al.*, "Characterisation of dynamic distortion in led light output for optical wireless communication," *Phot. Res. OSA*, vol. 9, no. 5, p. 0, 2021.

[5] S. Mardanikorani and J. Linnartz, "Capacity of the first-order low-pass channel with power constraint," in *Proc. of the 2018 Symp. on Inf. Theory and Sig. Proc. in the Benelux.* Twente University, May 2018, pp. 149–153.

[6] S. Mardanikorani, X. Deng, and J.-P. M. Linnartz, "Sub-carrier loading strategies for dco-ofdm led communication," *IEEE Trans. Comm.*, vol. 68, no. 2, pp. 1101–1117, 2019.

[7] X. Deng, S. Mardanikorani, G. Zhou, and J.-P. M. Linnartz, "Dc-bias for optical ofdm in visible light communications," *IEEE Access*, vol. 7, pp. 98 319–98 330, 2019.

[8] X. Deng *et al.*, "Modeling and analysis of transmitter performance in visible light communications," *IEEE Trans. Vehic. Technol.*, vol. 68, no. 3, pp. 2316–2331, 2019.

[9] K. Arulandu *et al.*, "Enhanced visible light communication modulator with dual feedback control," *IEEE Trans. Emerg. Sel. Topics Power Electron*, 2019.

[10] R. You and J. M. Kahn, "Upper-bounding the capacity of optical im/dd channels with multiple-subcarrier modulation and fixed bias using trigonometric moment space method," *IEEE Trans. Inf. Theory*, vol. 48, no. 2, pp. 514–523, 2002.

[11] J. Grubor and K.-D. Langer, "Efficient signal processing in ofdm-based indoor optical wireless links," *Journal of Networks*, vol. 5, no. 2, p. 197, 2010.

[12] S. Mardanikorani *et al.*, "Optimization and comparison of m-pam and optical ofdm modulation for optical wireless communication," *IEEE Open J. Commun. Soc.*, vol. 1, pp. 1721–1737, 2020.

[13] J.-P. M. Linnartz and X. Deng, "Continuous phase flip-ofdm in optical wireless communications," *Signal Processing*, vol. 182, p. 107963, 2021.

[14] S. C. J. Lee *et al.*, "Pam-dmt for intensity-modulated and direct-detection optical communication systems," *IEEE Photonics Technol. Lett.,*, vol. 21, no. 23, pp. 1749–1751, 2009.

[15] A. Alexeev *et al.*, "Multiple heat source thermal modeling and transient analysis of leds," *Energies*, vol. 12, no. 10, p. 1860, 2019.

[16] X. Deng, Y. Wu, A. Khalid, X. Long, and J.-P. M. Linnartz, "Led power consumption in joint illumination and communication system," *Optics express*, vol. 25, no. 16, pp. 18 990–19 003, 2017.

SWIPT waveform design: Phase Optimization for a Rectangular Power Pulse Superposed with CP-OFDM

Hussein Kassab Jérôme Louveaux
Université catholique de Louvain
ICTEAM
Louvain-la-Neuve, Belgium
hussein.kassab@uclouvain.be jerome.louveaux@uclouvain.be

Abstract

Simultaneous wireless information and power transfer (SWIPT) allows wireless power transfer (WPT) and wireless information transfer (WIT) to coexist based on shared resources. The non-linearity model of the rectifier in [1] makes the design of power signals very important since it shows that the harvested energy is directly related to the transmitted waveform. It also shows that high-PAPR (peak to average power ratio) signals provide better performance in terms of energy harvesting. In this work, we inherit the proposed SWIPT waveform design in [2] and [3], where a high peak modulated rectangular wave sent during the cyclic prefix of the orthogonal frequency-division multiplexing (CP-OFDM) system, and we add an extra phase to every none-zero sample in the rectangular pulse. We study the influence of this added phase on the energy harvesting performance. We leverage from the work in [1] by using its rectenna model and its corresponding calculations in order to obtain the expression of the energy harvested in terms of these phases. After calculating these expressions, the objective of this work is to optimize the phase of the rectangular power signal in order to maximize the harvested energy at the receiver. The results are illustrated throughout this work.

Keywords

SWIPT, WPT, WIT, energy harvesting, waveform design, interference, CP-OFDM, rectangular pulse, PAPR, rectenna model, phase optimization, rectangular pulse.

References

[1] B. Clerckx, “Wireless Information and Power Transfer: Nonlinearity, Waveform Design, and Rate-Energy Tradeoff,” IEEE Trans. Signal Process., vol. 66, no. 4, pp. 847–862, Feb. 2018.

[2] H. Kassab and J. Louveaux, “Simultaneous Wireless Information and Power Transfer using Rectangular Pulse and CP-OFDM,“ 2019 IEEE International Conference on Communications (ICC), Shanghai, China, pp. 1-6, 2019.

[3] H. Kassab et al., “Superposition of Rectangular Power Pulse and CP-OFDM Signals to Achieve Simultaneous Wireless Information and Power Transfer,“ in preparation.

Keeping up with the bits: tracking physical layer latency in millimeter-wave Wi-Fi networks

Alexander Marinšek Liesbet Van der Perre
KU Leuven
ESAT-WaveCore
Ghent Technology Campus, 9000 Ghent, Belgium
alexander.marinsek@kuleuven.be

Abstract

The wireless communications landscape is anticipated to offer new service levels following the introduction of the millimeter-wave (mmWave) spectrum to consumer electronics. With their broad bandwidths and corresponding multi-Gbps data rates, these mmWaves are a perfect fit for data hungry applications, such as streaming video to extended reality devices. However, the latter are also bound by maximal latency constraints as low as 1 ms. Understanding where such minuscule time delays lurk requires a close-up study of individual layers in the network stack. Starting from the bottom up, the present work describes an endeavor at uncloaking the origins of physical layer (PHY) latency in mmWave Wi-Fi networks. It proposes a newly designed simulation framework and sheds light on how any conventional laboratory can be turned into a virtual experiment setting, speeding up computation. A case study based on the IEEE 802.11ad standard demonstrates the framework's ability to track packet latency at the PHY-level and identify individual bottlenecks. In particular, it evaluates the impact of the number of LDPC decoding iterations on latency in short transmission sequences.

Keywords— Simulation framework, Wi-Fi, millimeter-wave, IEEE 802.11ad

1 Introduction

A growing number of latency sensitive applications are imposing ever stricter end-to-end (E2E) time delay constraints on wireless networks. The use cases range from more traditional ones such as voice over IP (VoIP) [1], to future connected vehicles, industry 4.0, and extended reality. The applications contained therein require extremely low network time delays. For example, coordinated driving [2], cooperative robots [3], and immersive media [4] may all impose E2E latency constraints as low as 1 ms. Moreover, they are starting to increasingly rely on computation offloading [5], giving rise to their second requirement: high data rates.

Within the broad range of mmWave frequencies, several gigahertz of bandwidth, centered around 60 GHz, have been allocated to unlicensed communications. One of the standards operating in the 60 GHz band is the IEEE 802.11ad mmwave Wi-Fi [6], also referred to as WiGig. Supporting 64QAM and high LDPC code rates, it manages to attain data rates of up to 8.085 Gbps while conforming to the 1.76 GHz channel bandwidth limitations. This reduces the transmission times of even the longest physical layer (PHY) payloads (262 KB) to well below 1 ms; however, the finite transfer rate of the wireless network is not the only factor determining E2E latency.

The IEEE 802.11ad PHY features several components, every one of which may either add a small delay to the propagation of data, or in some cases, become a bottleneck. For both analysis purposes and the conception of adequate latency mitigation strategies and solutions, a simulation framework has been developed.

1.1 State of the art simulation frameworks

Among the existing simulations tools, ns-3 is one of the most widespread publicly available network simulators. It is backed by a lively community, thanks to which it also supports the IEEE 802.11ad standard. Described in [7], the WiGig ns-3 model features a detailed MAC layer on top of an abstract PHY implementation. It is a versatile tool for higher-level studies, concerning beamforming and fast session transfer among others. It is also able to keep track of transmission delays and it can induce data corruption by leveraging one of the underlying PHY error models. However, it does not cover the PHY components in greater detail, meaning their contribution to latency is unknown.

A substitute for ns-3 that offers more control over the PHY is the set of tools contained within MATLAB and its Communications Toolbox. These offer direct access to PHY component implementations, such as the LDPC codec. The visualization and manual analysis of intermediate results is also simplified with the use of MATLAB's built-in debugging tools. While the toolbox offers control over individual components, some effort was made in the past to encapsulate them in an IEEE 802.11ad PHY simulation framework [8]. The framework resides in an open online repository and its functionality is verified against measurement data; yet, it does not support PHY latency tracking.

Another endeavor to provide reliable WiGig simulations was presented in [9], where the OPNET network simulator served as the foundation for the IEEE 802.11ad model. While the authors explicitly stated the timing results obtained during simulation, the values ranging up to 5 ms include MAC layer time delays as latency accumulation within the PHY remains unaddressed.

1.2 Contribution

With the progressively stricter latency requirements of future real-time applications, gaining better insight into latency accumulation across the network stack is growing in importance. As the foundation of the network stack, the PHY is also the first layer contributing to latency accumulation. The present work outlines the design principles of a fine-grained simulation framework for tracking latency in the IEEE 802.11ad PHY. Furthermore, it presents a potential usage of the framework for evaluating the impact of LDPC decoding on transmission latency. The contributions are, therefore, two-fold:

- Design of an open source IEEE 802.11ad PHY latency simulation framework in Python, made publicly accessible through an online repository [10].
- Evaluation of the contribution of iterative LDPC decoding to the total latency during the transmission of short payloads (100 KB).

The work is structured as follows. Section 2 describes the working principles of the simulation framework and presents a multi-processing approach to reducing execution times both on the local machine and by using otherwise idle remote hardware. Employing the framework on an example, section 3 demonstrates latency tracking in practice and evaluates the delays induced by the LDPC decoder. Finally, section 4 summarizes the main findings and outlines future research prospects.

2 Simulation framework

The key features associated with PHY latency accumulation are finite transmitter (TX) data rate and delays within the receiver (RX) digital baseband (DBB). Simulating data propagation through individual components, the simulation framework is not bound to a particular RX DBB implementation. Assuming basic programming knowledge, the components can be added, removed, or placed at different positions. Moreover, their latency performance figures are freely adjustable. The RX DBB components are represented by individual blocks and buffers within the simulation framework, the inner workings and invocation of which are outlined in the following two sections.

In providing an illustrative example, the present work implements an RX DBB that follows a particular design pattern. It assumes frequency domain equalization and, with minor exceptions, the RX DBB is loosely based on the one shown in Figure 1.

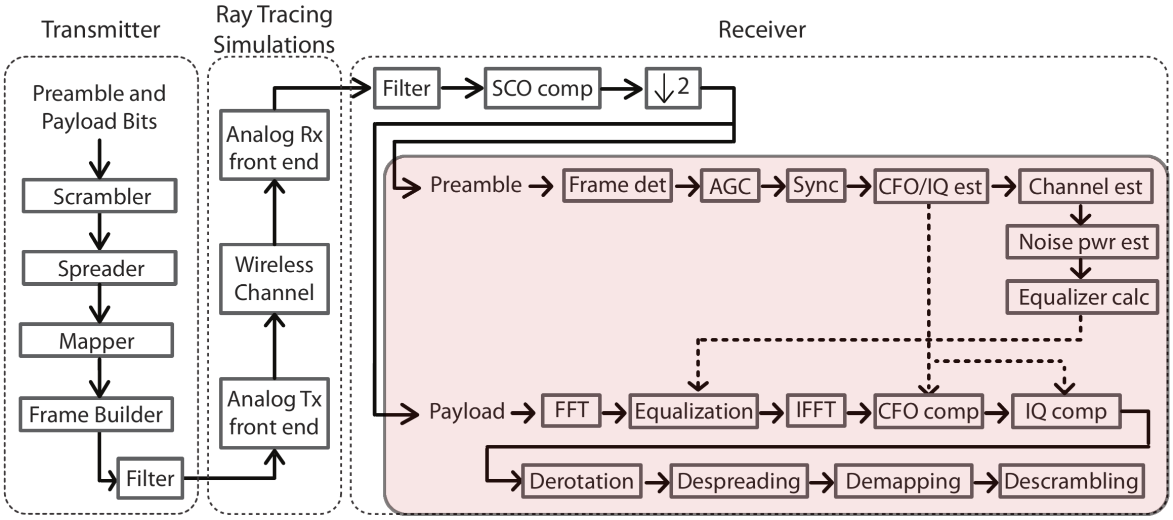


Figure 1: Overview of the simulated RX DBB (highlighted in red) within the PHY, as presented in [11].

2.1 Buffers, blocks, and events

The propagation of information through the RX DBB takes the form of data passing between the input buffers of individual blocks (components). If the block fails to process the data in time, then they start piling up in its input buffer. Several blocks also feature a second input buffer. Its function is to halt the block from processing the data before being allowed to do so. For example, the equalizer can only start processing the input data once channel estimation has concluded. Corresponding to Figure 1 and Figure 2, solid lines represent pathways leading to input buffers, while dashed lines show the propagation of flag inputs.

Every time data is passed to a buffer, an event is triggered. The simulation framework stores the event type (put, get, request) and the time at which it occurred. Timekeeping is achieved using the SimPy* Python package. The latter steps through simulated time at a sub-nanosecond resolution and enables time-based data propagation. The pace at which information is passed through the RX DBB is dictated by the performance figures of individual

*SimPy: https://simpy.readthedocs.io/

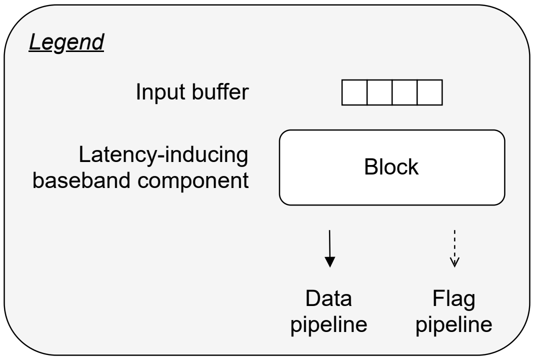


Figure 2: Illustration of a block, representing an individual RX DBB component.

components, sourced from state of the art literature. These are covered in the corresponding module in the simulation framework's online repository.

2.2 Reducing simulation time

Executing simulations for multiple payload sizes at different modulation and coding scheme (MCS) settings can present a time consuming task. Especially if only one core on the host machine is employed. Bridging such bottlenecks is often achieved using multithreading. However, being based on Python and tightly connected to SimPy, the simulation framework must instead rely on multiprocessing for parallel computing. Upon starting, the simulation is automatically split into concurrent processes, and as they execute, new ones are spawned based on the simulated parameter combinations. The size of the process pool is manually adjustable. As hinted by figure 3, the execution time is further reduced when multiple machines are used, such as otherwise idle computers in a classroom.

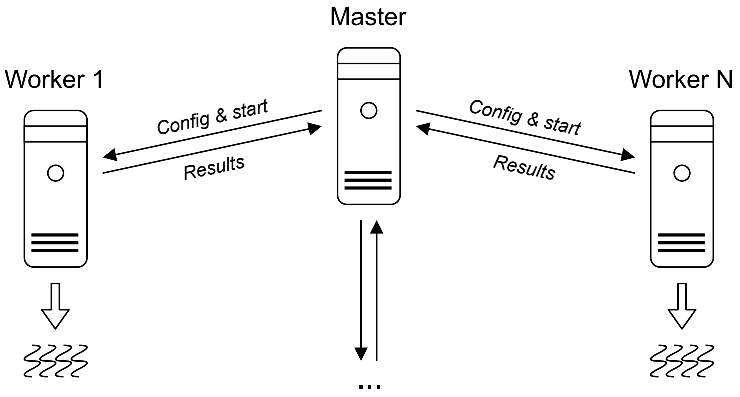


Figure 3: Example of applying multiprocessing and manual distributed computing for shorter execution times.

While process spawning is built into the framework, distributed computing must be set up manually. A typical workflow is to clone the simulation framework to all of the processing machines, configuring the input parameters (payload length, MCS), and starting the simulation in the background. Once finished, the results are manually aggregated before the data analysis is carried out. This is somewhat cumbersome compared to using purpose made distributed computing frameworks, yet, it reduces the number of dependencies and aids overall simplicity of implementation.

3 Simulation framework usage examples

The simulation framework includes detailed event logging in the background. It stores the exact time of every single block buffer input or output operation. Upon simulation execution, the logs are stored on the host machined, together with their corresponding metadata.

3.1 Latency accumulation in the physical layer

Illustrating the buffer events during a single transmission, Figure 4 analyzes how data propagate through the RX DBB and where they spend the most time. Observing buffer occupancy, there are two components where data are seemingly piling up: the demapper and the descrambler. This sudden increase in the item count is caused by the equalizer and decoder both outputting entire data blocks. These are then processed as individual symbols and bits by the demapper and descrambler. However, since the descrambler processes all of the data before the arrival of a new dataword, it is not a bottleneck. The same applies to the demapper, which, on the other hand, barely manages to keep up with the incoming data. Mildly reducing its throughput would already make it a bottleneck.

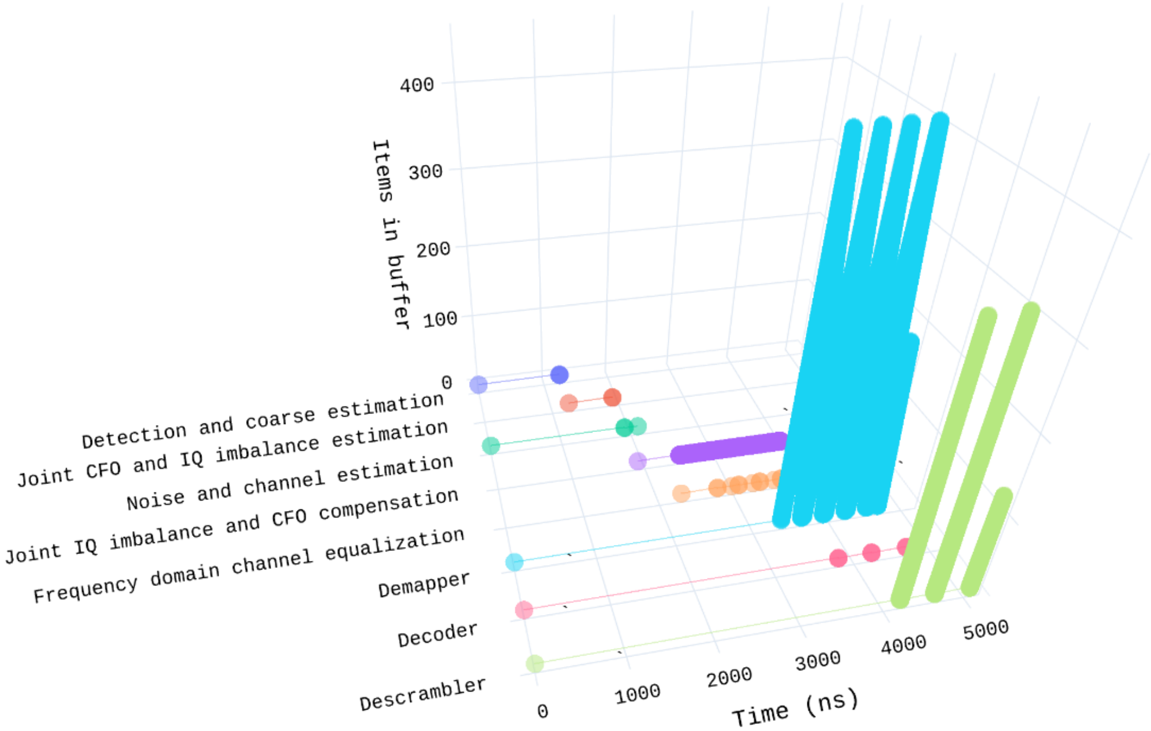


Figure 4: Example of a packet bearing 100 B of payload being transmitted using MCS 2.0 and setting the LDPC decoder iteration count to 10. The colored spheres represent buffer events, among which the first one is always the initial data request. Where offset from zero (e.g. Joint CFO and IQ imbalance estimation), there is an additional flag buffer, temporarily preventing the component from processing the incoming data.

Depending on the transmission parameters (MCS, payload length) and RX DBB configuration (component performance figures), the overview in Figure 4 might take an entirely different form.

3.2 Evaluating iterative LDPC decoder latencies

A practical example, where the simulation framework can provide valuable insight, is studying the effects of iterative decoding on the overall latency of the PHY. This is made possible by associating a time delay function with the decoder, dependent on the code rate and number of iterations. The latency results for short transmission sequences are summarized in Figure 5. For each modulation rate, the values obtained using code rate 0.5 feature the highest latency since they spawn the most overhead. Furthermore, the 0.5 and 0.625 code rate lines exhibit a similar slope because of the decoder defining a similar number of needed processing cycles for either of the two code rates [12]. The remaining two code rates, 0.75 and 0.8125, reduce the processing complexity and, therefore, the time delays diverge with the increasing number of iterations.

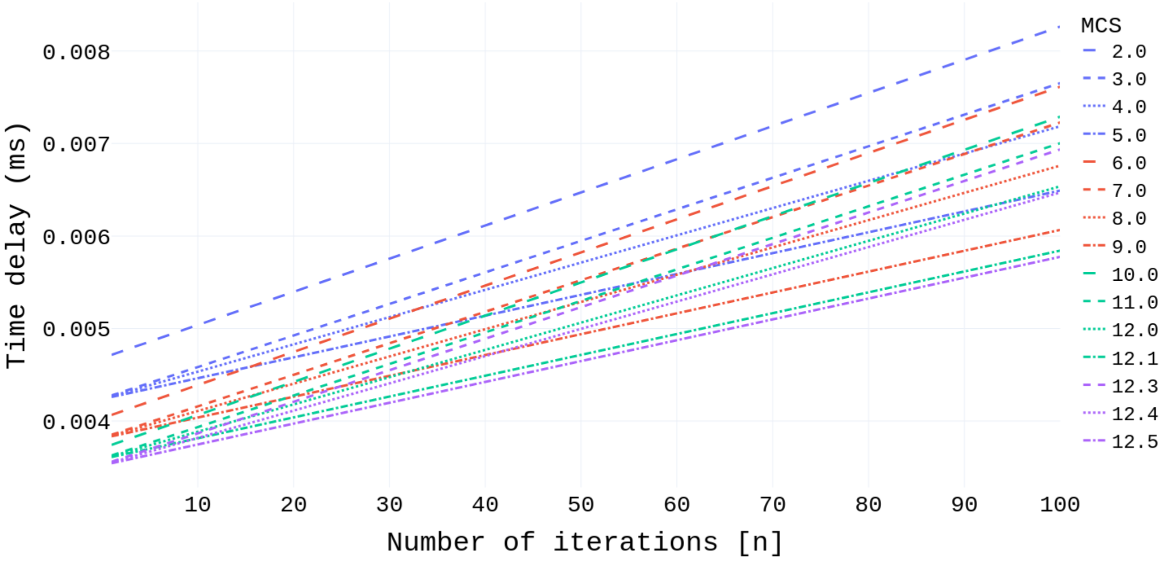


Figure 5: Time delay incurred in the PHY during the transmission and reception of PHY frames bearing 100 B payloads, dependent on the number of LDPC decoding iterations. Line colors represent modulation rate (BPSK, QPSK, 16QAM, 64QAM), while line types stand for different code rates (0.5, 0.625, 0.75, 0.8125).

From another perspective, the topmost and bottom-most lines are the constraints of a new latency region. It outlines the range of expected transmission time delays, regardless of the selected MCS, number of decoder iterations, or both of them.

4 Conclusion

The present work describes the design and usage of an IEEE 802.11ad PHY simulation framework, intended for PHY time delay analysis and the exploration of latency mitigation strategies to meet strict latency requirements. The core principles of how PHY components are reflected in the simulation framework blocks are outlined, and an approach to reducing execution times using parallel processing is discussed. The results demonstrate the simplicity of tracking data propagation through the PHY components and identifying potential bottlenecks. Finally, the evaluation of LDPC decoder latencies for small payload sizes highlights how the framework can be used in specific studies.

While the presented simulation framework manages to keep track of PHY time delays, it does not address data integrity. Since transmission errors are expected, the next step should focus on associating latency with other performance metrics, such as the bit error rate (BER), and evaluating the trade-offs between the two. Future work can also consider studying specific low-latency applications and the time delays of transmission sequences typically associated with them.

5 Acknowledgement

This project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No 861222 (MINTS MSCA-ITN).

The authors also wish to acknowledge Gilles Callebaut for providing support during different parts of debugging.

References

[1] M. Kassim, R. A. Rahman, M. A. A. Aziz, A. Idris, and M. I. Yusof, "Performance analysis of VoIP over 3G and 4G LTE network," en, in *2017 International Conference on Electrical, Electronics and System Engineering (ICEESE)*, Kanazawa: IEEE, Nov. 2017, pp. 37–41, ISBN: 978-1-5386-0908-8. DOI: 10.1109/ICEESE.2017.8298391. [Online]. Available: http://ieeexplore.ieee.org/document/8298391/.

[2] A. Kanavos, D. Fragkos, and A. Kaloxylos, "V2X Communication over Cellular Networks: Capabilities and Challenges," en, *Telecom*, vol. 2, no. 1, pp. 1–26, Jan. 2021, ISSN: 2673-4001. DOI: 10.3390/telecom2010001. [Online]. Available: https://www.mdpi.com/2673-4001/2/1/1.

[3] 5G Alliance for Connected Industries and Automation, *5G for Connected Industries and Automation*, Feb. 2019. [Online]. Available: https://5g-acia.org/wp-content/uploads/2021/04/WP_5G_for_Connected_Industries_and_Automation_Download_19.03.19.pdf.

[4] P. Lincoln, A. Blate, M. Singh, T. Whitted, A. State, A. Lastra, and H. Fuchs, "From Motion to Photons in 80 Microseconds: Towards Minimal Latency for Virtual and Augmented Reality," en, *IEEE Transactions on Visualization and Computer Graphics*, vol. 22, no. 4, pp. 1367–1376, Apr. 2016, ISSN: 1077-2626. DOI: 10.1109/TVCG.2016.2518038. [Online]. Available: http://ieeexplore.ieee.org/document/7383304/.

[5] A. Yousefpour, C. Fung, T. Nguyen, K. Kadiyala, F. Jalali, A. Niakanlahiji, J. Kong, and J. P. Jue, "All one needs to know about fog computing and related edge computing paradigms: A complete survey," en, *Journal of Systems Architecture*, vol. 98, pp. 289–330, Sep. 2019, ISSN: 13837621. DOI: 10.1016/j.sysarc.2019.02.009. [Online]. Available: https://linkinghub.elsevier.com/retrieve/pii/S1383762118306349.

[6] IEEE Computer Society, "Directional multi-gigabit (DMG) PHY specification," en, in *802.11-2016 - IEEE Standard for Information technology—Telecommunications and information exchange between systems Local and metropolitan area networks—Specific requirements - Part 11: Wireless LAN Medium Access Control (MAC) and Physical Layer (PHY) Specifications*, ISBN: 9781504436458, IEEE, Dec. 2016, pp. 2436–2496. DOI: `10.1109/IEEESTD.2016.7786995`. [Online]. Available: `https://ieeexplore.ieee.org/document/7786995/` (visited on 08/07/2020).

[7] H. Assasa and J. Widmer, "Implementation and Evaluation of a WLAN IEEE 802.11ad Model in ns-3," en, in *Proceedings of the Workshop on ns-3 - WNS3 '16*, Seattle, WA, USA: ACM Press, 2016, pp. 57–64, ISBN: 978-1-4503-4216-2. DOI: `10.1145/2915371.2915377`. [Online]. Available: `http://dl.acm.org/citation.cfm?doid=2915371.2915377`.

[8] J. Blumenstein, J. Milos, L. Polak, and C. Mecklenbrauker, "IEEE 802.11ad SC-PHY Layer Simulator: Performance in Real-world 60 GHz Indoor Channels," en, in *2019 IEEE Nordic Circuits and Systems Conference (NORCAS): NORCHIP and International Symposium of System-on-Chip (SoC)*, Helsinki, Finland: IEEE, Oct. 2019, pp. 1–4, ISBN: 978-1-72812-769-9. DOI: `10.1109/NORCHIP.2019.8906960`. [Online]. Available: `https://ieeexplore.ieee.org/document/8906960/`.

[9] C. Cordeiro, D. Akhmetov, and M. Park, "Ieee 802.11ad: Introduction and performance evaluation of the first multi-gbps wifi technology," en, in *Proceedings of the 2010 ACM international workshop on mmWave communications: from circuits to networks*, Chicago, Illinois, USA: Association for Computing Machinery, 2010, pp. 3–8. DOI: `10.1145/1859964.1859968`. [Online]. Available: `https://doi.org/10.1145/1859964.1859968`.

[10] A. Marinšek, *Phypy-802dot11ad/latency-simulator*, version v0.1, May 2021. DOI: `10.5281/zenodo.4749534`. [Online]. Available: `https://doi.org/10.5281/zenodo.4749534`.

[11] Z. Genc, W. V. Thillo, A. Bourdoux, and E. Onur, "60 GHz PHY Performance Evaluation with 3D Ray Tracing under Human Shadowing," en, *IEEE Wireless Communications Letters*, vol. 1, no. 2, pp. 117–120, Apr. 2012, ISSN: 2162-2337, 2162-2345. DOI: `10.1109/WCL.2012.022012.120033`. [Online]. Available: `http://ieeexplore.ieee.org/document/6155705/`.

[12] M. Li, F. Naessens, M. Li, P. Debacker, C. Desset, P. Raghavan, A. Dejonghe, and L. Van der Perre, "A processor based multi-standard low-power LDPC engine for multi-Gbps wireless communication," en, in *2013 IEEE Global Conference on Signal and Information Processing*, Austin, TX, USA: IEEE, Dec. 2013, pp. 1254–1257, ISBN: 978-1-4799-0248-4. DOI: `10.1109/GlobalSIP.2013.6737136`. [Online]. Available: `http://ieeexplore.ieee.org/document/6737136/`.

Spectrum Sensing in Mobile Cognitive Radio: A Bayesian Changepoint Detection Approach

Akpaki S. V. Chede[1,2], François Rottenberg[1], Michel Dossou[2] & Jérôme Louveaux[1]
Université catholique de Louvain[1], Université d'Abomey-Calavi[2]
ICTEAM-ELEN[1], EPAC-LETIA[2]
Belgium[1], Benin[2]

Abstract

Cognitive radio (CR) is the solution to the spectrum scarcity issue faced in wireless communications. Spectrum sensing is a crucial enabler for CR because it gives an insight into the free spectrum. In mobile environments, spectrum sensing is more critical as the spectrum occupancy becomes more dynamic with respect to time and frequency. It would be a good idea to exploit some knowledge about the mobility parameters to make better sensing decisions.

In this work, we consider a mobile CR scenario where some fixed primary users (PU) transmit in the same frequency band, each having its coverage area. A secondary user (SU), the CR node, is moving and continuously looking for an opportunity to use this band. We explore the possibility of using the SU mobility patterns to improve the spectrum sensing. We do that by using the Bayesian changepoint detection approach and assuming that the mobility parameters can be summarized in some a priori knowledge on the average time of spectrum change. Considering that in CR, the power of the signal to be detected is often unknown, we have introduced a low-complexity algorithm that does not rely on this knowledge, in a previous publication. The comparisons with existing algorithms in the literature have shown that the derived algorithm outperforms its non-Bayesian equivalent at low signal to noise ratio (SNR). The current work extends our previous one, presented in IEEE VTC-2021 Spring conference, in several ways. We go further in the comparisons by studying the computational complexity of the proposed algorithm. Moreover, we investigate different ways of computing the optimal detection threshold in practical scenarios of unknown-SNR.

1 Introduction

The ever-growing number of wireless devices, due to the Internet of Things (IoT) explosion, is dramatically leading to spectrum shortage. Dynamic and opportunistic spectrum access should be considered to face this problem. Cognitive radio is the enabling technology for dynamic spectrum access, as it allows a network node to be aware of its electromagnetic environment and adapt its transmission parameters accordingly [1]. This considerably increases the overall spectral efficiency. Spectrum sensing is the process by which the CR node knows the available frequency bands for transmission. As wireless devices may experience mobility, it is important to implement spectrum sensing algorithms that can to quickly detect changes in the spectrum occupancy when a device is moving.

In detection theory, sequential changepoint detection (or quickest detection) aims at detecting, as quickly as possible, a change in the distribution of the observed samples. It is, therefore, suitable for spectrum sensing in a mobile environment. Depending on the time of change (changepoint), there are two changepoint detection frameworks. On the one hand, the *minimax approach* assumes that the changepoint is deterministic and unknown. On the other hand, the *Bayesian approach* assumes that the changepoint is a random variable with a known prior distribution [2]. According to the literature, the optimal detection algorithms in both frameworks are the cumulative sum (CUSUM) and the Shiryaev algorithms. In practical scenarios, where the power of the signal to detect is not known, the generalized likelihood ratio (GLR) test could be applied to the optimal algorithms. It consists in replacing the unknown parameter of the distribution with its maximum likelihood estimate.

