Security Protocols and Resource Allocation for Fifth Generation Networks

Arsenia (ersi) Chorti

ETIS UMR 8051, CY Université, ENSEA, CNRS

Security Protocols and Resource Allocation for Fifth Generation Networks

Arsenia (Ersi) Chorti

Date of HDR Defence: 12th October 2020
Jury President: J.-M. Gorce,
Jury Members: M. Coupechoux, G. Rekaya Ben Othman, G. Fettweis,
K. Salamatian, C. Hollanti, I. Fijalkow, I. Andriyanova

I hereby declare that this thesis and the work reported herein was composed by and originated entirely from me, the following PhD students I co-supervise: M. Mitev, S. Skaperas, G. S. Nunez and M. Bello, and their theses's directors, unless otherwise stated. Information derived from the published and unpublished work of others has been acknowledged in the text and references are given in the list of sources.

Arsenia (Ersi) Chorti (2020)

Abstract

In my role as a research advisor, I strove to give to my students the opportunity to work on promising and far fetching – whenever possible – research topics in the general framework of fifth generation (5G) wireless. The works presented in this thesis reflect our studies in two areas of central importance for bringing 5G to life: wireless security and resource allocation.

With respect to security, novel challenges emerged in 5G with the Internet of things (IoT) paradigm and device to device (D2D) low latency communications. Novel verticals, such as haptics and vehicle to everything (V2X), require low complexity and low latency security mechanisms, particularly in the context of device authentication. In the present manuscript, lightweight solutions for device authentication using physical unclonable functions (PUF) and secret key generation (SKG) at the physical layer are presented.

Furthermore, as video content is responsible for more than 70% of the global IP traffic, it is important for content delivery infrastructures to rapidly detect and respond to changes in content popularity dynamics. In this thesis, we propose a flexible edge resource allocation approach leveraging unikernel and container technologies. The allocation of the edge server resources is driven by a real-time and low-complexity content popularity detector, implemented using off-line and on-line change point analysis. Variations of these algorithms have applications in intrusion detection in wireless sensor software defined networks, discussed next.

Finally, the potential use of non-orthogonal multiple access (NOMA) in the wireless uplink is considered. Early results on the performance comparison of NOMA vs orthogonal allocation schemes in asymptotic regimes, show that the gains in using NOMA carry on to the scenario of communications under statistical delay quality of service (QoS) constrains.

Dans mon rôle de co-encadrent de thèse, je me suis efforcé de donner à mes étudiants l'occasion de travailler sur des sujets de recherche prometteurs et fondamentaux dans le cadre général de communications sans fil de cinquième génération (5G). Les œuvres présentées dans cette thèse reflètent nos études dans deux domaines d'importance centrale pour la réalisation de la 5G : la sécurité et l'allocation des ressources.

En ce qui concerne la sécurité, de nouveaux défis sont apparus en 5G avec le paradigme de l'Internet des objets (IoT) et les communications device to device (D2D) à faible latence. Les nouvelles verticales, telles que l'haptique et les communications véhiculaires (V2X), nécessitent une faible complexité et des mécanismes de sécurité à faible latence, en particulier dans le contexte de l'authentification. Dans cette thèse, des solutions d'authentification de légère complexité en utilisant des fonctions physiques inclonables (PUF) et des générations de clés secrètes (SKG) à la couche physique sont présentées.

En outre, comme le contenu vidéo est responsable de plus de 70% du trafic IP mondial, il est important que les infrastructures de diffusion de contenu détectent et répondent rapidement aux changements de la dynamique de popularité du contenu. Dans cette thèse, nous proposons une approche flexible d'allocation des ressources qui tire parti des technologies unikernel et containers. L'allocation des ressources est entraînée par un détecteur de popularité de contenu en temps réel et à faible complexité, mis en œuvre à l'aide des analyses hors ligne et en ligne des points de changement. Des variantes de ces algorithmes ont des applications dans la détection d'intrusion dans les réseaux définis par les logiciels de capteurs sans fil, qui sont discutés ensuite.

Enfin, l'utilisation potentielle d'un accès multiple non orthogonal (NOMA) dans le lien ascendant sans fil est envisagée. Les premiers résultats de la comparaison des systèmes d'allocation NOMA par rapport aux schémas orthogonaux dans les régimes asymptotiques, montrent que les gains dans l'utilisation de NOMA se poursuivent dans le scénario des communications sous des contraintes statistiques de délai de qualité de service (QoS).

Acknowledgements

I would like to take this opportunity to thank wholeheartedly the *Telecom girls*, Inbar Fijalkow, Iryna Andriyanova, Veronica Belmega, Marwa Chafii, Laura Luzzi, and, also Mylène Pischella, Marine Moguen and Aymeric Histace.

To my family

Blessed are the cheesemakers.

- The Life of Brian (1979)

Contents

List of Tables

List of Figures

Nomenclature

List of Abbreviations

0-RTT Zero round trip time

3GPP The 3rd Generation Partnership Project

AE Authenticated encryption

B5G Beyond 5G

BF-AWGN Block fading additive white Gaussian noise

CRP Challenge-response pair

CSI Channel state information

EAP-TLS Extensible authentication protocol-transport layer security

IoT Internet of things

MiM Man in the middle

mMTC massive machine type communications

NB-IoT Narrow band IoT

OFDM Orthogonal frequency division multiplexing

PHY Physical layer

PKE Public key encryption

PLS Physical layer security

PUF Physical unclonable function

QoS Quality of service

RAN Radio access network

RSS Received signal strength

SKG Secret key generation

SNR Signal-to-noise ratio

STEK Session ticket encryption key

TLS Transport layer security

URLLC Ultra reliable low latency communication

V2X Vehicle-to-everything communication

Chapter 1

Activity Review

1.1 Motivation for Application for the HdR Diploma

With this thesis, I wish to submit my application for the Habilitation to Direct Research at CY - Cergy Paris Université. Currently, I am a Maître de Conférences at the Ecole Nationale Supérieure de l'Electronique et de ses Applications (ENSEA) in Cergy and in parallel I have a Visiting Research Fellow status at the Department of Electrical and Electronic Engineering of Princeton University in the USA and at the School of Computer Science and Electronic Engineering of the University of Essex in the UK.

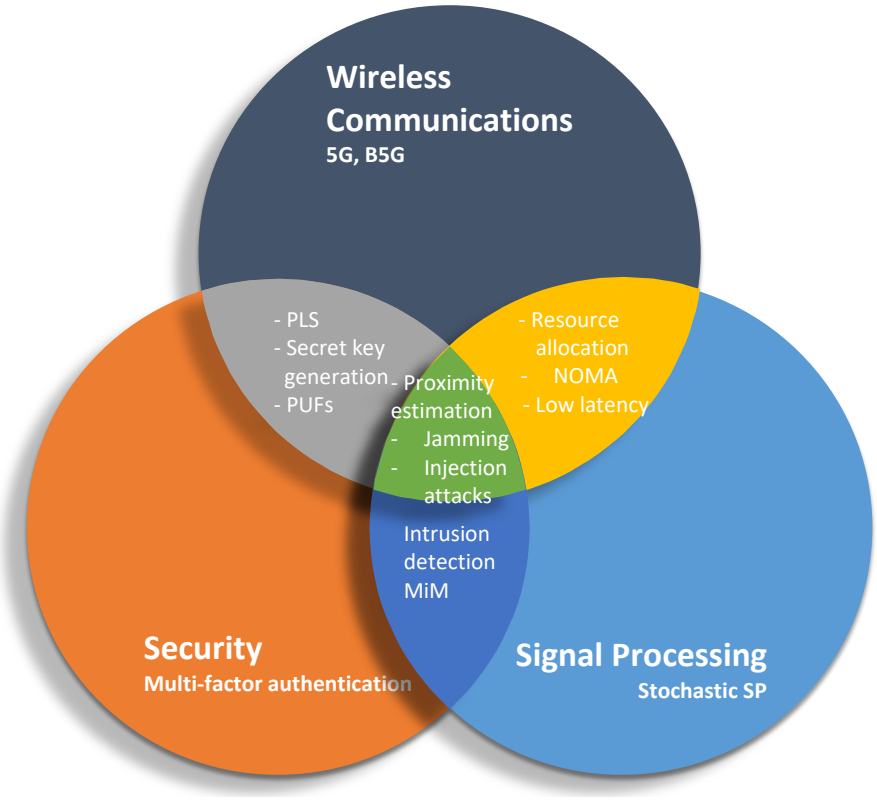

Figure 1.1: Recent research areas and topics

My current research activities relate to various topics in wireless communications and physical layer security with an emphasis on the proposal of low latency communication systems and the development of new security protocols for future generations of wireless. I actively work on topics in flexible numerology, non-orthogonal mutiple access (NOMA) and fast authentication protocols for

delay constrained systems using physical unclonable functions (PUFs) and RF fingerprinting. In this framework, with my current research team, that comprises four PhD students and two postdoctoral researchers, we investigate resource allocation in beyond fifth generation (B5G) leveraging NOMA, the efficient design of Slepian Wolf and Wyner Ziv reconciliation decoders at the short block-length, the development of zero-round-trip-time (0-RTT) authentication protocols using resumption keys generated from wireless fading coefficients, the analysis of the wireless channel secrecy capacity under statistical delay quality of service (QoS) constraints and the development of quick anomaly detection algorithms for software defined networks.

My research lies at the interface of wireless communications, signal processing and security studies, as depicted in Fig. 1.1; at this – not so-frequented – scientific crossroad, new engineering problems are encountered, a few of which will be discussed in later chapters of this thesis, along with proposed solutions.

With respect to my contribution as an academic teacher and supervisor, I have a long experience in teaching and supervising students in security, coding and wireless communications for more than 8 years in the UK and France. I have had the chance to teach a variety of courses both at the undergraduate and graduate level and contribute in teaching in the continuous education engineering track of ENSEA. I have taught to a variety of class sizes and have customarily received very positive feedback from my students, both in formal assessment and in face-to-face interaction. Furthermore, since last September I am acting as the liaison of the international mobility for ENSEA students towards the UK and have secured internships at Imperial College London, the University of York, etc.

In my academic employment I have had the opportunity to undertake a number of important administrative responsibilities. I currently head the research team ICI (information, communications, imaging) of the ETIS Lab that comprises 13 permanent faculty (3 PU, 8 MCF, 2 CNRS CR) and more than 18 research students and teaching fellows. In this role my aim is to help maintain and enhance the quality and quantity of the team's collective research output, its ability to attract research funding and good young researchers, increase further the team's visibility in the national and international level and ensure the team members work in a friendly and fertile learning environment. Furthermore, during my employment at the School of Computer Science and Electronic Engineering in the UK, I have acted in 2017 as the President of the Athena Swan Committee, steering 15 faculty and admin staff for the preparation of the department's gender equality and diversity charter.

With respect to my involvement in professional bodies, I am a member of the IEEE INGR Roadmap Security Workgroup, of the IEEE P1940 Standardization Workgroup on "Standard profiles for ISO 8583 authentication services" and have been a member of the IEEE Teaching Awards Committee for the last three years.

I feel that my overall experience is of sufficient standing to allow me to lead independent research and act as a stand-alone thesis advisor / director. A brief overview of my research and teaching activities the last 8 years is provided in Sections 1.4 and 1.5 respectively. First, I introduce myself to the reader through a detailed academic CV in Section 1.2. and an full list of my publications in Section 1.3.

1.2 Curriculum Vitae and List of Publications

In the following pages my full academic Curriculum Vitae is provided, including a full record of my publications in Section 1.3. The interested reader may also consult my web page and my google scholar page.

Dr. Arsenia Chorti

Address: Room 341, ENSEA, 6 Avenue du Ponceau, Cergy, FR
Telephone: +33 (0)769113367
e-mail: arsenia.chorti@ensea.fr achorti@princeton.edu

1.2.1 Current Position / Responsibilities (in chronological order)

Sep. 2017-present: **ENSEA (ETIS UMR8051) Associate Professor (MCF)** in Communications and Networking, Research Group: **4 PhD students, 2 postdocs**

Sep. 2017-Jul. 2020: **Member of** the IEEE Teaching Awards Committee

Sep. 2019 – present: **PEDR** (prime d'encadrement doctoral et recherche) – premium for excellence in supervision and research

Sep. 2019 – present: **Member of the IEEE P1940 Standardization Workgroup** on "Standard profiles for ISO 8583 authentication services"

April 2020-present: **Head of the Information, Communications and Imaging (ICI) Group of ETIS UMR 8051** (*Responsable d'équipe information, communications et imagerie*), comprising 3 Professors, 8 MCF, 2 CNRS Researchers, 2 Postdocs, 2 ATERs, 14 PhD students

Apr. 2020: Award of **CNRS delegation** (**half year travel sabbatical**) to visit Prof. H.V. Poor (Princeton University, NJ USA), Prof. T. Rappaport (NYU, NYC USA) and Dr. A. Barolo (Barkhausen Institute, Dresden DE) in Spring / Summer 2021

May 2020: **Member of the IEEE International Network Generations Roadmap (INGR) Security Workgroup** (pre-standardization workgroup for security in future networks)

June 2020: Elevated to **IEEE Senior Member**

1.2.2 Research Interests

My current research spans the areas of wireless security and beyond fifth generation (B5G) networks. I work on the design of security schemes for B5G, with a particular focus on physical layer security; my recent contributions concern fast authentication protocols using physical unclonable functions (PUFs) and secret key generation (SKG) from shared randomness, with proximity / localization as an extra authentication factor. Furthermore, I work on low latency communications, leveraging recent results on non-orthogonal multiple access (NOMA), investigate polynomial complexity algorithms for flexible numerology and eMBB – URLLC coexistence and joint PHY-MAC resource allocation optimization using the theory of the effective capacity. Recent contributions (since 2017) include:

- **Wireless security for B5G and Internet of things (IoT)** [J19], [J21], [C37], [C33], [S2]
- **Authentication protocols** leveraging PUFs, SKG and proximity estimation [BC3], [U1]
- **Resource allocation in 5G** using change point analysis [J17], [J20], [C32]
- **Anomaly detection** in software defined networks [C39], [J18], [S1], [U2]
- **Active attacks** in PHY [J15], [J16], [C36], [C28-C31] [BC2]
- **Low latency B5G communications**, non-orthogonal multiple access (**NOMA**), **NOMA-R**, **flexible numerology** for B5G [J22], [U3-U5]

1.2.3 Education

2000-2005 **Imperial College London:** Department of Electrical and Electronic Engineering
Ph.D. in Communications and Signal Processing
Thesis Title: *"The Impact of Circuit Nonlinearities and Noise in OFDM Receivers*", Supervisor: Mike Brookes, Scholarship awarded by I.K.Y. and Panasonic UK ltd.

1999-2000 **Université Pierre et Marie Curie – Paris VI**
MSc (D.E.A.) in Electronics
Dissertation Title: *"F.P.G.A. Implementation of Multi-Layer Perceptron Neural Network for Real-Time Applications in High Energy Physics",* Supervisor: Prof. Patrick Garda, I.K.Y. Scholar

1992-1998 **University of Patras:** Department of Electrical and Computer Science Engineering
M.Eng (Diploma) in Electrical Engineering
Dissertation Title: *"Development of User-Friendly Interface for the Testing of Nodal Cards of an Industrial Network"*, Supervisor: Prof. K. Koumbias

1.2.4 Academic Employment

- September 2017 – present: **ENSEA**, **Associate Professor in Communications and Networks**
- October 2013 – August 2017: **University of Essex**, School of Computer Science and Electrical Engineering, **Lecturer in Communications and Networks** and subsequently **Visiting Research Fellow** (ongoing)
- November 2012 – October 2013: **Foundation for Research and Technology Hellas (FORTH)**, Institute of Computer Science, **International Outgoing Fellow (IOF) Marie Curie Research Fellow**
- May 2011 – present: **Princeton University**, Dep. of Electrical Engineering, **IOF Marie Curie Fellow** and subsequently **Visiting Research Fellow**
- December 2008 – April 2011: **Middlesex University UK,** School of Engineering and Information Sciences, Dep. of Computer Communications, **Senior Lecturer in Communications and Networks**
- October 2007 – September 2009: **University College London (UCL),** Dep. of Electronic and Electrical Engineering, **Postdoctoral Research Fellow** and subsequently **Visiting Researcher**
- October 2006 – September 2007: **Technical University of Crete (TUC),** Department of Mineral Resources Engineering, Resources Detection and Identification Research Unit, **Postdoctoral Research Fellow**
- October 2005 – September 2006: **University of Southampton,** School of Electronics and Computer Science, Electronic Systems Design Group (ECS), **Postdoctoral Research Fellow**

1.2.5 Research Funding and Grants

Project proposals currently under review:

- **Principal Investigator (PI) of ANR PRCE project HERCULES** (enhancement measures in the security of beyond fifth generation networks) **2nd round AAPG 2020**, with the SME Montimage, K. Salamatian (LISTIC), I. Andriyanova (ETIS), A. Histace (ETIS) and F. Ghaffari (ETIS)
- **External collaborator** of project **LEON** (Intelligent Network Softwarization for the Internet of Things), ELIDEK, GR, with Dr. L. Mamatas
- **PI** project **PROCOPE PHC** (travel grant) to visit the Barkhausen Institute, DE in 2021-2022

Ongoing projects:

- **Co-investigator (co-I)** project **PHEBE** (Physical layer security for beyond fifth generation communications) with L. Wang (PI), L. Luzzi, M. Chafii, M. Le Treust**, Paris-Seine Excellence Initiative, 2020-2024**, 400,000€
- **PI** project **SAFEST with F. Jardel (NOKIA Bell Labs)** (Physical layer security for future generations wireless systems)**, DIM RFSI**, **2019-2020**, 27,500€
- **PI** project **eNiGMA** (Non-orthogonal multiple access techniques under security and delay constraints)**,** with I. Fijalkow, **Paris-Seine Excellence Initiative**, **2019-2021**, 110,000€
- **Co-I** project **ELIOT** (Enabling technologies for IoT), **ANR PRCI with Univ. Sao Paolo, Brazil**, with V. Belmega (PI), I. Andriyanova, I. Fijalkow, J. Lorandel, Role: **Leader of WP on IoT security**, **2019-2023**, ETIS: 390,420€ (total of 740 k€)

Past projects:

- **PI** SRV-ENSEA de l'Institut des Etudes Avancées Université Paris Seine, **2018-2019**: 3,000€
- **PI** SRV-ENSEA Institut des Etudes Avancées Université Paris Seine: **2017-2018** : 2,850€
- **PI** project **PHOTINO**, University of Essex, Research and Innovation Fund: **2014-2015**: £13,000

- **Co-I FP7 PEOPLE** Marie Curie IOF, project **APLOE with H.V. Poor (Princeton University)**, 245,448€, **2010-2013**
- PG Scholarship from the **State Scholarships Foundation of Greece–I.K.Y. 2000-2004**: £41,820

1.2.6 Teaching and Related Responsibilities at ENSEA (since 2017)

- 2019-present: **Responsible of student international mobility to the UK,** 2nd year MEng, 3rd year MEng, Erasmus programme with the UK
- 2019-present **Instructor in the MSc (M2R) module "Cryptography and Network Security", University Cergy Pontoise,** Master 2 Informatique et Ingénierie des Systèmes Complexes (IISC), specilization SIC (Signal, Information, Communications),
- 2018-present: **Responsible of the module** "**Network security**" 3rd year MEng, ENSEA
- 2018-present: **Responsible of module** "**Interconnexion réseaux**" 3rd year Cycle par Alternance, ENSEA
- 2017-present: **Responsible of the Option Internet of Things "Option IoT"**, 2nd year MEng, ENSEA
- 2017-present: **Instructor** "**IoT Security**", 2nd year MEng, ENSEA
- 2017-present: **Responsible of the module** "**Internetworking**", 3rd year MEng, ENSEA
- 2017-present: **Instructor** "**Wireless Communications**", 3rd year MEng, ENSEA
- 2017-present: **Lab instructor** in various courses, including Digital Communications, Internetworking, Signals and Systems, etc.

1.2.7 Research Supervision

Current supervision

- **PhD Student Mr. Miroslav Mitev: supervision @60%**, 25/4/2017-9/2020, "*Physical layer security for the Internet of things*", co-supervised with Dr. M. Reed, University of Essex, UK, Thesis VIVA (defence) scheduled for September 2020, publications: [J21], [C37], [C36], [C33], [P1], [U1]
- **PhD student Mr. Sotiris Skaperas: supervision @40%**, 1/9/2017-9/2020, "*Data analysis and forecasting models for flexible resource management in 5th generation networks*", co-supervised with Dr. L. Mamatas, University of Macedonia in Thessaloniki, GR, Thesis defence scheduled for September 2020, publications: [J20], [J17], [C32], [U5]
- **PhD student Mr. Gustavo Alonso Nunez Segura: supervision @35%**, 1/2/2019-projected to finish in 1/2022 (4-year thesis programme in Brazil), "Cooperative *Intrusion Detection System for Software Defined Wireless Sensor Networks*", co-supervised with Prof. Cintia Borges Margi, University of Sao Paolo, Brazil, publications: [J18], [C39], [C35], [S1], [U2]
- **PhD student Mr. Mouktar Bello: supervision @70%**, 1/11/2020-projected to finish in 10/2023, "*Meeting delay and security constraints in 6G wireless networks*", co-supervised with Prof. I. Fijalkow, ETIS/ENSEA, FR, publications: [C38], [U3, U4]
- **Postdoc Dr. Mahdi Shakiba Herfeh: supervision @100%**, 21/11/2019-20/5/2021 (fixed term 1.5 years), "*Physical layer security for IoT applications*", project ELIOT ANR PRCI, ETIS/ENSEA FR, publications: [BC3], [U1]
- **Postdoc Dr. Nasim Ferdosian: supervision @90%**, 1/1/2020-31/12/2021 (fixed term 2 years), "Non-orthogonal multiple access techniques under security and delay constraints", with Prof. I. Fijalkow, ETIS/ENSEA, FR, publications: [U5]

Past supervision

- **MSD (Master by Thesis – full year research project) student Cornelius Saiki: supervision @84%**, 1/9/2014-31/8/2015, "*A Novel Physical Layer Key Generation and Authenticated Encryption Protocol Exploiting Shared Randomness*", co-supervised with Prof. S. Walker, University of Essex, publications: [C27]

- **MSc (M2R) SIC student Gada Rezgui: supervision @50%**, 1/3/2017-31-8-2017, "*Energy Harvesting as a Means to Mitigate Jamming Attacks; a Game Theoretic Analysis"*, co-supervised with V. Belmega, ETIS/University of Cergy Pontoise, publications: [J16]
- **MSc (M2R) SIC student Rihem Nasfi: supervision @100%,** 1/11/2018-15/3/2019, Projet d'Initiation à la Recherche (PIR), *"Non-orthogonal multiple access networks under QoS delay constraints",* publications : [C34]
- **MSc (M2R) SIC student Gada Rezgui**, **supervision @50%**, 1/11/2016-15/3/2017, Projet d'Initiation à la Recherche (PIR), *"Secret Key Generation systems under Jamming Attacks via Game Theoretic Tools"*
- **MSc (M2R) IMD student Amani Gran**, **supervision @100%**, 1/11/2018-15/3/2019, Projet d'Initiation à la Recherche (PIR), *"IoT lightweight security"*
- **MSc (M2R) SIC student Fatiha Ait Larbi**, **supervision @100%**, 1/11/2018-15/3/2019, Projet d'Initiation à la Recherche (PIR), *"Cross-layer security protocol design"*
- **MSc (M2R) SIC student Mouad Nahri**, **supervision @100%**, 1/11/2019-15/3/2020, Projet d'Initiation à la Recherche (PIR), *"Flexible numerology for B5G"*
- **Other MSc/BSc supervision:** 5 MSc and 9 BSc dissertations at the University of Essex and more than 10 MSc and BSc dissertations at Middlesex University

1.2.8 Recruitment (Selection) Committees / Thesis Examiner

- May 2020: Recruitment Committee (**Comité de sélection**) **for a MCF** post at CY Cergy University on *Networks and Security*
- Sep. 2019: Recruitment Committee (**Comité de sélection**) **for a MCF** post at EISTI on *Cybersecurity*
- Jun. 2020: **Thesis Examiner (rapporteur),** A. Ben Hadj Fredj, Télécom ParisTech, supervisors Prof. G. Rekaya and Prof. J-C Belfiore, "Computations for Multiple Access Channels in Wireless Networks"
- Jan. 2019: **Thesis Reviewer,** L. Senigagliesi, Univ. Polytechnica delle Marche, supervisors Prof. L. Spalazzi and Prof. M. Baldi, "Information-theoretic security techniques for data communications and storage"
- Aug. 2014: **Thesis Examiner,** I. K. Musa, CSEE University of Essex UK, supervisor Prof. S. Walker, "Optimized Self-Service Resource Containers for Next Generation Cloud Delivery"

1.2.9 Workshop Organization / Keynotes / Tutorials

- **Tutorial** on "Statistical methods in physical layer security", **IEEE Statistical Signal Processing (SSP) Workshop**, July 2020, Rio de Janeiro, BR (*rescheduled to July 2021 due to COVID-19*)
- **Special Session Organizer**, "Selected topics on 6G security", **IEEE ISWCS**, Sep. 2020, Berlin, Germany (*rescheduled to Sep. 2021 due to COVID-19*)
- **Special Session Organizer**, "Statistical Methods for IoT", **IEEE SSP 2020**, Jul. 2020, Rio de Janeiro, Brazil (*postponed to July 2021 due to COVID-19*)
- **Training School Co-organizer** (with M. Chafii, S. Stanczak and R. Cavalcante), **"**Machine Learning for Communications", 3-4 Sep. 2020, Berlin (co-located with ISWCS, *rescheduled to Sep. 2021 due to Covid-19*)
- **Chair of the GdR ISIS Workshop** "Women in Communications, Information Theory and Signal Processing", May 19 2020 (*rescheduled to May 2021 due to Covid-19*)
- **Chair of the GdR ISIS Workshop** "Enabling ultra-reliability, low latency and massive connectivity", June 18 2020 (*virtual event due to Covid-19*)
- **Keynote IEEE PIMRC Workshop Security Public RATs**: "Practical examples of physical layer security", 4 Sep. 2016, Valencia, Spain
- **Chair** of the **workshop ACCESS** - Cutting edge topics in physical layer security, communications and distributed storage workshop, 11 May 2014, Aalborg, Denmark