In this paper, we study the use of changepoint detection for spectrum sensing in mobile cognitive radio. Assuming that the prior distribution of the changepoint could be, in practice, inferred from mobility parameters and that the SU does not know the PU signal power, we derived a low-complexity GLR test (termed as *LC GLR-Shiryaev*) for the Bayesian framework. The derived algorithm outperforms its non-Bayesian equivalent at low SNR. This work is an extension of our previous conference paper [3]. As a new contribution, we perform a complexity analysis of the derived algorithm and suggest two methods (adaptive and fixed) for setting the optimal threshold when the SNR is unknown. Finally, we conclude that both methods could give the same performance when a fixed threshold is set correctly. The rest of the paper is organized as follows. The system model is introduced in section 2. Section 3 reminds classical changepoint detection algorithms, including the LC GLR-Shiryaev. Then, in section 4, we carry out the complexity analysis and the optimal threshold setting. Numerical results are given in section 5 in order to compare the complexities of the simulated algorithms and evaluate the performance of the optimal threshold setting. Finally, Section 6 concludes the work and offers a discussion.

2 System Model

2.1 Scenario

We consider a scenario where some fixed PUs are transmitting in the same frequency band, each having its coverage area, as depicted in figure 1. They could be, for example, television (TV) transmitters in a TV broadcasting network or base stations in a mobile network. A mobile SU is looking for opportunities to use the spectrum and setting up a secondary communication. We assume that at the beginning, the SU is not in range of any PU. After a first spectrum sensing, the SU tunes to the frequency band and starts using it.

As it is moving, the spectrum occupancy may change and the SU may see the PU signal appearing unexpectedly. To detect this appearance as quickly as possible, the SU is continuously monitoring the frequency band.

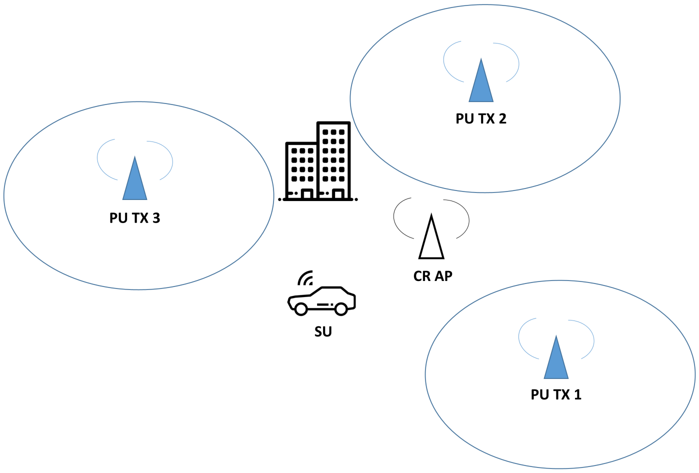


Figure 1: Mobile CR scenario

2.2 Signal model

The signal observed by the SU is denoted by $y[m]$, where m is the time sample of observation. When the PU is not active, $y[m] = w[m]$, where $w[m]$ is the Additive White Gaussian Noise (AWGN) sample. When the PU is active, $y[m] = s[m] + w[m]$, where $s[m]$ is the transmitted signal affected by the channel. We assume that $s[m]$ and $w[m]$ are independent circularly symmetric complex Gaussian variables, $s[m] \sim \mathcal{CN}(0, \sigma_s^2)$, $w[m] \sim \mathcal{CN}(0, \sigma_w^2)$, and independent of each other.

Initially, the SU is outside of the PUs' coverage areas, and the observed signal (only noise) follows a complex Gaussian distribution with mean 0 and variance σ_w^2. While the SU is moving, at an unknown time sample τ, it gets in range of a PU and the distribution changes to a complex Gaussian distribution with mean 0 and variance $\sigma_s^2 + \sigma_w^2$. As we use the changepoint detection approach, the SU observes each sample of the received signal sequentially. Considering all the received sequence $y_1^n = y[1], ..., y[n]$ up to the current time sample n, it tries to detect the possible change in the distribution by distinguishing between the following two hypotheses:

$$\begin{cases} \mathcal{H}_0: & y[m] = w[m], \qquad\qquad m = 1, ..., n \\ \mathcal{H}_1: & \exists\, \tau \in [0, n], \text{ such that} \\ & y[m] = w[m], \qquad\qquad m = 1, ..., \tau \\ & y[m] = s[m] + w[m], \quad m = \tau + 1, ..., n \end{cases}$$

where τ is the time sample of change (changepoint) and is assumed to be a random variable a priori geometrically distributed*. Moreover, we assume that the parameter p of this geometric prior could be inferred from the SU mobility parameters, with a certain degree of accuracy.

3 Review of existing changepoint detection algorithms

In this section, we introduce some classical changepoint detection algorithms belonging to the deterministic and Bayesian approaches. In general, the changepoint detection algorithm starts taking the observed signal samples. At each time sample n it computes a decision statistic C_n that is compared with a threshold α. When C_n exceeds α, an alarm is raised, and the algorithm stops. The stopping time is denoted by

$$T = \inf\{n \geq 1 : C_n \geq \alpha\}. \tag{1}$$

3.1 Deterministic approach

In the deterministic framework, the optimal algorithm is the well-known CUSUM algorithm, whose decision statistic is given by [4]

$$C_{n,\text{CUS}} = \max_{k \leq n} \sum_{m=k+1}^{n} \ln(L_m), \tag{2}$$

where $L_m = \frac{\sigma_w^2}{\sigma_s^2+\sigma_w^2} \exp\left(\frac{|y(m)|^2 \sigma_s^2}{\sigma_w^2\,(\sigma_s^2+\sigma_w^2)}\right)$ is the likelihood ratio of the m^{th} observed sample and k is the time sample from which the sum of the likelihood ratio has consistently increased.
In practical scenarios, the PU signal power σ_s^2 is not known by the SU, and the GLR-CUSUM algorithm is used. The decision statistic is defined as [4]

$$C_{n,\text{GLR-CUS}} = \max_{k \leq n} \sup_{\sigma_s^2} \sum_{m=k+1}^{n} \left(\frac{|y(m)|^2 \sigma_s^2}{\sigma_w^2\,(\sigma_s^2 + \sigma_w^2)} + \ln \frac{\sigma_w^2}{\sigma_s^2 + \sigma_w^2}\right), \tag{3}$$

where σ_s^2 is replaced by the value that maximizes the CUSUM decision statistic.

3.2 Bayesian approach

In the Bayesian approach, the changepoint τ is assumed to be random with a specific prior distribution [2]. Considering a geometric prior, the Shiryaev algorithm is the optimal algorithm in this approach. Its decision statistic is given by

$$C_{n,\text{SHI}} = \sum_{k=1}^{n} \prod_{m=k}^{n} \left(\frac{1}{1-p}\right) \frac{\sigma_w^2}{\sigma_s^2 + \sigma_w^2} \exp\left(\frac{|y(m)|^2 \sigma_s^2}{\sigma_w^2\,(\sigma_s^2 + \sigma_w^2)}\right). \tag{4}$$

*$P(\tau = l) = p(1-p)^l$, $l \geq 0$ where p is the probability of PU appearance at each time sample.

When σ_s^2 is unknown to the SU, the GLR approach can be applied to the Shiryaev algorithm and, as in the GLR-CUSUM case, the statistic $C_{n,\text{SHI}}$ has to be maximized, giving the GLR-Shiryaev statistic

$$C_{n,\text{GLR-SHI}} = \sup_{\sigma_s^2} \sum_{k=1}^{n} \prod_{m=k}^{n} \Big(\frac{1}{1-p}\Big) \frac{\sigma_w^2}{\sigma_s^2+\sigma_w^2} \exp\Big(\frac{|y(m)|^2\,\sigma_s^2}{\sigma_w^2\,(\sigma_s^2+\sigma_w^2)}\Big). \tag{5}$$

As this optimization problem is hard to solve analytically, due to non-convexity, we have introduced the so-called low-complexity (LC) GLR-Shiryaev algorithm [3], which maximizes each term of the sum instead of directly maximizing the sum. Its decision statistic is given by

$$\begin{aligned} C_{n,\text{LC GLR-SHI}} \simeq \sum_{k=1}^{n} \frac{1}{(1-p)^{n-k+1}} \Bigg(\Big(\frac{\sigma_w^2}{\tilde{\sigma}_{s,k}^2+\sigma_w^2}\Big)^{n-k+1} & \\ \exp\Big(\frac{\tilde{\sigma}_{s,k}^2}{\sigma_w^2\,(\tilde{\sigma}_{s,k}^2+\sigma_w^2)} \sum_{m=k}^{n} |y(m)|^2\Big)\Bigg), & \end{aligned} \quad \text{with } \tilde{\sigma}_{s,k}^2 = \Big\{\frac{\sum_{m=k}^{n}|y(m)|^2}{n-k+1} - \sigma_w^2\Big\}^{+}. \tag{6}$$

A numerical method for solving the GLR-Shiryaev algorithm is termed M-Shiryaev [5]. It assumes that the SU knows the M possible values of σ_s^2 (termed here as $\sigma_{s,i}^2$ for $i \in [1, M]$), and consists in running M parallel Shiryaev algorithms for each of these values. The PU appearance is declared when any one of the procedures stops:

$$\begin{aligned} C_{n,\text{SHI},i} = \sum_{k=1}^{n} \frac{1}{(1-p)^{n-k+1}} \Bigg(\Big(\frac{\sigma_w^2}{\sigma_{s,i}^2+\sigma_w^2}\Big)^{n-k+1} & \\ \exp\Big(\frac{\sigma_{s,i}^2}{\sigma_w^2\,(\sigma_{s,i}^2+\sigma_w^2)} \sum_{m=k}^{n} |y(m)|^2\Big)\Bigg) & \end{aligned} \quad \text{for } i \in [1, M] \tag{7}$$

A recursive implementation of M-Shiryaev [6] also exists and is given by

$$C_{n,\text{SHI},i} = (1 + C_{n-1,\text{SHI},i}) \Big(\frac{1}{1-p}\Big) \frac{\sigma_w^2}{\sigma_{s,i}^2+\sigma_w^2} \exp\Big(\frac{|y(n)|^2\,\sigma_{s,i}^2}{\sigma_w^2\,(\sigma_{s,i}^2+\sigma_w^2)}\Big), \quad \text{for } i \in [1, M]. \tag{8}$$

3.3 Comparison

In [3], we carried out some comparisons between the previously mentioned algorithms in several scenarios. We defined a new metric called the global penalty and given by

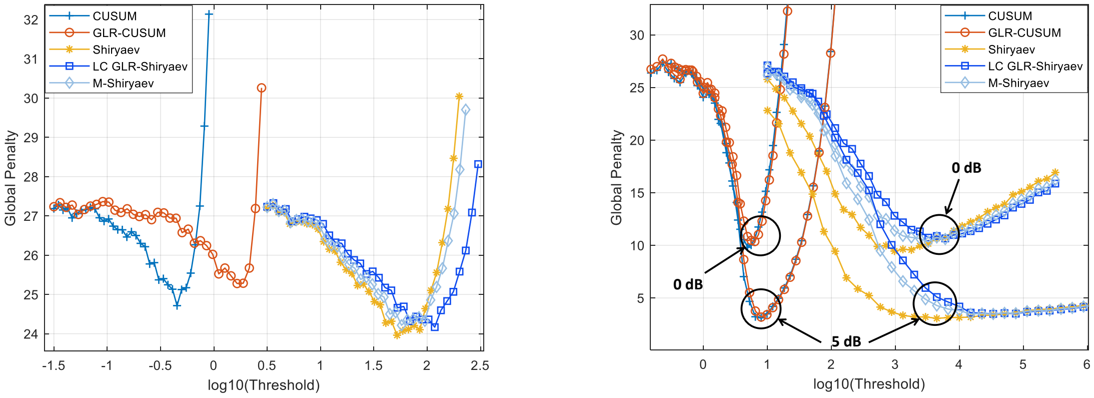
$$G_P = P_{int}\ A_{DD} + P_{wso}\ A_{FAD}, \tag{9}$$

where A_{DD} is the average detection delay, A_{FAD} is the average time between a false alarm and the changepoint, P_{int} and P_{wso} are their respective associated penalties. The global penalty jointly evaluates the costs of interference and spectrum waste induced by the changepoint detection algorithms. When comparing the various algorithms, the derived LC GLR-Shiryaev [3] has better performance (minimum global penalty) than its deterministic counterpart at low SNR, as shown in figure 2. This is due to the prior knowledge of the PU appearance.

It is also seen that LC GLR-Shiryaev and the numerical M-Shiryaev have comparable performances. Studying their computational complexities could give an insight into the best choice of algorithm. Moreover, it is suggested to choose the optimal threshold as the one that minimizes the global penalty. This threshold depends on p and the SNR. Thus, how to set the optimal threshold in an unknown-SNR case is an issue.

4 Complexity analysis and optimal threshold setting

In this section, we evaluate the computational complexity of the LC GLR-Shiryaev and M-Shiryaev algorithms for hardware implementation. Then, we suggest different methods to compute the optimal threshold for changepoint detection algorithms when the SNR is unknown.



(a) p = 0.005, SNR = -10 dB.

(b) p = 0.005, SNR = 0 dB & 5 dB.

Figure 2: G_P vs $log_{10}(\alpha)$ for $p = 0.005$ [3].

4.1 Computational complexity of algorithms

We express the complexity in terms of number of elementary operations needed to run the algorithms at each time sample n. The most used elementary operations are real summation (RS), real multiplication (RM), real division (RD), real exponential function (RE). A real exponentiation a^b can be expressed in terms RM assuming that fast exponentiation algorithms use at most $2\log_2(b)$ RM.

- The complexity of LC GLR-Shiryaev depends on both the computations of the $\tilde{\sigma}^2_{s,k}$ and $C_{n,\text{LC GLR-SHI}}$ (6). To compute each $\tilde{\sigma}^2_{s,k}$, we need $(2n-2k+2)$ RM, $(2n-2k+4)$ RS and 1 RD. Thus, the computation of the n $\tilde{\sigma}^2_{s,k}$ needs $\sum_{k=1}^{n}(2n-2k+2) = (n^2+n)$ RM, $\sum_{k=1}^{n}(2n-2k+4) = (n^2+3n)$ RS and n RD. As the term $\sum_{m=k}^{n}|y(m)|^2$ is already computed in $\tilde{\sigma}^2_{s,k}$, it can be reused in the computation of the decision statistic. Thus, the k^{th} term of $C_{n,\text{LC GLR-SHI}}$ needs $(4+4\log_2(n-k+1))$ RM, 7 RS, 3 RD and 1 RE. To compute and add the n terms, we need $\sum_{k=1}^{n}(4+4\log_2(n-k+1)) = (4n+2n\log_2(n))$ RM, $(7n+n-1)$ RS, $3n$ RD and n RE.
  In total, at each time sample n, LC GLR-Shiryaev needs $(n^2+5n+2n\log_2(n))$ RM, $(n^2+11n-1)$ RS, $4n$ RD and n RE.

- Following the same reasoning, the complexity of both implementations of M-Shiryaev is computed. At each time sample n, M statistics are computed. For the direct implementation (7), the terms $\sum_{m=k}^{n}|y(m)|^2$ and $\frac{1}{(1-p)^{n-k+1}}$ can be computed once and reused in the $M-1$ other statistics. In total, it needs $(n^2+(4M+1)n+(M+1)n\log_2(n))$ RM, $(n^2+(4M+3)n)$ RS, $((2M+1)n)$ RD and (nM) RE. The recursive implementation (8) needs $(5M+2)$ RM, $(3M+2)$ RS, $(2M+1)$ RD and M RE.

The complexity analysis can be summarized in table 1. Numerical comparisons of the complexities are shown in section 5.

4.2 Unknown-SNR optimal threshold setting

As mentioned before, the optimal threshold depends on the SNR value. In this section, we consider how to set the optimal threshold for algorithms which do not have the knowledge of the SNR. We suggest two ways to do that:

Table 1: Complexity analysis

	No. of RM	No. of RS	No. of RD	No. of RE
LC GLR-Shiryaev	$(n^2 + 5n$ $+2n\log_2(n))$	$(n^2 + 11n - 1)$	$4n$	n
M-Shiryaev (7)	$(n^2 + (4M + 1)n$ $+(M + 1)n\log_2(n))$	$(n^2 + (4M + 3)n)$	$((2M + 1)n)$	(nM)
M-Shiryaev (8)	$(5M + 2)$	$(3M + 2)$	$(2M + 1)$	M

- First, we try to estimate the SNR at each time sample n, based on the observation, and use the optimal threshold (for known SNR) corresponding to the estimated value. This is called *adaptive threshold setting*, as the estimated SNR may change at each time sample and the threshold is updated. We could reuse the SNR estimated in the decision statistic, to compute the threshold. Let us consider the LC GLR-Shiryaev statistic (6), where we compute the $\tilde{\sigma}^2_{s,k}$ that maximize each term because it is difficult to compute the optimal $\tilde{\sigma}^2_s$. Having the $\tilde{\sigma}^2_{s,k}$, how can we get an accurate estimate of σ^2_s? We make two proposals:
  - 1^{st} *proposal:* At every step n, we use the $\tilde{\sigma}^2_{s,k}$ that gives the highest term, as the most reliable estimate of σ^2_s.
  - 2^{nd} *proposal:* We average the current $\tilde{\sigma}^2_{s,k}$ that gives the highest term with the ones of the previous time samples, and use the average as an estimate of σ^2_s.
- Secondly, we suggest to compute numerically the optimal threshold that minimizes G_P in average, assuming a certain SNR distribution. This is called *fixed threshold setting.*

In the next section, we show the performance of each method.

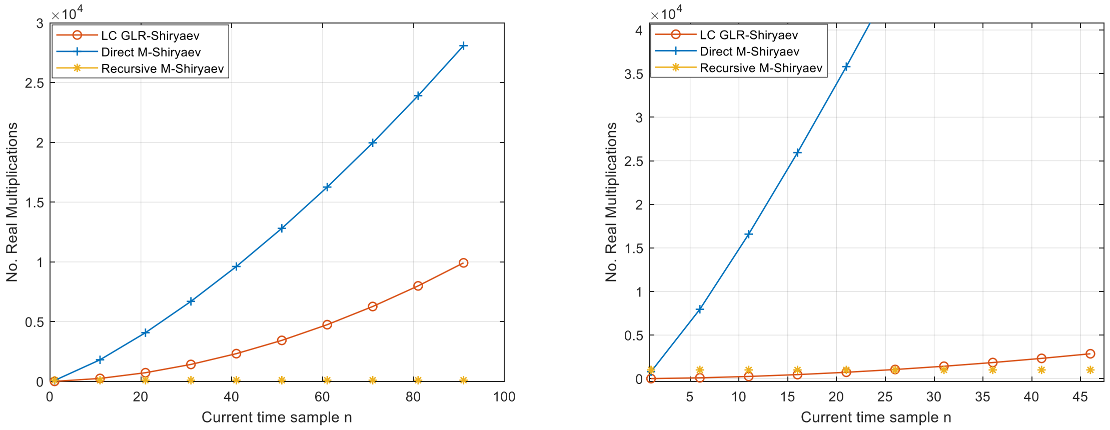
5 Simulation results & discussion

5.1 Complexity simulation

We show some numerical comparisons of the algorithms complexities. For a fixed value of M, we plot the complexity with respect to the number of time samples. We just focus on the number of real multiplications as the other operations have the same trend. As it can be seen in figure 3a, the complexities of LC GLR-Shiryaev and the direct M-Shiryaev grow quadratically ($O(n^2)$) when the number of samples increases, but the recursive M-Shiryaev has a constant complexity ($O(M)$). The direct M-Shiryaev is worse than LC GLR-Shiryaev when M is high but has a closer complexity when M is low. As shown in figure 3b, the only case when LC GLR-Shiryaev and the recursive M-Shiryaev have comparable complexity is when M is high and the number of observed samples is low. To conclude, the recursive M-Shiryaev has the lowest complexity but requires an accurate *a priori* knowledge of the SNR, otherwise the performance will degrade. LC GLR-Shiryaev has the advantage of not requiring any *a priori* knowledge but adds some complexity. A trade-off between complexity and performance should thus be considered when choosing an algorithm among LC GLR-Shiryaev and M-Shiryaev.

5.2 Optimal threshold performance

Here, we show the performance of the unknown-SNR threshold methods by using Monte Carlo simulations to compute the average global penalty at various SNR values. We use the following parameters : $p = 0.005$, SNR $= \frac{\sigma_s^2}{\sigma_w^2} \in \{-10, -5, 0, 5\}\, dB$, $P_{int} = 0.5$ and $P_{wso} = 0.14$.

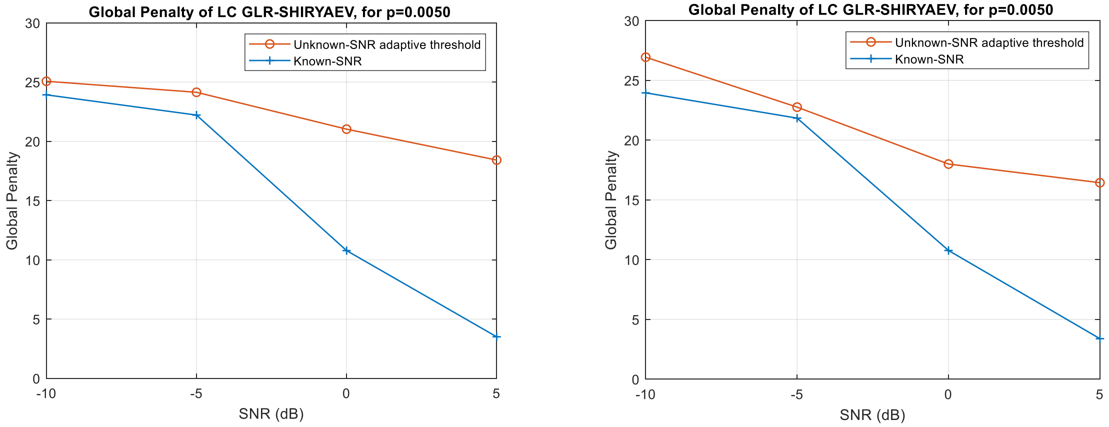


(a) Complexity for $M = 20$

(b) Complexity for $M = 200$

Figure 3: Complexity in terms of No. of Real Multiplications.

We first compute the global penalty for the adaptive threshold and compare it with the known-SNR case (the optimal one). As shown in figure 4, for the first proposal, the obtained global penalty is closer value to the minimal value (known-SNR) at low SNR, but deviates from the minimal value at high SNR. The second proposal improves a bit G_P for high SNR.

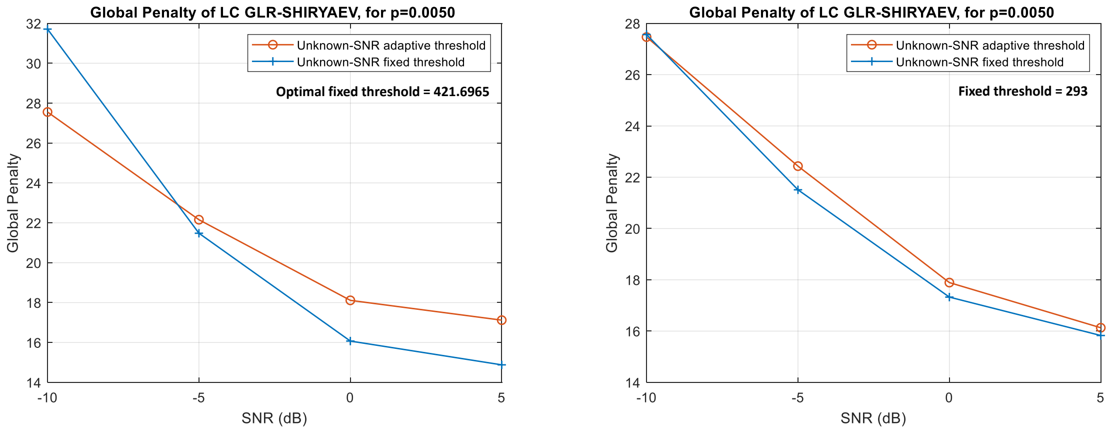


(a) 1^{st} proposal

(b) 2^{nd} proposal

Figure 4: Global penalty for unknown-SNR adaptive threshold.

We can see that adaptive threshold setting performs quite well for an unknown SNR method. The poor result at high SNR is due to the fact that a reliable estimate of the SNR cannot be obtained before the changepoint τ. So when the true SNR is low, the successive estimates are closer to it before τ so that G_P is closer to the optimal one. But when the true SNR is high, the estimates before τ are bad and this gives a poor global penalty.
Secondly, we assume a Rayleigh-distributed SNR to compute the optimal fixed threshold. We obtain the result shown in figure 5a, where the optimal fixed threshold method is better than adaptive threshold (2^{nd} proposal) in high SNR, but worse in low SNR. Moreover, it is shown, by varying the distribution parameters, that a fixed threshold can be correctly set to give the slightly same performance as the adaptive threshold. This is shown in shown in figure 5b.



(a) (b)

Figure 5: Global penalty for unknown-SNR adaptive vs fixed threshold.

6 Conclusion

In this paper, we investigate the use of Bayesian changepoint detection in mobile cognitive radio. We perform a complexity analysis of the previously derived LC GLR-Shiryaev and compare it with the complexity of the M-Shiryaev algorithm. While being less complex than the LC GLR-Shiryaev, the recursive implementation of M-Shiryaev needs an accurate a priori knowledge of the SNR, which is usually not available, to give the same performance. Moreover, we introduce an adaptive and a fixed threshold setting methods for the LC GLR-Shiryaev. Both methods give almost the same performance when the fixed threshold is set correctly. A potential direction for future work is to reduce the computational complexity of the LC GLR-Shiryaev.

References

[1] S. Haykin, “Cognitive radio: brain-empowered wireless communications,” in *IEEE Journal on Selected Areas in Communications*, vol. 23, no. 2, pp. 201-220, feb. 2005.

[2] A. Tartakovsky, I. Nikiforov, and M. Basseville, *Sequential Analysis.* Chapman and Hall/CRC, 1st ed., aug 2014.

[3] A. S. V. Chede, F. Rottenberg, M. Dossou, and J. Louveaux, “Use of Bayesian Changepoint Detection for Spectrum Sensing in Mobile Cognitive Radio,” in 93rd *IEEE Vehicular Technology Conference (VTC-Spring)*, Helsinki, Finland (Virtual Event), pp. 1-5, apr 2021.

[4] L. Lai, Y. Fan, and H. V. Poor, “Quickest detection in cognitive radio: A sequential change detection framework,” in *GLOBECOM - IEEE Global Telecommunications Conference*, pp. 2957–2961, 2008.

[5] J. Geng and L. Lai, “Bayesian quickest detection with unknown post-change parameter,” in *ICASSP, IEEE International Conference on Acoustics, Speech and Signal Processing - Proceedings*, vol. 2016-May, pp. 4169–4173, may 2016.

[6] J. Geng, E. Bayraktar, and L. Lai, “Multi-Chart Detection Procedure for Bayesian Quickest Change-Point Detection with Unknown Post-Change Parameters,” tech. rep., 2017.

Unclonable Encryption with Key Recycling using qubits

Daan Leermakers, Boris Škorić

Classically, the best confidentiality guarantee is provided by One-Time Pad (OTP) encryption. Provided that two hones parties Alice and Bob share a uniform secret string, they can exchange a message of the same length with information-theoretical security. An Eavesdropper (Eve) can save a copy of the ciphertext send over a classical channel. For the message to remain secure in the future, two conditions must be met:

1. The key is used only once.
2. The key is kept secret forever.

When using a quantum channel rather than a classical channel both of these conditions can be relaxed. Quantum Key Recycling (QKR) [1] and Unclonable Encryption (UE) [2] use quantum channel to encrypt a classical message with information theoretical security. A Quantum Key Recycling protocol allows Alice and Bob to safely re-use the encryption keys in a future communication. Unclonable Encryption guarantees the confidentiality of the message even when the keys leak some time in the future.

These two distinct protocols rely on the same property of quantum physics, the impossibility of cloning an unknown quantum state. It should therefore be possible to achieve both properties in a single protocol. We propose a protocol that achieves both Key Recycling and Unclonable Encryption while using only simple qubit preparation and measurements.

Two honest parties, Alice and Bob share a secret set of measurement bases, an authentication key, a hashing seed and some randomness used for padding. Alice uses an information-theoretically secure MAC function to authenticate her message and appends the resulting tag to the message. She then computes the inverse of a pairwise independent hash function using the shared hash seed. Finally she appends redundancy bits used for error correction and pads the structure of the redundancy bits with a shared secret. The result of her computation is a uniform string.

Alice encodes her uniform string into qubits according to the shared measurement bases. Bob can measure the qubits in the correct basis, perform the error correction, and hash the result to retrieve the message and the authentication tag. If the error correction was successful, Alice and Bob can re-use all their key material except for the pad that hides the structure of the error correction redundancy. The only classical communication in the protocol occurs when Bob sends a single authenticated bit to Alice indicating the success of the protocol.

We prove that our protocol has secure key re-use as well as Unclonable Encryption. These security properties are relevant in two scenarios: (i) the keys have not yet leaked and should be safe to re-use; (ii) all key material leaks. We prove secure key re-use as wel as UE by comparing the real protocol by an idealized protocol and bounding the diamond distance between the two protocols. By doing so we bound the probability that an attacker can distinguish between the real and the ideal situation when observing the protocol once. We find that the number of qubits required to achieve Unclonable Encryption is half the number of qubits required in the only other proven Unclonable encryption protocol [2] asymptotically.

References

[1] C.H. Bennett, G. Brassard, and S. Breidbart, *Quantum Cryptography II: How to re-use a one-time pad safely even if P=NP*, Natural Computing **13** (2014), 453–458. Original manuscript 1982.

[2] D. Gottesman, *Uncloneable encryption*, Quantum Information and Computation **3** (2003), no. 6, 581–602.

[3] D. Leermakers and B. Škorić, *Security proof for quantum key recycling with noise*, Quantum Information Processing **19** (2019).

NODE VARYING REGULARIZATION FOR GRAPH SIGNALS

Maosheng Yang, Mario Coutino, Elvin Isufi and Geert Leus

Delft University of Technology, Delft, The Netherlands
E-mails: m.yang-7@student.tudelft.nl; {m.a.coutinominguez-1, e.isufi-1, g.j.t.leus}@tudelft.nl

ABSTRACT

While regularization on graphs has been successful for signal reconstruction, strategies for controlling the bias-variance trade-off of such methods have not been completely explored. In this work, we put forth a node varying regularizer for graph signal reconstruction and develop a minmax approach to design the vector of regularization parameters. The proposed design only requires as prior information an upper bound on the underlying signal energy; a reasonable assumption in practice. With such formulation, an iterative method is introduced to obtain a solution meeting global equilibrium. The approach is numerically efficient and has convergence guarantees. Numerical simulations using real data support the proposed design scheme.

Index Terms— graph signal processing, bias-variance trade-off, graph regularization, graph signal denoising, minmax problems

1. INTRODUCTION

In this work, we focus on solving the following problem

$$\boldsymbol{\omega}^* := \underset{\boldsymbol{\omega}\in\mathcal{W}}{\arg\min}\ f_{\boldsymbol{\omega}}(\boldsymbol{y};\boldsymbol{\mu}), \tag{1}$$

where $\boldsymbol{\omega} \in \mathbb{R}^d$ is a regularization parameter for the loss function $f_{\boldsymbol{\omega}}(\cdot;\cdot)$ w.r.t the data $\boldsymbol{y} \in \mathbb{R}^n$ and an underlying unknown parameter $\boldsymbol{\mu} \in \mathbb{R}^q$. The regularization parameter $\boldsymbol{\omega}$ is within a convex set $\mathcal{W}$.

Problems of the form in (1) arise naturally in applications including, hyper-parameter tuning [1], biased estimators [2], image denoising [3], and signal reconstruction [4], to name a few. Though this formulation is simple, the dependency of $f_{\boldsymbol{\omega}}$ on the underlying unknown parameter $\boldsymbol{\mu}$ impedes a straightforward solution. For the common and simple case with a scalar regularization parameter ω, this dependency problem also happens and is well-studied in the literature, for instance, the author in [5] used SURE method to estimate the term involved with the unknown parameter $\boldsymbol{\mu}$, while in [6] based on subspace information criterion. In the context of Tikhonov regularization, methods based on like, the discrepancy principle [7,8], the L-curve criterion [9] and the generalized cross-validation [10], are used to select the regularization parameter.

However, these methods are designed for scalar regularization parameter selection. Instead of solving (1), we consider a worst case scenario and focus on its *minmax* formulation

$$\boldsymbol{\omega}^* := \underset{\boldsymbol{\omega}\in\mathcal{W}}{\arg\min}\ \max_{\boldsymbol{\mu}\in\mathcal{M}}\ f_{\boldsymbol{\omega}}(\boldsymbol{y};\boldsymbol{\mu}). \tag{2}$$

Here, set $\mathcal{M}$ defines the restrictions on the unknown parameter $\boldsymbol{\mu}$ that are often easy to get. For mathematical tractability, $\mathcal{M}$ is considered convex and compact. The study of (2) requires in general assumptions that could be violated in practice. Thus, it is always advisable to restrict this study to loss functions derived from particular problem instances. In our case, we analyze (2) for the problem of *graph signal reconstruction* [11–13]. The goal of this task is simple: *given a set of noisy observations taken over nodes of a graph, to reconstruct the graph signal using the underlying graph structure as prior knowledge.*

Different works have leveraged this problem for graph signal denoising [14–16], interpolation [17] and semi-supervised learning over graphs [18], where the most common regularizer is the so-called Tikhonov regularizer. The common assumption in all these works is that the regularization parameter is scalar. This, in turn, translates into a common factor applied by all nodes of the graph and fails to capture the signal detail in the neighborhood of a node. The specific local information of certain nodes cannot be taken into consideration.