- **Co-chair of** "2nd Women's Workshop on Communications and Signal Processing", 16-18 July 2014, Princeton NJ, US
- **Track chair of the IEEE Global Wireless Summit 2014**, 11-14 May 2014, Aaborg, Denmark
- **Chair** of the "Second International Conference on Communications, Connectivity, Convergence, Content and Cooperation", 11-14 May 2014, Aalborg, Denmark
- **Chair** of the "WirelessVITAE, 10-13 May 2014, Aalborg, Denmark

1.2.10 Editor / Reviewer / Selected TPCs

- 2020- present: **Associate Editor** of the **IEEE Open Journal on Signal Processing** (OJSP)
- Sep. 2019-present: **Lead Guest Editor, EURASIP JWCN Special Issue** "Physical layer security solutions for 5G-and-beyond", Editors: S. Tomasin, H.V. Poor, M. Baldi, S. El Ruayheb, X. Wang, to appear in 2020
- 2018-2019: **Executive Editor Transactions on Emerging Telecommunications Technologies** (ETT), *Wiley*
- 2017-2019: **Executive Editor** of **Internet Technology Letters** (ITL), *Wiley*
- **Reviewer**: IEEE Transactions (Trans.) on Information (Inf.) Forensics and Security, Elsevier Computers and Security, IEEE Trans. on Wireless Communications (Commun.)., IEEE Trans. Signal Processing, IEEE Trans. Vehicular Technologies, IEEE JSAC, IEEE Wireless Commun. Letters (L.), IEEE Commun. L., Trans. on Emerging Telecom Tech. (ETT), Eurasip JWCN, IEEE Trans on Commun., …
- **TPCs**: more than 30 TPCs, indicatively IEEE GLOBECOM 2015, 2016, 2017, 2018, 2019, 2020, IEEE ICC 2014, 2015, 2016, 2018, 2019, 2020, IEEE WCNC 2016, 2019 (executive member), …

1.2.11 Selected Invited Talks (after 2016)

- July **2019**, "Physical layer security in delay constrained applications", **NOKIA Bell Labs**, FR
- May **2019**, "Physical layer security in delay constrained applications", **Barkhausen Institute**, Dresden DE
- May **2019**, "Physical layer security in delay constrained applications", **ICS FORTH**, GR
- October **2017,** "Emerging security paradigms", **Thales**, FR
- March **2017,** "Physical layer security for future networks", **British Telecom**, Adastral Park, UK
- June **2016,** "Practical examples of physical layer security", **Summer Research Institute, EPFL**, CH

1.2.12 Past Administrative Responsibilities and Outreach Activities

- 2016-2017: **President** of the Committee for Gender Equality and Diversity *Athena Swan*, Univ. Essex, UK
- 2016-2017: **Vice-president** "Research Student Progress and Management Committee", Univ. Essex, UK
- 2015-present: **Fellow of the Higher Education Academy**, UK (professional title in pedagogical training)
- 2014-2015: **Organizer** of student recruitment activities "Visit Days", Univ. Essex, UK

1.3 Publication List

1.3.1 Books [B] / Book Chapters [BC]

(supervised students and postdocs appear underlined)

BC3 M. Shakiba Herfeh, **A. Chorti**, V.H. Poor, *A Review of Recent Results on Physical Layer Security*, to appear in Springer Nature 2020;

BC2 **A. Chorti**, *A Study of Injection and Jamming Attacks in Wireless Secret Sharing Systems*, (Proc. 2nd Workshop Communication Security, WCS 2017), Lect. Notes in Elect. Eng., vol 447, pp. 1-14, Springer;

BC1 **A. Chorti**, C. Hollanti, J.-C. Belfiore, H.V. Poor, *Physical Layer Security: A Paradigm Shift in Data Confidentiality*, Springer, Lecture Notes in Electrical Engineering - Physical and Data-Link Security Techniques for Future Communication Systems, vol. 358, pp. 1-15, Sep. 2015;

B **A. Chorti**, *The Impact of Circuit Nonlinearities and Noise in OFDM Receivers*, Feb. 2010, Verlag

1.3.2 Refereed International Journals [J]

(supervised students and postdocs appear underlined)

J22 M. Pischella, **A. Chorti**, I. Fijalkow, "On the Performance of NOMA-Relevant Strategies Under Statistical Delay QoS Constraints", *IEEE Wireless Commun. Letters*, in print (early access);

J21 M. Miroslav, **A. Chorti**, M.J. Reed, L. Musavian, "Authenticated Secret Key Generation in Delay Constrained Wireless Systems", *EURASIP J Wireless Com Network*, vol. 122, Jun. 2020;

J20 S. Skaperas, L. Mamatas, **A. Chorti**, "Real-Time Algorithms for the Detection of Changes in the Variance of Video Content Popularity", *IEEE Access*, vol. 8, pp: 30,445-30,457, Feb. 2020;

J19 W. Yu, **A. Chorti**, L. Musavian, V.H. Poor, Q. Ni, "Effective Secrecy Capacity for a Downlink NOMA Network", *IEEE Trans. Wireless Commun.*, vol. 18, no 12, pp: 5,673-5690, Dec. 2019;

J18 G.A. Nunez Segura, C. B. Margi, **A. Chorti**, "Understanding the Performance of Software Defined Wireless Sensor Networks Under Denial of Service Attack", *Open Journal of Internet of things (OJIOT)*, Vol.5, no 1, pp:59-68 Aug. 2019 (published in the OJIOT as a special issue);

J17 S. Skaperas, L. Mamatas, **A. Chorti**, "Real-Time Video Content Popularity Detection Based on Mean Change Point Analysis", *Access*, vol.7 pp: 142,246-142,260, Jul. 2019;

J16 G. Rezgui, E.V. Belmega, **A. Chorti**, "Mitigating Jamming Attacks Using Energy Harvesting", IEEE Wireless Commun. Let., vol. 8 no 1, pp: 297-300, Feb. 2019;

J15 E.V. Belmega, **A. Chorti** "Protecting Secret Key Generation Systems against Jamming: Energy Harvesting and Channel Hopping Approaches", *IEEE Trans. Inf. Forensics Security*, vol. 12, no 11, pp: 2611-2626, Nov. 2017;

J14 D. Karpuk, **A. Chorti**, "Perfect Secrecy in Physical-Layer Network Coding Systems from Structured Interference", *IEEE Trans. Inf. Forensics Security*, vol. 11, no 8, pp. 1875-1887, Aug. 2016;

J13 **A. Chorti**, K. Papadaki, H.V. Poor, "Optimal power allocation in block fading channels with confidential messages", *IEEE Trans. Wireless Commun.*, vol. 14, no 9, pp. 4708-4719, Sep. 2015;

J12 **A. Chorti**, S. Perlaza, Z. Han, H.V. Poor, "On the resilience of wireless multiuser networks to passive and active eavesdroppers", *IEEE Journal of Selected Areas in Commun.*, vol. 31 no 9, pp. 1850-1863, Sep. 2013;

J11 **A. Chorti**, M. Brookes, "On the effect of Voigt profile oscillators on OFDM systems", *IEEE Trans. Circuits Syst. II*, vol. 58, no 11, pp. 768-772, Nov. 2011;

J10 G. Spiliopoulos, D.T. Hristopulos, M.P. Petrakis, **A. Chorti**, "A multigrid method for the estimation of geometric anisotropy in environmental data from sensor networks", Elsevier *Computers and Geosciences*, vol. 37, no 3, pp. 320-330, Mar. 2011;

J9 **A. Chorti**, M. Brookes, "Performance Analysis of COFDM and DAB Receivers in narrow-band and tonal interference", Springer *Telecommunication Systems J.*, vol. 46, no 2, pp. 181-190, 2011.

J8 Y. Kanaras, **A. Chorti**, M. Rodrigues, I. Darwazeh, "A fast constrained sphere decoder for ill conditioned communication systems", *IEEE Commun. Let.*, vol. 14, no 11, pp. 999-1001, Nov. 2010.

J7 **A. Chorti**, "How to model the near-to-the-carrier regime and the lower knee frequency of real RF oscillators", *J. Electrical Computer Eng.*, vol. 2010, article ID 537132, Oct. 2010.

J6 **A. Chorti**, D.T. Hristopulos, "Non-parametric identification of anisotropic correlations in spatially distributed data sets", *IEEE Trans. Signal Proces*, vol. 56, no 10, pp. 4738-4751, Oct. 2008.

J5 D. Karantzas, **A. Chorti**, N.M. White, C.J. Harris, "Teaching old sensors new tricks: archetypes of intelligence", *IEEE Sensors J.*, Special Issue on Intelligent Sensing", invited paper, vol. 7, no 5, pp. 868-881, May 2007.

J4 **A. Chorti**, D. Karatzas, N.M. White and C.J. Harris, "Intelligent Sensors in Software: The Use of Parametric Models for Phase Noise Analysis", *Int. J. Inf. Process.*, vol. 1, no. 2, June 2007.

J3 **A. Chorti**, D. Karantzas, N.M. White and C.J. Harris, "Use of the extended Kalman filter for state dependent drift estimation in weakly nonlinear sensors", *Sensors Let.*, vol. 4, no 4, pp. 377-379, Dec. 2006.

J2 **A. Chorti**, M. Brookes, "A spectral model for RF oscillators with power-law phase noise, *IEEE Trans. Circuits Syst. I*", vol. 53, no 9, pp. 1989-1999, Sep. 2006.

J1 **A. Chorti**, M. Brookes, "On the effects of memoryless nonlinearities on M-QAM and DQPSK OFDM Signals", *IEEE Trans. Microw. Theory Techn.*, vol. 54, no 8, pp. 3301-3315, Aug. 2006.

1.3.3 Refereed International Conference Proceedings [C]

(supervised students and postdocs appear underlined)

C39 G.A. Nunez Segura, S. Skaperas, **A. Chorti**, L. Mamatas, C. Borges Magri, "Denial of Service Attacks Detection in Software-Defined Wireless Sensor Networks", Proc. *IEEE Int. Conf. Commun. (ICC) Worskhop on SDN Security*, Dublin UK, 7-11 Jun. 2020;

C38 B. Mouktar, W. Yu, **A. Chorti**, L. Musavian, "Performance Analysis of NOMA Uplink Networks under Statistical QoS Delay Constraints", Proc. IEEE Int. Conf. Commun. (ICC), Dublin UK, 7-11 Jun. 2020;

C37 M. Mitev, **A. Chorti**, M.J. Reed "Subcarrier Scheduling for Joint Data Transfer and Key Generation Schemes in Multicarrier Systems", Proc. IEEE Int. Global Commun. Conf. (GLOBECOM), Hawaii US, 9-13 Dec. 2019;

C36 M. Mitev, **A. Chorti**, E.V. Belmega, M.J. Reed "Man-in-the-Middle and Denial of Service Attacks in Wireless Secret Key Generation", Proc. *IEEE Global Commun. (GLOBECOM)*, Hawaii US, 9-13 Dec. 2019;

C35 G.A. Nunez Segura, C. B. Margi, **A. Chorti** , "Understanding the Performance of Software Defined Wireless Sensor Networks Under Denial of Service Attack", Proc. Int. Workshop on Very Large IoT (VLIoT) 2019, Los Angeles, US, 30th Aug. 2019 (*invited paper);

C34 R. Nasfi, **A. Chorti**, "Performance Analysis of the Uplink of a Two User NOMA Network under QoS Delay Constraints", Proc. *IEEE Int. Conf. on Ubiquous and Future Networks (ICUFN)* 2018, Zagreb, Croatia, 2-5 July 2019;

C33 M. Mitev, **A. Chorti**, M.J. Reed "Optimal Resource Allocation in Joint Secret Key Generation and Data Transfer Schemes", Proc. *IEEE Int. Conf. Wireless Commun. Mobile Comput. (IWCMC)*, Tangiers Morocco, 24-28 June 2019;

C32 S. Skaperas, L. Mamatas, **A. Chorti**, "Early Video Content Popularity Detection with Change Point Analysis", Proc. *IEEE Int. Global Commun. (GLOBECOM)*, Abu Dhabi, UAE, 6-11 December 2018;

C31 E.V. Belmega, **A. Chorti**, "Energy Harvesting in Secret Key Generation Systems under Jamming Attacks", Proc. *IEEE Int. Conf. Commun. (ICC)*, Paris, France, May 2017;

C30 **A. Chorti**, "Secret Key Generation in Rayleigh Block Fading AWGN Channels under Jamming Attacks", Proc. *IEEE Int. Conf. Commun. (ICC)*, Paris France, May 2017;