To tackle this issue and improve the reconstruction performance, we consider $\boldsymbol{\omega}$ to be a vector of regularization parameters where each entry is associated to a node of the graph. We call it *node varying regularization* since each node is regularized differently. The proposed generalization matches naturally the form in (2) by particularizing the loss function to the mean squared error. We provide a gradient descent-based algorithm to find the optimal node varying regularization parameters and show its superior performance compared to the scalar regularization.

In the following, we formalize the problem of graph signal reconstruction in Section 2. In Section 3, we develop the node varying regularization problem, while in Section 4 we focus on the minmax design of the regularization parameter. Numerical results with synthetic and real data corroborate our theory in Section 5 and the paper conclusions are drawn in Section 6.

2. GRAPH SIGNAL RECONSTRUCTION

Let $\boldsymbol{y} \in \mathbb{R}^n$ be a vector of measurements taken over an undirected graph of n nodes, where entry $y_i := [\boldsymbol{y}]_i$ is the measurement collected on the ith node. The node measurements are of the form

$$y_i = \mu_i + \epsilon_i, \tag{3}$$

where μ_i is the mean of the ith measurement and ϵ_i is an i.i.d. random Gaussian variable distributed as $\mathcal{N}(0, \sigma_i)$. That is, the measurement vector follows the distribution $\mathcal{N}(\boldsymbol{\mu}, \boldsymbol{\Sigma})$ with covariance matrix $\boldsymbol{\Sigma} := \operatorname{diag}(\sigma_1, \ldots, \sigma_n)$. Further, let $\boldsymbol{L} \in \mathbb{R}^{n\times n}$ be the *graph Laplacian* capturing the connectivity between nodes, defined as $\boldsymbol{D} - \boldsymbol{A}$, where $\boldsymbol{D}$ is the degree matrix and $\boldsymbol{A}$ is the adjacency matrix. The Laplacian is symmetric for undirected graphs .

The problem of graph signal reconstruction consists of estimating the noise-free signal $\boldsymbol{\mu}$ from the noisy measurements $\boldsymbol{y}$. A common

This work has been submitted to EUSIPCO, 2020, currently under review. This research is supported in part by the ASPIRE project (project 14926 within the STW OTP programme), financed by the Netherlands Organization for Scientific Research (NWO). Mario Coutino is partially supported by CONACYT. Maosheng Yang's study is supported by Microelectronic-MSc track scholarship from TU Delft.

approach to solve this problem is to explore the prior information about the signal behavior over the graph [14,15]. If we assume the signal changes slowly over, we can consider the Tikhonov regularization problem

$$\hat{\boldsymbol{\mu}}_\omega := \underset{\boldsymbol{x}\in\mathbb{R}^n}{\arg\min}\ \|\boldsymbol{y}-\boldsymbol{x}\|_2^2 + \omega\cdot\boldsymbol{x}^\top\boldsymbol{L}\boldsymbol{x}, \tag{4}$$

where scalar $\omega > 0$ is the regularization parameter that controls the signal's *smooth variation* over the network topology [15]. Problem (4) is convex and has the closed-form solution

$$\hat{\boldsymbol{\mu}}_\omega = (\boldsymbol{I}+\omega\boldsymbol{L})^{-1}\boldsymbol{y} := \boldsymbol{H}_\omega\boldsymbol{y} \tag{5}$$

where we defined $\boldsymbol{H}_\omega := (\boldsymbol{I}+\omega\boldsymbol{L})^{-1}$ to ease notation. The mean square error (MSE) of the estimate in (5) is

$$\text{MSE}(\hat{\boldsymbol{\mu}}_\omega) = \text{tr}\big((\boldsymbol{I}-\boldsymbol{H}_\omega)^2\boldsymbol{\mu}\boldsymbol{\mu}^\top\big) + \text{tr}\big(\boldsymbol{H}_\omega^2\boldsymbol{\Sigma}\big). \tag{6}$$

By minimizing the MSE over ω, we can find the optimal regularization parameter ω by solving

$$\omega^* := \underset{\omega}{\arg\min}\ \text{MSE}(\hat{\boldsymbol{\mu}}_\omega). \tag{7}$$

Expression (7) is in the form (1) with $f(\boldsymbol{y};\boldsymbol{\mu}) := \text{MSE}(\hat{\boldsymbol{\mu}}_\omega)$, where now $\boldsymbol{\mu}$ is the unknown noise-free measurement to be estimated. To deal with the dependency on the unknown parameter $\boldsymbol{\mu}$ of the MSE, the work in [14] substituted the MSE cost in (7) with an upper bound. Although this approach can serve to design the regularization parameter ω through order-matching, it falls short (i) in instances where each node is weighted differently, since in (4) all nodes are weighted with a common ω; and (ii) to provide a method without spectral knowledge as its optimal design requires full eigendecomposition of the Laplacian matrix. Therefore, in the following, we propose a generalized graph-based penalizer able to capture node heterogeneity while allowing for a computationally efficient design.

3. NODE VARYING GRAPH SIGNAL RECONSTRUCTION

To address the problem of node heterogeneity, we consider a parameter vector $\boldsymbol{\omega}\in\mathbb{R}^n$, where entry ω_i is associated to node i. The node varying equivalent to (4) can be written as

$$\hat{\boldsymbol{\mu}}_{\boldsymbol{\omega}} := \underset{\boldsymbol{x}\in\mathbb{R}^n}{\arg\min}\ \|\boldsymbol{y}-\boldsymbol{x}\|_2^2 + \boldsymbol{x}^\top\text{diag}(\boldsymbol{\omega})\boldsymbol{L}\text{diag}(\boldsymbol{\omega})\boldsymbol{x}. \tag{8}$$

The term $\boldsymbol{x}' = \text{diag}(\boldsymbol{\omega})\boldsymbol{x}$ can be seen as each node weighting accordingly its own signal before computing the signal variation $\boldsymbol{x}'^\top\boldsymbol{L}\boldsymbol{x}'$. Further, note that by setting $\boldsymbol{\omega}=\omega\boldsymbol{1}$, problem (8) specializes to (4).

Define $\boldsymbol{S}_{\boldsymbol{\omega}} := \text{diag}(\boldsymbol{\omega})\boldsymbol{L}\text{diag}(\boldsymbol{\omega}) = \boldsymbol{\omega}\boldsymbol{\omega}^\top\odot\boldsymbol{L}$ as an edge dependent matrix [19], which is positive semidefinite by Schur product theorem [20, p. 14, Thm. VII] and shares the same support with $\boldsymbol{L}$. Problem (8) is convex by construction since $\text{diag}(\boldsymbol{\omega})\boldsymbol{L}\text{diag}(\boldsymbol{\omega})$ is positive semidefinite. By setting the gradient of (8) to zero, the optimal closed-form solution for (8) is

$$\hat{\boldsymbol{\mu}}_{\boldsymbol{\omega}} := (\boldsymbol{I}+\boldsymbol{S}_{\boldsymbol{\omega}})^{-1}\boldsymbol{y}. \tag{9}$$

The MSE of the estimate in (9) is now given by

$$\text{MSE}(\hat{\boldsymbol{\mu}}_{\boldsymbol{\omega}}) := \text{tr}\big((\boldsymbol{I}-(\boldsymbol{I}+\boldsymbol{S}_{\boldsymbol{\omega}})^{-1})^2\boldsymbol{\mu}\boldsymbol{\mu}^\top\big) + \text{tr}\big((\boldsymbol{I}+\boldsymbol{S}_{\boldsymbol{\omega}})^{-2}\boldsymbol{\Sigma}\big). \tag{10}$$

It consists of the squared norm of the bias as the first term and the variance as the second term. Likewise the scalar counter part (6), the $\text{MSE}(\hat{\boldsymbol{\mu}}_{\boldsymbol{\omega}})$ depends on the unknown parameter $\boldsymbol{\mu}$. To tackle this dependency and design the regularization parameter $\boldsymbol{\omega}$, we depart from approaches of the form (7) and consider a minmax formulation as in (2). Beside tackling the dependency on the underlying parameter $\boldsymbol{\mu}$, the minmax formulation also avoids working with upper bounds.

4. MINMAX PARAMETER DESIGN

The minmax formulation for the optimal regularization parameter design is

$$\hat{\boldsymbol{\omega}} := \underset{\boldsymbol{\omega}\in\mathcal{W}}{\arg\min}\ \max_{\boldsymbol{\mu}\in\mathcal{M}}\ f_{\boldsymbol{\omega}}(\boldsymbol{y};\boldsymbol{\mu}), \tag{11}$$

where $f_{\boldsymbol{\omega}}(\boldsymbol{y};\boldsymbol{\mu}) := \text{MSE}(\hat{\boldsymbol{\mu}}_{\boldsymbol{\omega}})$ and where $\mathcal{W}$ and $\mathcal{M}$ are two sets to be specified in the sequel. In a practical setting, no much information is available about the unknown parameter $\boldsymbol{\mu}$; however its energy (norm) is typically bounded. For instance, an energy bound on the measurements $\boldsymbol{y}$ will simply impose a (may not tight) bound on $\boldsymbol{\mu}$; or if the signal-to-noise ratio (SNR) is available, through the knowledge of the noise power, we can obtain a bound on the signal power. Hence, it is reasonable to consider that $\boldsymbol{\mu}$ lies within an ℓ_2-norm ball with radius μ_*, i.e., $\mathcal{M} := \{\boldsymbol{\mu} : \|\boldsymbol{\mu}\|_2 \le \mu_*\}$. Set $\mathcal{M}$ meets all assumptions required by minmax problems, i.e., it is convex and compact. In addition, to preserve the convexity of problem (8), we only require the regularizer parameters to be within the real set, $\mathcal{W}=\mathbb{R}^n$, so that $\text{diag}(\boldsymbol{\omega})\boldsymbol{L}\text{diag}(\boldsymbol{\omega})$ is positive semi-definite.

Before studying the details of (11), let us first analyze the $\text{MSE}(\hat{\boldsymbol{\mu}}_{\boldsymbol{\omega}})$ expression in (10). We observe that only the first term depends on $\boldsymbol{\mu}$. This term captures the squared norm of the bias of the estimator, $\|\text{bias}(\hat{\boldsymbol{\mu}}_{\boldsymbol{\omega}})\|^2$ and can be written in the quadratic form

$$\|\text{bias}(\hat{\boldsymbol{\mu}}_{\boldsymbol{\omega}})\|^2 = \text{tr}(\tilde{\boldsymbol{S}}_{\boldsymbol{\omega}}\boldsymbol{\mu}\boldsymbol{\mu}^\top) = \boldsymbol{\mu}^\top\tilde{\boldsymbol{S}}_{\boldsymbol{\omega}}\boldsymbol{\mu}, \tag{12}$$

where $\tilde{\boldsymbol{S}}_{\boldsymbol{\omega}} = (\boldsymbol{I}-(\boldsymbol{I}+\boldsymbol{S}_{\boldsymbol{\omega}})^{-1})^2$ is a positive definite matrix that depends on $\boldsymbol{\omega}$. For a fixed $\boldsymbol{\omega}$, the bias term accepts a simple maximization when $\boldsymbol{\mu}$ is restricted to set $\mathcal{M}$. That is, it suffices to find the eigenvector $\boldsymbol{\mu}^*$ corresponding to the largest eigenvalue of $\tilde{\boldsymbol{S}}_{\boldsymbol{\omega}}$, more specifically

$$\boldsymbol{\mu}^* := \underset{\boldsymbol{\mu}\in\mathcal{M}}{\arg\max}\ \boldsymbol{\mu}^\top\tilde{\boldsymbol{S}}_{\boldsymbol{\omega}}\boldsymbol{\mu} = \mu_*\lambda_{\max}(\tilde{\boldsymbol{S}}_{\boldsymbol{\omega}}), \tag{13}$$

where $\lambda_{\max}(\tilde{\boldsymbol{S}}_{\boldsymbol{\omega}})$ denotes the largest eigenvalue of $\tilde{\boldsymbol{S}}_{\boldsymbol{\omega}}$ and μ_* is the energy upper bound. The following result ensures that the solution of (13) can be obtained efficiently.

Proposition 1. *Let $\tilde{\boldsymbol{S}}_{\boldsymbol{\omega}}$ be given as above, then the maximizer of* (13) *is the eigenvector of $\boldsymbol{S}_{\boldsymbol{\omega}}$ related with $\lambda_{\max}(\boldsymbol{S}_{\boldsymbol{\omega}})$.*

Proof. From their definition, both $\tilde{\boldsymbol{S}}_{\boldsymbol{\omega}}$ and $\boldsymbol{S}_{\boldsymbol{\omega}}$ are positive semidefinite, thus $\boldsymbol{S}_{\boldsymbol{\omega}}+\boldsymbol{I}\succeq\boldsymbol{I}$. As the ith eigenvalue of $\tilde{\boldsymbol{S}}_{\boldsymbol{\omega}}$ is given as $\lambda_i(\tilde{\boldsymbol{S}}_{\boldsymbol{\omega}}) := (1-(1+\lambda_i(\boldsymbol{S}_{\boldsymbol{\omega}}))^{-1})^2$, we conclude that $\max_i\lambda_i(\tilde{\boldsymbol{S}}_{\boldsymbol{\omega}}) = \max_i\lambda_i(\boldsymbol{S}_{\boldsymbol{\omega}})$.

□

Hence, the solution of (13) can be computed efficiently through power iteration using the matrix $\boldsymbol{S}_{\boldsymbol{\omega}}$ which enjoys the sparsity of the graph Laplacian, instead of $\tilde{\boldsymbol{S}}_{\boldsymbol{\omega}}$.

Since the inner maximization step can be solved exactly, the outer minimization can be performed with an iterative first-order method [21]. A first-order based method for solving the minmax problem (11) has the update

$$\boldsymbol{\omega}_{t+1} = \boldsymbol{\omega}_t - \eta_t\nabla_{\boldsymbol{\omega}}f_{\boldsymbol{\omega}_t}(\boldsymbol{y};\boldsymbol{\mu}_t^*), \tag{14}$$

where η_t is the step size at iteration t and $\nabla_{\boldsymbol{\omega}}f_{\boldsymbol{\omega}_t}$ is the gradient of f w.r.t $\boldsymbol{\omega}$ evaluated at $\boldsymbol{\omega}_t$. Here, $\boldsymbol{\mu}_t^* := \mu_*\boldsymbol{v}_t$ with $\boldsymbol{v}_t$ being the normalized eigenvector related with the largest eigenvalue of $\boldsymbol{S}_{\boldsymbol{\omega}_t} = \text{diag}(\boldsymbol{\omega}_t)\boldsymbol{L}\text{diag}(\boldsymbol{\omega}_t)$. With derivations in appendix[1], the

[1] Available online at this_link

closed-form expression of the gradient $\nabla_{\boldsymbol{\omega}} f_{\boldsymbol{\omega}_t}(\boldsymbol{y};\boldsymbol{\mu}_t^*)$ is

$$\nabla_{\boldsymbol{\omega}} f_{\boldsymbol{\omega}_t}(\boldsymbol{y};\boldsymbol{\mu}_t^*) = \mathrm{diag}^{-1}\Big([-4(\boldsymbol{I}+\boldsymbol{S}_{\boldsymbol{\omega}_t})^{-2}(\boldsymbol{\mu}_t^*\boldsymbol{\mu}_t^{*\top}+\boldsymbol{\Sigma}) + 4(\boldsymbol{I}+\boldsymbol{S}_{\boldsymbol{\omega}_t})^{-1}\boldsymbol{\mu}_t^*\boldsymbol{\mu}_t^{*\top}](\boldsymbol{I}+\boldsymbol{S}_{\boldsymbol{\omega}_t})^{-1}\mathrm{diag}(\boldsymbol{\omega}_t)\boldsymbol{L}\Big). \quad (15)$$

Although it seems that evaluating (15) requires the inversion of some matrices, these operations can be implemented efficiently as the solution to symmetric diagonal dominant (SDD) systems [22, 23]. Algorithm 1 summarizes the minmax procedure for solving (11).

To study the theoretical guarantees of this Algorithm, we introduce the following definition.

Definition 1 (FNE). *A point $(\boldsymbol{\omega}^*;\boldsymbol{\mu}^*)$ is a first-order Nash equilibrium (FNE) of the game* (11) *if*

$$\langle \nabla_{\boldsymbol{\omega}} f(\boldsymbol{\omega}^*;\boldsymbol{\mu}^*), \boldsymbol{\omega}-\boldsymbol{\omega}^*\rangle \geq 0,\ \forall\ \boldsymbol{\omega}\in\mathcal{W} \quad (16)$$

and

$$\langle \nabla_{\boldsymbol{\mu}} f(\boldsymbol{\omega}^*;\boldsymbol{\mu}^*), \boldsymbol{\mu}-\boldsymbol{\mu}^*\rangle \leq 0,\ \forall\ \boldsymbol{\mu}\in\mathcal{M} \quad (17)$$

where $\langle\cdot,\cdot\rangle$ denotes the inner product.

This definition guarantees first-order necessary optimality conditions for the objective function (for each player). Hence, they are necessary conditions to guarantee the so-called first-order Nash equilibrium [24], i.e.,

$$f(\boldsymbol{\omega}^*;\boldsymbol{\mu}) \leq f(\boldsymbol{\omega}^*;\boldsymbol{\mu}^*) \leq f(\boldsymbol{\omega};\boldsymbol{\mu}^*),\ \forall\ \boldsymbol{\omega}\in\mathcal{W},\ \forall\boldsymbol{\mu}\in\mathcal{M}. \quad (18)$$

As Algorithm 1 is an iterative method and in a practical setting it always has a numerical tolerance, in the following, the notion of approximate-FNE is introduced.

Definition 2 (Approximate FNE). *A point $(\boldsymbol{\omega}^*;\boldsymbol{\mu}^*)$ is an ϵ-first-order Nash equilibrium (ϵ-FNE) of the game* (11) *if*

$$\mathcal{X}(\boldsymbol{\omega}^*;\boldsymbol{\mu}^*) \leq \epsilon \text{ and } \mathcal{Y}(\boldsymbol{\omega}^*;\boldsymbol{\mu}^*) \leq \epsilon, \quad (19)$$

where

$$\begin{aligned}\mathcal{X}(\boldsymbol{\omega}^*;\boldsymbol{\mu}^*) := -\min_{\boldsymbol{\omega}} &\langle \nabla_{\boldsymbol{\omega}} f(\boldsymbol{\omega}^*;\boldsymbol{\mu}^*), \boldsymbol{\omega}-\boldsymbol{\omega}^*\rangle \\ &\text{s.t. } \boldsymbol{\omega}\in\mathcal{W}, \|\boldsymbol{\omega}-\boldsymbol{\omega}^*\|\leq 1,\end{aligned} \quad (20)$$

and

$$\begin{aligned}\mathcal{Y}(\boldsymbol{\omega}^*;\boldsymbol{\mu}^*) := \max_{\boldsymbol{\mu}} &\langle \nabla_{\boldsymbol{\mu}} f(\boldsymbol{\omega}^*;\boldsymbol{\mu}^*), \boldsymbol{\mu}-\boldsymbol{\mu}^*\rangle \\ &\text{s.t. } \boldsymbol{\mu}\in\mathcal{M}, \|\boldsymbol{\mu}-\boldsymbol{\mu}^*\|\leq 1.\end{aligned} \quad (21)$$

This definition is based on the first-order optimality measure of the objective of each variable. Such a condition guarantees that each variable cannot improve their objective function by using first-order information, providing a both theoretical and numerically meaningful stopping criteria i.e., convergence criterion. Now, we are ready to state the following result regarding the convergence of Algorithm 1.

Proposition 2 (Convergence). *Let the problem* (13) *have a unique solution for each $\boldsymbol{\omega}_t$. Then, Algorithm 1 is guaranteed to converge to an ϵ-FNE of the game* (11) *for $T, K\to\infty$.*

Algorithm 1 Iterative First-Order Method for MinMax Game (11)

```
Input: $\mu^*$: signal energy bound; $T$: number of gradient descent iterations; $K$: number of power iterations; $\eta$: step size; $\boldsymbol{\Sigma}$: noise covariance matrix; $\boldsymbol{L}$: graph Laplacian matrix;
Output: optimized regularization parameter $\boldsymbol{\omega}$
Initialization : $\boldsymbol{\omega}_0 = \omega_0\mathbf{1}$, $\boldsymbol{\mu}_0 = \mathbf{1}$
for $t = 0$ to $T-1$ do
    $\boldsymbol{S}_{\boldsymbol{\omega}_t} = \mathrm{diag}(\boldsymbol{\omega}_t)\boldsymbol{L}\mathrm{diag}(\boldsymbol{\omega}_t)$
    for $k = 0$ to $K-1$ do
        $\mathbf{z}_k = \boldsymbol{S}_{\boldsymbol{\omega}_t}\boldsymbol{\mu}_k$
        $\boldsymbol{\mu}_{k+1} = \mu^*\mathbf{z}_k/\|\mathbf{z}_k\|_2$
    end for
    $\boldsymbol{\mu}_t^* = \boldsymbol{\mu}_K$, the largest eigenvector of $\boldsymbol{S}_{\boldsymbol{\omega}_t}$
    Compute $\nabla_{\boldsymbol{\omega}} f_{\boldsymbol{\omega}_t}$ as (15) by substituting $\boldsymbol{\Sigma}$ and $\boldsymbol{\mu}_t^*$
    $\boldsymbol{\omega}_{t+1} = \boldsymbol{\omega}_t - \eta\nabla_{\boldsymbol{\omega}} f_{\boldsymbol{\omega}_t}(\boldsymbol{y};\boldsymbol{\mu}_t^*)$
end for
return $\boldsymbol{\omega} = \boldsymbol{\omega}_T$
```

Proof. (Sketch.) When the stated condition holds, then Danskin's theorem [25] holds, i.e.,

$$\nabla_{\boldsymbol{\omega}} \max_{\boldsymbol{\mu}\in\mathcal{M}} f_{\boldsymbol{\omega}}(\boldsymbol{y};\boldsymbol{\mu}) = \nabla_{\boldsymbol{\omega}} f_{\boldsymbol{\omega}}(\boldsymbol{y};\boldsymbol{\mu}^*) \quad (22)$$

with $\boldsymbol{\mu}^* = \arg\max\ f_{\boldsymbol{\omega}}(\boldsymbol{y};\boldsymbol{\mu})$ and for $K\to\infty$, the power method finds the exact solution to (13). The rest of the proof specializes the result in [21, Thm. 3.4] □

Although the uniqueness condition for problem (13) might seem restrictive, this behaviour is typically observed in practice. Furthermore, even when this is not the case, we can consider a proximal term $\|\boldsymbol{\mu}-\bar{\boldsymbol{\mu}}\|_2$ in (13) to guarantee the convergence to an ϵ-FNE; see [21] for further technical details. Next, we use the proposed method to design a robust worst-case regularizer for node varying graph signal reconstruction.

5. NUMERICAL EXPERIMENTS

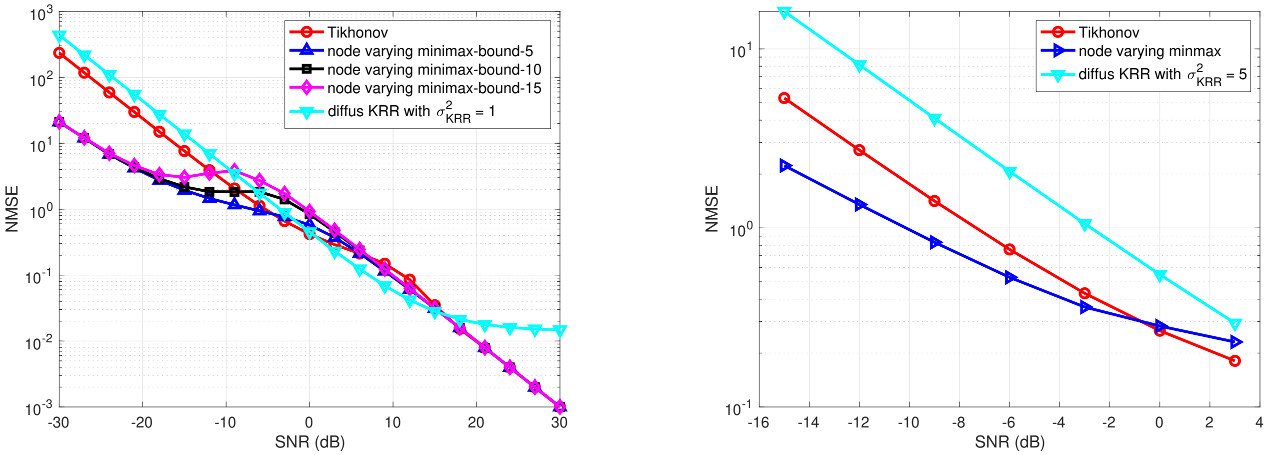
In this section, we corroborate the proposed method with synthetic data on Erdős–Rényi graphs and with real data from the Molene weather dataset[2]. We first obtained an optimal regularization parameter $\boldsymbol{\omega}^*$ by solving the minmax problem (11) with Algorithm 1. Then, we used the found regularization parameter to reconstruct the signal as in (9). We compared our method with two other state-of-the-art approaches:

i) the standard Tikhonov based denoising (4). Based on the bias-variance trade-off study and scaling law in [14], we optimally set the regularization parameter $\omega = \mathcal{O}(\sqrt{\theta/(\lambda_2\lambda_n)})$, where $\theta = \sqrt{1/\text{SNR}}$, and λ_2 and λ_n are the smallest and the largest non-trivial eigenvalues of the graph Laplacian, respectively.

ii) Diffusion kernel-based ridge regression with parameter $\sigma_{KRR}^2 = 1$ (diffusion kernel parameter) and regularization parameter $c = 10^{-4}$, which is studied well in kernel-based graph signal reconstruction [13].

We measured the performance through the normalized mean squared error (NMSE), which is defined as $\text{NMSE} = \|\boldsymbol{\mu}-\hat{\boldsymbol{\mu}}\|^2/\|\boldsymbol{\mu}\|^2$. In our experiments, we analyzed under different signal-to-noise ratios (SNRs). The true signal is corrupted with white Gaussian noise to

[2] https://donneespubliques.meteofrance.fr/donneeslibres/Hackathon/RADOMEH.tar.gz



(a) Synthetic data.

(b) Molene weather dataset.

Fig. 1: Normalize mean squared error comparison of different methods as a function of the signal-to-noise ratio.

yield an SNR given by SNR $= 10\log_{10}(\|\boldsymbol{\mu}\|^2/(n\sigma^2))$ with n being the number of graph nodes and σ^2 the noise variance.

Synthetic data. We built an Erdős–Rényi graph of $n = 50$ nodes with a connection probability of 0.5. We generated a deterministic smooth graph signal which has an ideal low-pass graph frequency content with bandwidth 20 [15]. We observed the ℓ_2-norm $\|\boldsymbol{\mu}\|_2 = 4.47$. For the gradient descent based method in Algorithm 1, we set the number of iterations $T = 100$, $K = 30$, and the step size $\eta = 0.002$. We then initialized $\boldsymbol{\omega} = \omega_0\mathbf{1}$ with ω_0 being the optimal Tikhonov regularization parameter, and $\boldsymbol{\mu} = \mathbf{1}$. To evaluate the recovery performance in different noisy situations, we considered an SNR in the range $[-30, 30]$ dB. To evaluate the effect of the energy bound on the reconstruction performance, we considered the energy bound μ_* to have three values $\{5, 10, 15\}$, which are all above the true energy. Our results are averaged over 100 noise realizations and 50 graph realizations for a total of 5000 Monte-Carlo runs.

From Fig. 1a, the performance of our minmax formulation stands out. Specifically, with any energy bound –whether a stricter one ($\mu_* = 5$) or a looser one ($\mu_* = 15$)– the proposed method gives better results compared with the other contenders in the low-SNR regimes. In the medium-SNR range, our method with a loose energy bound will behave worse than Tikhonov, but this difference becomes negligible when the energy bound gets tighter. Finally, at high-SNR regime all methods reach a similar performance except for the diffusion kernel-based method that may have a bias. This trend shows that our method generalizes the Tikhonov regularization and indicates that local node detail is more important in low SNR settings.

Molene weather data. This dataset consists of 744 hourly temperature recordings collected in January 2014 over 32 cities in the region of Brest, France. We built the graph from the coordinates of the stations by connecting all the neighbours in a given radius with a weight $\mathbf{W}_G(i, j) = \exp\{-kd^2(i, j)\}$, where $d(i, j)$ is the Euclidean distance between stations i and j, and parameter k is five. We removed the average value of weather data over time and space. For this experiment, we artificially added noise and considered an SNR in the interval -15 to 3 dB. We set the energy bound to the true one plus a trivial deviation, which is a reasonable assumption based on historical same-day recordings. For the diffusion kernel method, we here modified the parameter σ^2_{KRR} to be 5 for a better performance. The other parameters remain the same as in the former experiment.

Fig. 1b shows the performance of the three different methods. This result shows that the proposed node varying approach should be considered in harsher settings. When the SNR improves (i.e., the data matches better the true one) regularization is less needed as it biases the results. However, likewise in the synthetic dataset, our method yields a superior performance in low SNR regimes.

6. CONCLUSIONS

In this paper, we proposed a node varying regularizer for reconstructing graph signals. The method considers a vector of regularizer parameters where each entry is associated to a specific node. As such, it generalizes state-of-the-art regularizers which consider the same scalar parameter for all nodes. To design the regularization that minimizes the MSE, we develop a minmax approach that tackles the dependency issue on the unknown parameter. By levering results from the first-order Nash equilibrium, we provide a gradient-descent algorithm to obtain the optimal regularization parameter vector with convergence guarantees. Numerical results with synthetic and real data corroborate our findings and show the proposed method outperforms the state-of-the-art methods, especially when the signal-to-noise ratio is low.

7. REFERENCES

[1] D. J. Toal, N. W. Bressloff, and A. J. Keane, “Kriging hyper-parameter tuning strategies,” *AIAA journal*, vol. 46, no. 5, pp. 1240–1252, 2008.

[2] P. Stoica and R. L. Moses, “On biased estimators and the unbiased cramér-rao lower bound,” *Signal Processing*, vol. 21, no. 4, pp. 349–350, 1990.

[3] A. Buades, B. Coll, and J.-M. Morel, “A review of image denoising algorithms, with a new one,” *Multiscale Modeling & Simulation*, vol. 4, no. 2, pp. 490–530, 2005.

[4] J. Haupt and R. Nowak, "Signal reconstruction from noisy random projections," *IEEE Transactions on Information Theory*, vol. 52, no. 9, pp. 4036–4048, 2006.

[5] S. Kay and Y. C. Eldar, "Rethinking biased estimation [lecture notes]," *IEEE Signal Processing Magazine*, vol. 25, no. 3, pp. 133–136, 2008.

[6] M. Sugiyama and H. Ogawa, "Optimal design of regularization term and regularization parameter by subspace information criterion," *Neural networks : the official journal of the International Neural Network Society*, vol. 15, no. 3, p. 349â361, April 2002. [Online]. Available: https://doi.org/10.1016/s0893-6080(02)00022-9

[7] O. Scherzer, "The use of morozov's discrepancy principle for tikhonov regularization for solving nonlinear ill-posed problems," *Computing*, vol. 51, no. 1, pp. 45–60, 1993. [Online]. Available: https://doi.org/10.1007/BF02243828

[8] S. W. Anzengruber and R. Ramlau, "Morozovs discrepancy principle for tikhonov-type functionals with nonlinear operators," *Inverse Problems*, vol. 26, no. 2, p. 025001, Dec 2009.

[9] P. C. Hansen and D. P. O'Leary, "The use of the l-curve in the regularization of discrete ill-posed problems," *SIAM Journal on Scientific Computing*, vol. 14, no. 6, pp. 1487–1503, 1993. [Online]. Available: https://doi.org/10.1137/0914086

[10] G. H.Golub and U. von Matt, "Generalized cross-validation for large-scale problems," *Journal of Computational and Graphical Statistics*, vol. 6, no. 1, pp. 1–34, 1997. [Online]. Available: https://amstat.tandfonline.com/doi/abs/10.1080/10618600.1997.10474725

[11] V. Cevher, P. Indyk, L. Carin, and R. G. Baraniuk, "Sparse signal recovery and acquisition with graphical models," *IEEE Signal Processing Magazine*, vol. 27, no. 6, pp. 92–103, 2010.

[12] S. Segarra, A. G. Marques, G. Leus, and A. Ribeiro, "Reconstruction of graph signals through percolation from seeding nodes," *IEEE Transactions on Signal Processing*, vol. 64, no. 16, pp. 4363–4378, 2016.

[13] D. Romero, M. Ma, and G. B. Giannakis, "Kernel-based reconstruction of graph signals," *IEEE Transactions on Signal Processing*, vol. 65, no. 3, pp. 764–778, Feb 2017.