C29 **A. Chorti**, "Optimal Signalling Strategies and Power Allocation for Wireless Secret Key Generation Systems in the Presence of a Jammer", Proc. *IEEE Int. Conf. Commun. (ICC)*, Paris, France, May 2017;

C28 **A. Chorti**, "Overcoming limitations of secret key generation in block fading channels under active attacks", Proc. *IEEE 17th Int. Workshop Signal Process. Advances Wireless Commun. (SPAWC)*, pp. 1-5, Jul. 2016 (*invited paper);

C27 C. Saiki, **A. Chorti**, "A novel authenticated encryption protocol exploiting shared randomness", Proc. *IEEE Commun. Network Security (CNS)*, 2nd Workshop on Physical Layer methods for Wireless Security, pp. 651-656, Sep. 2015;

C26 **A. Chorti**, M.M. Molu, D. Karpuk, C. Hollanti, A. Burr, "Strong secrecy in wireless network coding systems with M-QAM modulators", Proc. *IEEE Int. Conf. Commun. China (ICCC)*, pp. 181-186, Oct. 2014;

C25 **A. Chorti**, K. Papadaki, H.V. Poor, "Optimal power allocation in block fading Gaussian channels with causal CSI and secrecy constraints", Proc. *IEEE Global Commun. (GLOBECOM)*, pp. 752-757, Dec. 2014;

C24 S.M. Perlaza, **A. Chorti**, H.V. Poor, Z. Han, "On the trade-offs between networks state knowledge and secrecy", Proc. *IEEE Int. Symp. Wireless Personal Multimedia Commun. (WPMC)*, pp. 1-6, Jun. 2013;

C23 **A. Chorti**, K. Papadaki, P. Tsakalides, H.V. Poor, "The secrecy capacity of block fading multiuser wireless networks", Proc. *IEEE Int. Conf. Adv. Tech. Commun. (ATC)*, pp. 247-251, Oct. 2013, (*best paper award);

C22 S.M. Perlaza, **A. Chorti**, H.V. Poor and Z. Han, "On the impact of network-state knowledge on the feasibility of secrecy", Proc. *IEEE Int. Symp. Inf. Theory (ISIT)*, pp. 2960-2964, Istanbul, Turkey, Jul. 2013;

C21 **A. Chorti**, S. Perlaza, Z. Han, H.V. Poor, "Physical layer security in wireless networks with passive and active eavesdroppers", Proc. *IEEE Global Commun. (GLOBECOM)*, Anaheim, USA, 3-7 Dec. 2012;

C20 **A. Chorti**, "Helping interferer physical layer security strategies for M-QAM and M-PSK systems", Proc. *IEEE CISS* 2012, Princeton NJ, USA, 21-23 Mar. 2012;

C19 **A. Chorti** and V. Poor, "Achievable secrecy rates in physical layer security systems with a helping interferer", Proc. *IEEE Int. Conf. Comp. Netw. Commun. (ICNC)*, Maui, HI, Feb. 2012;

C18 **A. Chorti** and V. Poor, "Faster than Nyquist interference assisted secret communication for OFDM systems, *IEEE Asilomar*, Pacific Grove, CA, US, 4-7 Nov. 2011, (*invited paper);

C17 **A. Chorti**, "Masked M-QAM OFDM: Encryption of OFDM signals through faster than Nyquist signalling", Proc. *IEEE MCECN Global Commun. (GLOBECOM)*, Miami, US, 6-10 Dec. 2010;

C16 **A. Chorti**, Y. Kanaras, M. Rodrigues, I. Darwazeh, "Joint channel equalization and detection of spectrally efficient FDM signals", Proc. *IEEE Personal Indoor Multimedia Radio Commun. (PIMRC)*, Istanbul, Turkey, 26-29 Sep. 2010

C15 Y. Kanaras, **A. Chorti**, M. Rodrigues, I. Darwazeh, "A new quasi-optimal detection algorithm for a non-orthogonal spectrally efficient FDM system", Proc. *Int. Symp. Commun. Inf. Tech. (ISCIT)*, Incheon, Korea, 28-30 Sep. 2009;

C14 Y. Kanaras, **A. Chorti**, M. Rodrigues, I. Darwazeh, "An Overview of Optimal and sub-Optimal Detection Techniques for a Non Orthogonal Spectrally Efficient FDM", Proc. *LCS/NEMS*, London UK, 3-4 Sep. 2009;

C13 **A. Chorti**, Y. Kanaras, "Masked M-QAM OFDM: A simple approach for enhancing the security of OFDM systems", *IEEE Personal Indoor Multimedia Radio Commun. (PIMRC)*, Tokyo, Japan, 13-16 Sep. 2009;

C12 D.T. Hristopulos, M.P. Petrakis, G. Spiliopoulos, **A. Chorti**, "Non-parametric estimation of geometric anisotropy from environmental sensor network measurements", Proc. *StatGIS2009*, Milos, Greece, 17-19 Jun. 2009;

C11 Y. Kanaras, **A. Chorti**, M. Rodrigues, and I. Darwazeh, "Spectrally efficient FDM signals: bandwidth gain at the expense of receiver complexity", *IEEE Int. Conf. Commun. (ICC)*, Dresden, Germany, 13-17 Jun. 2009;

C10 Y. Kanaras, **A. Chorti**, M. Rodrigues, I. Darwazeh, "A near optimum detection for a spectrally efficient non orthogonal FDM system", Proc. *InOWo'08*, Hamburg Germany, 27-28 Aug. 2008;

C9 D.T. Hristopulos, **A. Chorti**, G. Spiliopoulos, E. Petrakis, "Systematic detection of anisotropy in spatial data obtained from environmental monitoring networks", *EGU2008*, Vienna, Austria, 13-18 Apr. 2008;

C8 Y. Kanaras, **A. Chorti**, M. Rodrigues, I. Darwazeh, "A combined MMSE-ML detection for a Gram-Schmidt orthogonalized FDM system", Proc. *IEEE BROADNETS*, London, UK, Sep. 2008;

C7 Y. Kanaras, **A. Chorti**, M. Rodrigues, I. Darwazeh, "Sub-optimum detection techniques for a bandwidth efficient multi-carrier communication system", *Mutli-Strand Conf.*, Milton, UK, 6-7 May 2008.

C6 **A. Chorti**, D.T. Hristopulos, "Automatic detection of spatial anisotropy in environmental data sets", Proc. *StatGIS2007*, Klagenfurt, Austria, Oct. 2007.

C5 A. Moustakas, **A. Chorti** and D.T. Hristopulos, "Geostatistical analysis of tree size distributions in the southern Kalahari, obtained from remotely sensed data", Proc. *SPIE Europe Remote Sensing*, Florence, Italy, 17-20 Sep. 2007.

C4 **A. Chorti** and M. Brookes, "Resolving near carrier spectral infinities due to 1/f phase noise in oscillators", Proc. *IEEE Int. Conf. Acoustics Speech Signal Process. (ICASSP)*, vol. 3, pp. III 1005-III 1008, Hawaii, USA, 15-18 Apr. 2007

C3 **A. Chorti**, D. Karatzas, N.M. White, C.J. Harris, "Intelligent sensors in software: the use of parametric models for phase noise analysis", Proc. *IEEE Int. Conf. Intelligent Sensing Inf.*, Bangalore, India, 15-18 Dec. 2006.

C2 **A. Chorti**, B. Granado, B. Denby, P. Garda, "Une architecture electronique temps reel pour les reseaux connexionnistes en physique des hautes energies", *NSI2000*, Toulouse FR, May 2000.

C1 **A. Chorti**, B. Granado, B. Denby and P. Garda, "An electronic system for the simulation of neural networks with real time constraints", Proc. *ACAT*, Chicago, U.S., Dec. 2000.

1.3.4 Posters

P2 M. Mitev, **A. Chorti**, M.J. Reed, "Physical layer security in wireless networks with active eavesdroppers", *Munich Workshop on Coding and Cryptography (MWCC) 2018*, (*invited poster, Germany, 10-11 April 2018;

P1 **A. Chorti**, "Optimal resource allocation in secure multi-carrier systems", *1st IEEE Women's Workshop Commun. Signal Proc.*, Banff, (*invited poster), Canada, 13-15 Jul. 2012.

1.3.5 In Preparation [U] / Submitted [S]

U1 M. Mitev, M. Shakiba Herfeh, **A. Chorti**, M.J. Reed, "Multi-factor lightweight authentication for the Internet of Things", IEEE Trans. Inf. Forensics Security, in preparation;

U2 G. A. Nunez Segura, **A. Chorti**, C. Borges Magri, "Multimetric centralized and decentralized intrusion detection in software defined networks", IEEE Internet of Things Journal, in preparation;

U3 M. Bello, W. Yu, M. Pischella, **A. Chorti**, I. Fijalkow, L. Musavian, "A Review of DL/UL Multiple Access Enabling Low-Latency Communications", IEEE Access, in preparation;

U4 M. Bello, **A. Chorti**, I. Fijalkow, W. Yu, L. Musavian, "Performance Analysis of NOMA Uplink Networks under Statistical QoS Delay Constraints", IEEE Trans. Communications, in preparation;

U5 N. Ferdosian, **A. Chorti**, S. Skaperas, L. Mamatas, "Unleashing the Potential of Flexible Numerology by Resolving Conflicts", IEEE Trans. Wireless Communications, in preparation;

S1 G.A. Nunez Segura, **A. Chorti**, C. Borges Magri, "Multimetric Online Intrusion Detection in Software-Defined Wireless Sensor Networks", in review IEEE Globecom 2020;

S2 **A. Chorti**, V.H. Poor, "What Physical Layer Security Can Do for 6G", accepted in IEEE Globecom 2020 Tutorials.

1.4 Recent Research Results

1.4.1 Motivation on studying physical layer security and resource allocation for 5G Systems

Physical Layer Security

The goal of physical layer security (PLS) [1–3] is to make use of the properties of the physical layer – including the wireless communication medium and / or the transceiver hardware – to enable critical security aspects. In particular, PLS can be employed to provide i) node (device) authentication, ii) message authentication, iii) message confidentiality through the use of secrecy encoders, and, iv) key management and distribution solutions through symmetric secret key generation from shared randomness. Furthermore, proposals for intrusion detection and counter-jamming at PHY have recently emerged [4]; indeed these two topics emerge as important research areas in B5G systems, particularly in the industrial Internet of things (IoT) and the mmWave era.

PLS has been explicitly mentioned in the first white paper on 6G: "The strongest security protection may be achieved at the physical layer". Importantly, it is stated as an enabling technology in the IEEE International Network Generations Roadmap 1st Edition 2019 in the Chapters on "Security" (Section 1.1 pp. 1-2) and on "Massive MIMO" (Section 4.3 pp. 8-9). The increasing interest in PLS has been stimulated by many practical needs. Notably, many critical IoT networks require ultra-low latency communications ($<$ 1msec), e.g., in autonomous driving and vehicle to everything (V2X) applications, telemedicine and haptics. However, standard authentication often requires significant processing time. We note in passing that in the Third Generation Partnership Project (3GPP) technical report "Study on the Security of URLLC" [5], all aspects related to low latency (fast) authentication remain open and no solutions have so far been standardized. An added complication is due to hardware limitations of low-end sensors and their ineptest to execute sophisticated security protocols such as the IPSec or the DTLS.

A further challenge comes from quantum computing, which has seen significant progress after massive investment by companies such as Google, Intel and IBM to build prototypes with more than 50 qubits. In October 2019 Google published in the journal "Nature" their quantum computer experiments showing they have achieved quantum supremacy for a particular set of problems [6]. In this aspect, PLS, that relies upon information-theoretic security proofs, could resist quantum computers, unlike corresponding asymmetric key schemes relying on the (unproven) intractability in polynomial time of certain algebraic problems. Even state-of-the-art elliptic curve cryptography (ECC) schemes, that require substantially shorter keys than RSA or Diffie Hellman (DH) schemes, are still considerably more intensive computationally than their PLS counterparts and are not post-quantum.

As a result, the study of novel PLS based solution for 5G and B5G security is highly pertinent. Related proposals using physical unclonable functions [7] and secret key generation from shared randomness [8] are included in this thesis.

Resource Allocation

The roll-out of fifth-generation (5G) mobile networks and the forthcoming 6G will bring about fundamental changes in the way we communicate, access services and entertainment.With respect to the latter, the multi-fold increase in the service data rates will provide users with ultra high resolution in video-streaming, multi-media and virtual reality, offering immersive experiences. To this end, it is important for Edge content delivery infrastructures to rapidly detect and respond to changes in content popularity dynamics. For flexible and highly adaptive solutions, the capability for quick resource (re-)allocation should be driven by early (*real-time*) and low-complexity content popularity detection schemes. In this thesis, we study aspects of low-complexity detection of changes in video content popularity in real-time, addressed as a statistical change point (CP) detection problem [9], breaking completely new ground compared to earlier works that relied upon prediction models [10], [11].