[14] P. Chen and S. Liu, "Bias-variance tradeoff of graph laplacian regularizer," *IEEE Signal Processing Letters*, vol. 24, no. 8, pp. 1118–1122, Aug 2017.

[15] D. I. Shuman, S. K. Narang, P. Frossard, A. Ortega, and P. Vandergheynst, "The emerging field of signal processing on graphs: Extending high-dimensional data analysis to networks and other irregular domains," *IEEE Signal Processing Magazine*, vol. 30, no. 3, pp. 83–98, May 2013.

[16] J. Pang, G. Cheung, A. Ortega, and O. C. Au, "Optimal graph laplacian regularization for natural image denoising," in *2015 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*. IEEE, 2015, pp. 2294–2298.

[17] S. K. Narang, A. Gadde, and A. Ortega, "Signal processing techniques for interpolation in graph structured data," in *2013 IEEE International Conference on Acoustics, Speech and Signal Processing*. IEEE, 2013, pp. 5445–5449.

[18] A. Gadde, A. Anis, and A. Ortega, "Active semi-supervised learning using sampling theory for graph signals," in *Proceedings of the 20th ACM SIGKDD international conference on Knowledge discovery and data mining*, 2014, pp. 492–501.

[19] M. Coutino, E. Isufi, and G. Leus, "Advances in distributed graph filtering," *IEEE Transactions on Signal Processing*, vol. 67, no. 9, pp. 2320–2333, May 2019.

[20] J. Schur, "Bemerkungen zur theorie der beschränkten bilinearformen mit unendlich vielen veränderlichen." *Journal für die reine und angewandte Mathematik*, vol. 140, pp. 1–28, 1911. [Online]. Available: http://eudml.org/doc/149352

[21] M. Nouiehed, M. Sanjabi, T. Huang, J. D. Lee, and M. Razaviyayn, "Solving a class of non-convex min-max games using iterative first order methods," in *Advances in Neural Information Processing Systems*, 2019, pp. 14 905–14 916.

[22] J. A. Kelner, L. Orecchia, A. Sidford, and Z. A. Zhu, "A simple, combinatorial algorithm for solving sdd systems in nearly-linear time," in *Proceedings of the forty-fifth annual ACM symposium on Theory of computing*, 2013, pp. 911–920.

[23] D. A. Spielman and S.-H. Teng, "Nearly-linear time algorithms for graph partitioning, graph sparsification, and solving linear systems," in *Proceedings of the thirty-sixth annual ACM symposium on Theory of computing*, 2004, pp. 81–90.

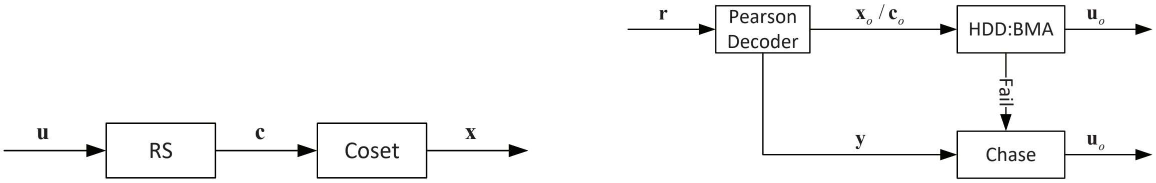
[24] C. Jin, P. Netrapalli, and M. I. Jordan, "Minmax optimization: Stable limit points of gradient descent ascent are locally optimal," *arXiv preprint arXiv:1902.00618*, 2019.

[25] J. M. Danskin, "The theory of max-min, with applications," *SIAM Journal on Applied Mathematics*, vol. 14, no. 4, pp. 641–664, 1966. [Online]. Available: https://doi.org/10.1137/0114053

Decoding of Concatenated Codes for Noisy Channels With Unknown Offset

Renfei Bu Jos Weber
Delft University of Technology
Applied Mathematics Dept., Optimization Group
R.Bu@tudelft.nl J.H.Weber@tudelft.nl

In communication and storage systems, noise and interference are not the only disturbances during the data transmission, sometimes the error performance is also seriously degraded by offset mismatch. We consider a simple channel such that the received signal is distorted by noise and offset mismatch, that is, $\mathbf{r} = \mathbf{x}+\mathbf{v}+b\mathbf{1}$, where $\mathbf{x} = (x_1, x_2, \ldots, x_n)$ is the transmitted codeword from a codebook, $\mathbf{v} = (v_1, v_2, \ldots, v_n) \in \mathbb{R}^n$ is the noise vector, where the v_i are independently normally distributed with mean 0 and standard deviation σ, b is a real number representing the channel offset, $\mathbf{1}$ is the real all-one vector $(1, \ldots, 1)$ of length n, and $\mathbf{r} \in \mathbb{R}^n$ is the received vector. Minimum modified Pearson distance (MMPD) detection has been proposed [1] as an alternative to minimum Euclidean distance (MED) detection to counter the effects of offset mismatch. A major concern, however, is the fact that the evaluation of MMPD is an exhaustive search over all candidate codewords which is infeasible for large codes. Various block codes have been proposed [2] to get good performance for channels with both noise and offset if the MMPD detection is used.



(a) Block diagram of the RS-Coset encoder

(b) Block diagram of the decoding algorithm

Figure 1: Block diagram of concatenated scheme.

In this work, a concatenated coding scheme is proposed for noisy channels with unknown offset mismatch. The concatenation is between a Reed-Solomon (RS) code and a certain coset of a binary block code proposed in [2]. The two codes are chosen according to a rule that the inner code is of a short length such that an exhaustive search of MMPD is for relatively small codes. The encoder block diagram of RS-Coset codes is shown in Figure 1(a). A message vector $\mathbf{u}$ is encoded to a RS codeword $\mathbf{c}$. Then it will be converted into a binary sequence and mapped to a codeword $\mathbf{x}$ in a coset of a block code. A novel soft decoding algorithm shown in Figure 1(b) for the concatenated scheme is proposed. The MMPD detection is used to decode the inner code that guarantees immunity to channel offset mismatch. Its output will be given to a two-stage hybrid decoding algorithm for the outer RS code. The first stage of decoding carries out an algebraic hard-decision decoding (HDD) algorithm, such as the Berlekamp-Massey algorithm (BMA), to the outer RS code. If the HDD declares a successful decoding, then the algorithm outputs the decoded codeword and terminates the decoding process. Otherwise, the decoding is continued with the second stage of decoding, using the reduced test-pattern Chase algorithm. The reliability information $\mathbf{y}$ used in Chase algorithm is obtained from a subtraction between the received vector and an estimated offset, where a dynamic threshold estimation [3] is explored.

As an example, the performance of the proposed concatenation scheme that implements $(7, 3, 5)$ RS code as the outer code and the coset of binary $(6, 3, 3)$ code as the

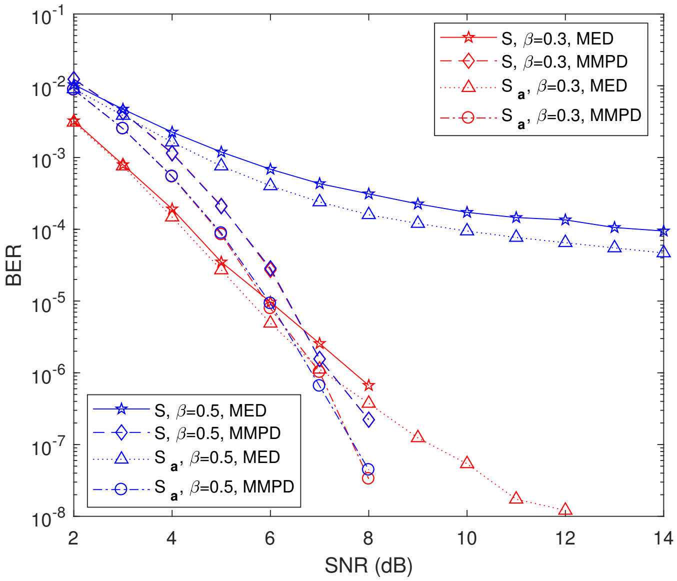
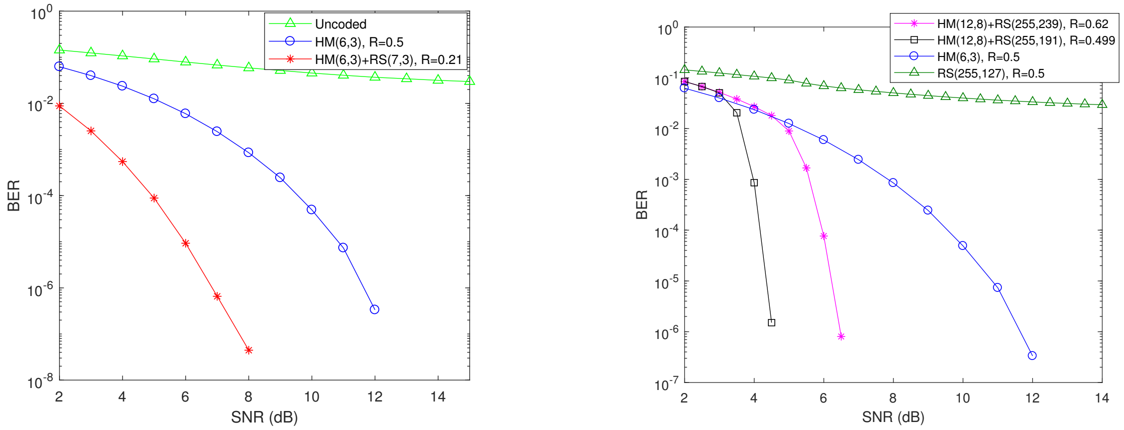


Figure 2: BER performances of (7,3,5) RS code concatenated with shortened (6,3,3) Hamming code or its coset, with the inner soft decision decoder based on two criteria – MED and MMPD – over channels with Gaussian noise and offset, where standard deviations of the offset are $\beta = 0.5$ (blue) and 0.3 (red).

inner code is evaluated. Binary $(6, 3, 3)$ code is a shortened version of $(7, 4, 3)$ Hamming code and a coset vector $\mathbf{a} = (1, 0, 0, 0, 0, 0)$ is considered. In Figure 2, we show bit error rate (BER) performances of RS-Coset codes versus signal to noise ratio (SNR (dB)) over channels with Gaussian noise and offset mismatch. We also assume that standard deviations of the offset are $\beta = 0.3$ (red curves) and $\beta = 0.5$ (blue curves). Let us first compare the performance of the inner soft-decision decoders based on two criteria, specifically, MED and MMPD. The MED detection has worse performance when the offset is larger, while the scheme using MMPD remains the same for any offset as we expected. We conclude that the MMPD detection is immune to the offset mismatch and achieves considerable performance improvements, particularly when the offset is large compared to the noise. We also observe that with the Hamming coset code as inner code instead of Hamming code itself, the simulated results of the concatenated scheme have been improved. The MMPD of the shortened Hamming code $(6, 3, 3)$ and its coset code is the same, however, the average number of neighbors with the minimum distance has decreased after introducing the coset of code. In addition, the introduction of the coset increases the MED detection's resistance of the offset mismatch. Thus, the performances of both MED and MMPD detection have improved by using the coset of block codes as proved in [2].

Figure 3(a) compares the proposed concatenated scheme with the Hamming coset code and uncoded scheme. Simulation is carried out over channels with $\beta = 0.5$. The Hamming coset code is decoded using the MMPD scheme. By referring to the channel raw BER illustrated by Curve Uncoded, the proposed concatenated scheme achieves a significant gain in BER over a wide range of SNR. Furthermore, compared with the Hamming coset code, it can be observed that at BER $= 10^{-4}$, the gain of the concatenated scheme corresponds to the decrease of the system required SNR from around 9 dB (corresponding to the HM(6,3)) to 5 dB (corresponding to the HM(6,3)+RS(7,3)). Thus, we achieve more than 4 dB SNR improvement of achieving a BER $= 10^{-4}$ with the proposed RS-Coset codes.

Our example code has a code rate of 0.21, which is lower than the non-concatenated



(a) BER performances of $(7,3,5)$ RS code concatenated with coset shortened Hamming (6,3,3), only coset Hamming (6,3,3), and uncoded case.

(b) BER performances of different coding schemes.

Figure 3: Performance evaluation over channels with Gaussian noise and offset, where the standard deviation of the offset is $\beta = 0.5$.

scheme. Here, we have other two concatenation schemes with higher code rates: coset shortened Hamming (12,8) concatenated with (255,191) RS with a code rate of 0.499; coset shortened Hamming (12,8) concatenated with (255,239) RS with a code rate of 0.62. Observe from Figure 3(b), the concatenation codes with a higher rate are shown to have worse BER performance. However, the concatenated scheme with a code rate of 0.62 still performs better than the Hamming coset code, which has a code rate of 0.5. What's more, (255,127) RS code is simulated under the same channel condition, whose code rate is 0.5. We can see that its performance is seriously distorted by the offset mismatch without the help of the inner MMPD decoder. The simulation results show the considerable coding gain compared with the non-concatenated codes with an even higher code rate over noisy channels with offset mismatch. Thus we conclude that the RS-Coset codes with the proposed decoding scheme can achieve great coding gain and also maintain the immunity to offset mismatch.

References

[1] K. A. S. Immink and J. H. Weber, "Minimum Pearson Distance Detection for Multilevel Channels with Gain and/or Offset Mismatch," IEEE Transactions on Information Theory, Vol. 60, no. 10, pp. 5966-5974, Oct. 2014.

[2] J. H. Weber, R. Bu, K. Cai and K. A. S. Immink, "Binary Block Codes for Noisy Channels with Unknown Offset," IEEE Transactions on Communications, vol. 68, no. 7, pp. 3975-3983, 2020.

[3] K. A. S. Immink, K. Cai and J. H. Weber, "Dynamic Threshold Detection Based on Pearson Distance Detection," IEEE Transactions on Communications, vol. 66, no. 7, pp. 2958-2965, Jul. 2018.

A Primer on Techtile: An R&D Testbed for Distributed Communication, Sensing and Positioning

Gilles Callebaut Jarne Van Mulders Geoffrey Ottoy Liesbet Van der Perre
ESAT-DRAMCO, Ghent Technology Campus, KU Leuven
9000 Ghent, Belgium
gilles.callebaut@kuleuven.be

Abstract

The Techtile measurement infrastructure is a multi-functional, versatile testbed for new communication and sensing technologies relying on fine-grained distributed resources. The facility enables experimental research on hyper-connected interactive environments and validation of new wireless connectivity, sensing and positioning solutions. It consists of a data acquisition and processing equipment backbone and a fabric of dispersed edge computing devices, Software-Defined Radios (SDRs), sensors, and LED sources. These bring intelligence close to the applications and can also collectively function as a massive, distributed resource. Furthermore, the infrastructure allows exploring more degrees and new types of diversity, i.e., scaling up the number of elements, introducing '3D directional diversity' by deploying the distributed elements with different orientations, and 'interface diversity' by exploiting multiple technologies and hybrid signals (RF, acoustic, and visible light).

Keywords— Testbed, 6G, Software-Defined Radio, Precision Time Protocol, Power-over-Ethernet, RadioWeaves

1 Techtile – A Teaser

As envisioned in [1], new wireless access infrastructures will be integrated in existing structures, bringing them in closer proximity of devices and eventually becoming truly ubiquitous. The Techtile infrastructure (Figure 1) is the first testbed capable of evaluating this RadioWeaves concept. To support this, the measurement room or –more general– the testbed, hosts 140 detachable tiles of equal size on the walls, floor, and ceiling. A versatile set of equipment can be mounted on these panels, effectively embedding the electronics into the room.

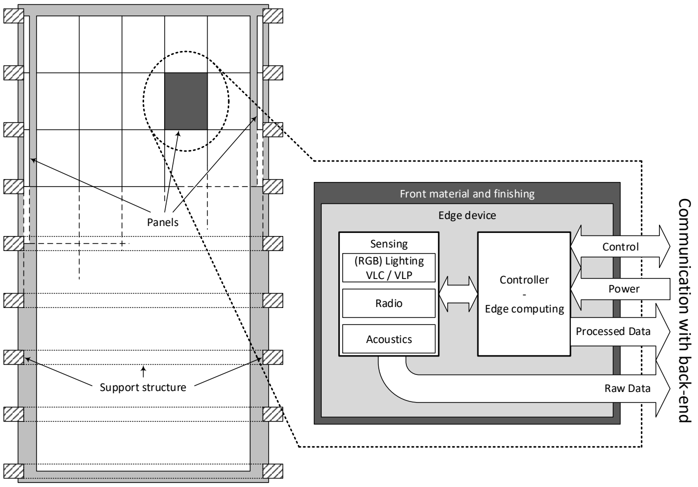


Figure 1: Left: Support structure (top view) – Right: Tile with embedded edge device. The support structure is designed to support 140 detachable panels.

The main technical challenges, in both the testbed and future distributed infrastructures, are the scalability and synchronization. In contrast to other testbeds [2]–[4], this testbed does not utilize dedicated connections for communication and synchronization to each processing point, i.e., –in our case– tile. The conventional approach to time synchronization requires all cables to have the same length, making deployment cumbersome and not scalable. In order to support high-speed connections between the tiles and the central processing, multiple hierarchies of processing must be organized to aggregate the high number of separate connections. To tackle these issues, all tiles are connected, powered (PoE++ IEEE802.3bt) and time synchronized (PTP IEEE 1588) over Ethernet. By default, each tile has a processing unit, an SDR and a power supply, as depicted in Figure 2. This base configuration can be extended with custom solutions to support other use cases. To facilitate other research activities, all developed equipment and software is open-source available.[*]

In this paper we further provide more details on how the testbed is established in a modular way, and its features. The article is structured as follows. First, the support structure is discussed, focusing on the modular design of the detachable tiles. Thereafter, the backbone of the testbed is elaborated. It consists of a central processing unit or server and Ethernet connections to all tiles. The on-tile equipment is introduced in Section 4 with a focus on the default setup. In Section 5, we discuss the rover supporting automated 3D sampling of the room, thereby reducing the time-consuming and labor-intensive task of conducting experiments manually. Lastly, the foreseen research and development activities are summarized in Section 6. The list is by no means exhaustive and we welcome other R&D activities and applications from industry and academic.

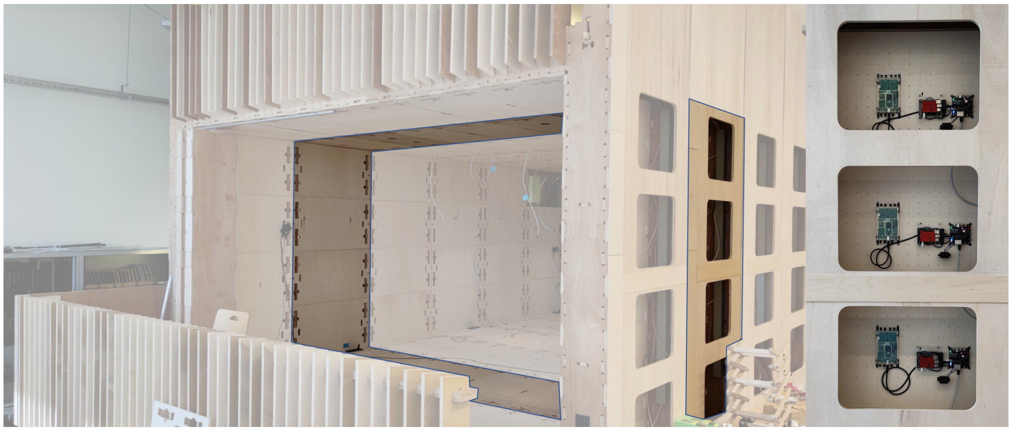


Figure 2: Left: The Techtile support structure – Right: The back of three tiles, equipped with the default setup, i.e., a software-defined radio (USRP B210), processing unit (Raspberry Pi 4) and power supply with Power-over-Ethernet. Each tile is connected to the central unit with an Ethernet cable, providing both power and data.

2 Techtile Construction – Modular and Open

The Techtile support structure, illustrated in Figure 2, is based on the WikiHouse[†] concept. WikiHouse is an open-source building system. Such constructions are made of standardized wooden parts. Building further on the modular design, the walls, floor and ceiling of the structure are comprised of 140 detachable tiles. The two walls of the building support 28 tiles each. The ceiling and ground floor support 42 and 52 tiles, respectively. The room is 4 m by 8 m and is 2.4 m.

A tile has a size of 120 cm by 60 cm. To mount the electronics, the tile features a 5 cm by 5 cm grid. The grid is made of M3 inserts, allowing conveniently installing components with standardized M3 screws. In addition to attaching equipment and custom electronics to the tile, we allow designing your own tile for specific applications. For instance, Visible Light Communication (VLC) and Visible Light Positioning (VLP) systems using LED fixtures can permanently be mounted in the ceiling. We also support attaching equipment to the inside of the room by means of overlays. An overlay has the same grid structure as a tile and is mounted on top of the tile. This keeps the aesthetics, while still supporting attaching equipment inside. As the infrastructure is mainly designed for embedding electronics inside walls, we emphasize on only using the overlays if absolutely required. By using overlays, research demanding a Line-of-Sight (LoS), e.g., visible light and acoustics, are supported. Overlays are more flexible compared to designing your own tile, as overlays can be installed on all tiles.

[*] github.com/techtile-by-dramco

[†] www.wikihouse.cc

3 Techtile Backbone – Everything over Ethernet

The backbone of the infrastructure consists of a central server which is connected to all tiles over Ethernet. By means of Power over Ethernet (PoE) midspans, 90 W of power is supplied to each tile. Furthermore, the Ethernet switches support the IEEE-1588 Precision Time Protocol (PTP), enabling high-accuracy clock distribution to all connected devices. Hence, the testbed provides communication, synchronization and power over Ethernet. As everything is connected over Ethernet, the developed system is easily scalable and is flexible in the manner in which devices are added and removed from the network.

3.1 Central Processing and Networking

The central server (Dell PowerEdge R7525) has 512 GB RAM, two NVIDIA Tesla T4 16 GB Graphical Processing Units (GPUs) and two AMD 7302 3 GHz Central Processing Units (CPUs), running Ubuntu Server 20.04 LTS.

Apache Kafka is used as a communication and processing platform between the server and the tiles. It follows the publish/subscribe paradigm. Events and streams are organized and stored in topics. Events can be thought of as files residing in a topic acting as a folder. Besides communication, Kafka support stream processing, database integrations, searching and querying data, management, logging and monitoring, making it very flexible to support future applications. Furthermore, data is generated and processed following a consumer/producer scheme, decoupling the systems. This allows to easily let the same data be processed by different clients and to dynamically plug-in and shutdown consumers.

At the networking side, the Dell S4148 switch is used. It features 48 10 Gbit Ethernet ports with IEEE 1588v2 support. In Techtile, four switches are deployed with a total of 192 connections. A Dell Virtual Edge Platform (VEP), running VyOS, handles routing and firewall.

3.2 Power-over-Ethernet

In order to keep the cable management practical, power to the tiles is provisioned with PoE technology, where both data and power go over the same Ethernet cable. With the introduction of the latest standard, IEEE 802.3bt, a power of maximum 90 W can be provided. While this maximum is not required for the default setup, implementing the latest standard ensures a generic solution, supporting also high-power applications.

Our PoE architecture consists of a Power Sourcing Equipment (PSE) and 140 Powered Devices (PDs), supporting all PoE standards, i.e., IEEE 802.3af (PoE), IEEE 802.3at (PoE+) and recent 802.3bt (PoE++). To provision the PoE, the PD-96XXGC series midspans from Microchip is selected as PoE injectors, supporting also the latest 802.3bt PoE++ standard. Each midspan has a 10/100/1000 Mbps data rate pass through. In total a PoE budget of approximately 9 kW is available.

3.3 Synchronization-over-Ethernet

In order to provide time synchronization for all the tiles, PTP IEEE 1588v2 is used. This protocol achieves a clock accuracy in the sub-microsecond range –and depending on the network, configuration and version even sub-nanosecond. The protocol follows a master-slave architecture. The root time reference is hold by the grandmaster clock and distributed to the other clocks in the network. Precision Time Protocol supports both L2 and UDP transport. It has an operation similar to Network Time Protocol (NTP), where the master and slave exchange messages to determine the path delays and correct their clocks accordingly. In PTP, different profiles are defined, each having different configurations and requirements. Such profiles are tailored for specific application and are, for example, used in Time-Sensitive Networking (TSN) [5]. Different clocks are defined based on their capabilities. For instance, transparent clocks are network devices which alter the timestamps in the packets to remove the time spent in these devices, effectively making them transparent to the PTP protocol. A boundary clock on the other hand, serves as both master and slave. It functions as a master at some connections, while synchronizing to a master on another port. One of the ongoing work is to tune the configuration of the protocol to achieve nanosecond accuracy and determine the impact on the processing capabilities of the Raspberry Pi (RPi) and the load on the network.

3.4 Data Acquisition System

The data acquisition system will provide panels with a synchronized analog data acquisition channel, able to sample at 1.25 MS/s, a resolution of 16 bit, and for a total number of 192 channels. This sample frequency and number of bits is sufficient to sample almost every sensor which allows us to move the state-of-the-art in sensor fusion and federation and underpins our aim for multi-modal sensing and positioning research. It includes sampling of full continuous audio streams (i.e., microphones recording audible or ultrasonic sound) and sensors registering high-bandwidth visible light communication. The high number of synchronized channels allows for setting up truly dispersed architectures, so large and distributed arrays can be obtained.

The data acquisition system also has 48 synchronized 16-bit DAC output channels able to steer a variety of actuators. For example an array of speakers, ranging into the deep ultrasonic range, can be implemented in research on transmit beam forming and backscattering for fully passive mobile devices. Additional possibilities are the generation of modulated signals and multiple access coding schemes to drive power LEDs in VLC applications.

The combination of the data-acquisition system and the edge devices forms the infrastructure that is required to validate new concepts based on distributed nodes. By using the fully synchronized sensing capabilities of the central data acquisition and processing infrastructure, a baseline performance can be established. This baseline performance can act as a reference for more realistic scenario's where distributed edge devices collaborate in a loosely coupled (and varying) configuration.

4 On-Tile – Sensors, Radios, Processing, and Other Cool Stuff

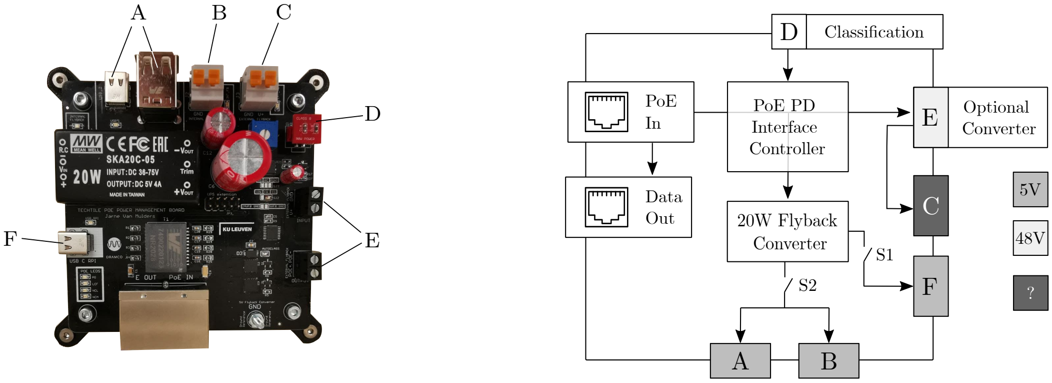
Each tile can accommodate a diversity of sensors and actuators, transmitters and receivers for radio - and other waves, host processing resources to add distributed intelligence in the environments, and potential other cool stuff that creative researcher may want to experiment with. In this section we elaborate on the custom PoE board, the SDR and the RPi. The PoE board delivers power to the tile and the RPi processes the SDR signals and serves as a platform for edge computing.

4.1 PoE Board

The standard PoE Hardware Attached on Top (HAT) for the Raspberry Pi, or other derivative forms, are readily available but do not support the latest PoE version. For this reason, a 802.3bt supported PD circuit was developed, as shown in Figure 3. The system is based on the ON Semiconductor NCP1096 which is the PoE-PD interface controller. On top of the 802.3bt support, the board features several connectors and voltages to power different devices and sensors, as illsutrated in Figure 3b. PoE distributes 48 V over Ethernet.

Connectors and Usage. The 48 V output is directly available on the board through terminal block C. In case other voltages are required, an external DC/DC converter can be connected through terminal blocks E, and the ouput is available on terminal block C. The other output connections are connected to the internal flyback converter providing 5 V with a maximum power of 20 W. We adopted two bistable switches. One switch (S1) is dedicated to the USB-C connector for the RPi, the other (S2) is used for the other USB-C connector, the two USB-A connectors and terminal block B. These switches prevent that the peripherals –and especially– the RPi will boot-up every time there is a power cycle. Furthermore, we can ensure that not all systems are powered, saving energy and mitigating any damage due to unintentional power outages.

Power Classification. The PoE PD interface controller supports 8 different PoE classes, which are programmable with resistors. The classification is adjustable by means of replacing the classification board and consists of two resistors and a connector (Figure 3a [D]). The connector makes it convenient to switch the class rapidly although a power cycle is required. The most regular classes: 3 (13 W), 4 (25.5 W) and 8 (71.3 W–90 W) are provided. Besides a fixed power class, the class assignment is negotiated automatically. For instance, when only the RPi and Universal Software Radio Peripheral (USRP) is connected, a power class of 3 is sufficient. However, as the class is determined at the beginning of power delivery –and fixed afterwards–, it is crucial that sufficient energy is extracted during start-up. Because the RPi and USRP power consumption and the tile setup is variable, it is not recommended to use autoclass, rather select the desired class in advance. Furthermore, as we have a finite power budget, knowing the consumption of each tile ensures that we adhere to the power constraint. Not respecting the agreed power, leads unfortunately to a power disconnection by the PSE and results in an abrupt power shutdown, causing potential damage and loss of measurement data.



(a) Custom PD board. The different connectors are labeled and elaborated in (b).

(b) This block diagram represents the connection between the connectors, PoE controller and optional DC/DC converter. The output voltages available on the connectors are illustrated by different shades of gray. The optional DC/DC converter attached to E, can output different voltages on connector C. The switches S1 and S2 can be toggled at anytime.

Figure 3: Custom Power over Ethernet Powered Device (PD)

4.2 Wireless RF Communication through Software-Defined Radio

The testbed hosts a fabric of distributed SDRs. Each tile is equipped with one USRP B210, featuring two RF channels. The B210 has a maximum transmit power of 20 dBm. This SDR supports up to 56 MHz of real-time bandwidth through the AD9361 direct-conversion transceiver. The B210 can operate over a frequency range of 70 MHz to 6 GHz, thereby covering most licensed and unlicensed bands. The B210 hosts an open and reprogrammable Spartan6 XC6SLX150 Field-Programmable Gate Array (FPGA). The baseband signal is processed by the host, i.e., RPi 4, using USB 3.0. GNURadio, supported by the B210, enables adopting and designing a high range of protocols and standards, e.g., IEEE 802.11 (Wi-Fi), Bluetooth, LoRaWAN, provided by the open-source community. The B210 can be fed with an external 10 MHz clock and Pulse Per Second (PPS) for synchronized operation. The reference clock is used to generate all data clocks, sample clocks and local oscillators. In addition, an external PPS signal can be used for time synchronization between the SDRs. In this manner, it is possible to coherently transmit and receive on all SDRs.

4.2.1 Architecture

A simplified architecture of the USRP 210 is depicted in Figure 4. It illustrates the connection with the host through the USB interface. The external time and frequency reference is used in the FPGA and radio-frequency integrated circuit (RFIC). The RF frontend consists of mixers, filter, oscillators and amplifiers translating the signal from the RF domain to an intermediate frequency (IF). In receive mode, the complex baseband of that IF signal is sampled and clocked into the FPGA. National Intstruments has published the stock FPGA in open-source, and provides, o.a. digital down-conversion and filters for decimation. The resulting I/Q samples are transferred to the host though USB 3.0. The B210 features 2 RX/TX connections and 2 RX connections. The RX/TX connections are switched by an RF switch, thereby using only two identical transmit chains and two received chains sharing the same local oscillator for each TX or RX chain.

4.2.2 Processing the Baseband Signals

The baseband signals are processed by the Raspberry Pi 4 and transfered over USB 3.0. Depending on the application, the distributed signals are aggregated to the central server for further processing. By having both local processing and central processing, different techniques regarding local, edge and cloud processing can be studied. Designing RF communication systems with the B210 is facilitated by the open-source USRP Hardware Driver (UHD) from National Instruments. This library is supported in GNURadio and OpenBTS. In order to coherently transmit and receive, i.e., at the same time, over multiple USRPs, we output 10 MHz and PPS with a custom Printed Circuit Board (PCB). The 10 MHz clock is derived from the 54.0 MHz oscillator on the RPi and tuned with the on-board phase-locked loop (PLL). By using a device tree overlay in Linux, the 10 MHz signal is always present and independent on the state of the Operating System (OS). The PPS is generated

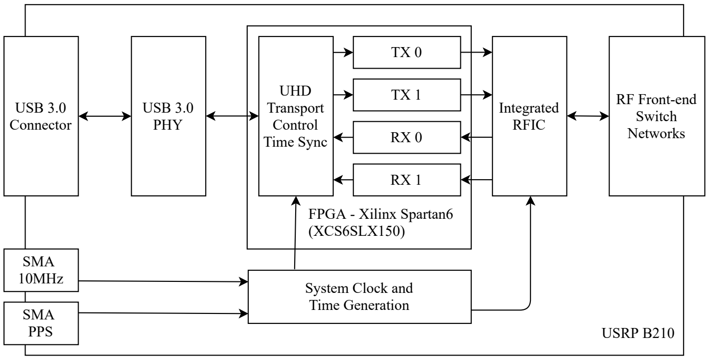


Figure 4: Simplified architecture the B210.

in a loadable kernel module. The PCB, installation procedure and drivers are available in open-source[‡].

4.3 Edge Processing

The default setup contains a RPi capable of edge computing. To latest RPi model, i.e., RPi 4, is adopted. It has a rich feature set, such as 8 GB of LPDDR4-3200 SDRAM and a powerfull processor, i.e., Quad core Cortex-A72 (ARM v8) 64-bit SoC 1.5GHz, tailored for high computing tasks. For dedicated and more computation-intensive applications, custom edge computing platforms can be used, e.g., NVIDIA Jetson Nano, Google Coral and Intel NCS.

To speed-up read/write operations, we adopted a solid state drive (SSD) opposed to the standard micro-SD card storage. Besides the speed improvement, the SSD is less sensitive to power failure and has a higher capacity to price ratio. Furthermore, SD cards are designed to store data, while SSDs are tailored for a high number of read/write operations, resulting from running an OS. These characteristics make SSD a more robust and suitable technology.

5 Automated 3D Sampling – a Rover in the House

To speed-up measurements, reduce labor-intensive tasks and mitigate human-errors, we have designed a rover to perform automated 3D sampling inside the testbed. The rover consists of a baseplate, hosting the processing unit and batteries, and a scissor lift to move the measurement equipment to the required height.[§]

Baseplate. The baseplate contains a Raspberry Pi running Robot Operating System (ROS). It controls the wheels, reads-out the sensors and performs navigation. By adopting ROS, the development of the rover is sped-up and is more easily extendable for future applications. ROS consists of software packages dedicated for building robots, containing the necessary drivers and algorithms. The communication, employed in ROS, between nodes, i.e., processing entities, is handled by an asynchronous publish/subscribe message passing system. In this manner, ROS forces the developer to decouple different parts of the system. For synchronized communication, the request/response pattern is used.

A 16 cells battery pack, in a 4S4P configuration, powers the system. The battery voltage ranges from 10 V to 16.2 V depending on the charge state. The battery back, based on Lithium Cobalt Oxide (LCO) technology, has a capacity of 170 W h and can handle up to 480 W peak power. Besides powering the rover, the battery pack provides power to the peripherals and measurement equipment in order to conduct the experiments. When the battery is low, the rover is able to navigate to its charging station, recharge and resume the measurements.

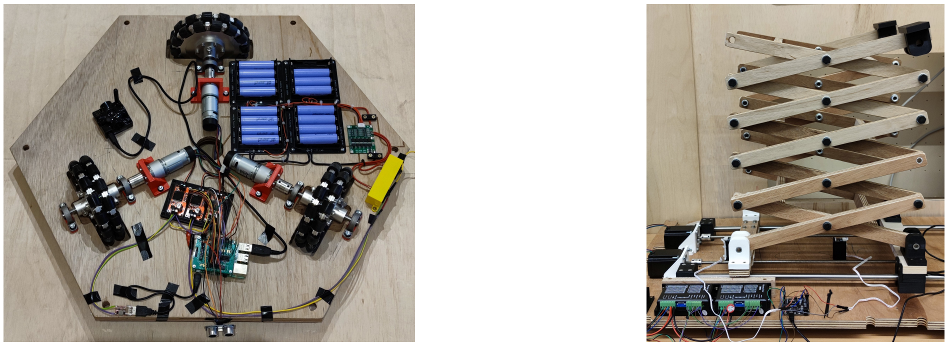
Navigation. Marvelmind indoor RTLS is used as a positioning system. It has a precision of 2 cm [6]. Ultrasonic beacons are used to acquire the position of the mobile beacon (mounted on the baseplate). Four beacons are fixed in the Techtile infrastructure. The position is determined by trilateration. To filter out outliers, we make use of a Kalman filter. The recursive algorithm uses consecutive inputs to estimate the current state. Based on the uncertainties of the inputs and the uncertainties of the previous estimates, the current state is approximated. Besides localization, the rover features obstacle detection to mitigate crashing

[‡] `github.com/techtile-by-dramco/raspberrypi-sync`

[§] This section contains work realized by two students, Arne Reyniers and Jonas De Schoenmacker, in the context of their master thesis.

into objects in the room. Ultrasonic sensors mounted on the sides of the rover's baseplate are used and have a resolution of 3 mm and a range of 2 cm to 400 cm.

Scissor lift. The baseplate of the rover can navigate through the room in 2D space. By mounting the scissor lift (Figure 5b) on top of the baseplate, the rover is able to measure in 3D space. The lift has a range of 55 cm to 185 cm. The equipment can be mounted on top of the lift. Power is supplied to the equipment from the battery pack. The scissor lift can be controlled with AT-commands. The height of the lift is determined by the VL53L1X Time-of-Flight (ToF) sensor using a 940 nm invisible Class 1 laser. An advertised accuracy and ranging error of ± 20 mm in both ambient light and dark light conditions.



(a) Baseplate hosting the omnidirectional wheels, Controller (Raspberry Pi running ROS), battery pack and Marvelmind beacon.

(b) Scissor lift having a range of 55-185 cm.

Figure 5: Rover to perform automated 3D sampling.

6 Research and Development Activities - Inviting Creative Experiments

The TechTile experimental facility has been designed to enable a wide range of experiments. We foresee, a.o., the following research and development activities: i) development of next-generation Internet-of-Things solutions, ii) experimental validation of beyond-5G communication, e.g., RadioWeaves, iii) positioning and sensing based on acoustic signals, iv) wireless charging and v) visible light communications and positioning.

Internet-of-Things. We develop the next generation IoT solutions that should accommodate a massive number of low power devices and upgrade them to get smarter, leveraging AI approaches. We conceive auto-discovery, self-diagnostics, and recovery approaches, to ultimately achieve zero e-waste impact.

(Beyond) 5G communication. We pursue experimental validation of (distributed) multi-antenna systems to support huge number of (low power) devices and provide ultra-reliable low latency connectivity. We will investigate cell-free operation and Large Intelligent Surfaces, new paradigms that may serve 6G systems.

Acoustic sensing and indoor positioning. We sense acoustic signals for a better understanding of the environment. Through hybrid RF-acoustic signaling we pursue positioning of low energy devices with unprecedented accuracy.

Secure connected devices. We exploit physical features of propagation and directivity to increase the security of connected devices, and enable private local networking solutions.

Wireless charging. We investigate whether/how devices can get charged without the need for cables, and eventually 'on their spot'.

Visible Light Communication and positioning. We develop communication and positioning systems that modulate LED light sources without impacting their illumination functionality. They complement conventional RF systems where the interference poses aggravating problems.

Acknowledgments

This project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No. 101013425.

The realization of the Techtile infrastructure is made possible thanks to the generous support of a bequest granted to the Science, Engineering and Technology Group of the KU Leuven. We have received hardware from both Niko and On Semiconductor. We also thank our equipment suppliers National Instruments, Dell and Würth Electronics for their special support for this pioneering development.

References

[1] L. Van der Perre, E. G. Larsson, F. Tufvesson, L. De Strycker, E. Bjornson, and O. Edfors, "Radioweaves for efficient connectivity: Analysis and impact of constraints in actual deployments," eng, Matthews, MB, IEEE, 2019, pp. 15–22.

[2] J. Vieira, S. Malkowsky, K. Nieman, Z. Miers, N. Kundargi, L. Liu, I. Wong, V. Öwall, O. Edfors, and F. Tufvesson, "A flexible 100-antenna testbed for massive mimo," in *2014 IEEE Globecom Workshops (GC Wkshps)*, 2014, pp. 287–293. DOI: 10.1109/GLOCOMW.2014.7063446.

[3] À. O. Martínez, J. Ø. Nielsen, E. De Carvalho, and P. Popovski, "An experimental study of massive mimo properties in 5g scenarios," *IEEE Transactions on Antennas and Propagation*, vol. 66, no. 12, pp. 7206–7215, 2018. DOI: 10.1109/TAP.2018.2871881.

[4] C. W. Shepard, R. Doost-Mohammady, R. E. Guerra, and L. Zhong, "Demo: Argosv3: An efficient many-antenna platform," in *Proceedings of the 23rd Annual International Conference on Mobile Computing and Networking*, ser. MobiCom '17, Snowbird, Utah, USA: Association for Computing Machinery, 2017, pp. 501–503, ISBN: 9781450349161. DOI: 10.1145/3117811.3119863. [Online]. Available: https://doi.org/10.1145/3117811.3119863.

[5] N. Finn, "Introduction to time-sensitive networking," *IEEE Communications Standards Magazine*, vol. 2, no. 2, pp. 22–28, 2018. DOI: 10.1109/MCOMSTD.2018.1700076.

[6] R. Amsters, E. Demeester, N. Stevens, Q. Lauwers, and P. Slaets, "Evaluation of low-cost/high-accuracy indoor positioning systems," in *Proceedings of the 2019 International Conference on Advances in Sensors, Actuators, Metering and Sensing (ALLSENSORS), Athens, Greece*, 2019, pp. 24–28.

Robust Local Differential Privacy

Milan Lopuhaä-Zwakenberg
Eindhoven University of Technology

Jasper Goseling
University of Twente and CWI

Setting

Users which have data $X = (S, U)$ from a finite alphabet $\mathcal{X} = \mathcal{S} \times \mathcal{U}$ that an aggregator is interested in, but users do not wish to disclose information about sensitive data S.[1] Therefore, users release a sanitised version $Y = \mathcal{Q}(X)$ of X, such that Y retains as much information about X as possible, measured by $\mathrm{I}(X;Y)$, without leaking too much about S. This leakage is measured via the differential privacy-like metric introduced in [2]. This has the advantage of being a worst-case metric, offering privacy for every user, not just the average user. This metric depends on the actual distribution P^* of the data. Since this distribution might not be available to the aggregator, we demand privacy for a subset $\mathcal{F}$ of the space $\mathcal{P}_{\mathcal{X}}$ of probability distributions on $\mathcal{X}$. This leads to

Definition 1 (Robust Local Differential Privacy)**.** *Let $\varepsilon \geq 0$ and $\mathcal{F} \subset \mathcal{P}_{\mathcal{X}}$. We say that $\mathcal{Q}$ satisfies $(\varepsilon, \mathcal{F})$-RLDP if for all $s \in \mathcal{S}$, all values y of Y, and all $P \in \mathcal{F}$ we have*

$$\mathbb{P}_{X\sim P}(Y = y | S = s) \leq \mathrm{e}^{\varepsilon}\mathbb{P}_{X\sim P}(Y = y | S = s'). \quad (1)$$

The aggregator can choose $\mathcal{F}$. Increasing $\mathcal{F}$ protects against more attackers, but typically comes at a utility cost. We consider the following $\mathcal{F}$: There is a public database $\vec{x}$ of entries drawn independently from P^*. From this database the aggregator obtains an empirical distribution $\hat{P}$. Then $\mathcal{F}$ is a χ^2-confidence set around $\hat{P}$ for a given confidence level $1-\alpha$. This situation is depicted in Figure 1. We aim to answer the following research question:

Problem 1. *Given ε, $\vec{x}$, and α, find a protocol $\mathcal{Q}$ that maximises $\mathrm{I}_{X\sim\hat{P}}(X;Y)$ while satisfying $(\varepsilon, \mathcal{F})$-RLDP.*

Polyhedral optimisation

We approximate $\mathcal{F}$ by an enveloping polyhedron $\mathcal{D}$. This allows us to use techniques from polyhedral duality to describe the set $\{\mathcal{Q} : \mathcal{Q}$ satisfies $(\varepsilon, \mathcal{D})$-RLDP$\}$ as a polyhedron itself. Since the utility $\mathrm{I}(X;Y)$ is convex in $\mathcal{Q}$, its maximum is attained at a vertex of this polyhedron, so we can find a solution by solving a vertex enumeration problem.

Low-complexity mechanisms

For large alphabets vertex enumeration is computationally infeasible, and low-complexity solutions to Problem 1 are required. We introduce two types of protocols, Independent Reporting (IR) and Conditional Reporting (CR), that rely on applying standard Local Differential Privacy (LDP) protocols to S and U separately. We show that in the low privacy regime using Generalised Randomised Response (GRR) as LDP protocol maximises utility. This allows us to find optimal IR and CR protocols, using only a 1-dimensional optimisation.

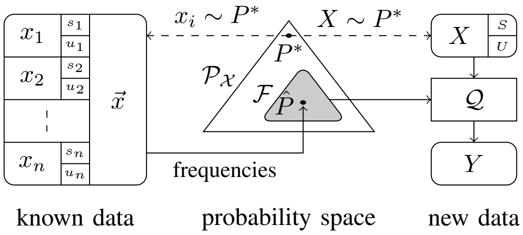


Fig. 1: Overview.

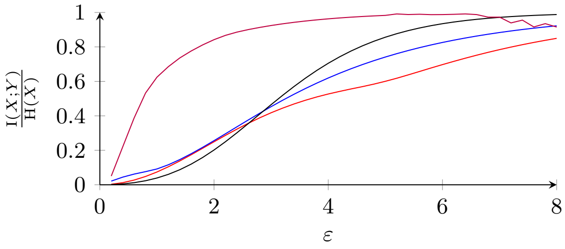


Fig. 2: Experiments on synthetic data for $|\mathcal{S}| = |\mathcal{U}| = 3$. (— Polyh.Opt., — GRR, — IR, — CR)

Experiments

We test the validity of these methods in experiments on synthetic data. Figure 2 shows the average utility over 200 randomly generated probability distributions. We compare them to the standard ε-LDP protocol GRR, which certainly satisfies RLDP. We see that for low ε, polyhedral optimisation offers significantly higher utility than the other methods, and the low-complexity methods also outperform the baseline. Experiments on real-world data, also on larger input alphabets where the polyhedral method is computationally infeasible, confirm these results.

References

[1] M. Lopuhaä-Zwakenberg and J. Goseling, "The privacy-utility tradeoff of robust local differential privacy," *arXiv:2101.09139*, 2021.

[2] M. Lopuhaä-Zwakenberg, H. Tong, and B. Škorić, "Data sanitisation for the privacy funnel with differential privacy guarantees," *arXiv:2008.13151*, 2020.

[1] This talk is based on [1].

Distributed Estimation of Common Signal Components across the Nodes of a Sensor Network

Charles Hovine Alexander Bertrand
KULeuven
Department of Electrical Engineering (ESAT), STADIUS
Kasteelpark Arenberg 10, B-3001 Leuven
charles.hovine@esat.kuleuven.be alexander.bertrand@esat.kuleuven.be

Abstract

A wireless sensor network (WSN) consists of a collection of sensor nodes, which are equipped with processing and wireless communication facilities to share data between each other. In this context, two paradigms are considered when applying array processing methods. Centralized fusion relies on collecting the network-wide observations in a fusion center (FC) where they are jointly processed, at the cost of large bandwidth and processing requirements at the FC, while distributed processing relies on the nodes collaboratively solving a task without any single node accessing the full network-wide observations. As the value of many array processing methods often depends on the presence of correlation between the channels of interest, the nodes can save bandwidth by identifying nodes whose channels correlate with their own (i.e. whose sensors observe common latent phenomena) and solve the given task by only communicating with those nodes. However, identifying correlated components is hard to realize in a distributed context, in particular between node pairs that do not share a direct wireless link. We introduce a distributed algorithm for estimating the signal subspace that (on average) is closest to the pairwise intersections between any two of the per-node sensor signal subspaces. In order to facilitate an efficient data fusion, we assume the WSN has (or can be pruned to) a tree-topology. As opposed to a centralized algorithm where all the sensor signals are transmitted to an FC, the per-node bandwidth and processing requirements are independent of the network-size and only depend on the number of neighbors per node and a chosen compression parameter. By construction, our algorithm converges to the solution of the so-called "maximum variance" (MAXVAR) formulation of the generalized canonical correlation anlalysis (GCCA) problem in which observations of every node act as a separate "view" in the GCCA formulation. Therefore, even though our work is formalized within a WSN context, it can be used as a generic distributed MAXVAR algorithm in other application contexts as well.

This project has received funding from the European Research Council (ERC) under the European Union's Horizon 2020 research and innovation programme (grant agreement No. 802895) and from the Flemish Government under the "Onderzoeksprogramma Artificiële Intelligentie (AI) Vlaanderen" programme.

A Comparative Study of Vision and AES in FHE Setting

Dilara Toprakhisar
Mathematics and Computer Science
Eindhoven University of Technology
5612 AZ Eindhoven, the Netherlands
d.toprakhisar@student.tue.nl

Tomer Ashur
imec-COSIC
KU Leuven
3001 Leuven, Belgium

Mathematics and Computer Science
Coding Theory and Cryptology
Eindhoven University of Technology
5612 AZ Eindhoven, the Netherlands
t.ashur@tue.nl

Abstract

The design of traditional block ciphers such as AES calls for efficient software and hardware implementations due to their domain. However, the increasing number of applications that use advanced cryptographic protocols such as multi-party computation (MPC) or zero-knowledge (ZK) proofs shifts the focus of optimization to a different metric: arithmetic complexity.

In this work, we present an understanding of the expected behaviors of arithmetization-oriented ciphers in an FHE setting. We compare AES as a representation of traditional block ciphers and Vision as a representation of arithmetization-oriented ciphers.

1 Introduction

Block ciphers are the building blocks of many encryption schemes and hash functions. Block ciphers such as AES and 3DES are traditionally designed to be efficient in hardware and software implementations as their design goal prioritizes efficiency in addition to security. Therefore, the design of traditional ciphers focuses on running time, gate count, or memory/power consumption [1]. As advanced cryptographic protocols such as multi-party computation (MPC) or zero-knowledge (ZK) proofs are becoming more frequently used in applications, different design constraints are becoming important. The design of block ciphers that are used in such applications is focused on arithmetic complexity. The arithmetic complexity of a cipher is defined by the number of non-linear arithmetic operations and such ciphers are called arithmetization-oriented ciphers. Arithmetization-oriented ciphers deviate from traditional ciphers in terms of relevant attacks and security analysis as the target applications are different.

Advanced cryptographic protocols such as ZK proofs, MPC, and fully homomorphic encryption operate on mathematical objects applying algebraic operations. Converting computations into a set of algebraic operations is called arithmetization.

The Marvellous design strategy defines a set of decisions to be taken when designing arithmetization-oriented ciphers. The design principles are led by security, simplicity, robustness, and efficiency. In the Marvellous design strategy, non-procedural computations, algebraic complexity, and algebraic attacks are taken into account. Minimizing the running time or energy/memory consumption are not the main optimization metrics as in traditional ciphers. In terms of efficiency, optimization metrics are defined specific to the advanced cryptographic protocol. AIR or R1CS constraints are decisive for ZK proofs, number of multiplications and number of communication rounds are decisive for MPC [1]. While statistical attacks are the main focus in the design of traditional block ciphers, arithmetization-oriented ciphers make attacks that manipulates simple polynomials possible.

The motivation of this work is to serve as a first exploration of the expected behaviours of arithmetization-oriented ciphers in an FHE setting. For that aim, this work compares AES as a representation of traditional block ciphers and Vision as a representation of the Marvellous design strategy.

Outline. This paper is organized as follows. Section 2 recalls the basics of AES and Vision. In Section 3, we describe the test environment as well as implementation specific details. In Section 4, we present our findings and results, and in Section 5, a discussion is held about our results followed by a conclusion.

2 Preliminaries

2.1 Advanced Encryption Standard (AES)

AES operates on a 128-bit plaintext as a sequence of operations. These operations form a round and are repeated a certain number of times. The number of rounds in an AES execution depends on the length of the key. In this work, AES-128 is used, which has 10 rounds.

Each round of AES consists of the following four steps:

- SubBytes: a fixed permutation is applied to each byte of the state. This permutation is chosen in such a way that small differences in the input will result in arbitrary differences in the output;
- ShiftRows: row_i is shifted to the left by i positions;
- MixColumns: a linear bijection is applied to the state. The four bytes in each column are used to generate one column of the resulting state;
- AddRoundKey: the round key is added to the state using a bit-wise XOR.

As an exception, prior to the first round, an AddRoundKey operation is applied and in the last round, the MixColumns operation is removed.

2.2 Vision

Vision is designed following the Marvellous design strategy [1] to operate over binary fields with its native field $\mathbb{F}_{2^n}$.Vision's state is an element of $\mathbb{F}_{2^n}^m$ which has m field elements, whereas AES state is an element of $\mathbb{F}_{2^8}^{4x4}$. Vision's round function consists of two steps where each step employs three layers: S-box, linear and key addition. The S-box layer applies an inversion operation to each of the state elements followed by an $\mathbb{F}_2$ linearized affine polynomial. The affine polynomial in the S-box layer of the second step is chosen such that it has a low degree and its inverse has a high degree. The inverse of this affine polynomial is used in the S-box layer of the first step. The linear layer is the multiplication of the state with an MDS matrix.

3 Experimental Setup

For our benchmarks, we use an existing implementation of a leveled homomorphic encryption that can evaluate the AES circuit presented by Gentry et al. [3] built on top of the HElib library. The implementation is based on the BGV cryptosystem [2]. Additionally, we describe a working implementation of a leveled homomorphic encryption that can evaluate a Vision circuit. This implementation is built on top of the HElib library using the same optimizations proposed by Gentry et al. The optimizations include a method for implementing the Brakerski-Gentry-Vaikuntanathan modulus switching transformation on ciphertexts using CRT representation and an alternative Brakerski-Vaikuntanathan key-switching method that does not necessarily decrease the norm of the ciphertext vector. Many of the optimizations aim to decrease the number of conversions between coefficient and evaluation representations of polynomials.

For a fixed state size, we can choose the order of the base field and the dimension of the state for Vision. A vision instance with 16 elements over $\mathbb{F}_{2^8}$ has the same state size as the state of AES. Figure 1 compares the running times of one round of AES and Vision for variable state elements implemented with toy parameters. According to our experiments, when 128 bit state is used, AES performs 60% faster than Vision.

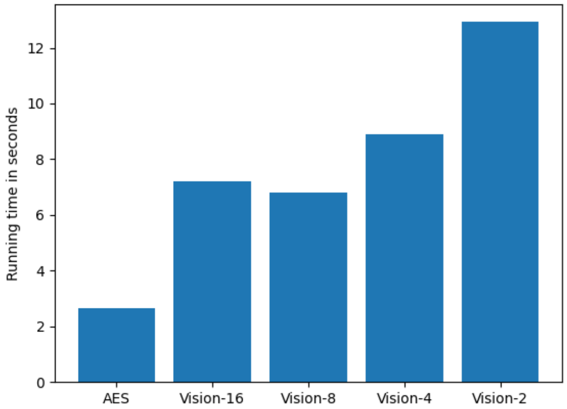


Figure 1: Running time comparison of AES and Vision for one round. Vision-X is a Vision instance with a state consisting of X field elements.

Table 1 shows the running times of each operation of a Vision's state for different state dimensions. As the dimension of the state decreases, the running time of the inversion operation and the computation of the 1^{st} affine polynomial increases.

Table 1: Running times in seconds for each operation in a Vision's round for variable state elements with toy parameters

State	Inversion	Affine Polynomial-1	Matrix Multiplication	Key Addition	Affine Polynomial-2	Total
Vision-16/$\mathbb{F}_{2^8}^{16}$	1.19	0.6006	2.008	0.0003	0.19	7.2
Vision-8/$\mathbb{F}_{2^{16}}^{8}$	1.79	1.18	0.92	0.003	0.21	6.81
Vision-4/$\mathbb{F}_{2^{32}}^{4}$	2.3	3.47	0.393	0.003	0.182	8.89
Vision-2/$\mathbb{F}_{2^{64}}^{2}$	2.853	7.106	0.145	0.003	0.193	12.93

4 Findings and Results

We reproduced our experiments for the full encryption with the parameters that provides more than 128 bits of security. Due to large parameter sizes, experiments are done in an environment that runs Ubuntu Server 18.04 LTS with 3 TB RAM and 4 x Intel(R) Xeon(R) Gold 6136 CPU @ 3.00GHz.

Table 2 shows the running times of AES and Vision variations for fixed state size of 128 bits. Due to the differences in multiplication levels and the sizes of the field elements, different ring polynomials are used to obtain given security levels.

Table 2: Total running times in minutes for AES and Vision for a fixed state size of 128 bits

Algorithm	State	Parameters	Number of rounds	Total running time
AES	$\mathbb{F}_{2^8}^{16}$	Cyclotomic polynomial: 53261 Length of the modulus chain (bits): 820 Security level (bits): 136.256	10	1.93
Vision	Vision-16 $\mathbb{F}_{2^8}^{16}$	Cyclotomic polynomial: 163455 Length of the modulus chain (bits): 1550 Security level (bits): 128.529	10	26.83
Vision	Vision-8 $\mathbb{F}_{2^{16}}^{8}$	Cyclotomic polynomial: 148733 Length of the modulus chain (bits): 2000 Security level (bits): 158.132	10	37.07
Vision	Vision-4 $\mathbb{F}_{2^{32}}^{4}$	Cyclotomic polynomial: 164737 Length of the modulus chain (bits): 2300 Security level (bits): 179.373	10	63.3
Vision	Vision-2 $\mathbb{F}_{2^{64}}^{2}$	Cyclotomic polynomial: 164737 Length of the modulus chain (bits): 3150 Security level (bits): 126	13	76.9

In these experiments, the two bottlenecks we faced are parameter selection and the running time of the affine polynomial. For Vision, the ring polynomial is chosen as $\Phi_m(X)$ that factors modulo 2 into degree-d polynomials such that *degree of the field element*$|d$ which limits the parameter selection. On the other hand, ring polynomial selection is easier for AES as it factors modulo 2 into degree-d polynomials such that $8|d$. This difference makes it possible for AES to get the same security level as Vision, but with a smaller ring polynomial, therefore creating a gap between running times. The second bottleneck is that the degree of the affine polynomial that is used in the first step of Vision round is linearly increasing as the size of the field elements increase.

5 Conclusion

Our experiments show that in an FHE setting, AES performs 90% faster than Vision when more than 128 bits of security is obtained. In order to make the gap between the running times of AES and Vision smaller, one can improve the Vision design and try to reduce the number of multiplications. It is important to make sure the improvements does not jeopardize the security of the design.

Acknowledgements Tomer Ashur is an FWO post-doctoral fellow under Grant Number 12ZH420N.

References

[1] Abdelrahaman Aly et al. "Design of Symmetric-Key Primitives for Advanced Cryptographic Protocols". In: *IACR Trans. Symmetric Cryptol.* 2020.3 (2020), pp. 1–45. DOI: 10.13154/tosc.v2020.i3.1-45. URL: https://doi.org/10.13154/tosc.v2020.i3.1-45.

[2] Zvika Brakerski, Craig Gentry, and Vinod Vaikuntanathan. "(Leveled) fully homomorphic encryption without bootstrapping". In: *Innovations in Theoretical Computer Science 2012, Cambridge, MA, USA, January 8-10, 2012*. Ed. by Shafi Goldwasser. ACM, 2012, pp. 309–325. DOI: 10.1145/2090236.2090262. URL: https://doi.org/10.1145/2090236.2090262.

[3] Craig Gentry, Shai Halevi, and Nigel P. Smart. "Homomorphic Evaluation of the AES Circuit". In: *IACR Cryptol. ePrint Arch.* 2012 (2012), p. 99. URL: http://eprint.iacr.org/2012/099.

Improved Bit-mappings for IEEE 802.11 LDPC Codes with Applications to Amplitude Shaping

Yunus Can Gültekin Alex Alvarado Frans M. J. Willems
Information and Communication Theory Lab
Department of Electrical Engineering
Eindhoven University of Technology, The Netherlands
y.c.g.gultekin@tue.nl

Extended Abstract

Binary low-density parity-check (LDPC) codes are linear block codes which can be represented by their $(n-k)$-by-n parity-check matrix $\boldsymbol{H}$. Here, k is the length of the binary information sequence $\boldsymbol{u} = (u_1, u_2, \dots, u_k)$, and n is the length of the binary codeword $\boldsymbol{c} = (c_1, c_2, \dots, c_n)$. Each codeword $\boldsymbol{c}$ in the code $\mathcal{C}$ satisfies $\boldsymbol{cH}^T = \boldsymbol{0}$, i.e., the code $\mathcal{C}$ is the null space of $\boldsymbol{H}$. Equivalently, a valid codeword has to satisfy $n-k$ parity-check equations. In Fig. 1 (left), parity-check matrix of the $[n = 7, k = 4, (n-k) = 3]$ Hamming code is shown as an example [1, p. 15].

Any linear block code can be represented by a Tanner graph. A Tanner graph is a bipartite graph that consists of n variable nodes and $n-k$ check nodes. There is a connection between the i^{th} variable node and the j^{th} check node if $\boldsymbol{H}_{ji} = 1$. The Tanner graph that corresponds to the parity-check matrix in Fig. 1 (left) is shown in Fig. 1 (right). Here, circles represent variable nodes, i.e., coded bits in $\boldsymbol{c}$, squares represent check nodes, i.e., parity-check equations.

$$\boldsymbol{H} = \begin{bmatrix} 1 & 1 & 0 \\ 1 & 0 & 1 \\ 0 & 1 & 1 \\ 1 & 1 & 1 \\ 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix}^T$$

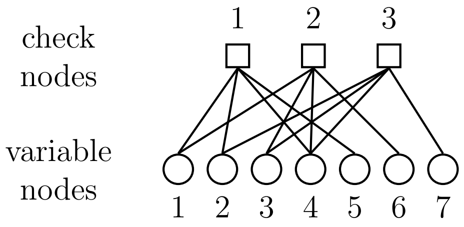


Figure 1: The Tanner graph and the corresponding parity-check matrix.

The degree of a node in a Tanner graph is defined as the cardinality of the set of nodes that it is connected to. Thus, the degree of the i^{th} variable node is given by the sum of all elements of the i^{th} column of $\boldsymbol{H}$. As an example, there are three degree-1, three degree-2, and one degree-3 variable nodes in the Tanner graph in Fig. 1. A coded bit is a part of several parity-check equations, and this number is given by the degree of the corresponding variable node.

In bit-interleaved coded modulation (BICM), the n-bit output of a binary forward error correction (FEC) code is typically mapped to an N-tuple of 2^m-ary amplitude-shift keying (2^m-ASK) symbols. This mapping is achieved using a binary labeling strategy that assigns an m-bit binary label $(B_1, B_2, \dots, B_m)$ to each symbol, e.g., the binary reflected Gray code (BRGC) [2, Defn. 2.10]. Here, we assumed that $N = mn$.

For BICM systems, the error probability is not identical for different bit levels, and each bit is not "protected" equally against errors. This phenomenon is called unequal error protection (UEP). As an example, we show in Fig. 2 the bit error rate (BER) of different bit levels for uniform and shaped 16-ASK with BRGC. Here, shaping is realized using enumerative sphere shaping [3, 4]. For uniform 16-ASK, the protection is higher for more significant bit levels. However, for shaped 16-ASK, the most protected bit level is B_2. This is because the ratio $\Pr\{B_i = 0\}/\Pr\{B_i = 1\}$ is the largest for $i = 2$, an effect observed earlier in [5, Sec. III-C].

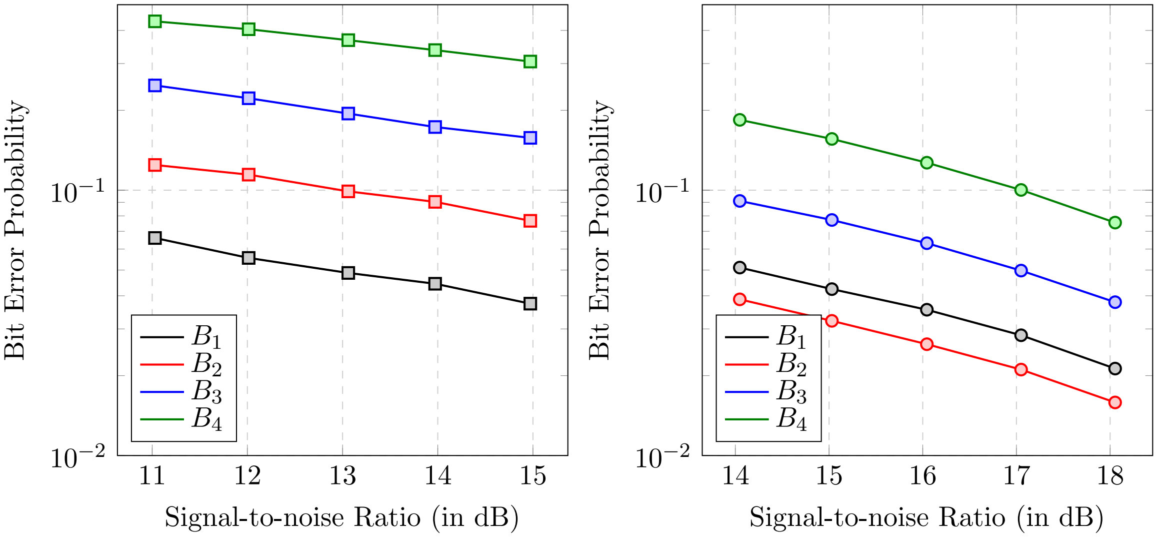


Figure 2: Pre-FEC, hard-detection BER for uniform (left) and shaped (right) 16-ASK with 4 and 3.5 bits of entropy, respectively.