Furthermore, novel exciting use cases were introduced in 5G in the context of ultra-reliable low latency communications (URLLC) and massive machine type communications (mMTC); the new industrial revolution, dubbed as Industry 4.0, along with emerging verticals in telemedicine, smart agriculture, etc., will bring about automation and intelligence to levels never seen before.

As 5G is required to support a large variety of services, novel solutions to enable higher resource efficiency are sought; amongst the various possible solutions, in this thesis we study non-orthogonal multiple access (NOMA) because of its advantages over conventional orthogonal multiple access (OMA) schemes in terms of spectral efficiency [12], cell-edge throughput [13], and energy efficiency [14], rendering it an attractive solution in particular for the mMTC uplink scenario.

Additionally, to account for medium access (MAC) sub-layer latency, we use the theory of the *effective capacity* [15], which can serve in wireless networks to provide *statistical* delay guarantees. The pertinence of the theory of the effective capacity as a suitable metric results from the fact that in the wireless MAC, due to small scale fading and shadowing, it is *inherently* impossible to provide hard delay guarantees.

In the following, a brief presentation of my principal past contributions over the last 7 years is given in reverse chronological order, to emphasize more recent results. Section 1.4.2 offers an outline of recent results in the area of resource allocation using NOMA, the theory of the effective capacity and CP analysis, while results in the area of PLS are described in Section 1.4.3.

1.4.2 Results in Resource Allocation

NOMA and Effective Capacity

Related Contributions: [J22], [J19], [C38], [C34]

In our works a flexible delay quality of service (QoS) model was employed using the theory of large deviations (Gärtner-Ellis theorem [16]) that allows defining the metric of the effective capacity (EC) on block fading additive white Gaussian noise (BF-AWGN) channels. The EC denotes the maximum constant arrival rate that can be served by a given service process, while guaranteeing a required statistical delay provisioning and is closely related to the concept of the effective bandwidth [17]. In order to capture the impact of link layer (MAC) delays in the secrecy capacity of wireless BF-AWGN channels, we introduced a novel metric, referred to as the "effective secrecy rate" (ESR); the ESR represents the maximum constant arrival rate that can be *securely* served (with perfect secrecy), on the condition that the required delay constraint can be statistically satisfied.

In more detail, in [J19] a novel approach was introduced to study the achievable delay-guaranteed secrecy rate, focusing on the downlink of a NOMA network with one base station, multiple single-antenna NOMA users and an eavesdropper. Two possible eavesdropping scenarios were considered; an internal, unknown, eavesdropper in a purely antagonistic network and an external eavesdropper in a network with trustworthy peers. For a purely antagonistic network with an internal eavesdropper, the only receiver with a guaranteed positive ESR was proved to be the one with the highest channel gain. The ESR in the high signal to noise ratio (SNR) regime was shown to approach a constant value irrespective of the power coefficients, while the strongest user was shown to achieve a higher ESR when it had a distinctive advantage in terms of channel gain with respect to the second strongest user. For a trustworthy NOMA network with an external eavesdropper, a lower bound and an upper bound on the ESR were proposed and investigated for an arbitrary legitimate user. For the lower bound, a closed-form expression was derived in the high SNR regime. For the upper bound, the analysis showed that if the external eavesdropper could not attain any channel state information (CSI), the legitimate NOMA user at high SNRs would always achieve positive ESR. Simulation results numerically validated the accuracy of the derived closed-form expressions and verified the analytical results given in the theorems and lemmas.

Furthermore, in [J22], [C38], [C34], we turned our attention to NOMA uplink networks. We provided performance analyses in asymptotic regimes (low and high SNR) and also proposed a novel multiple access (MA) scheme referred to as NOMA-Relevant (NOMA-R). In NOMA-R, a flexible MA

scheme is proposed based on the requirement that any user will opt for NOMA only when there is a rate gain associated. We have shown that NOMA-R outperforms both NOMA and OMA in terms of sum rates achievable in all SNR regions. Importantly, using the theory of the effective capacity we demonstrated that the NOMA-R strategy is more favorable when the target delay-bound violation probabilities are more stringent, especially for weak NOMA users.

Resource Allocation Using Change Point Analysis

Related Contributions: [J17], [J20], [C32]

In [J17], [J20] and [C32] we developed novel algorithms for the real-time detection of changes in the mean and the variance of content popularity. Approaching the problem statistically, we efficiently combined off-line and on-line non-parametric CUSUM procedures. The use of non-parametric CUSUM allowed us to avoid making assumptions about the underlying statistics of the popularity of any particular content, with the additional benefit of reduced computational cost. For the detection of changes in the mean we divided the algorithm in two phases. The first phase was an extended retrospective (off-line) procedure with an improved binary segmentation step and was used to adjust on-line parameters, based on historical data of the particular video. The second phase integrated a modified trend indicator to the sequential (on-line) procedure, to reveal the direction of a detected change. We provided extensive simulations, using real data, that demonstrated the performance of the first phase of our algorithm. We also provided proof-of-concept results that highlighted the efficiency of the overall algorithm.

The approach of combining off-line and on-line CP algorithms was also employed in [J20] for the detection of changes in the variance. However, a major difference concerned the choice of the underlying test statistic, as unlike in the case of the mean, tracking changes in the variance is inherently a nonlinear estimation problem. To develop the test statistic we proposed three different approaches: i) a non-parametric approach, ii) a parametric approach using an autoregressive moving average (ARMA) model, and, iii) a parametric approach using a nonlinear generalised autoregressive conditional heteroskedasticity (GARCH) model. Our studies using synthetic data indicated that the ARMA parametric approach did not generalize well. Due to this fact, we only performed experiments on real data using the non-parametric and the GARCH approaches. We concluded that both can equally well identify large deviations in the variance and that in the general case the non-parametric approach can provide quicker detection of CPs in the datasets studied in this work. In the future, we will develop joint detectors for the mean and the variance of video content popularity.

1.4.3 Results in PLS

The Role of PLS in 6G Security

Related contributions: [J21], [BC3], [C37], [C33], [C27]

With the emergence of 5G low latency applications, such as haptics and V2X, low complexity and low latency security mechanisms are needed. Promising lightweight mechanisms include physical unclonable functions (PUF) and secret key generation (SKG) at the physical layer from wireless fading coefficients, as considered in [J21], [C37], [C33]. In this framework we proposed a zero-round-trip-time (0-RTT) authentication protocol combining PUF for fast authentication and generation of resumption keys using SKG. Furthermore, a novel authenticated encryption (AE) scheme using SKG and standard symmetric key block ciphers for encryption and message authentication – first proposed in [C27] – was enhanced in [J21]. Aiming at a fast PHY protocol we proposed the pipelining of the AE SKG process and the encrypted data transfer at PHY in order to reduce latency. Looking at various alternatives to implement the pipelining at PHY, we investigated a "parallel" SKG approach for multi-carrier systems (e.g., using orthogonal frequency division multiplexing (OFDM) as in LTE and 5G new radio). In the parallel approach a subset of the subcarriers was used for SKG and the rest for encrypted data transmission (using the keys generated on the subset of SKG subcarriers). The optimal solution

to the respective PHY resource allocation problem was identified under security, power and delay constraints, by formulating the subcarrier scheduling as a subset-sum 0-1 knapsack optimization [18] . A heuristic algorithm of linear complexity was proposed and shown to incur negligible loss with respect to the optimal dynamic programming solution [J21], [C37], [C33]. The proposed mechanisms, have the potential to pave the way for a new breed of latency aware PHY security protocols with an emphasis on URLLC and IoT emerging systems.

Finally, the main lines of application of PLS in 6G systems were reviewed in [BC3], starting with node authentication, moving to the information theoretic characterization of message integrity, and finally, discussing message confidentiality both in the SKG and from the wiretap channel point of view. The aim of this review was to provide a comprehensive roadmap on important relevant results by the authors and other contributors and discuss open issues on the applicability of PLS in 6G systems.

Anomaly Detection in Software Defined Networks

Related contributions: [C39], [J18]

Software-defined networking (SDN) is a promising technology to overcome many challenges in wireless sensor networks (WSN), particularly with respect to flexibility and reuse. Conversely, the centralization and the planes' separation turn SDNs vulnerable to new security threats in the general context of distributed denial of service (DDoS) attacks. State of-the-art approaches to identify DDoS do not always take into consideration restrictions in typical WSNs e.g., computational complexity and power constraints, while further performance improvement is always a target. The objective of the works in [J18], [C39] was to propose a lightweight but very efficient DDoS attack detection approach using CP analysis. Our approach was shown to have a high detection rate, while its complexity grows linearly with the observed time series length, rendering it suitable for WSNs. We demonstrated the performance of our detector in software-defined WSNs of 36 and 100 nodes with varying attack intensity (the number of attackers ranging from 5% to 20% of nodes).

We used CP detectors to monitor anomalies in two metrics: the data packets delivery rate and the control packets overhead. Our results showed that as the intensity of the attack increased, our approach could achieve a detection rate close to 100% and that, importantly, the type of the attack could also be inferred. As an extension of this work, we will look into distributed anomaly detection by allowing clusters of nodes to act on local early detection systems. A trade-off to be studied will concern the cluster size versus the speed of the detection while maintaining the ability to localize the source of the anomaly.

Shielding PLS Against Active Attacks

Related contributions: [BC2], [J16], [J15], [J12], [C36], [C31], [C30], [C29], [C28], [C24], [C22], [C21]

SKG schemes have been shown to be vulnerable to DoS attacks in the form of jamming and to man in the middle attacks implemented as injection attacks. In [BC2] and [C36], a comprehensive study on the impact of correlated and uncorrelated jamming and injection attacks in wireless SKG systems was presented. First, two optimal signaling schemes for the legitimate users were proposed and the impact of injection attacks as well as counter-measures were investigated. Finally, it was demonstrated that the jammer should inject either correlated jamming when imperfect channel state information (CSI) regarding the main channel was at their disposal, or, uncorrelated jamming when the main channel CSI was completely unknown.

As jamming attacks represent a critical vulnerability for wireless SKG systems, in [J15], [C31], [C30], [C29], [C28] two counter-jamming approaches were investigated for SKG systems: first, the employment of energy harvesting (EH) at the legitimate nodes to turn part of the jamming power into useful communication power, and, second, the use of channel hopping or power spreading in BF-AWGN channels to reduce the impact of jamming.[1] In both cases, the adversarial interaction between the pair

[1]We note in passing that spreading / hopping can be directly implemented with a standard inverse fast Fourier

of legitimate nodes and the jammer was formulated as a two-player zero-sum game and the Nash and Stackelberg equilibria (NE and SE) were characterized analytically and in closed form. In particular, in the case of EH receivers, the existence of a critical transmission power for the legitimate nodes allowed the full characterization of the game's equilibria and also enabled the complete neutralization of the jammer. In the case of channel hopping vs. power spreading techniques, it was shown that the jammer's optimal strategy was always power spreading while the legitimate nodes should only use power spreading in the high signal-to-interference ratio (SIR) regime. In the low SIR regime, when avoiding the jammer's interference becomes critical, channel hopping is optimal for the legitimate nodes. Numerical results demonstrated the efficiency of both counter-jamming measures.

Furthermore, in [J16] the novel proposal of using EH as a counter-jamming measure for point-to-point communication was investigated on the premise that part of the harmful interference could be harvested to increase the transmit power. We formulated the strategic interaction between a pair of legitimate nodes and a malicious jammer as a zero-sum game. Our analysis demonstrated that the legitimate nodes were able to neutralize the jammer. However, this policy was not necessarily a Nash equilibrium and hence was sub-optimal. Instead, harvesting the jamming interference could lead to relative gains of up to 95%, on average, in terms of Shannon capacity, when the jamming interference was high.

Finally, in our earlier works [J12], [C24], [C22], [C21] the resilience of wireless multiuser networks to passive (interception of the broadcast channel) and active (interception of the broadcast channel and false feedback) eavesdroppers was investigated. Stochastic characterizations of the secrecy capacity (SC) were obtained in scenarios involving a single transmitter (base station) and multiple destinations. The expected values and variances of the SC along with the probabilities of secrecy outage were evaluated in the following cases: (i) in the presence of passive eavesdroppers without any side information; (ii) in the presence of passive eavesdroppers with side information about the number of eavesdroppers; and (iii) in the presence of a single active eavesdropper with side information about the behavior of the eavesdropper. This investigation demonstrated that substantial secrecy rates are attainable on average in the presence of passive eavesdroppers as long as minimal side information is available. On the other hand, it was further found that active eavesdroppers could potentially compromise such networks unless statistical inference was employed to restrict their ability to attack. Interestingly, in the high SNR regime, multiuser networks were shown to become insensitive to the activeness or passiveness of the attack.

PLS Encoders and Secrecy Enhancement in Collaborative Networks

Related contributions: [J14], [J13], [C23], [C19], [C18]

Physical layer network coding (PNC) has been proposed for future generations of wireless networks. In [J14], we investigated PNC schemes with embedded perfect secrecy by exploiting structured interference in relay networks with two users and a single relay. In a practical scenario where both users employed finite and uniform signal input distributions, we established upper bounds (UB) on the achievable perfect secrecy rates and made these explicit when pulse amplitude modulation (PAM) modems were used, while our results extend straightforwardly to quadrature amplitude modulation (QAM) modems. We then described two simple, explicit encoders that could achieve perfect secrecy rates close to these UBs with respect to an untrustworthy relay in the single antenna and single relay setting. Lastly, we generalized our system to a MIMO relay channel where the relay had more antennas than the users and studied optimal precoding matrices which satisfied a required secrecy constraint. Our results established that the design of PNC transmission schemes with enhanced throughput and guaranteed data confidentiality was feasible.

Finally, in [J13] the optimal power allocation that maximizes the SC of BF-AWGN networks with causal CSI, M-block delay tolerance and a frame based power constraint was examined. In particular,

transform (IFFT) transmitter employed in OFDM systems; spreading requires no change in the canonical OFDM transmitter, while hopping requires setting all but one of the IFFT inputs to zero.

the SC maximization was formulated as a dynamic program. First, the SC maximization without any information on the CSI was studied; in this case the SC was shown to be maximized by equidistribution of the power budget, denoted as the " blind policy ". Next, extending earlier results on the capacity maximization of BF-AWGN channels without secrecy constraints, transmission policies for the low SNR and the high SNR regimes were proposed. When the available power resources were very low the optimal strategy was a "threshold policy". On the other hand, when the available power budget was very large a "constant power policy" was shown to maximize the frame secrecy capacity. Subsequently, a novel universal transmission policy was introduced, denoted as the "blind horizon approximation" (BHA), by imposing a blind policy in the horizon of unknown events. Through numerical results, the novel BHA policy was shown to outperform both the threshold and constant power policies as long as the mean channel gain of the legitimate user was distinctively greater than the mean channel gain of the eavesdropper. Furthermore, the secrecy rates achieved by the BHA compared well with the secrecy rates of the secure waterfilling policy in the case of acausal CSI feedback to the transmitter.

1.5 Recent Teaching Activities

I have had the opportunity to teach for over 7 consecutive years in France and the UK, a variety of courses from cryptography and network security, to networking and wireless communications. A detailed description of my teaching record is presented in reverse chronological order in the following Sections.

1.5.1 Overview of Teaching Activities in France (ENSEA)

Since September 2017 I have been engaged with teaching at ENSEA, giving courses both in French and in English. I have taught in the second and third year of the engineering track of ENSEA, as well as in the continuing education track (cycle par alternance). Furthermore, since September 2019 I am responsible of the students' international mobility towards the UK.

Engineering Track (teaching in English)

I have been teaching in the third year specialization "Networks and Telecommunications" the modules of "Network Security" (module responsible), "Internetworking" (module responsible) and "Wireless Communications". In parallel I am teaching Cryptography in the M2 MSc of ETIS SIT (Systèmes, Information, Télécommunications), whose syllabus mirrors in great extent that of "Network Security". A brief presentation of the courses is given below:
(1) Network Security / Cryptography: 10 hours of lectures. Topics covered include:

- Data Confidentiality: perfect secrecy, semantic security, block ciphers, DES, 3DES, AES;
- Data Integrity: message authentication codes (MAC), authenticated encryption;
- Key management using a trusted third party;
- Public key encryption, Diffie Hellman, El Gamal, RSA;
- Digital signatures, digital certificates, public key infrastructure;
- SSL / TLS.

(2) Internetworking: 16 hours of lectures, 24 hours of lab work (on GNS3), 6 hours of seminars (classes)

- IP protocol, DHCP, ARP, ICMP, NAT;

- Routing protocols: RIP, OSPF, BGP, Mobile IP, Dynamic Source Routing, Reverse Path Forwarding, Multicasting;
- Quality of Service: Integrated Services, Differentiated Services, MPLS;
- Congestion control, TCP Tahoe, TCP Reno, TCP Vegas, Fast-TCP.

(3) Wireless Communications: 6 hours of lectures, 4 hours of lab work, 4 hours of seminars (classes)

- Signal space, maximum a posteriori detection, maximum likelihood detection;
- Design of communication system, power / bandwidth limited systems, digital modulations;
- Narrowband fading channel models and channel capacity;
- Waterfilling algorithm, adaptive QAM.

Furthermore, immediately after my recruitment at ENSEA I was tasked with developing the second year option on "Internet of things" (IoT Option: 36 hours of lectures, 28 hours of lab work in total). I have engaged with the FIT IoT-lab of INRIA in Saclay and secured three related lab sessions with remote access to the FIT-IoT lab. In the IoT option, typically 2 instructors from the industry (Nokia, Huawei or Orange) give a number of lectures on topics related to low power wide area networks (LPWAN), 3GPP standards (NB-IoT, MTC), vehicular IoT, wireless sensor networks. In the IoT option I give 6 hours of lectures on
(4) IoT security and 4 hours of lab work, covering the following topics:

- Background concepts,introduction to DTLS and IPSec;
- Introduction to blockchains for IoT;
- RFID authentication;
- Jamming attacks (primarily through lab work).

Student satisfaction in the IoT Option has been strong with an average 4/5 in the first year, bringing it amongst the best ranked second year options with a consistently high demand.

Continuous Education Track (teaching in French)

Additionally, I am teaching at the final year of the "cycle par alternace" of ENSEA in the specialization "Réseaux et Télécommunications" the module "Interconnexion et Administration des Réseaux ", mirroring a reduced syllabus of the topics covered in the engineering track module "Internetworking". The module consists of 10 hours of lectures, 16 hours of lab work and 10 hours of seminars (classes).

1.5.2 Overview of Teaching Activities in the UK

Since July 2015 I have been a Fellow of the Higher Education Academy (FHEA) of the UK. FHEA is a professional title in higher education that is recognized (and currently required) by academic institutions in the United Kingdom. To become FHEA, I followed the courses offered at the University of Essex as part of the CADENZA program, between October 2014 and March 2015. Then, I prepared my teaching portfolio which included: (i) factual aspects concerning my teaching experience, (ii) in-depth familiarization with recognized teaching theories, showcased in a pedagogical thesis reflecting my teaching experience. The evaluation of my portfolio by the HEA took place in May 2015 and I obtained the title of a FHEA the following July. In addition, student satisfaction in the courses I have taught has been particularly strong. Notably, in the Student Evaluation of Teaching (SET) for the year 2014-2015 I scored a perfect 5/5 for my teaching of the course CE702 Digital Communications at the University of Essex.

University of Essex

I served as a Lecturer at the University of Essex, School of Computer Science and Electronic Engineering between 2013-2017. From 2014 to 2017 I was responsible for the module **"CE702 - Digital Communications"** of the MSc in Advanced Communication Systems. The 12-week module included weekly lectures and seminars (classes). Throughout the semester, 2 different assignments were given. Topics covered include:

- Systems and signals, channel coding, modulation, OFDM, MIMO;
- Multiple access methods: TDD, FDD, TDMA, FDMA, CDMA, OFDMA;
- Wireless multipath channels, equalization;
- Antennas, satellite communication;
- Optical networks, wavelength division multiplexing (WDM), dense WDM (DWDM).

In January 2015 I became the responsible of the module **"CE823 – Network Security and Cryptographic Principles"** of the MSc in Computer Networks and Security and of the optional third-year BSc module CE324 (with the same description). The 12 weeks long module included weekly lectures and lab sessions. Throughout the semester, 2 different assignments were given. The course syllabus is described below:

- Data Confidentiality: perfect secrecy, semantic security, stream ciphers, block ciphers, DES, 3DES, AES;
- Data Integrity: message authentication codes (MAC), authenticated encryption;
- Key management using a trusted third party, Kerberos protocol;
- Public key encryption, Diffie Hellman, El Gamal, RSA;
- Digital signatures, digital certificates, public key infrastructure;
- SSL / TLS, HTTPS, SSH, IPSec, DNSSec;
- Denial of service (DoS), intrusion detection, firewalls;
- Security of wireless networks, WEP, WPA, WPA2.

After my maternity leave (Sep. 2015-Jun. 2016) I became responsible for the reorganization of the module **"CE740 Mobile Communications"** of the MSc Computer Networks and Security and of the MSc in Electronic Engineering. The 12-weeks long module consisted of weekly lectures. Throughout the semester, 3 different assignments are given. Topics covered included:

- Routing: routing for static networks (Dijkstra algorithm), dynamic source routing for ad-hoc networks, clustering, data aggregation;
- MAC: Static access methods (TDMA, FDMA, CDMA, OFDMA), random access (Aloha, Slotted Aloha, CSMA, CSMA / CD, CSMA / CA), MACA protocol, scatternets, piconets, master-slave protocols, management power management / wake-up patterns, infrastructure networks and ad-hoc networks, 802.11;
- Physical layer: wireless channel, capacity, waterfilling, diversity, modulation, OFDM, Direct Sequence Spread Spectrum, Frequency Hopping Spread Spectrum, MIMO, 5G.

In addition, while at the University of Essex I supervised 3 MSc students in their projects, one of which obtained an MSc by research dissertation (year-long research project, related conference paper [C27]). Finally, I supervised 8 BSc projects, one of which was awarded the best departmental project award on "Securing DNS on Android".

Middlesex University

Between January 2009 and April 2011, I held the position of Senior Lecturer at Middlesex University, Department of Computer Communications. I was in charge of the course **"CCM4820 - Digital Transmission Techniques"** of the MSc in Telecommunications Engineering. In this context, I designed and developed my own course on digital communications. Teaching at the master's level for the first time was a great experience for me. I was able to explore all aspects of the management of teaching a course: the creation of the syllabus, the choice of the textbook, the employment and supervision of teaching assistants, the development of tutorials, additional exams and materials, preparation and presentation of courses. The course lasted 12 weeks during one semester, and included 2 hour lectures per week, weakly seminars and weakly lab sessions. Throughout the semester, 3 different assignments were given. The typical class size ranged from 25 to 80 students, and gradually increased over the years. Topics covered included:

- Stochastic signals, systems and processes, spectrum;
- Source coding: entropy, Huffman encoders, Lempel-Ziv encoders;
- Channel coding: block and convolution encoders, introduction to Turbo encoders;
- Digital modulation, OFDM systems;
- Multiple access techniques, TDMA, FDMA, CDMA, OFDMA;
- Introduction to MIMO systems, multi-path wireless channels, equalization;
- Introduction to optical systems.

In addition, I have supervised more than 25 master students in their projects, in a variety of subjects and areas of research, including physical layer security, network security, detection of anomalies in networks.

1.6 Research Supervision

My supervision activities at the PhD level include 2 students that have scheduled thesis defences for September 2020 and two further that are ongoing. In detail:

1.6.1 PhD theses to be defended in September 2020:

PhD Student Mr. Miroslav Mitev
Supervision @60% for the period 25/4/2017-9/2020
Thesis title: "Physical layer security for the Internet of things".
Student co-supervised with Dr. M. Reed, Senior Lecturer at University of Essex, UK.
Thesis VIVA (defence) scheduled for September 2020.
Publications from thesis: [J21], [C37], [C36], [C33], [P1], [U1].
M. Mitev is registered at the Ecole Doctorale of CY University and was the thesis director between April-August 2017 before joining ENSEA.

PhD student Mr. Sotiris Skaperas
Supervision @40% for the period 1/9/2017-9/2020
Thesis title: "Data analysis and forecasting models for flexible resource management in 5th generation networks".
Co-supervised with Dr. L. Mamatas, Assistant professor at the University of Macedonia, GR.
Thesis defence scheduled for September 2020.
Publications from thesis: [J20], [J17], [C32], [U5].

1.6.2 Ongoing theses

PhD student Mr. Gustavo Alonso Nunez Segura
Supervision @35% started on 1/2/2019.
Thesis title: "Cooperative Intrusion Detection System for Software Defined Wireless Sensor Networks".
Co-supervised with Dr. Cintia Borges Margi, Associate Professor at the University of Sao Paolo, BR.
Publications from thesis: [J18], [C39], [C35], [S1], [U2].

PhD student Mr. Mouktar Bello
Supervision @70% started on 1/11/2020.
Title: "Meeting delay and security constraints in 6G wireless networks".
Co-supervised with Prof. I. Fijalkow, ETIS/ENSEA, FR.
Publications from thesis: [C38], [U3, U4].

1.6.3 Current Postdoctoral Students

- Postdoc Dr. Mahdi Shakiba Herfeh: supervision @100%, 21/11/2019-20/5/2021 (fixed term 1.5 years), "Physical layer security for IoT applications", project ELIOT ANR PRCI, ETIS/ENSEA FR, publications: [BC3], [U1].
- Postdoc Dr. Nasim Ferdosian: supervision @90%, 1/1/2020-31/12/2021 (fixed term 2 years), "Non-orthogonal multiple access techniques under security and delay constraints", with Prof. I. Fijalkow, ETIS/ENSEA, FR, publications: [U5].

1.7 Structure of the Rest of the Thesis

This thesis is structured around my most recent publications (dating within the last two years) with the PhD students I supervise.