UEP observed in Fig. 2 implies that the performance of an LDPC code depends on how the coded bits, i.e., variable nodes, are "connected" to different bit levels. A coded bit with a corresponding variable node with a higher degree is a part of a higher number of parity checks and can be considered to play a more important role during the decoding procedure. Thus, connecting variable nodes with higher degrees to more protected bit levels can improve the performance.

In Fig. 3, the performance of uniform BICM and probabilistic amplitude shaping (PAS) [6] is shown for 16-ASK and the corresponding BRGC. As the FEC code, the rate-5/6 and 3/4 648-bit systematic LDPC codes from the IEEE 802.11 standard [7] are used for uniform and shaped signaling, respectively. Both the regular bit-mapping and the heuristically modified bit-mapping explained above are simulated. In regular bit-mapping, bits at the output of the FEC encoder are consecutively gathered into groups of m and mapped to 2^m-ASK symbols. In modified bit-mapping, bits at the output of the FEC encoder are re-ordered before ASK mapping such that the ones that correspond to higher degree variable nodes are mapped to a more protected bit level according to Fig. 2. The transmission rates of the uniform and the shaped schemes are 3.33 and 2.5 bit/1D, respectively. We see that at a BER of 10^{-5}, modified bit mapping provides 0.2 and 0.1 dB gains over the regular mapping for uniform and shaped signaling, respectively. We believe the reason why the gain is smaller for shaped signaling is due to a constraint imposed by PAS: all parity bits should be used as the sign bit level B_1 [6, Fig. 5]. Accordingly, we were only able to change the mapping for variable nodes that correspond to systematic bits. A future research direction can be to investigate the effect of this constraint by using less restrictive shaping approaches as in [8, 9].

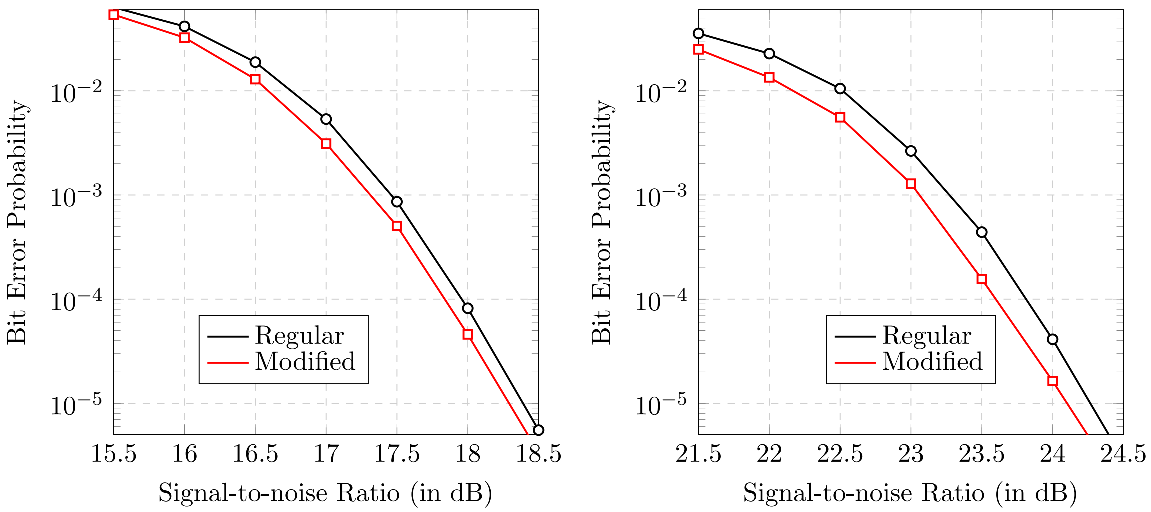


Figure 3: Post-FEC BER for shaped (left) and uniform (right) signaling.

References

[1] T. Richardson and R. Urbanke, *Modern Coding Theory.* Cambridge, UK: Cambridge University Press, 2008.

[2] L. Szczecinski and A. Alvarado, *Bit-Interleaved Coded Modulation: Fundamentals, Analysis, and Design.* Chichester, UK: John Wiley & Sons, 2015.

[3] Y. C. Gültekin, F. M. J. Willems, W. J. van Houtum, and S. Şerbetli, “Approximate enumerative sphere shaping,” in *Proc. IEEE Int. Symp. Inf. Theory*, Vail, CO, USA, June 2018, pp. 676–680.

[4] Y. C. Gültekin, W. J. van Houtum, A. Koppelaar, and F. M. J. Willems, “Enumerative sphere shaping for wireless communications with short packets,” *IEEE Trans. Wireless Commun.*, vol. 19, no. 2, pp. 1098–1112, Feb. 2020.

[5] F. Steiner and G. Kramer, “Optimization of bit mapping and quantized decoding for off-the-shelf protograph LDPC codes with application to IEEE 802.3ca,” in *Proc. Int. Symp. on Turbo Codes and Iterative Inf. Process.*, Hong Kong, China, Dec. 2018.

[6] G. Böcherer, F. Steiner, and P. Schulte, “Bandwidth efficient and rate-matched low-density parity-check coded modulation,” *IEEE Trans. Commun.*, vol. 63, no. 12, pp. 4651–4665, Dec. 2015.

[7] *IEEE Std. for Inform. Technol.-Telecommun. and Inform. Exchange Between Syst. LANs and MANs-Part 11: Wireless LAN MAC and PHY Spec.*, IEEE Std. 802.11-2016, Dec. 2016.

[8] M. Pikus and W. Xu, “Bit-level probabilistically shaped coded modulation,” *IEEE Commun. Lett.*, vol. 21, no. 9, pp. 1929–1932, Sep. 2017.

[9] Y. C. Gültekin, W. J. van Houtum, A. Koppelaar, and F. M. J. Willems, “Partial enumerative sphere shaping,” in *Proc. IEEE Veh. Technol. Conf. (Fall)*, Honolulu, HI, USA, Sep. 2019.

Throughput of Distributed MIMO Optical Wireless Communication system with multiple power-constrained LED emitters

T. E. Bitencourt Cunha[1], X. Deng[1,2], J. P. M. G. Linnartz[1,3]
[1]Eindhoven University of Technology, 5600 MB Eindhoven, The Netherlands
[2]South China Normal University, Guangzhou 510006, China
[3]Signify (Philips Lighting) Research, 5656 AE Eindhoven, The Netherlands
{t.e.bitencourt.cunha, x.deng, J.P.linnartz}@tue.nl, j.p.linnartz@signify.com

Post Conference Abstract

The exponential increase in the number of multimedia-capable and internet-connected devices is leading to a huge demand for mobile data traffic. In this scenario, the currently available radio frequency (RF) spectrum is unable to handle all connected devices. The necessity to support higher and higher data rates and to support the huge number of devices is challenging research for the development of new technologies capable to drive communication using frequencies outside the RF band [1-3]. In the development of a new technology, which has an impact worldwide, characteristics such as low cost and high energy efficiency are very important to accelerate mass-market adoption. While mobile data traffic grows, light-emitting diodes (LEDs) are quickly replacing old light sources around the world. This has attracted the research community's interest for designing LED-based optical wireless communication (OWC) systems using this existing infrastructure. In general, OWC systems have strong qualities including no interference with existing RF systems, no licensing requirements, higher security, and no interference from adjacent rooms once light wave does not go through the wall. However, the limited modulation bandwidth of LEDs of some MHz is a bottleneck for high data rate transmission [1-5]. Moreover, LEDs exhibit a limited linear range.

To achieve high data rates using LEDs, modulate signals beyond the 3-dB bandwidth becomes attractive. However, doing this blindly means that too much signal power can be wasted on the subcarriers that are subject to unfavourable channel conditions and it will probably generate high levels of inter-symbol interference (ISI). In this context, orthogonal frequency division multiplexing (OFDM) techniques [3-7] are particularly interesting as they can exploit the frequency roll-off by adaptive bit loading [4] and they are a strong tool against ISI. In parallel, the use of multiple LEDs and multiple photodetectors (PDs) in multiple-input multiple-output (MIMO) techniques have been also seen as a key tool. MIMO is well known in RF systems, as they can be used either, to increase the overall data rate, or to improve the reliability of transmission [8-10]. Now MIMO is applied in OWC systems under new constraints imposed by the optical channel. The joint use of MIMO and OFDM has huge value once it allows to joint optimize the distribution of modulation power among LEDs and their frequency response.

Many power loading strategies for LED-based MIMO OWC systems have been described in the literature considering a total average power constraint, where power can be freely exchanged between LEDs [8-10]. However, this is not suitable for practical implementation because each LED and also its own power amplifier are linear devices only in a limited range, i.e., the output power is linear between 0 and p_{max} when the input current swings between 0 and i_{max}. Moreover, the amount of power into an LED shall be limited to avoid overheating. Thus, in practice, the MIMO OWC power loading problem should consider a per-LED power constraint instead of a total average power constraint.

In this work, we explore the throughput of a distributed MIMO OWC system with multiple power-constrained LED emitters. The power loading coefficients are computed for maximizing the achievable rate by using the software package CVX for solving convex problems [11]. We compare the performance of the system using the CVX solutions for the case with multiple power-constrained LED emitters with the performance of the system using the CVX solutions for the case where LEDs are allowed to exchange power between them but under a total power constraint. Moreover, we also consider a uniform power loading strategy, where power is uniformly distributed over the LEDs and over OFDM subcarriers. In the obtained results, it is possible to observe that the system with multiple power-constrained LEDs achieved a lower throughput than the system under a constraint on the total power. Due to the low-pass behavior of the LEDs, spread uniformly the total average modulation power over the entire modulation bandwidth of the LEDs leads to a power wasting on the higher frequencies and, consequently, it achieves lower throughput than the other strategies. Looking closely at the CVX solutions for the system with power-constrained LED emitters, it is also possible to conclude that LEDs closer to the user are able to use higher modulation bandwidth than LEDs farther more distant. This is explained by the combination between path loss and low-pass behavior, which results in the strong channels being from LEDs and photodiodes that are closer to each other. Thus, for distant LEDs, it is more efficient to use lower frequency subcarriers.

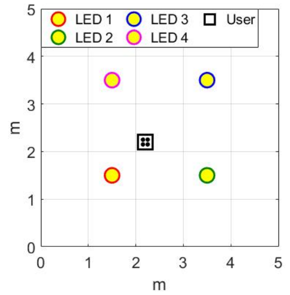


Figure 1 Geometric illustration of the first scenario. The LEDs are considered to have a 3-dB bandwidth of 20 Mhz.

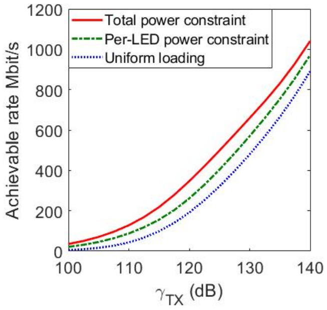


Figure 2 Achievable rate of the 4x4 LED-based MIMO OWC system.

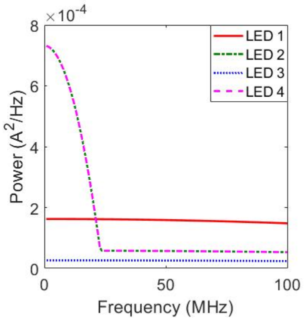


Figure 3 Power distribution on each LED of a 4x4 LED-based MIMO OWC system considering γ_{TX} =130 dB.

References

[1] J. -P. M. G. Linnartz, X. Deng, A. Alexeev and S. Mardanikorani, "Wireless Communication over an LED Channel," in IEEE Communications Magazine, vol. 58, no. 12, pp. 77-82, December 2020, doi: 10.1109/MCOM.001.2000138.

[2] H. Haas, L. Yin, Y. Wang and C. Chen, "What is LiFi?," in *Journal of Lightwave Technology*, vol. 34, no. 6, pp. 1533-1544, 15 March15, 2016.

[3] X. Wu, M. D. Soltani, L. Zhou, M. Safari and H. Haas, "Hybrid LiFi and WiFi Networks: A Survey," in IEEE Communications Surveys & Tutorials.

[4] S. Mardanikorani, X. Deng and J. M. G. Linnartz, "Sub-Carrier Loading Strategies for DCO-OFDM LED Communication," in IEEE Transactions on Communications, vol. 68, no. 2, pp. 1101-1117, Feb. 2020.

[5] S. Mardani, X. Deng and J. Linnartz, "Efficiency of Power Loading Strategies for Visible Light Communication," *2018 IEEE Globecom Workshops (GC Wkshps)*, Abu Dhabi, United Arab Emirates, 2018, pp. 1-6.

[6] X. Deng, S. Mardanikorani, G. Zhou and J. M. G. Linnartz, "DC-Bias for Optical OFDM in Visible Light Communications," in *IEEE Access*, vol. 7, pp. 98319-98330, 2019.

[7] X. Li, J. Vucic, V. Jungnickel and J. Armstrong, "On the Capacity of Intensity-Modulated Direct-Detection Systems and the Information Rate of ACO-OFDM for Indoor Optical Wireless Applications," in IEEE Transactions on Communications, vol. 60, no. 3, pp. 799-809, March 2012.

[8] L. An, H. Shen, J. Wang, Y. Zeng and R. Ran, "Energy Efficiency Optimization for MIMO Visible Light Communication Systems," in *IEEE Wireless Communications Letters*, vol. 9, no. 4, pp. 452-456, April 2020.

[9] Panta, J., Saengudomlert, P., Sterckx, K.L. and Pham, A.T. (2020), Performance optimisation of indoor SVD-based MIMO-OFDM optical wireless communication systems. IET Optoelectron., 14: 159-168.

[10] Y. Zhai, H. Chi, J. Tong and J. Xi, "Capacity Maximized Linear Precoder Design for Spatial-Multiplexing MIMO VLC Systems," in IEEE Access, vol. 8, pp. 63901-63909, 2020

[11] M. Grant and S. Boyd, “CVX: Matlab software for disciplined convex programming,” Version 2.2. [Online]. Available: http://cvxr.com/cvx/. Mar. 2021.

AnonImMed: An Open-Source Tool for Fast Anonymization of Medical Images

Sander R. Klomp[12], Rob G.J. Wijnhoven[2], Peter H.N. de With[1]
[1]Eindhoven University of Technology, [2]ViNotion B.V.

I. Introduction

In the past few years, several datasets that include privacy-sensitive information have been taken offline due to privacy concerns and new laws such as the GDPR regulation in Europe. Medical images contain patient information that is regulated even more strictly. Deep learning has become the standard for Computer Aided Diagnosis (CAD) based on medical images and requires large amounts of data to achieve good performance. The current standard in medical imaging is to print patient text directly over the measurement image, sometimes even partly occluding tissue pixels. This privacy-sensitive patient data is not required for general medical imaging research. Thus, anonymizing these medical images enables researchers to use this data for research purposes.

This paper describes a method to automatically detect and remove text from medical images at high processing speeds. We base our method on the EAST text detector [1] and make the following four contributions: (1) open-source implementation of the anonymization tool; (2) method to generate large amounts of synthetic training text; (3) multiple optimizations to improve the processing speed of anonymization.

II. Method and results

Our method consists of two steps. First, we improve the detection accuracy of the default EAST text detector. Second, we improve the computation time of the anonymization system.

A. Improving detection accuracy. To improve the performance of the standard EAST text detector for medical images, we fine-tune the pretrained detector with both synthetic and real image data. (1) We create a synthetic dataset by generating text with random font, size, color, and JPEG compression artefacts. Prior probabilities of the generated synthetic text match expected properties of target medical image overlay texts. That text is then added images of the ICDAR dataset. (2) A small number of images with hand-labeled bounding boxes around medical text are added to the dataset.

TABLE I
Speed comparison of optimizations, 1920×1080 pix, RTX 2080Ti.

Engine	**FP acc.**	**EAST opt.**	**DALI**	**JPEG**	**Mask mode**	**Multiproc**	**Batch**	**msec**	**FPS**
PyT	fp32						1	263	3.8
PyT	fp32	✓					1	189	5.3
PyT	fp16	✓					1	152	6.6
TRT	fp16	✓					1	127	7.9
TRT	fp16	✓	✓	✓			1	74.6	13.4
TRT	fp16	✓	✓	✓			1	52.6	19.0
TRT	fp16	✓	✓	✓	✓		1	48.1	20.8
TRT	fp16	✓	✓	✓	✓		4	47.8	20.9
TRT	fp16	✓	✓	✓	✓	✓	4	29.3	**34.1**

Fine-tuning on the combined dataset improves the performance significantly. Our evaluation is based on esophageal images. Because no public evaluation datasets are available for medical text detection, a quantitative evaluation has not yet been performed. We have found that in general our system results in similar precision, but obtains a much higher recall (less missed detections). We have found that the resulting text detector generalizes well to other types of medical images with similar text overlays.

B. Improving detection computation time. The starting point for our work is a public PyTorch (PyT) implementation of EAST[1] and improve its computation time using several optimizations. The impact of each optimization is shown in Table I. We use the optimized inference engine TensorRT (TRT), parallel post-processing using multi-processing (CPU), GPU batch processing and reduce floating-point accuracy (32 to 16 bit). The DALI[2] pipeline improves data loading and image preprocessing. Anonymized images are stored using LibJpegTurbo[3] compression. We propose “Mask mode” as our novel optimization method that does not output bounding boxes around text and thereby avoids the computation of the non-maxima suppression. Instead, we directly use the probability output map of the detector as a mask to black-out text detections. When applying all optimizations, a processing throughput of 34.1 frames/second is obtained on an RTX 2080 Ti GPU. This relates to a factor of 9 speedup when compared to the naive PyTorch implementation.

III. Conclusions

We have proposed a tool to anonymize text on medical images, based on the EAST text detector. To obtain high detection accuracy, we have extended the training set with synthetic images and some images of the target domain. We have performed several optimizations to improve the computation time of the anonymizer. The final optimized system obtains a throughput of 34.1 frames per second without a notable loss in detection accuracy. We provide the tool as open-source software, enabling the use to other researchers that want to use medical image datasets, but are currently unable to due to privacy constraints.

References

[1] X. Zhou, C. Yao, H. Wen, Y. Wang, S. Zhou, W. He and J. Liang *EAST: An Efficient and Accurate Scene Text Detector*, CVPR 2017, pp.2642-2651.

[1] https://github.com/SakuraRiven/EAST
[2] https://developer.nvidia.com/dali
[3] https://github.com/libjpeg-turbo/libjpeg-turbo

Orthogonal Time Frequency Space Modulation: Implementation and Constraints

Franz Lampel
Department of Electrical Engineering
Information and Communication Theory (ICT) Lab
TU Eindhoven, the Netherlands
f.lampel@tue.nl

Frans M.J. Willems
Department of Electrical Engineering
Information and Communication Theory (ICT) Lab
TU Eindhoven, the Netherlands
f.m.j.willems@tue.nl

I. Abstract

Motivated by challenges encountered in wireless communication over time-frequency (TF) dispersive channels such Doppler dispersion, equalization, or multi-user MIMO, a new modulation technique termed orthogonal time frequency space (OTFS) was recently introduced in [1]. The driving idea behind OTFS is to utilize the delay-Doppler (DD) domain to represent information-carrying symbols. Compared to the TF domain, utilized by orthogonal frequency division multiplexing (OFDM), the DD is time-invariant. OTFS exploits the channel's time-invariance in the DD domain and outperforms OFDM in several aspects, especially in high mobility scenarios.

When OTFS was introduced first in [1], it was presented as an overlay technique for existing OFDM systems, as depicted in Fig 1. The overlay technique is based on the so-called symplectic finite Fourier transform (SFFT) and the inverse SFFT (ISFFT). The ISFFT and SFFT allow transforming the symbols in the DD domain to the TF domain and vice versa. Initially intended to allow for the reuse of existing hardware, the overlay technique later became the standard formulation for OTFS. Recent works on OTFS, such as [2], use the overlay technique and pulse-shaped OFDM to study the input-output relation for OTFS.

The overlay technique represents one possible way to obtain the time domain signal from the DD representation. A more fundamental connection between the DD domain and the time domain is provided by the so-called Zak transform as pointed out in [3]. In [4], a complete treatment of OTFS based on the Zak transform is presented. In our recent work [5], we presented an OTFS transmitter and receiver based on the discrete Zak transform (DZT) and the inverse DZT.

The DZT is a signal transform also associated with Weyl-Heisenberg frames (WHF) [6]. Moreover, pulse-shaped OFDM can be cast in the framework of WHF. In this work, we use WHFs and the DZT to formally establish DD and TF domain connections through the SFFT and ISFFT. Moreover, we show that the overlay technique is equivalent to the DZT/IDZT if rectangular transmit and receive pulses for the OFDM systems are considered. Nonrectanglar pulses, on the other hand, result in nonconstant channel gains in the DD domain. The DZT defines constraints on the transmit and receive pulse pair that ensures undistorted recovery of the transmit symbols. This constraint has been widely neglected in the literature so far. Therefore, our work fills a gap in the literature on OTFS and opens a new way of studying OTFS as an OFDM overlay.

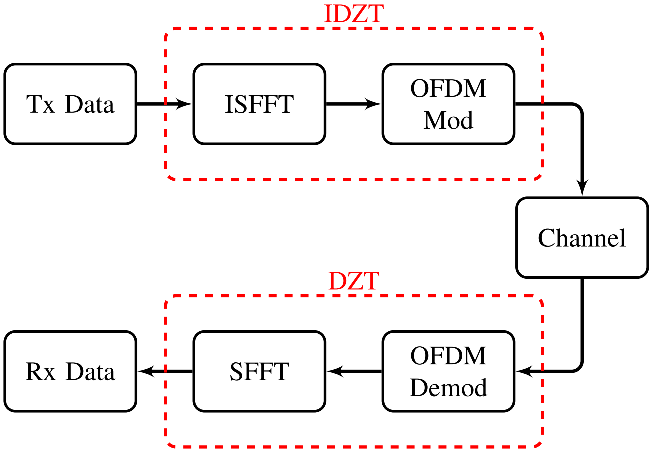


Fig. 1. OTFS as an overlay system for OFDM. The ISFFT is used to transform the symbols defined in the DD domain to the TF domain, which allows the implementation of OTFS in existing OFDM systems. The ISFFT combined with the OFDM modulation coincides with the IDZT in OFDM with rectangular pulses. Equivalently, the OFDM demodulation and the SFFT resembles the DZT.

References

[1] R. Hadani, S. Rakib, M. Tsatsanis, A. Monk, A. J. Goldsmith, A. F. Molisch, and R. Calderbank, "Orthogonal time frequency space modulation," in *2017 IEEE Wireless Communications and Networking Conference (WCNC)*, 2017, pp. 1–6.

[2] P. Raviteja, Y. Hong, E. Viterbo, and E. Biglieri, "Practical pulse-shaping waveforms for reduced-cyclic-prefix otfs," *IEEE Transactions on Vehicular Technology*, vol. 68, no. 1, pp. 957–961, Jan. 2019.

[3] R. Hadani, S. Rakib, S. Kons, M. Tsatsanis, A. Monk, C. Ibars, J. Delfeld, Y. Hebron, A. J. Goldsmith, A. F. Molisch, and R. Calderbank, "Orthogonal time frequency space modulation," 2018.

[4] S. K. Mohammed, "Derivation of OTFS modulation from first principles," *IEEE Trans. Veh. Technol. (preprint)*, 2021.

[5] F. Lampel, A. Alvarado, and F. M. Willems, "Orthogonal time frequency space modulation: A Zak transform approach," *IEEE Transaction on Wireless Communication*, to be submitted.

[6] H. Bölcskei and F. Hlawatsch, "Discrete zak transforms, polyphase transforms, and applications," *IEEE Transactions on Signal Processing*, vol. 45, no. 4, pp. 851–866, Apr. 1997.

On the SNOW Stream Ciphers

Liliya Kraleva Chaoyun Li
imec-COSIC
Dept. Electrical Engineering (ESAT), KU Leuven
Leuven, Belgium
liliya.kraleva@esat.kuleuven.be chaoyun.li@esat.kuleuven.be

Abstract

SNOW 3G is the core of the UEA2 & UIA2 algorithms in the 3GPP standard. Our work revisits previous results on chosen IV differential cryptanalysis and adjusts some results within the original steps of the SNOW algorithm. Further, we extend the analysis with 2 rounds and compare the algorithms of SNOW 2/ $2^{\oplus}$/ 3G / $3G^{\oplus}$ with respect to the number of active Sboxes.

1 Introduction

SNOW 3G is a word-oriented synchronous stream cipher. It is the core of the integrity and confidentiality algorithms UEA2 and UIA2 of the 3GPP UMTS system, which are identical to the algorithms 128-EEA1 and 128-EIA1 in LTE. The previous version SNOW 2.0 is developed by Ekdahl and Johansson [4]. Recently, a new version, namely SNOW-V [5] has been proposed for 5G.

We briefly recall some cryptanalysis of SNOW 2.0 and SNOW 3G. Linear distinguishing attacks and fast correlation attacks against SNOW 2.0 and have been investigated in [6, 11, 8, 13]. To strengthen the resistance against algebraic attacks [1], the SNOW 3G has been proposed. Recently, Yang *et al.* propose distinguishing and key recovery attacks on SNOW 3G based on vectorized linear approximations [12].

With respect to differential attacks there are few results on SNOW 3G, assuming chosen IV scenarios. In [2] Biryukov *et al.* show a 13-round multiset distinguisher with low complexity. Further, they extend similar distinguisher on SNOW $3G^{\oplus}$ and recover the key for up to 18 initialization rounds. Recently, a differential attack improving the results from [2] is proposed in [7] .

In this paper we revisit the attack in [7] and present adjusted results according to the original details of the algorithm. We show that the adjusted attacks can in fact cover one round less than the authors presented. Further, we consider the composition of Sboxes in order to extend the analysis with 2 more rounds. Additionally, we compare the SNOW 2.0 and SNOW 3G versions with respect to differential cryptanalysis and use the metric of active Sboxes to measure the diffusion of the FSM. We show that for up to 12 initialization rounds the differences behave identically for the two versions and produce keystream words with the same difference. Finally, we give a lower bound on the number of active Sboxes over 2 rounds.

Section 2 introduces the algorithm details. The adjusted results compared to [7] are given in Section 3. We further give observations on how to extend the analysis in Section 4. The branch number of a transformation is introduced in section 5, together with our motivation about the minimal number of active Sboxes for 2 rounds of the algorithm. Then we compare the SNOW versions and provide tables for easier visualization. Finally, some conclusions and future work are presented in Section 6.

2 The algorithms of SNOW 2 and SNOW 3G

SNOW 3G consist of a linear feedback shift register (LFSR) with 16 32-bit registers denoted s_i and a finite state machine(FSM) as a nonlinear component with three 32-bit registers denoted R_1, R_2, R_3. The feedback word of the LFSR is recursively computed as

$$s_{15}^t = \alpha^{-1} \cdot s_{11}^{t-1} + s_2^{t-1} + \alpha \cdot s_0^{t-1},$$

where α is the root of $x^4 + \beta^{23}x^3 + \beta^{245}x^2 + \beta^{48}x + \beta^{239}$ over $GF(2^8)[x]$ and β is the root of $x^8 + x^7 + x^5 + x^3 + 1$ over $GF(2)[x]$.

At time t, the FSM outputs

$$F^t = (s_{15}^{t-1} + R_2^{t-1}) \oplus R_2^{t-1},$$

after which the registers are updated as

$$R_1^t = R_2^{t-1} + (R_3^{t-1} \oplus s_5^{t-1}), \quad R_2^t = S_1(R_1^{t-1}), \quad R_3^t = S_2(R_2^{t-1}).$$

The 32-bit Sboxes S_1 and S_2 are composed of a MixColumn operation and four 8-bit Sboxes S_R and S_Q applied to each byte, respectively. Note that S_R is the AES S-box. Let $w = (w_0, w_1, w_2, w_3)$ be a 32-bit word and let $r = (r_0 r_1, r_2, r_3)$ be the output of $S_1(w)$. Then

$$\begin{pmatrix} r_0 \\ r_1 \\ r_2 \\ r_3 \end{pmatrix} = \begin{pmatrix} x & x+1 & 1 & 1 \\ 1 & x & x+1 & 1 \\ 1 & 1 & x & x+1 \\ x+1 & 1 & 1 & x \end{pmatrix} \cdot \begin{pmatrix} S_R(w_0) \\ S_R(w_1) \\ S_R(w_2) \\ S_R(w_3) \end{pmatrix},$$

where x is the root of $x^8 + x^4 + x^3 + x^1 + 1$ over $GF(2^8)$. Correspondingly, for $r = S_2(w)$ we have

$$\begin{pmatrix} r_0 \\ r_1 \\ r_2 \\ r_3 \end{pmatrix} = \begin{pmatrix} y & y+1 & 1 & 1 \\ 1 & y & y+1 & 1 \\ 1 & 1 & y & y+1 \\ y+1 & 1 & 1 & y \end{pmatrix} \cdot \begin{pmatrix} S_Q(w_0) \\ S_Q(w_1) \\ S_Q(w_2) \\ S_Q(w_3) \end{pmatrix},$$

where y is the root of $y^8 + y^6 + y^5 + y^3 + 1$ over $GF(2^8)$.

The cipher works in two modes - an initialization mode and a key-stream mode. In the initialization mode, first the state is initialized by the 128 bit key and the 128 bit IV as:

$$\begin{array}{llll} s_{15} = k_3 + IV_0, & s_{14} = k_2, & s_{13} = k_1, & s_{12} = k_0 + IV_1 \\ s_{11} = k_3 \oplus 1, & s_{10} = k_2 \oplus 1 \oplus IV_2, & s_9 = k_1 \oplus 1 \oplus IV_3, & s_8 = k_0 \oplus 1 \\ s_7 = k_3, & s_6 = k_2, & s_5 = k_1, & s_4 = k_0 \\ s_3 = k_3 \oplus 1, & s_2 = k_2 \oplus 1, & s_1 = k_1 \oplus 1, & s_0 = k_0 \oplus 1 \end{array}$$

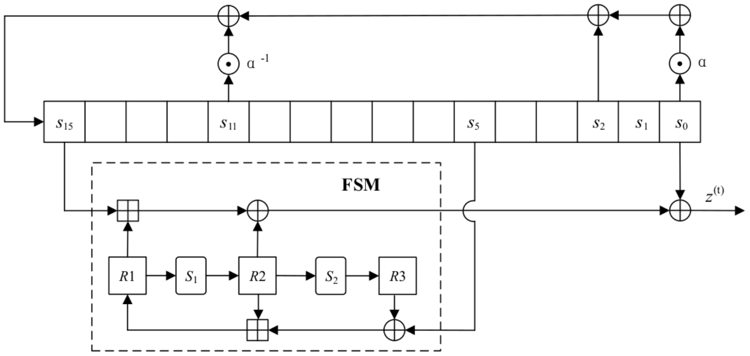


Figure 1: Diagram of the key-stream mode of SNOW3G.

During the initialization mode the output of the FSM is fed to the LFSR as $s_{15}^{t+1} = \alpha^{-1} \cdot s_{11}^{t} \oplus s_2^t \oplus \alpha \cdot s_0^t \oplus F$. After the initialization, the FSM and LFSR are clocked once more in keystream mode and the output of the FSM is discarded. Finally, the first keystream is produced and for each next keystream the stream cipher works in a keystream mode as shown in Figure 1.

SNOW 2 has the same LFSR, initialization and key generation as SNOW 3G, however the FSM has one less register. Therefore, we have

$$S_{15} = \alpha^{-1}S_{11} \oplus S_2 \oplus \alpha S_0 \oplus F,$$
$$F = (S_{15} + R_1) \oplus R2$$
$$R_1 = R_2^{t-1} + S_5^{t-1}, \quad R_2 = S_1(R_1^{t-1}),$$

where S_1 is the same as in SNOW 3G. Finaly, the keystream

$$z = F \oplus S_0.$$

The algorithms of SNOW $2^{\oplus}/3\mathrm{G}^{\oplus}$ are the same, but with all modular additions substituted with XORs.

3 Differential attack revisited

The authors of [7] merge the initialization rounds and the step performed before producing the keystream. This is not necessary here since the analysis is fully feasible when keeping the original structure of the algorithm. In this section we show the difference between the two initialization algorithms and propose adjusted results.