In Chapter 2, novel authentication protocols using PUFs and SKG proposed by Miroslav Mitev, myself and Martin Reed are presented. This Chapter focuses on works presented in [J21] and [C33] and [C37] and includec contributions by Dr. L. Musavian.

Next, in Chapter 3 a novel, real-time and non-parametric detector for changes in the mean value of content popularity is discussed, reflecting [J17] and [C32] with Sotiris Skaperas and Lefteris Mamatas. Additionally, the application of the same detector for intrusion detection in a software defined network is demonstrated, showcasing part of our contributions with Gustavo Nunez and Cintia Borges Magri in [J18] and [C35].

Chapter 4 includes some of our early results with Mouktar Bello, Wenjuan Wu and Leila Musavian on the performance analysis of NOMA uplink networks under statistical delay constraints, published in [C38].

Finally, my perspectives for future research in 6G technologies are presented in Chapter 5.

References

[1] A. Chorti, K. Papadaki, and H. V. Poor. Optimal power allocation in block fading channels with confidential messages. *IEEE Trans. Wireless Commun.*, 14(9):4708–4719, Sep. 2015.

[2] A. Chorti, S. M. Perlaza, Z. Han, and H. V. Poor. On the resilience of wireless multiuser networks to passive and active eavesdroppers. *IEEE J. Sel. Areas Commun.*, 31(9):1850–1863, Sep. 2013.

[3] Arsenia Chorti, Camilla Hollanti, Jean-Claude Belfiore, and H. Vincent Poor. Physical layer security: A paradigm shift in data confidentiality. *Lecture Notes in Electrical Engineering*, 358, 01 2016.

[4] G. Rezgui, E.V. Belmega, and A. Chorti. Mitigating jamming attacks using energy harvesting. *IEEE Wireless Commun. Lett.*, 8:297–300, 2019.

[5] Third Generation Partnership Project (3GPP). TR 33.825 Study on the Security of URLLC, 2019.

[6] F. Arute, K. Babbush, and *et al.* Quantum supremacy using a programmable superconducting processor. *Nature*, 574:505–510, 2019.

[7] G. E. Suh and S. Devadas. Physical unclonable functions for device authentication and secret key generation. In *2007 44th ACM/IEEE Design Automation Conference*, pages 9–14, June 2007.

[8] E. V. Belmega and A. Chorti. Protecting secret key generation systems against jamming: Energy harvesting and channel hopping approaches. *IEEE Trans. Inf. Forensics Security*, 12(11):2611–2626, Nov 2017.

[9] Michèle Basseville, Igor V Nikiforov, et al. *Detection of abrupt changes: theory and application*, volume 104. Prentice Hall Englewood Cliffs, 1993.

[10] Alexandru Tatar, Marcelo Dias De Amorim, Serge Fdida, and Panayotis Antoniadis. A survey on predicting the popularity of web content. *J. Internet Services Appl.*, 5(1):8, Dec. 2014.

[11] Gabor Szabo and Bernardo A Huberman. Predicting the popularity of online content. *Commun. ACM*, 53(8):80–88, Aug. 2010.

[12] Z. Ding, Y. Liu, J. Choi, Q. Sun, M. Elkashlan, C. I, and H. V. Poor. Application of non-orthogonal multiple access in LTE and 5G networks. *IEEE Commun. Mag.*, 55(2):185–191, Feb. 2017.

[13] Z. Ding, M. Peng, and H. V. Poor. Cooperative non-orthogonal multiple access in 5G systems. *IEEE Commun. Lett.*, 19(8):1462–1465, Aug. 2015.

[14] F. Fang, H. Zhang, J. Cheng, and V. C. M. Leung. Energy efficiency of resource scheduling for non-orthogonal multiple access wireless network. In *Proc. IEEE Int. Conf. Commun. (ICC)*, Kuala Lumpur, Malaysia, May 2016.

[15] D. Wu and R. Negi. Effective capacity: a wireless link model for support of quality of service. *IEEE Trans. Wireless Commun.*, 2(4):630–643, 2003.

[16] Po-Ning Chen. Generalization of Gärtner-Ellis theorem. *IEEE Trans. Inf. Theory*, 46:2752–2760.

[17] D. N. C. Tse and S. V. Hanly. Linear multiuser receivers: effective interference, effective bandwidth and user capacity. *IEEE Trans. Inf. Theory*, 45:641–657.

[18] D.P. Williamson and D.B. Shmoys. *Approximation Algorithms.* Cambridge University Press, 2011.

Chapter 2

Security Protocols for Internet of Things Applications

2.1 Introduction

With the emergence of 5G low latency applications, such as haptics and V2X, low complexity and low latency security mechanisms are needed. Promising lightweight mechanisms include physical unclonable functions (PUF) and secret key generation (SKG) at the physical layer, as considered in this Chapter. In this framework we propose i) a zero-round-trip-time (0-RTT) resumption authentication protocol combining PUF and SKG processes; ii) a novel authenticated encryption (AE) using SKG; iii) pipelining of the AE SKG and the encrypted data transfer in order to reduce latency. Implementing the pipelining at PHY, we investigate a *parallel* SKG approach for multi-carrier systems, where a subset of the subcarriers are used for SKG and the rest for data transmission. The optimal solution to this PHY resource allocation problem is identified under security, power and delay constraints, by formulating the subcarrier scheduling as a subset-sum $0-1$ knapsack optimization. A heuristic algorithm of linear complexity is proposed and shown to incur negligible loss with respect to the optimal dynamic programming solution. All of the proposed mechanisms, have the potential to pave the way for a new breed of latency aware security protocols.

2.2 Contributions and Chapter Organization

Many standard cryptographic schemes, particularly those in the realm of public key encryption (PKE), are computationally intensive, incurring considerable overheads and can rapidly drain the battery of power constrained devices [1], [2], notably in Internet of things (IoT) applications [3]. For example, a 3GPP report on the security of ultra reliable low latency communication (URLLC) systems notes that authentication for URLLC is still an open problem [4]. Additionally, traditional public key generation schemes are not *quantum secure* – in that when sufficiently capable quantum computers will be available they will be able to break current known PKE schemes – unless the key sizes increase to impractical lengths.

In the past years, physical layer security (PLS) [5–9] has been studied as a possible alternative to classic, complexity based, cryptography. As an example, signal properties as in [10], can be exploited to generate opportunities for confidential data transmission [11, 12]. Notably, PLS is explicitly mentioned as a 6G enabling technology in the first white paper on 6G [13]: "The strongest security protection may be achieved at the physical layer." In this work, we propose to move some of the security core functions down to the physical layer, exploiting both the communication radio channel and the hardware, as unique entropy sources.

Since the wireless channel is reciprocal, time-variant and random in nature, it offers a valid, inherently secure source that may be used in a key agreement protocol between two communicating

parties. The principle of secret key generation (SKG) from correlated observations was first studied in [14] and [15]. A straightforward SKG approach can be built by exploiting the reciprocity of the wireless fading coefficients between two terminals within the channel coherence time [16] and the contributions in this Chapter build upon this mechanism. This is pertinent to many forthcoming B5G applications that will require a strong, but nevertheless, lightweight security key agreement; in this direction, PLS may offer such a solution, or, complement existing algorithms. With respect to authentication, physical unclonable functions (PUFs), firstly introduced in [17] (based on the idea of physical one-way functions [18], [19]), could also enhance authentication and key agreement in demanding scenarios, including (but not limited to) device to device (D2D) and tactile Internet. We note that others also point to using physical layer security to reduce the resource overhead in URLLC [20].

A further advantage of PLS is that it is information-theoretic secure [21], *i.e.*, it is not open to attack by future quantum computers, and, it requires lower computation costs. In this work, we will discuss how SKG from shared randomness [22] is a promising alternative to PKE for key agreement. However, unauthenticated key generation is vulnerable to man in the middle (MiM) attacks. In this sense, PUFs, can be used in *conjunction* with SKG to provide authenticated secret key agreement. As summarised in [19], the employment of PUFs can decrease the computational cost and play a pivotal role in reducing the authentication latency in constrained devices.

In this study we introduce the joint use of PUF authentication and SKG in a zero-round-trip-time (0-RTT) [23, 24] approach, allowing to build quick authentication mechanisms with forward security. Further, we develop an authenticated encryption (AE) primitive [25–27] based on standard SKG schemes. To investigate a fast implementation of the AE SKG we propose a pipelined (*parallel*) scheduling method for optimal resource allocation at the physical layer (PHY) (*i.e.*, by optimal allocation of the subcarriers in 5G resource blocks).

Next, we extend the analysis to account for statistical delay quality of service (QoS) guarantees, a pertinent scenario in B5G. The support of different QoS guarantee levels is a challenging task. In fact, in time-varying channels, such as in wireless networks, determining the exact delay-bound depending on the users' requirements, is impossible. However, a practical approach, namely the effective capacity [28], can provide statistical QoS guarantees, and, can give delay-bounds with a small violation probability. In our work, we employ the effective capacity as the metric of interest and investigate how the proposed pipelined AE SKG scheme performs in a delay-constrained scenario.

The system model introduced in this work assumes that a block fading additive white Gaussian noise (BF-AWGN) channel is used with multiple orthogonal subcarriers. In our *parallel* scheme a subset of the subcarriers is used for SKG and the rest for encrypted data transfer. The findings of this study are supported by numerical results, and the efficiency of the proposed *parallel* scheme is shown to be greater or similar to the efficiency of an alternative approach in which SKG and encrypted data transfer are sequentially performed.

To summarize, the contributions of this Chapter are as follows:

1. We combine PUF authentication and SKG for resumption key agreement in a single 0-RTT protocol.

2. We develop an AE SKG scheme.

3. We propose a fast implementation of the AE SKG based on pipelining of key generation and encrypted data transfer. This *parallel* approach is achieved by allocation of the PHY resources, *i.e.*, by optimal scheduling of the subcarriers in BF-AWGN channels.

4. We propose a heuristic algorithm of linear complexity that finds the optimal subcarrier allocation with negligible loss in terms of efficiency.

5. We numerically compare the efficiency of our *parallel* approach with a *sequential* approach where SKG and data transfer are performed sequentially. This comparison is performed in two delay

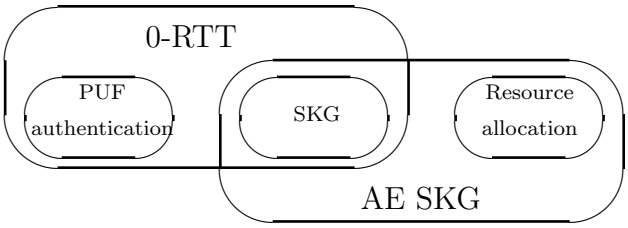

Figure 2.1: Roadmap of contributions.

scenarios:

- When a relaxed QoS delay constraint is in place;
- When a stringent QoS delay constraint is in place.

A roadmap of the Chapter's contributions is shown in Fig. 2.1.

2.2.1 Threat Model

In this work we assume a commonly used adversarial model with an active man-in-the-middle attacker (Eve) and a pair of legitimate users (Alice and Bob). For simplicity, we assume a rich Rayleigh multipath environment where the adversary is more than a few wavelengths away from each of the legitimate parties. This forms the basis of our hypothesis that the measurements of Alice and Bob are uncorrelated to the Eve's measurements.

2.2.2 Notation

Random variables are denoted in italic font, e.g., x, vectors and matrices are denoted with lower and upper case bold characters, e.g., $\mathbf{x}$ and $\mathbf{X}$, respectively. Functions are printed in a fixed-width teletype font, e.g., `F`. All sets of vectors are given with calligraphic font $\mathcal{X}$ and the elements within a set are given in curly brackets e.g. $\{\mathbf{x}, \mathbf{y}\}$, the cardinality of a vector or set is defined by vertical lines e.g., $|\mathbf{x}|$ or $|\mathcal{X}|$. Concatenation and bit-wise XORing are represented as $[\mathbf{x}||\mathbf{y}]$ and $\mathbf{x} \oplus \mathbf{y}$, respectively. We use H to denote entropy, I mutual information, $\mathbb{E}$ expectation and $\mathbb{C}$ the set of complex numbers.

2.2.3 Chapter Organization

The rest of the Chapter is organized as follows: related work is discussed in Section 2.3 followed by the general system model introduced in Section 2.4. The use of PUF authentication is illustrated in Section 2.4.1, the baseline SKG in Section 2.4.2; next, in Sections 2.4.3 and 2.4.4 we present an AE scheme using SKG and a resumption scheme to build a 0-RTT protocol. Subsequently, we evaluate the optimal power and subcarrier allocation at PHY considering both the long term average rate in Section 2.5 and the effective rate in Section 2.6. In Section 2.7, the efficiency of the proposed approach is evaluated against that of a sequential approach, while conclusions are presented in Section 2.8.

2.3 Related Work

This work assumes the use of PUF-based authentication with SKG. PUFs are hardware entities based on the physically unclonable variations that occur during the production process of silicon. These unique and unpredictable variations allow the extraction of uniformly distributed binary sequences. Due to their unclonability and simplicity, PUFs are seen as lightweight security primitives that can offer alternatives to today's authentication mechanisms. Furthermore, employing PUFs can eliminate

the need of non-volatile memory, which reduces cost and complexity [29]. Common ways of extracting secret bit sequences are through measuring jitter on oscillators, delays on gates, or, observing the power up behavior of a silicon.