Algorithms 1 and 2 describe the two implementations by taking t initialization rounds. According to Algorithm 1, the value of the register s_{15}, used in the first keystream word is calculated as $s_{15}^t = \alpha^{-1}s_{11}^{t-1} \oplus s_2^{t-1} \oplus \alpha s_0^{t-1} \oplus F^t$, leading to a keystream

$$z^0 = F^{t+1} \oplus s_0^t = (\alpha s_{11}^{t-1} \oplus s_2^{t-1} \oplus \alpha s_0^{t-1} \oplus F^t) + R1^t) \oplus R2^t \oplus s_0^t.$$

Alternatively, using Algorithm 2 leads to the value $s_{15}^{*(t+1)} = \alpha^{-1}s_{11}^t \oplus s_2^t \oplus \alpha s_0^t$. Then the keystream is

$$z^{*(0)} = F^{t+2} \oplus s_0^{t+1} = (\alpha s_{11}^t \oplus s_2^t \oplus \alpha s_0^t) + R1^{t+1}) \oplus R2^{t+1} \oplus s_0^{t+1}$$

Algorithm 1 Key generation as in [7]

```
for t=1,2,... t do
    $F^t \leftarrow clockFSM$
    $ClockLFSR\_InitMode$
for $i = 0, 1, \dots n$ do
    $F^{t+1+i} \leftarrow clockFSM$
    $z^i \leftarrow F^{t+1+i} \oplus s_0^{t+i}$
    $clockLFSR\_KeystreamMode$
```

Algorithm 2 Key generation as in the specification [10] and this paper

```
for round=1,2,... t do
    $F^t \leftarrow clockFSM$
    $ClockLFSR\_InitMode$
$F^{t+1} \leftarrow clockFSM$
$ClockLFSR\_KeystreamMode$
for $i = 0, 1, \dots n$ do
    $F^{t+2+i} \leftarrow clockFSM$
    $z^i \leftarrow F^{t+2+i} \oplus s_0^{t+1+i}$
    $clockLFSR\_KeystreamMode$
```

Now let us consider IV and IV', such that $IV_0' = IV_0 \oplus a$, where $a = 0x80000000$. Then if we load the state using the two IVs with the same key, we can follow how the difference a propagates for different number of initialization rounds. Table 1 shows how this difference propagates through the state for SNOW 3G, including the differences for s_{15} and s_{15}^*. Instead of computing the keystream difference

$$\Delta(0x????0000 + a) \oplus \Delta S_1(a) = 0x????||A||A$$

over 12 initialization rounds with Algorithm 1, by using Algorithm 2 we would have

$$\Delta(0x00??8000 + a) \oplus \Delta S_1(a) = 0x????||(A \oplus 0x80)||A$$

over 11 initialization rounds, where $A = \Delta S_R(0x80)$ is one of the 127 possible output differences of the S_R Sbox when the input difference is $0x80$, as shown in table 6 in [7].

In all further analysis in this paper we refer only to values obtained by Algorithm 2.

4 Observations on more rounds

As the registers of the FSM are updated by applying the Sboxes S_1 and S_2, it is natural to look into their compositions. In this section we analyse such compositions for SNOW 3G$^{\oplus}$.

The difference in register R_3 after 13 rounds is $R_3^{13} = \Delta S_2 S_1(a)$, so we look into the possible output difference of $S_2 S_1(a)$, when the input difference of S_1 is a. From the specifications of the two Sboxes we have the following difference after S_1:

$\Delta(r_0^1, r_1^1, r_2^1, r_3^1) = M_1 \cdot \Delta(0,0,0,0x80)^T = (\Delta S_R(0x80), \Delta S_R(0x80), 3\Delta S_R(0x80), 2\Delta S_R(0x80))$.

Further, when we apply S_2 we have

$$\Delta(r_0^2, r_1^2, r_2^2, r_3^2) = M_2 \cdot (\Delta S_R(0x80), \Delta S_R(0x80), 3\Delta S_R(0x80), 2\Delta S_R(0x80)),$$

where $\Delta r_0^2 = 2S_Q(A) \oplus 3S_Q(A) \oplus S_Q(2A) \oplus S_Q(3A)$ and $A = \Delta S_R(0x80)$. Corresponding equations are derived for $\Delta r_1^2, \Delta r_2^2$ and Δr_3^2. We generated difference tables for the output bytes of the compositions $\Delta S_2 S_1(a)$ and $\Delta S_1(a \oplus S_1(a))$. Those are the compositions involved in the FSM registers at round 13 and 14, and involved in the keystream after 13 initialization rounds.

Table 1: The propagation of the difference a for SNOW 3G/2.0, where (s_{15}, z) and $(s_{15}*, z^*)$ are based on Algorithms 1 and 2 respectively, and $a = 0x80000000; b = 80800000; c = ??000000; d = ??800000; e = ??000000; f = ??800000; g = ??008000; h = ????8000; i = ????0000; j = ????AA; k, l =$ unknown.

SNOW 3G

r	s0	s1	s2	s3	s4	s5	s6	s7	s8	s9	s10	s11	s12	s13	s14	s15	s15*	R1	R2	R3	z	z^*
11	0	0	0	0	a	a	a	a	a	b	c	d	e	f	g	h	00??8000	a	0	0	????8000	$????(A \oplus 0x80)A$
12	0	0	0	a	a	a	a	a	b	c	d	e	f	g	h	i	00??0000	a	$\Delta S_1(a)$	0	$????AA$	?
13	0	0	a	a	a	a	a	b	c	d	e	f	g	h	i	j	8???8000	$\Delta(a + \Delta S_1(a))$	$\Delta S_1'(a)$	$\Delta S_2 S_1(a)$	?	?
14	0	a	a	a	a	a	b	c	d	e	f	g	h	i	j	k	80????80	$\Delta((a \oplus S_2 S_1(a)) + R_2^{13})$	$\Delta S_1(a + S_1(a))$	$\Delta S_2 S_1'(a)$	?	?
15	a	a	a	a	a	b	c	d	e	f	g	h	i	j	k	l	d0????17	$\Delta((a \oplus S_2 S_1'(a)) + R_2^{14})$	$\Delta S_1(R_1^{14})$	$\Delta S_2(R_2^{14})$	?	?

SNOW 2.0

11	0	0	0	0	a	a	a	a	a	b	c	d	e	f	g	h	00??8000	a	0	-	????8000	$????(A \oplus 0x80)A$
12	0	0	0	a	a	a	a	a	b	c	d	e	f	g	h	i	00??0000	a	$\Delta S_1(a)$	-	$????AA$	?
13	0	0	a	a	a	a	a	b	c	d	e	f	g	h	i	j	8???8000	$\Delta(a + \Delta S_1(a))$	$\Delta S_1'(a)$	-	?	?
14	0	a	a	a	a	a	b	c	d	e	f	g	h	i	j	k	80????80	$\Delta(a + \Delta S_1'(a))$	$\Delta S_1(\Delta(a + \Delta S_1(a))$	-	?	?
15	a	a	a	a	a	b	c	d	e	f	g	h	i	j	k	l	d0????17	$\Delta(b + \Delta S_1(a + S_1(a)))$	$\Delta S_1(\Delta(a + \Delta S_1'(a))$	-	?	?

As we can see from Table 2, for SNOW 3G$^{\oplus}$ the keystream difference after 12 rounds is $0x00800000 \oplus a \oplus S_1(a) \oplus S_1'(a) = ????||X||X$, where X is the XOR of the outputs of two S_R Sboxes with input difference a. According to the difference table of S_R for input difference a (Table 6 in [7]), there are between 28 and 35 possible combinations, or candidate values, for the input values of S_R. Therefore, there are in average 31 possible values for the last byte of the keystream word, compared to 256 values if we have to guess.

The keystreams after 13 initialization rounds is

$$z = S_{15}^* \oplus R_1 \oplus R_2 \oplus S_0 = 0x80808000 \oplus \Delta S_1(a \oplus \Delta S_1(a)) \oplus a \oplus \Delta S_1'(a) \oplus \Delta S_2 S_1(a).$$

Since from the tables we know the possible values for $\Delta S_1(a \oplus S_1(a))$, $\Delta S_1'(a)$ and $\Delta S_2 S_1(a)$, combining them we have N possible values for z. Following the key-recovery approach of [7], by running the algorithm again with $IV'' \neq IV$ we make a new candidate set of values. Finally, we intersect them in order to obtain the actual value.

5 Active Sboxes

When applying the chosen IV differential attack on the versions of SNOW 3G and SNOW 2.0, we notice that the differences propagate identically for the first 13 rounds. This is not surprising since their algorithms are so similar, however it brings up the question how much more secure is SNOW 3G in terms of differential cryptanalysis? To answer this question we looked into the diffusion of the state, which can be measured by active Sboxes in the FSM.

We consider the branch number of the matrixes M_1 and M_2 used in S_1 and S_2 in order to evaluate the minimum number of active 8-bit Sboxes over 2 rounds. This metric is typically used in the security analysis of block ciphers, for example the AES [3].

The branch number provides a lower bound on the minimum byte weight of any two-round trail. For a matrix multiplication, such as $M \cdot (x_0, x_1, x_2, x_3)^T = (y_0, y_1, y_2, y_3)^T$, the branch number β is the minimum number of nonzero values among $x_0, x_1, x_2, x_3, y_0, y_1, y_2, y_3$.

Table 2: The propagation of the difference a for SNOW 3G$^{\oplus}$/2.0$^{\oplus}$, where (s_{15}, z) and $(s_{15}*, z^*)$ are based on Algorithms 1 and 2 respectively, and $a = 80000000; b = 80800000; c = 80008000; d = 80808000; e = ????AA; f = ????(A \oplus 0x80)A$.

SNOW 3G$^{\oplus}$

r	s0	s1	s2	s3	s4	s5	s6	s7	s8	s9	s10	s11	s12	s13	s14	s15	s15*	R1	R2	R3	ks	ks*
11	0	0	0	0	a	a	a	a	a	b	a	b	a	b	c	d	808000	a	0	0	00808000	$????(A \oplus 0x80)A$
12	0	0	0	a	a	a	a	a	b	a	b	a	b	c	d	0	800000	a	$\Delta S_1(a)$	0	$????AA$	$????XX$
13	0	0	a	a	a	a	a	b	a	b	a	b	c	d	0	e	80808000	$a \oplus \Delta S_1(a)$	$\Delta S_1'(a)$	$\Delta S_2 S_1(a)$	$????A'A'$	?
14	0	a	a	a	a	a	b	a	b	a	b	c	d	0	e	f	80800080	$a \oplus \Delta S_1'(a) \oplus \Delta S_2 S_1(a)$	$\Delta S_1(a \oplus S_1(a))$	$\Delta S_2 S_1'(a)$	?	?
15	a	a	a	a	a	b	a	b	a	b	c	d	0	e	f	?	d0???17	$b \oplus R_2^{14} \oplus R_3^{14}$	$\Delta S_1(R_1^{14})$	$\Delta S_2(R_2^{14})$	?	?

SNOW 2.0$^{\oplus}$

11	0	0	0	0	a	a	a	a	a	b	a	b	a	b	c	d	808000	a	0	-	00808000	$????(A \oplus 0x80)A$
12	0	0	0	a	a	a	a	a	b	a	b	a	b	c	d	0	800000	a	$\Delta S(a)$	-	$????AA$	$????XX$
13	0	0	a	a	a	a	a	b	a	b	a	b	c	d	0	e	80808000	$a \oplus \Delta S_1(a)$	$\Delta S_1'(a)$	-	$????A'A'$	?
14	0	a	a	a	a	a	b	a	b	a	b	c	d	0	e	f	80800080	$a \oplus \Delta S_1'(a)$	$\Delta S_1(a \oplus S_1(a)$	-	?	?
15	a	a	a	a	a	b	a	b	a	b	c	d	0	e	f	?	d0???17	$b \oplus \Delta S_1(a \oplus S_1(a))$	$\Delta S_1'(a \oplus S_1'(a))$	-	?	?

The branch number ranges from 2 (i.e., there is no diffusion) to the total number of bytes in the state plus 1. Formally described, having a difference between two inputs $a' = a \oplus b$ and w_t being the weight, we have:

Definition 5.1 *[3] The* differential branch number *of a linear transformation $\lambda : \lambda(a) \oplus \lambda(b) = \lambda(a \oplus b)$ is given by*

$$\mathfrak{B}_d(\lambda) = \min_{a' \neq 0}\{w_t(a') + w_t(\lambda(a'))\}$$

Let us consider each update of the LFSR and the FSM as one round of the algorithm. The matrices M_1 and M_2 are used in AES and are proven to have branch number 5, implying that if one byte is nonzero, after the matrix multiplication all 4 bytes will be nonzero. Since an 8-bit Sbox is applied on each byte before the matrix multiplication, then we have one active Sbox before and 4 active Sboxes after, giving a minimum of 5 active Sboxes for 2 rounds. Let us now look into the FSM and how the two Sboxes are applied. For the second register of SNOW 3G we have $R_2^{t+2} = S_1(R_2^t + (R_3^t \oplus s_5^t)$. This means that the bound derived from the branch number can be reached starting from two rounds after the difference enters the FSM. This is the case for SNOW 2.0 as well, since then we have $R_2^{t+2} = S_1(R_2^t + s_5^t)$. Figure 2 illustrates how the registers are connected and how the Sboxes are activated for SNOW 3G. Theoretically, it is possible the difference of R_2^t to be canceled by the addition of $R_3^t \oplus s_5^t$, leading to difference 0 for the register R_1^{t+1}. In this case the number of active Sboxes will be 0 in the following round.

Observation 5.1 *Over two rounds of SNOW 2.0 we have at least 5 active Sboxes starting from 2 rounds after the difference enters the FSM, unless the difference is canceled locally.*

Note that for the chosen IV difference here, the first round with nonzero difference of R_1 is round 11, however the active Sboxes are still 0. In round 12 and 13 we have one active Sbox each and in round 14 we have 4, since the input is the output of R_2^{12} added to s_5^{12}. We

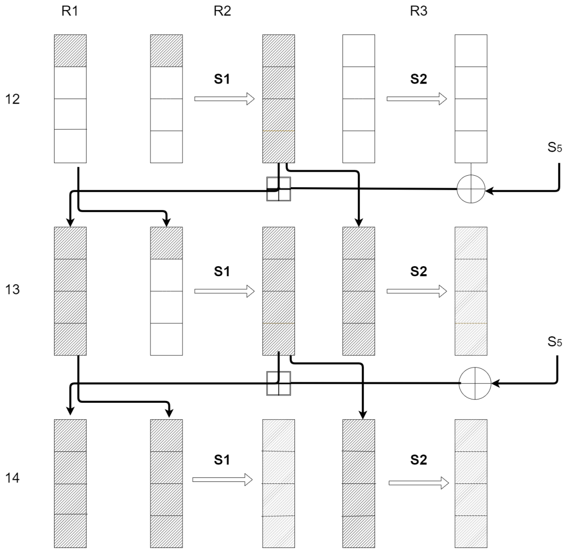


Figure 2: Three consecutive updates of the registers of SNOW 3G according to the chosen difference $a = 0x80000000$. The dark gray blocks indicate active Sboxes, while the light grey blocks indicate that at least one of them will be activated.

start counting the active Sboxes over two rounds from round 13, when the difference has reached all registers and one round has passed. Then over round 13 and 14 we have 5 active Sboxes for the chosen difference a. For any other 2 rounds we have at least 5 active Sboxes, unless we have 0 active Sboxes for one of the transitions.

Observation 5.2 *For two rounds of SNOW 3G we have at least 10 active Sboxes starting from 2 rounds after the difference enters the FSM, unless the difference is canceled locally.*

Again, the chosen difference a enters the first register at round 11, so over rounds 13 and 14 we have at least 10 active Sboxes, 5 from S_1 and at least 5 from S_2. This observation is valid for any other 2 consecutive rounds after round 13, unless we have 0 active Sboxes for one of the transitions.

The same observations are valid for the versions of SNOW 2.0$^{\oplus}$ and SNOW 3G$^{\oplus}$.

6 Conclusion and future work

In this work we adjusted the results of previous study without considering the tweak for easier analysis and extended the analysis by considering the compositions of Sboxes. Further, we looked into the diffusion of the FSM for several SNOW versions in terms of active Sboxes per 2 rounds. Even though they have the same security level until certain number of rounds, after that the diffusion of SNOW 3G is twice as fast as SNOW 2.0.

As a future work we will extend the analysis of composition of Sboxes in order to predict the keystream difference and use the results to perform key-recovery attacks over more rounds. Another research direction is to use the kyestream difference of several consecutive keystream words in order to gain information on the UEA2, since this algorithm uses SNOW 3G as one-time pad.

Acknowledgements. This work was supported by CyberSecurity Research Flanders with reference number VR20192203.In addition, this work was partially supported by the Research Council KU Leuven, C16/18/004, through the IF/C1 on New Block Cipher Structures and by the Flemish Government through FWO Project Locklock G0D3819N. The authors are funded by an FWO PhD and an FWO postdoc fellowship, respectively.

References

[1] O. Billet, H. Gilbert: "Resistance of SNOW 2.0 Against Algebraic Attacks", CT-RSA 2005, 19-28, 2005.

[2] A. Biryukov, D. Priemuth-Schmid, B. Zhang: "Multiset Collision Attacks on Reduced-Round SNOW 3G and SNOW 3G (+)", ACNS 2010: 139-153, 2010

[3] J. Daemen, V. Rijmen: "The Design of Rijndael: AES - The Advanced Encryption Standard", Information Security and Cryptography, Springer 2002

[4] P. Ekdahl, T. Johansson: "A new version of the stream cipher SNOW", SAC 2002, pp. 47-61, 2003

[5] P. Ekdahl, T. Johansson, A. Maximov, J. Yang: "A new SNOW stream cipher called SNOW-V", IACR Trans. Symmetric Cryptol. 2019(3): 1-42, 2019

[6] J.-K. Lee, D.-H. Lee, S. Park: "Cryptanalysis of SOSEMANUK and SNOW 2.0 using linear masks", ASIACRYPT 2008, pp. 524-538, 2008

[7] S. Ma, J. Guan: "Differential attacks on reduced-round SNOW 3G and SNOW 3G(+)", IET Inf. Secur., 14(5):587–594, 2020.

[8] K. Nyberg, J. Wall en: "Improved linear distinguishers for SNOW 2.0", FSE 2006, pp. 144-162, 2006

[9] Specification of the 3GPP Confidentiality and Integrity Algorithms UEA2 & UIA2. Document 5: Design and Evaluation Report, 2006

[10] Specification of the 3GPP Confidentiality and Integrity Algorithms UEA2 & UIA2. Document 2: SNOW3G Specification, 2006. https://www.gsma.com/aboutus/wp-content/uploads/2014/12/snow3gspec.pdf

[11] D. Watanabe, A. Biryukov, C. De Cannière: "A Distinguishing Attack of SNOW 2.0 with Linear Masking Method", SAC 2003, pp. 222-233, 2003

[12] J. Yang, T. Johansson, A. Maximov: "Vectorized linear approximations for attacks on SNOW 3G", IACR Trans. Symmetric Cryptol. 2019(4): 249-271, 2019

[13] B. Zhang, C. Xu, W. Meier: "Fast Correlation Attacks over Extension Fields, Large-Unit Linear Approximation and Cryptanalysis of SNOW 2.0", CRYPTO (1) 2015, pp. 643-662, 2015

Deep Adaptive Beamforming for Photoacoustic Computed Tomography

Joep van de Weem Ben Luijten Ruud van Sloun
Technical University Eindhoven
Dept. Electrical Engineering, signal Processing Systems Group
De Groene Loper 19, 5612 AP Eindhoven
J.C.W.v.d.weem@student.tue.nl W.M.B.luijten@tue.nl R.J.G.v.Sloun@tue.nl

Abstract

Photoacoustic Tomography (PACT) is an emerging field in Biomedical Imaging used for e.g. breast cancer detection and brain lesion detection. Current conventional reconstruction techniques use Delay and Sum (DAS) based backprojection algorithms to reconstruct the image. This paper proposes an end-to-end deep learning approach for the reconstruction of PACT images consisting of two separate neural networks. The first network suppresses sidelobes using neural network-aided adaptive beamforming, and the second extends the bandwidth of the bandlimited data using a bandwidth extension network. From the results of these networks, it can be concluded that the bandwidth extension network proves powerful in obtaining the underlying intensity map. The combined use with the beamforming network shows good results in enhancing the contrast near the center of the image as well as finding vessel structures in in vivo images that where not visible before.

1 Background

Photoacoustic Computed Tomography (PACT) is a relatively new imaging modality that takes advantage of the natural optical contrast of biological tissues and combining this with a high ultrasonic resolution. This modality mainly relies on filtered backprojection or delay and sum beamforming to obtain images. The limited amount of ultrasound transducers together with the bandlimitations of conventional piezoelectric transducers cause limitations in acquiring the underlying map of scatterers.

Here we propose a Deep Beamforming method similar to the ABLE method proposed by Luiten et al. [1] In which we try to find the ideal weights of each channel for each pixel data-adaptively based on a Neural network together with a neural extension of the sensors' bandwidth

2 Methods

In conventional Delay and Sum (DAS) beamforming the output is a uniformly weighted sum of the delayed received signals. In this case, the delay is based on the distance between the source and the individual sensors and is usually estimated by using a fixed speed of sound value. After aligning the received channel signals, these are weighted following a predetermined apodization scheme and summed to yield the final image. The downside of such a method is the inherent design choice between resolution and contrast. Adaptive beamforming algorithms such as the Capon beamformer overcome these issues through data-adaptive weighing of the components at each point in the image.

Here we propose an end-to-end deep learning solution that both optimizes the beamformer apodization data-adaptively and extends the limited bandwidth of the

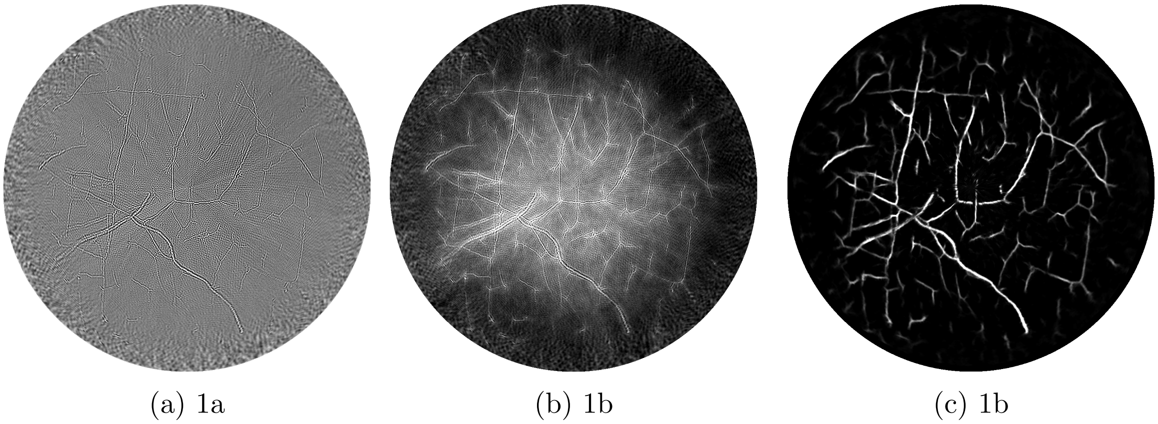

(a) 1a (b) 1b (c) 1b

Figure 1: Results of the beamforming network with a) being regular delay and sum beamforming b) being the output of the beamforming network and c) being the output of the bandwidth extension network.

ultrasonic receivers. The apodization is determined using a network that is based on recent developments in Ultrasonography and is used to minimize the sidelobes and improves the contrast and resolution, effectively tackling the limited view problem. The limited bandwidth problem is tackled using a Unet to find the underlying intensity map by expanding the bandwidth of the band-limited sensors.

For the beamforming network, Due to heavy memory limitations in reconstruction, a convolutional approach to the ABLE beamformer is proposed consisting of multiple convolutional layers based solely on the sensor dimension. The networks are trained using 450 images and tested with 50 images generated using K-wave [2]. The beamforming network is also trained on in vivo images that are beamformed using a capon beamformer for improved generalization.

3 Results

As can be seen in Figure 1. the beamforming network shows great improvements in resolution as compared to regular beamforming. The bandwidth extension however shows similar results as compared to using a UNET on regular delay and sum beamformed data.

References

[1] B. Luijten et al., “Adaptive Ultrasound Beamforming using Deep Learning,” arXiv:1909.10342 [eess], Sep. 2019

[2] B. E. Treeby and B. T. Cox, “k-Wave: MATLAB toolbox for the simulation and reconstruction of photoacoustic wave fields,” J Biomed Opt, vol. 15, no. 2, p. 021314, Apr. 2010, doi: 10.1117/1.3360308.

Tracklet-based Vessel Re-identification for Multi-Camera Vessel-Speed Enforcement

H.G.J. Groot[1] M.H. Zwemer[1,2] E. Bondarev[1]
R.G.J. Wijnhoven[2] P.H.N. de With[1]

[1]Eindhoven University of Technology, Dept. of EE, VCA Research Group
P.O. Box 513, 5600 MB Eindhoven, the Netherlands
[2]ViNotion B.V., Daalakkersweg 2-58, 5641 JA Eindhoven, the Netherlands
emails: {H.G.J.Groot, M.Zwemer, E.Bondarev}@tue.nl,
rob.wijnhoven@vinotion.nl, P.H.N.de.With@tue.nl

Abstract

In crowded waterways, maritime traffic is bound to speed regulations for safety reasons. Although several speed measurement techniques exist for road traffic, such systems are not available for vessels. This paper proposes a new approach for tracklet-based re-identification (re-ID) as a solution for vessel-speed enforcement. For evaluation, the Vessel-reID dataset is used that we introduced in previous work [2]. The core of the tracklet re-ID approach is based on a novel Tracklet-based Querying Procedure as a more effective alternative to the Common Querying Procedure (CQP) found in popular re-ID datasets [7, 8]. The existing procedure randomly selects a single image from the whole query-vessel trajectory (in one camera view). This is improved by (1) detecting a set of most representative images per tracklet of a query-vessel, and by (2) raising the matching accuracy based on accumulating the gallery similarity scores for all images in the set. In the experimental validation, we adopt two well-known person re-ID algorithms, TriNet [3] and MGN [6], since most re-ID literature focuses on person re-ID. Results show a significant increase in performance by applying the tracklet-based approach instead of CQP: a gain of 5.6% and 8.1% Rank-1 for MGN and TriNet, respectively.

1 Introduction

For maritime traffic, speed regulations are important, since speeding vessels can produce waves with a high water displacement that can be dangerous to other waterway users, especially small boats or swimmers. Road-traffic speed-control systems are typically implemented by applying a radar system to detect the speeding vehicles in combination with a camera system to record their license plates. Systems that measure the speed of road users over a longer trajectory also exist and are typically implemented by applying license-plate detection in two sufficiently distant cameras. However, for maritime traffic, none of these systems are suitable, since (1) vessels do not have well-defined license plates or other standardized visual registration markers, and (2) radar systems are too expensive and have difficulties with irregularly maneuvering vessels.

The overall visual vessel appearance of maritime traffic is often unique, because most vessels are of different vessel type, have a different bow or cabin, or can have distinguishing details such as flags or buoys. Therefore, visual re-identification of vessels between various camera locations is theoretically possible. To this end, we introduce a novel speed-enforcement system of vessels by applying video-based vessel re-identification (re-ID). Since this system measures the speed of vessels over a long

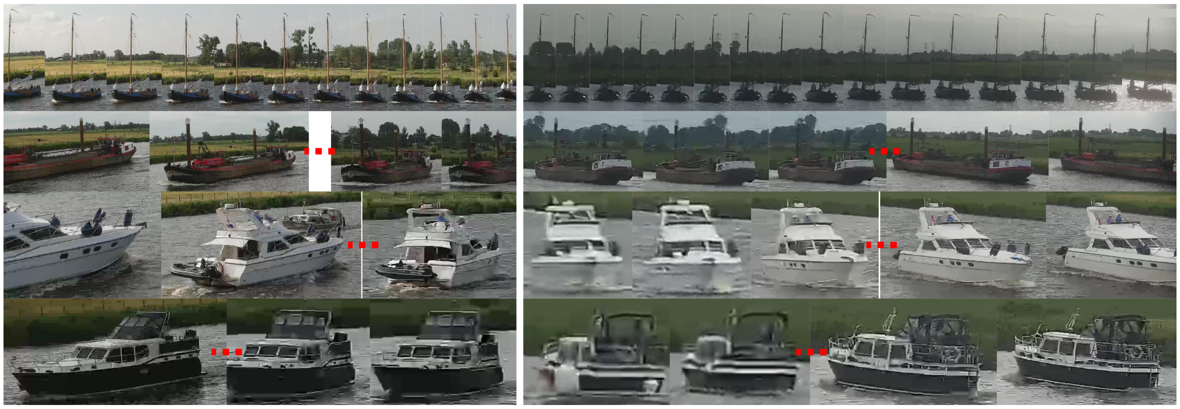


Figure 1: Visual example of several vessel trajectories in Camera 1 (left) and Camera 2 (right) of the Vessel-reID dataset, where each row shows a vessel instance in our dataset. Some images are skipped for visualization (denoted by dotted red line). Image from [2].

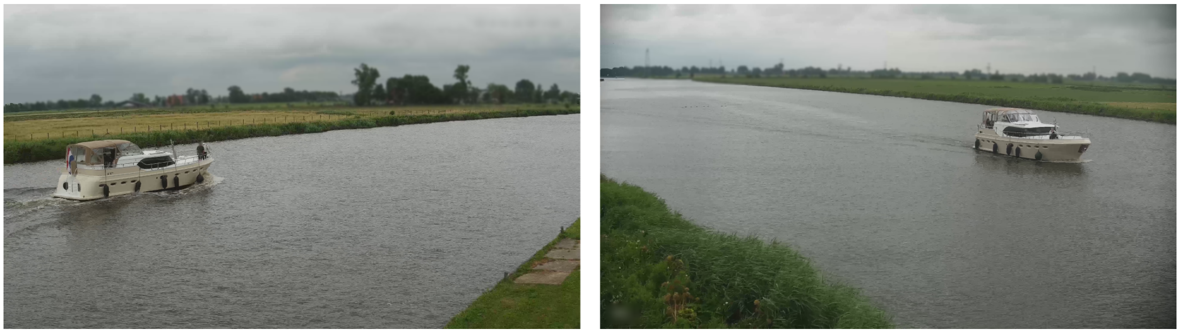


Figure 2: Example images of the same vessel appearing in Camera 1 (left) and 2 (right). Image adopted from [2]

trajectory, this will inherently lead to an accurate speed measurement. However, visual vessel re-ID poses several challenges, such as fluctuating weather and lighting conditions. Moreover, depending on the camera angle and position constraints, it should deal with varying poses of the object with respect to the camera.

This work builds further on our previous research [2], where we have introduced a novel Vessel-reID dataset to investigate whether visual-based vessel re-ID is feasible in practice. This Vessel-reID dataset is captured with two cameras placed several kilometers apart and contains image cutouts of all the detected vessels. In total, there are 2,474 unique vessels, where several example cutouts are shown in Figure 1 for both cameras. These image cutouts are automatically determined by cropping and selecting the bounding-box detections that a Single Shot Multibox Detector (SSD) [5, 9] produces on the full-frame camera footage, see Figure 2. Since all popular re-ID datasets use this image cutout format [7, 8], our Vessel-reID dataset is suitable to evaluate re-ID techniques on vessels. Contrary to these datasets though, we will test our technique on new unseen data during other weather conditions, instead of mixing the data of all captured conditions.

In this paper, we improve the querying procedure commonly applied in re-ID literature. Furthermore, as compared to our previous work [2], we incorporate an improved re-ID model. These two contributions further enhance re-ID performance and application feasibility. We refer to our initial work [2] for the details of creating the Vessel-reID dataset and focus on re-ID in this paper.

The main contributions of this paper are as follows. First, a novel Tracklet-based

Querying Procedure is introduced as an alternative to the querying procedure commonly applied in nearly all re-ID literature. Second, it is shown that the Tracklet-based Querying Procedure significantly improves baseline performance, independent of the re-ID model used.

The remainder of the paper is structured as follows. Section 2, explains important related work on re-identification. Section 3 introduces the system overview for the proposed Tracklet-based Querying Procedure. Section 4 discusses the experimental validation of the procedure. The paper finalizes in Section 5 with the conclusions.

2 Related work

Most re-ID literature focuses on person re-ID, where persons are tracked when they travel from one camera view to another. Recently, re-ID algorithms have also been evaluated on the road vehicle object class, while achieving high performance [1]. This shows that re-ID learning networks do generalize well to other domains, such as vessel detection. These techniques often adopt the use of Convolutional Neural Networks (CNNs), since their introduction clearly made re-ID algorithms more mature [4]. Initially, pairwise verification CNNs were popular, where the networks trained using image pairs of samples from either similar or different persons [4]. By applying the contrastive loss, these networks learn to embed the visual properties of the persons into meaningful feature vector descriptions. Verification CNNs achieve this by increasing the feature distance between images of different persons, while decreasing the distance between images of the same person. However, at present nearly all re-ID algorithms are metric-embedding CNNs [3, 1, 6]. These are trained by sets of 3 images, so-called triplets. A triplet consists of an anchor image, an image of the same person (as in the anchor), and an image of a different person. The training is guided by applying the corresponding triplet loss. Together with a more-effective mining strategy, this type of CNN is popular and offers state-of-the-art performance [3]. The TriNet network [3] proves this concept and still continues to perform well.

Another often exploited technique is to focus on image partitions, for instance on a person's head, body and legs. Here, the Multiple Granularity Network (MGN) [6] devotes 3 separate ResNet-50 branches to different partitionings. Furthermore, MGN also introduces a training strategy that involves both the triplet loss and the cross-entropy loss. The combination of both contributions results in a significant performance gain. The majority of related work, including [3, 1, 6], utilize the ResNet-50 network as a backbone architecture.

3 System overview

As depicted in Figure 3, the re-ID system aims to find a vessel of the same identity as the query-vessel in the available gallery. The Query-vessel tracklet shown in the figure can originate from any camera and represents a newly detected vessel when it appears. The gallery consists of all images of all vessels previously seen by the camera network. All images from both the Query-vessel tracklet and the gallery, are embedded before matching commences in the depicted Tracklet-based matching component. The two Feature embedding components are responsible for creating those embeddings. We experimentally validate our Tracklet-based matching component with two well-known person re-ID algorithms, TriNet [3] and MGN [6]. The Feature embedding component is thus either TriNet or MGN.

The core of the proposed re-ID system, i.e. the Tracklet-based matching component, is presented in Figure 4. From now on, this component is referred to as the Tracklet-based Querying Procedure. The key property of this procedure is that it utilizes several images from the query tracklet instead of just one. More exactly, it selects and

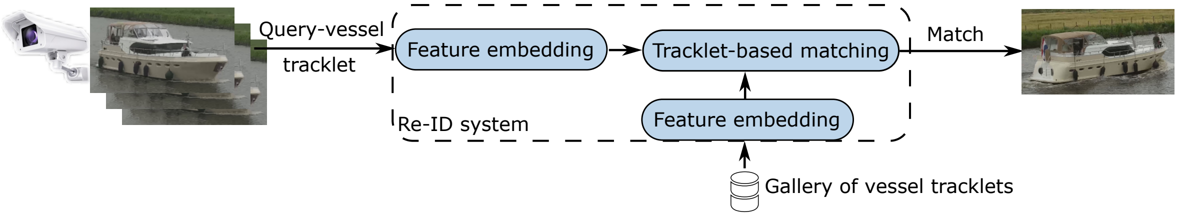


Figure 3: System overview of the re-ID system. The Query-vessel tracklet may originate from any camera and the Tracklet-based matching component is illustrated in Figure 4.

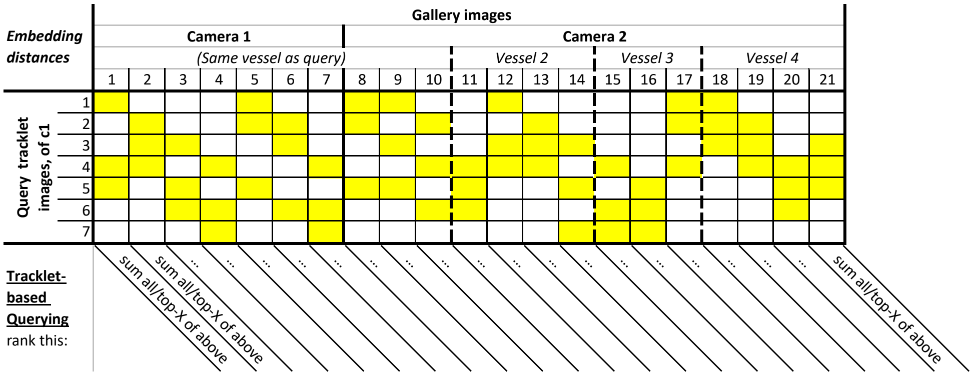


Figure 4: Visualization of our novel Tracklet-based Querying Procedure. The yellow boxes indicate the top-3 query tracklet images per gallery image (i.e. per column). For conciseness, there is only a single vessel visualized in the gallery of Camera 1 (c1).

exploits only the most representative images of the query tracklet. To achieve this, the procedure contains the following processing steps. First, the Euclidian distance between every embedded image in the query tracklet and every embedded image in the gallery is determined. This results in the matrix as depicted in Figure 4. Note that this Euclidian distance is effectively a similarity score, where smaller means more similar. Second, for every gallery image (for every column shown), all similarity scores are accumulated into a single (accumulated) similarity score. The accumulation happens by summing only the Top-X scores in that column, i.e. the X smallest distances in the column. These correspond to the most representative images of the query tracklet. The end result of this step is the horizontal vector illustrated underneath the matrix in Figure 4. Third, the vessel ID corresponding to the gallery image with the smallest accumulated similarity score defines the match. This process is called ranking. Since the gallery per definition also contains the query tracklet, the accumulated similarity scores corresponding to the query tracklet are excluded during ranking. Otherwise, the re-ID system would always be correct, since the embedding distance towards the same image is zero.

In line with the most popular (person) re-ID datasets [7, 8], we adopt the so-called closed-set approach. This means that each vessel tracklet in the test set is alternatingly used as the query, while all vessel tracklets are in the gallery. Several examples of the applied images are shown in Figure 1. During matching, re-ID algorithms typically apply the Common Querying Procedure, which is also used here. This procedure would replace the Tracklet-based matching component in Figure 3 and is substantially different from the proposed Tracklet-based Querying Procedure as explained above. In the Common Querying Procedure, a single image from the whole query-vessel tracklet (in one camera view) is selected randomly as the query-vessel representative. During ranking, the gallery image closest to this single representing image (in embedding

Table 1: Re-ID performance on our Vessel-reID dataset, when applying the novel querying procedure. The best performance a re-ID model achieves is indicated in bold. Results reported as mean (std.dev.) over 10 cycles.

	Accuracy on test set [%]			
	TriNet		**MGN**	
Description	**Rank-1**	**mAP**	**Rank-1**	**mAP**
Common Querying Procedure	55.9 (±1.4)	49.7 (±1.0)	68.9 (±0.9)	62.6 (±0.6)
Tracklet-based Querying				
All	61.9 (±1.4)	55.0 (±1.0)	73.9 (±0.7)	68.1 (±0.5)
Top-10	63.7 (±1.1)	56.3 (±0.8)	73.8 (±0.9)	68.0 (±0.6)
Top-15	**64.0** (±**1.2**)	**56.5** (±**0.9**)	74.3 (±0.9)	68.2 (±0.6)
Top-25	63.8 (±1.2)	56.4 (±0.9)	**74.5** (±**0.9**)	**68.3** (±**0.5**)

space) defines the most likely match. Hence, in Figure 4, only a single random row of the matrix would be incorporated during ranking. However, this single random image is likely not always a good representative of the query vessel. For instance, it occurs often that the query-vessel is temporarily occluded by another vessel, e.g. a vessel that travels in the opposite direction. Therefore, we propose to apply the Tracklet-based Querying Procedure as a more effective alternative. It should be noted that the Common Querying Procedure is implied by the popular datasets and their definition of the query set.

4 Experiments

To evaluate the proposed querying procedure, we use Rank-1 and mAP (mean Average Precision). The Rank-1 is most important because this metric indicates best how the overall system will perform in practice, i.e. in an open-set case. The adopted TriNet and MGN re-ID models are trained 10 times, to determine the resulting re-ID performance on the test set of each training cycle and to report mean and standard deviations on the results. These results are presented in Table 1. For the Tracklet-based Querying Procedure, the per-gallery-image similarity scores are accumulated by summing only the Top-X scores of that gallery image, as described in Section 3. For the experimental validation, we choose X=*all* (all query tracklet images are summed), and X=10, 15, and 25 (also indicated in Table 1).

Evidently, the proposed Tracklet-based Querying Procedure is significantly beneficial for re-ID, starting with a performance gain of 5.0% Rank-1 (73.9%), when summing all query-tracklet images for MGN. Further improvement is observed with the Top-15 (74.3%) and Top-25 (74.5%) matching methods. Clearly, using a multi-image query is preferred over a single-image query. Moreover, when using multiple but not all images, the matching method can filter out the least representative images of the query vessel. This explains why the *Top-X* approach shows higher performance than the *All* approach. Intuitively, when a query-vessel is occluded in some of its tracklet images, those samples can be removed with the *Top-X* approach.

Finally, when comparing TriNet with MGN, it is clear that the gain in performance is substantial and independent of the applied re-ID model. Ultimately, the improvements are 8.1% and 5.6% Rank-1, for TriNet and MGN, respectively. The gain is somewhat lower for MGN and this can be explained by the baseline performance of MGN. As this is significantly higher, +13.0% Rank-1 compared to TriNet, it is also harder to improve this performance. Furthermore, although the difference is well within the observed spread between training cycles, it is remarkable that Top-15 performs better than Top-25 for TriNet, while for MGN this is vice versa.

5 Conclusions

This paper has presented a novel tracklet-based approach for vessel re-ID. In contrast to common re-ID datasets, our Vessel-reID dataset provides the full tracklet for each query-vessel. This helps to assess the impact of the tracklet-based procedure on re-ID performance. In comparison with the querying procedure of existing re-ID literature, the proposed tracklet-based alternative exploits most images from the query-vessel tracklet instead of only a single image. Moreover, since the tracklet-based alternative solely uses the most representative images of the tracklet, it is less susceptible to temporal occlusions. Experimental evaluation reveals that the Tracklet-based Querying Procedure significantly outperforms the Common Querying Procedure, elevating performance from 68.9% Rank-1 to 74.5% Rank-1 on our Vessel-reID dataset. For the intended vessel-speed enforcement application, re-ID is applied to measure the vessel speeds. For this, the final re-ID result is attractive in two ways. First, the obtained Rank-1 score presents a feasible performance for a practical system, since the performance level supports direct law enforcement. Second, the result is still conservative, because we have tested on new unseen data during other weather conditions, instead of using data with mixed conditions.

References

[1] H. Chen, B. Lagadec, and F. Bremond. Partition and reunion: A two-branch neural network for vehicle re-identification. In *CVPR Workshops*, 2019.

[2] H. G. J. Groot., M. H. Zwemer., R. G. J. Wijnhoven., Y. Bondarev., and P. H. N. de With. Vessel-speed enforcement system by multi-camera detection and re-identification. In *Proceedings of the 15th International Joint Conference on Computer Vision, Imaging and Computer Graphics Theory and Applications - Volume 5: VISAPP,*, pp. 268–277. INSTICC, SciTePress, 2020. doi: 10.5220/0008911202680277

[3] A. Hermans*, L. Beyer*, and B. Leibe. In Defense of the Triplet Loss for Person Re-Identification. *arXiv preprint arXiv:1703.07737*, 2017.

[4] S. Karanam, M. Gou, Z. Wu, A. Rates-Borras, O. Camps, and R. J. Radke. A systematic evaluation and benchmark for person re-identification: Features, metrics, and datasets. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 41(3):523–536, March 2019. doi: 10.1109/TPAMI.2018.2807450

[5] W. Liu, D. Anguelov, D. Erhan, C. Szegedy, S. Reed, C.-Y. Fu, and A. C. Berg. Ssd: Single shot multibox detector. In *ECCV*, pp. 21–37. Springer, 2016.

[6] G. Wang, Y. Yuan, X. Chen, J. Li, and X. Zhou. Learning discriminative features with multiple granularities for person re-identification. In *2018 ACM Multimedia Conference on Multimedia Conference*, pp. 274–282. ACM, 2018.

[7] L. Zheng, L. Shen, L. Tian, S. Wang, J. Wang, and Q. Tian. Scalable person re-identification: A benchmark. In *Proceedings of the IEEE International Conference on Computer Vision*, pp. 1116–1124, 2015.

[8] Z. Zheng, L. Zheng, and Y. Yang. Unlabeled samples generated by gan improve the person re-identification baseline in vitro. In *Proceedings of the IEEE International Conference on Computer Vision*, pp. 3754–3762, 2017.

[9] M. H. Zwemer, R. G. Wijnhoven, and P. H. N. de With. Ship detection in harbour surveillance based on large-scale data and cnns. In *Proceedings of the 13th International Joint Conference on Computer Vision, Imaging and Computer Graphics Theory and Applications - Volume 5: VISAPP,*. INSTICC, INSTICC, Funchal, Madeira, Portugal, 2018. doi: 10.5220/0006541501530160

Reducing Transmit Power by Optimizing the Location of Distributed Antenna Arrays

Yi-Hang Zhu, Laura Monteyne, Gilles Callebaut, François Rottenberg, Liesbet Van der Perre
KU Leuven Department of Electrical Engineering (ESAT) DRAMCO
Ghent Technology Campus, 9000 Ghent, Belgium
yihang.zhu@kuleuven.be

Abstract

Unlike the conventional centralized antenna system (CAS), which has all the antennas located in one place, a distributed antenna system (DAS) consists of interconnected antenna arrays which are geographical distributed over an area. Previous work has demonstrated the advantage of the DAS in terms of capacity and power efficiency [1]. The existing studies also show that the performance of the DAS can be further improved by optimizing the placement for its arrays [2, 3]. This work extends the state-of-the-art, by including both mutual coupling and antenna gain. We optimize the array placement to minimize the total transmit power while maintaining a minimum required received signal power. The problem configuration is illustrated in Figure 1. The impact of different factors, including mutual coupling, cell size, and signal coherence, on the optimal array placement and the transmit power is investigated. Our results demonstrate that (1) the mutual coupling effect has a much larger impact on the DAS's optimal array placement compared to the other two factors, (2) the energy saving by using the optimized DAS compared to the unoptimized DAS and the CAS is around 3 dB and 11 dB, respectively when the path-loss exponent equals two and using patch antennas. Furthermore, our analysis provides a guideline for determining the array placement for the DAS in practical scenarios.

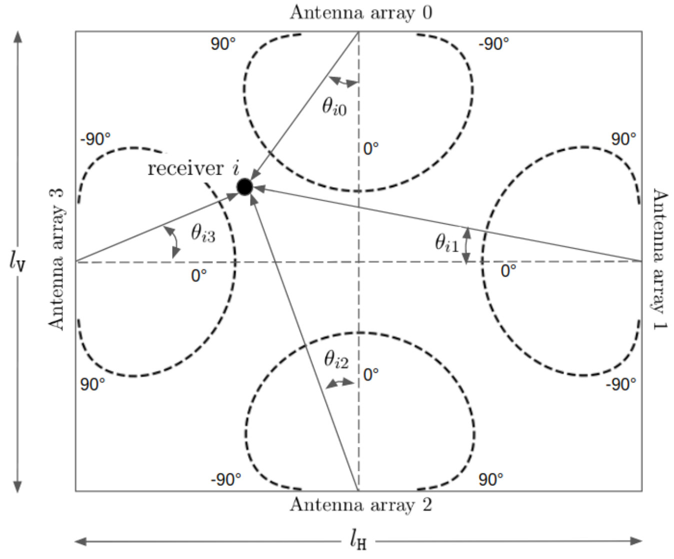


Figure 1: Problem configuration, where the dashed curves represent the gain patterns

References

[1] Firouzabadi, S. and Goldsmith, A., 2011, June. Optimal placement of distributed antennas in cellular systems. In 2011 IEEE 12th International Workshop on Signal Processing Advances in Wireless Communications (pp. 461-465). IEEE.

[2] Han, L., Tang, Y., Shao, S. and Wu, T., 2010, May. On the design of antenna location for OSTBC with distributed transmit antennas in a circular cell. In 2010 IEEE International Conference on Communications (pp. 1-5). IEEE.

[3] Zhang, Y. and Dai, L., 2020. On the Optimal Placement of Base Station Antennas for Distributed Antenna Systems. IEEE Communications Letters, 24(12), pp.2878-2882.

Predicting co-existence of sources through self-supervised Contrastive Predictive Coding to reveal underlying structures during sleep

I.Huijben, L. Hermans, M. van Gilst, R. van Sloun, S. Overeem

Sleep staging is concerned with assigning a label to each 30 seconds of biophysiological data, acquired during someone's sleep. These labels follow the AASM guidelines [1], and the resulting representation is called a hypnogram. In recent years, machine learning developments have moved the field of sleep medicine to find automatic sleep staging algorithms in order to alleviate the burden of manual scoring. Inter-rater disagreement between multiple human scorers is typically present due to (among others) data ambiguities, but state-of-the-art algorithms have reported on par performance with consensus scorings from multiple scorers [2].

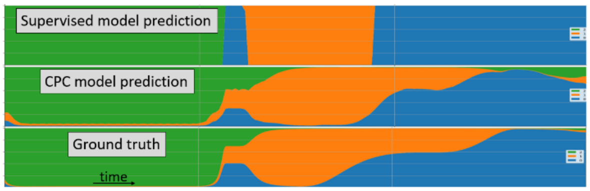
The authors of [3] propose the hypnodensity plot; a visual representation that does not show the most prevalent sleep stage, but a probabilistic mixture of stages at each 30 seconds. The predicted probabilities were found to correspond well with inter-rater disagreement. Co-existence of characteristics of multiple sleep stages was typically found around transitions, in which the hypnogram would switch from one (discrete) state to another. We argue that this co-existence is not necessarily only expected near transitions, as sleep is an inherently continuous biological process. We thus aim to develop a model that predicts this co-existence of characteristics, rather than the inter-rater disagreement, as done in [3].

Predicting co-existence is a blind-source separation problem. Conventional approaches (e.g. NMF or ICA) make specific assumptions on the sources, and can not guarantee that each predicted source corresponds to (exactly) one of the classes of interest. The authors of [3] leverage a convolutional neural network (therewith circumventing these drawbacks), trained in a supervised fashion using the (hard/discrete) sleep stage predictions of human scorers. Final 'soft'/probabilistic predictions are gathered using the softmax output of the model. We hypothesize that, by training a model to mimic discrete sleep stages, the model is primed to learn hard decision boundaries. Therefore, the softmax probabilities are thus expected to provide skewed probabilities with a too high probability being assigned to the most prevalent class.

We propose to train a convolutional neural encoder in a self-supervised fashion, using recently proposed Contrastive Predictive Coding (CPC) [4], in order to learn encodings that are not encouraged to learn hard decision boundaries. To finally get a probabilistic prediction over sleep stages, we freeze the encoder network and train a supervised linear classifier (with softmax activation) on top of this frozen latent space, using the discrete labels. Even though these labels are thus still used for this second phase of training, the entire encoder model is frozen, and therefore uninfluenced by the discrete labels.

As the ground truth co-existence of different sources (or sleep stages) is unknown for data acquired during someone's sleep, we generate a toy dataset in order to validate the model. Each datapoint comprises linear combinations of three different sinusoidals, of which the envelopes are modelled as a smoothened square wave (with random phase) over time. One linear combination thus yields a measurement with moments when either one, two or all three sinusoidals are concurrently present (i.e. envelope >0).

We train our encoder model both supervised (with the hard labels), and unsupervised using the InfoNCE loss as used in CPC. After training a supervised classifier on top of the frozen CPC encoding, we compare the soft predictions of the two models. As expected, the supervised predictions showed much harder decision boundaries than the CPC model. The figure shows one example of the test set (each color denotes a source). The average Kullback-Leibler divergence (a metric that compares probability distributions) over 25 randomly generated test samples was found to be 11.2 and 0.04 for the supervised and unsupervised (CPC) model, respectively.



[1] Iber (2007). The AASM manual for the scoring of sleep and associated events.
[2] Fiorillo *et al.* (2019). Automated sleep scoring: A review of the latest approaches.
[3] Stephansen *et al.* (2018). Neural network analysis of sleep stages enables efficient diagnosis of narcolepsy.
[4] Oord *et al.* (2018). Representation learning with contrastive predictive coding.

High resolution fast 3D imaging with deep learning based adaptive beamforming

Boudewine W. Ossenkoppele[1], Ben Luijten[2], Deep Bera[4], Nico de Jong[1,3], Ruud J.G. van Sloun[2,5] and Martin D. Verweij[1,3]

[1]Laboratory of Medical Imaging, Department of Imaging Physics, Delft University of Technology, Delft, The Netherlands, [2]Department of Electrical Engineering, Eindhoven University of Technology, Eindhoven, The Netherlands, [3]Biomedical Engineering, Thorax Center, Erasmus MC, Rotterdam, The Netherlands, [4]Philips Research, Bangalore, India,[5]Philips Research, Eindhoven, The Netherlands

Abstract

Introduction

Realizing fast volumetric ultrasound imaging, facilitated trough plane- or divergent transmit fields, often comes at the expense of spatial resolution. Additionally, micro-beamforming techniques, needed to achieve a manageable data rate and channel count for matrix probes, further reduce the resolution. Data processing techniques aiming to improve resolution in 3D are required to allow for such limitations in data streaming.

Minimum variance beamforming (MVBF) can improve resolution, compared to conventional delay-and-sum (DAS) beamforming. However, the high computational cost hampers real time applications, especially in 3D imaging. Recently, deep learning-based beamformers have been employed to improve resolution in 2D images. Inspired by MVBF, adaptive beamforming by deep learning ABLE [1] yielded a resolution comparable to MVBF at low computational cost using a neural network (NN) to optimize apodization weights. Similarly, the Delay-And-Neural-Network (DANN) [2] method employs a NN on aligned per-channel data, but utilizes simulations to generate a high resolution target. Recently, we used simulations to create a high-resolution training target for large 3D arrays, that is not limited by the capability of current beamformers. Here, we propose a modified ABLE beamformer, trained on the high-resolution targets.

Methods

We considered a real adult matrix TEE probe with a split transmit receive design. Each volumetric image is a (weighted) combination of 85 steered transmit-receive cycles [3]. The 2048 received signals are micro-beamformed in 4x4 sub-arrays to form 128 channels. The real transducer and a virtual transducer with 2.3 times the number of receive elements were simulated in Field II [4] for 3 volumes with point scatterers. All 4608 channels of the larger array were used to create DAS target voxels. The micro-beamformed and subsequently focused data of the array formed the input to the ABLE network. The ABLE architecture was modified to increase the receptive field. Resolution is evaluated on simulations and a tissue mimicking phantom.

Results and discussion

ABLE improves resolution compared to DAS in simulations (Fig. 1). In the CIRS phantom, the FWHM of the point spread function improved from 7.5° for DAS to 3.6° for ABLE without decreasing the GCNR, showing that the resolution

improvement generalizes to in vitro imaging.

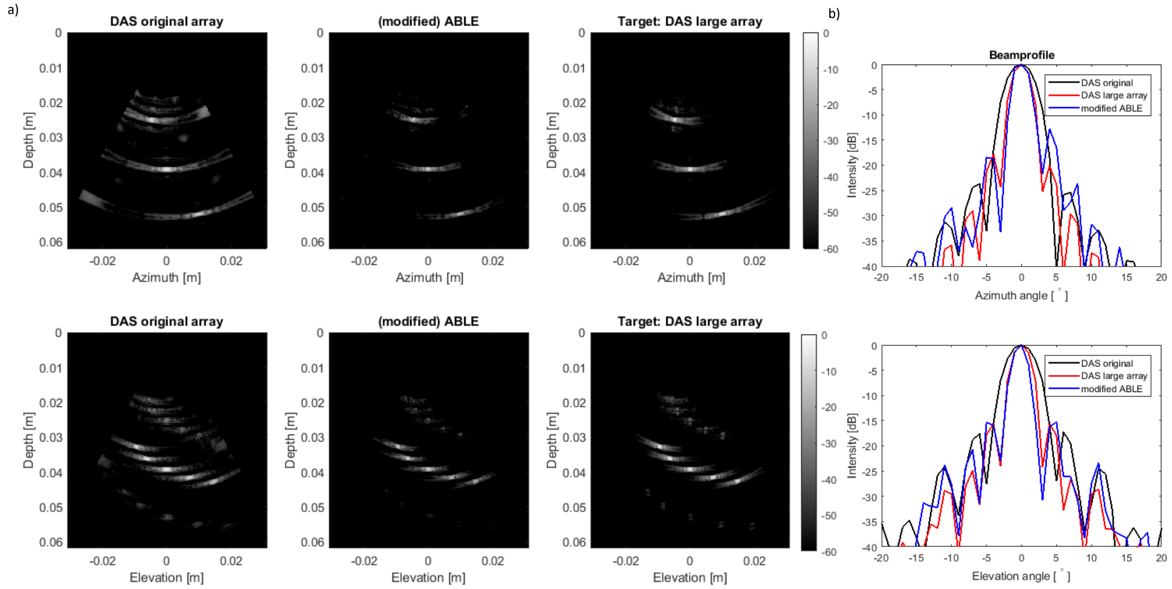


Figure 1: a) Images of a simulated phantom with 15 point scatterers, reconstructed with the original array and DAS (left) the original array and modified ABLE (middle) and the larger array and DAS (right). b) Point spread function around the middle scatterer.

References

[1] B. Luijten et al., ”Adaptive Ultrasound Beamforming using Deep Learning,” IEEE Trans. Med. Imaging, December 2020, doi: 10.1109/TMI.2020.3008537.

[2] F. Vignon et al. ”Resolution Improvement with a Fully Convolutional Neural Network Applied to Aligned Per-Channel data,” 2020 IEEE International Ultrasonics Symposium (IUS) doi: 10.1109/IUS46767.2020.9251482.

[3] D. Bera et al., “Fast Volumetric Imaging Using a Matrix Transesophageal Echocardiography Probe with Partitioned Transmit–Receive Array,” Ultrasound Med. Biol., vol. 44, no. 9, pp. 2025–2042, 2018.

[4] J. A. Jensen, “Field: A program for simulating ultrasound systems,” Med. Biol. Eng. Comput., vol. 4, no. 1, pp. 351–353, 1996.

Quantum signatures with smaller keys

Boris Škorić

We introduce a variant of Gottesman-Chuang quantum signatures [1] in which nonbinary symbols are signed instead of bits. The public keys are fingerprinting states, just as in [1], but we allow for multiple ways to reveal the private key partially. This reduces the number of qubits expended per message bit. We give a security proof and we present numerical results that show how the improvement in public key size depends on the message length. (See [2] for the full paper.)

Gottesman and Chuang [1] (GC01) introduced *quantum digital signatures*. GC01 is based on the fact that state preparation is an asymmetric operation. The private key is the random classical information that Peggy puts into a pair of qudits. The two qudits are the public key. Peggy signs a classical bit by revealing the content of one of the two qudits. Verification consists of a projective measurement. GC01 is information-theoretically secure, in contrast to classical schemes, which need computational hardness assumptions.

We enable Peggy to 'open' a public key in multiple ways. Our public-key qudits are fingerprinting states [3, 4] $\frac{1}{\sqrt{d}}\sum_{j=0}^{d-1}(-1)^{x_j}|j\rangle$, where d is the dimension of the Hilbert space and $x \in \{0,1\}^d$ is the embedded string. We write $(d-\ell)S = d$, with S the alphabet size. In order to sign a symbol in $\{0,\ldots,S-1\}$, Peggy reveals x except for a block of $d-\ell$ bits. Security against forgery is derived from the fact that $d-\ell$ is sufficiently large to make guessing x infeasible given the revealed bits. To further improve efficiency, our scheme uses the idea suggested in [1] to work with codewords instead of repeated public keys. Peggy 'opens' a set of qudits according to the symbols that make up a codeword; Victor counts the number of correctly opened qudits, which allows him to distinguish between a correct signature and a forgery.

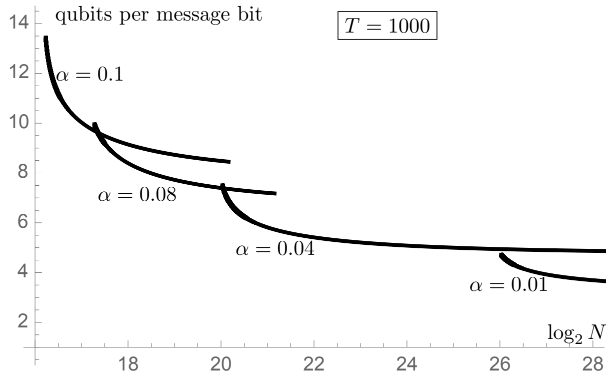


Figure 1: *Number of spent qubits per signed message bit versus the codeword length N (in bits), for $T = 1000$ recipients of the public key, and various values of $\alpha = 1/S$. False accept rate 10^{-12}; False reject rate 10^{-9}. In each curve the dimension d is varied. The curve for GC01 is almost flat, at ≈14 qubits per message bit.*

References

[1] D. Gottesman and I.L. Chuang. Quantum digital signatures, 2001. `https://arxiv.org/abs/quant-ph/0105032`.

[2] B. Škoricć. Quantum digital signatures with smaller public keys, 2020. `https://arxiv.org/abs/2012.15493`.

[3] H. Buhrman, R. Cleve, J. Watrous, and R. de Wolf. Quantum fingerprinting. *Physical Review Letters*, 87(16):167902, 2001.

[4] D. Gavinsky and T. Ito. Quantum fingerprints that keep secrets. *J. Quantum Inf. Comput.*, 13:583–606, 2013.

Performance analysis of IEEE 802.15.4 Networks under various Interference patterns

A. A. van der Pijl C. Papatsimpa
Eindhoven University of Technology
Dept. Electrical Engineering, Signal Processing Systems Group
Groene Loper 3, Eindhoven, the Netherlands
a.a.v.d.pijl@student.tue.nl c.papatsimpa@tue.nl

Wireless sensor networks (WSNs) have become ubiquitous over the last decade as they have enabled a wide range of new applications ranging from industrial monitoring to lighting control and building automation. In order to address the demands for low power and low data rate connectivity of low-cost devices, several protocols and standards have been established, among which, IEEE 802.15.4/ZigBee has received considerable attention. It finds its way particularly in sizeable lighting control networks.

IEEE 802.15.4 operates in the license free 2.4 GHz ISM band, which can be exploited by multiple networks and technologies at the same time. This medium sharing is known to result in co-existence issues among the collocated protocols. In particular, numerous experimental studies [1, 4] show that Wi-Fi interference can heavily impair IEEE 802.15.4 throughput IEEE 802.11 (Wi-Fi) transmission power is much higher than ZigBee and has 10-20 times shorter access timing. This higher transmit power and its impatience give IEEE 802.11 nodes access priority, thereby causing unfairness to the ZigBee nodes. Thus, analyzing the performance of the WSN subject to external interference is becoming increasingly critical, as the usage of the 2.4 GHz ISM band is growing.

Co-existence issues between the ZigBee and external interference, in particular, Wi-Fi, have also been studied in the past both experimentally [3], [2] and with the use of analytical models [5]. Yet, existing studies use the traffic load or clear channel rate as the only metric that characterizes Wi-Fi interference. Unfortunately, this metric is unable to distinguish between bursty traffic (which is characteristic in Wi-Fi) and a channel that has periodic traffic pattern with the same average traffic load. As this paper shows, this leads to major inaccuracies. It appears a necessity to consider both the statistical characteristics of external interference and the characteristics of the ZigBee network.

In our previous work [6], we have developed a model based on mean-field analysis to analyze the performance of a large-scale WSN under external interference, where node interaction may give rise to a complex behavior and patterns that cannot be found considering the single node model, but emerge by their interaction. Our theoretical work showed that the type of external interference traffic, the typical traffic mix of different frame types and their temporal characteristics, plays a significant role on the ZigBee network performance. We showed that for the same clear channel rate, the performance of the network can vary hugely, depending on the distribution of interference, i.e., the distribution of busy and idle periods. We now perform a large-scale simulation analysis to validate the outcomes of this theoretical analysis.

In particular, we use the OMNET simulator to compare the effects of periodic and Poisson generated traffic patterns on network performance. Our results show that under the same clear channel rate, the pattern of the interference traffic has a major influence on the network performance. In fact, in terms of success probability, randomness and burstiness of external interference tend to lead to better performance. It appears that the availability of large gaps plays a significant role on the network performance as during the wide gaps, ZigBee nodes can transmit a larger number of packets. We find these results very relevant towards a more realistic understanding of the performance of large IoT networks in the ISM band.

References

[1] J. H. Hauer, V. Handziski, and A. Wolisz. Experimental study of the impact of WLAN interference on IEEE 802.15.4 body area networks. In *Wireless Sensor Networks*, pages 17–32, Berlin, Heidelberg, 2009. Springer.

[2] Yu-Kai Huang and Ai-Chun Pang. A Comprehensive Study of Low-power Operation in IEEE 802.15.4. In *Proceedings of the 10th ACM Symposium on Modeling, Analysis, and Simulation of Wireless and Mobile Systems*, MSWiM '07, pages 405–408, New York, NY, USA, 2007. ACM.

[3] Mikko Kohvakka, Mauri Kuorilehto, Marko Hännikäinen, and Timo D. Hämäläinen. Performance analysis of ieee 802.15.4 and zigbee for large-scale wireless sensor network applications. In *Proceedings of the 3rd ACM International Workshop on Performance Evaluation of Wireless Ad Hoc, Sensor and Ubiquitous Networks*, PE-WASUN '06, pages 48–57, New York, NY, USA, 2006. ACM.

[4] R. Musaloiu-E and A. Terzis. Minimising the effect of wifi interference in 802.15.4 wireless sensor networks. *Int. J. Sen. Netw.*, 3(1):43–54, December 2008.

[5] C. Papatsimpa and J. P.M.G. Linnartz. Coexistence performance model for a large ZigBee lighting sensor network in the 2.4 GHz ISM band. *Proceedings of the 2017 IEEE 14th International Conference on Networking, Sensing and Control, ICNSC 2017*, pages 465–470, 2017.

[6] Mahmoud Talebi, C. Papatsimpa, and Jean-Paul M.G. Linnartz. Dynamic performance analysis of ieee 802.15.4 networks under intermittent wi-fi interference. In *2018 IEEE 29th Annual International Symposium on Personal, Indoor and Mobile Radio Communications (PIMRC)*, pages 1–7, 2018.