Numerous PUF architectures have been proposed for IoT applications in the literature. A few of these architectures are: arbiter PUF [30], ring oscillator PUF [17], transient effect ring oscillator PUF [31], static random-access memory PUF [32], hardware embedded delay PUF [33] and more [34]. Utilising these basic properties, many PUF-based authentication protocols have been proposed, both for unilateral authentication [35, 36] and mutual authentication [29, 36–38]. A comprehensive survey on lightweight PUF authentication schemes is presented by Delvaux *et al.* [39].

On the other hand, due to the nature of propagation in the shared, free-space medium, wireless communications remain vulnerable to different types of attacks. Passive attacks such as eavesdropping or traffic analysis can be performed by anyone in the vicinity of the communicating parties; to ensure confidentiality, data encryption is vital for communication security. The required keys can be agreed at PHY using SKG. In this case, all pilot exchanges need to take place over the coherence time of the channel[1], during which Alice and Bob can observe highly correlated channel states that can be used to generate a shared secret key between them. SKG has been implemented and studied for different applications such as vehicular communications [42, 43], underwater communications [44], optical fiber [45], visible light communication [46] and more as summarized in [47]. The key conclusion from these studies is that SKG shows promise as an important alternative to current key agreement schemes.

Widely used sources of shared randomness used for SKG are the received signal strength (RSS) and the full channel state information (CSI) [48]. In either case, it is important to build a suitable pre-processing unit to decorrelate the signals in the time / frequency and space domains. As an example, some recent works have shown that the widely adopted assumption [49] that a distance equal to *half* of the wavelength (which at 2.4 GHz is approximately 6 cm [50]) is enough for two channels to decorrelate, may not hold in reality [40]. Other works show that the mobility can highly increase the entropy of the generated key [51, 52] while an important issue with the RSS-based schemes is that they are open to predictable channel attacks [40, 53]. These important issues need to be explicitly accounted for in actual implementations, but fall outside the scope of the present Chapter. We note in passing that pilot randomization can be employed to overcome limitations related to channel predictability [54].

2.4 Node Authentication Using PUFs and SKG

In this Section we present a joint physical layer SKG and PUF authentication scheme. To the best of our' knowledge this is the first work that proposes the utilization of the two schemes in conjunction. As discussed in Section 2.3, many PUF authentication protocols have been proposed in the literature, with even a few commercially available [55, 56]. We do not look into developing a new PUF architecture or a new PUF authentication protocol, instead, we look at combining existing PUF mechanisms with SKG. In addition, we develop an AE scheme that can prevent tampering attacks. To further develop our hybrid crypto-system we propose a resumption type of authentication protocol, inspired by the 0-RTT authentication mode in the transport layer security (TLS) 1.3 protocol. The resumption protocol is important as it significantly reduces the use of the PUF to the initial authentication, thus, overcoming the limitation of a PUFs' challenge response space [34, 57].

[1]The coherence time corresponds to the interval during which the multipath properties of wireless channels (channel gains, signal phase, delay) remain stable [40–42]. It is inversely proportional to the Doppler spread, which on the other hand, is a dispersion metric that accounts for the spectral broadening caused by the user's mobility (for more details and derivation please see [41]).

2.4.1 Node Authentication Using PUFs

As discussed in Section 3.9.3, for security against MiM attacks, the SKG needs to be protected through authentication. While existing techniques, such as the extensible authentication protocol-transport layer security (EAP-TLS), could be used as the authentication mechanism, these are computationally intensive and can lead to significant latency [58, 59].

This leads to the motivation to seek lightweight authentication mechanisms that can be used in conjunction with SKG. Such a mechanism that is achieving note within the research community uses a PUF. A typical PUF-based authentication protocol consists of two main phases, namely *enrolment phase* and *authentication phase* [60–64]. During the *enrolment phase* each node runs a set of challenges on its PUF and characterizes the variance of the measurement noise in order to generate side information. Next, a verifier creates and stores a database of all challenge-response pairs (CRPs) for each node's PUF within its network. A CRP pair in essence consists of an authentication key and related side information. Within the database, each CRP is associated with the ID of the corresponding node.

Later, during the *authentication phase* a node sends its ID to the verifier requesting to start a communication. Receiving the request, the verifier checks if the received ID exists in its database. If it does, the verifier chooses a random challenge that corresponds to this ID and sends it to the node. The node computes the response by running the challenge on its PUF and sends it to the verifier. However, the PUF measurements at the node are never exactly the same due to measurement noise, therefore, the verifier uses the new PUF measurement and the side information stored during the enrollment to re-generate the authentication key. Finally, the verifier compares the re-generated key to the one in the CRP and if they are identical the authentication of the node is successful. A simple approach to prevent replay attacks consists in deleting a CRP from the verifier database once it is used, but more elaborate schemes can also be built.

In summary, the motivation for using a PUF authentication scheme in conjunction with SKG is to exclude all of the computationally intensive operations required by EAP-TLS, which use modulo arithmetic in large fields. Measurements performed on current public key operations within EAP-TLS on common devices (such as IoT) give average authentication and key generation times of approximately 160 ms in static environments and this can reach up to 336 ms in high mobility conditions [65].

On the other hand, PUF authentication protocols have very low computational overhead and require overall authentication times that can be less than 10 ms [61, 66]. Furthermore, our key generation scheme, proposed in Section 2.4.2, requires just a hashing operation and (syndrome) decoding. Hashing mechanisms such as SHA256 performed on an IoT device require less than 0.3ms [66, 67]. Regarding the decoding, if we assume the usage of standard LDPC or BCH error correcting mechanisms, even in the worst-case scenario with calculations carried out as software operations, the computation is trivial compared to the hashing and requires less computational overhead [68].

2.4.2 SKG procedure

The SKG system model is shown in Fig. 2.2. This assumes that two legitimate parties, Alice and Bob, wish to establish a symmetric secret key using the wireless fading coefficients as a source of shared randomness. Throughout our work a rich Rayleigh multipath environment is assumed, such that the fading coefficients rapidly decorrelate over short distances [16]. Furthermore, Alice and Bob communicate over a BF-AWGN channel that comprises N orthogonal subcarriers. The fading coefficients $\mathbf{h} = [h_1, \ldots, h_N]$, are assumed to be independent and identically distributed (i.i.d), complex circularly symmetric zero-mean Gaussian random variables $h_j \sim \mathcal{CN}(0, \sigma^2), j = 1, \ldots, N$. Although in actual multicarrier systems neighbouring subcarriers will typically experience correlated fading, in the present work this effect is neglected as its impact on SKG has been treated in numerous contributions in the past [69–71] and will not enhance the problem formulation in the following Sections.

The SKG procedure encompasses three phases: *advantage distillation*, *information reconciliation*, and *privacy amplification* [14], [15] as described below:

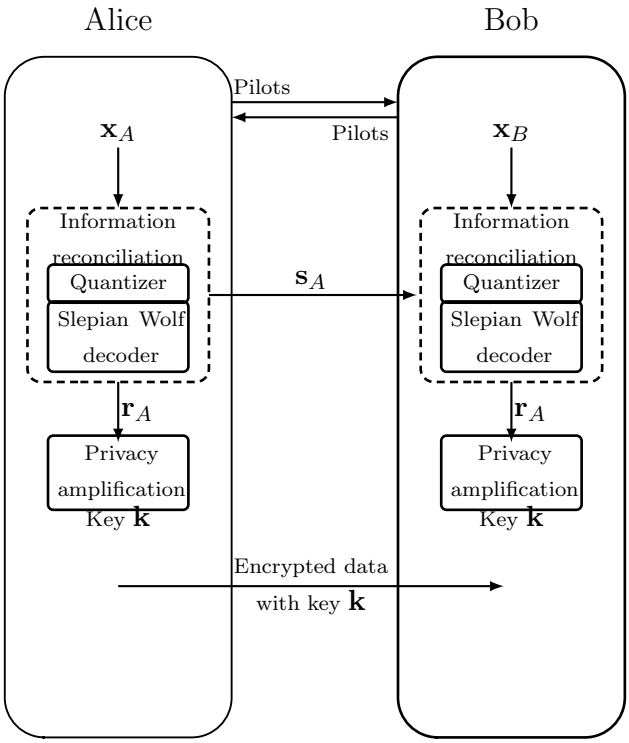

Figure 2.2: Secret key generation between Alice and Bob.

1) Advantage distillation: This phase takes place during the coherence time of the channel. The legitimate nodes sequentially exchange constant probe signals with power P on all subcarriers[2], to obtain estimates of their reciprocal CSI. We note in passing that the pilot exchange phase can be made robust with respect to injection type of attacks (that fall in the general category of MiM) as analyzed in [54]. Commonly, the received signal strength (RSS) has been used as the source of shared randomness for generating the shared key, but it is possible to use the full CSI [72]. At the end of this phase, Alice and Bob obtain observation vectors $\mathbf{x}_A = [x_{A,1}, \ldots, x_{A,N}], \mathbf{x}_B = [x_{B,1}, \ldots, x_{B,N}]$, respectively, so that:

$$\mathbf{x}_A = \sqrt{P}\mathbf{h} + \mathbf{z}_A, \tag{2.1}$$

$$\mathbf{x}_B = \sqrt{P}\mathbf{h} + \mathbf{z}_B, \tag{2.2}$$

where $\mathbf{z}_A$ and $\mathbf{z}_B$ denote zero-mean, unit variance circularly symmetric complex AWGN random vectors, such that $(\mathbf{z}_A, \mathbf{z}_B) \sim \mathcal{CN}(\mathbf{0}, \mathbf{I}_{2N})$. On the other hand, Eve observes $\mathbf{x}_E = [x_{E,1}, \ldots, x_{E,N}]$ with:

$$\mathbf{x}_E = \sqrt{P}\mathbf{h}_E + \mathbf{z}_E. \tag{2.3}$$

Due to the rich Rayleigh multipath environment, Eve's channel measurement $\mathbf{h}_E$ is assumed uncorrelated to $\mathbf{h}$ and $\mathbf{z}_E$ denotes a zero-mean, unit variance circularly symmetric complex AWGN random vector $\mathbf{z}_E \sim \mathcal{CN}(\mathbf{0}, \mathbf{I}_N)$.

2) Information reconciliation: At the beginning of this phase the observations $x_{A,j}, x_{B,j}$ are quantized to binary vectors[3] $\mathbf{r}_{A,j}$, $\mathbf{r}_{B,j}$ $j = 1, \ldots, N$ [73–75], so that Alice and Bob distill $\mathbf{r}_A = [\mathbf{r}_{A,1}|| \ldots ||\mathbf{r}_{A,N}]$ and $\mathbf{r}_B = [\mathbf{r}_{B,1}|| \ldots ||\mathbf{r}_{B,N}]$, respectively. Due to the presence of noise, $\mathbf{r}_A$ and $\mathbf{r}_B$ will differ. To reconcile discrepancies in the quantizer local outputs, side information needs to be exchanged via a public channel. Using the principles of Slepian Wolf decoding, the distilled binary vectors can be

[2] An explanation of the optimality of this choice under different attack scenarios is discussed in [22].

[3] Note that each observation can generate a multi-bit vector at the output of the quantizer.

expressed as

$$\mathbf{r}_A = \mathbf{d} + \mathbf{e}_A, \tag{2.4}$$
$$\mathbf{r}_B = \mathbf{d} + \mathbf{e}_B, \tag{2.5}$$

where $\mathbf{e}_A, \mathbf{e}_B$ are error vectors that represent the distance from the common observed (codeword) vector $\mathbf{d}$ at Alice and Bob, respectively.

Numerous practical information reconciliation approaches using standard forward error correction codes (e.g., LDPC, BCH, etc.,) have been proposed [16], [72]. As an example, if a block encoder is used, then the error vectors can be recovered from the syndromes $\mathbf{s}_A$ and $\mathbf{s}_B$ of $\mathbf{r}_A$ and $\mathbf{r}_B$, respectively. Alice transmits her corresponding syndrome to Bob so that he can reconcile $\mathbf{r}_B$ to $\mathbf{r}_A$. It has been shown that the length of the syndrome $|\mathbf{s}_A|$ is lower bounded by $|\mathbf{s}_A| \geq H(\mathbf{x}_A|\mathbf{x}_B) = H(\mathbf{x}_A, \mathbf{x}_B) - H(\mathbf{x}_B)$ [15]. This has been numerically evaluated for different scenarios and coding techniques [74,76–78]. Following that, the achievable SKG rate is upper bounded by $I(\mathbf{x}_A; \mathbf{x}_B|\mathbf{x}_E)$.

3) Privacy amplification: The secret key is generated by passing $\mathbf{r}_A$ through a one-way collision resistant *compression* function i.e., by hashing. Note that this final step of privacy amplification, is executed locally without any further information exchange. The need for privacy amplification arises in order to suppress the entropy revealed due to the public transmission of the syndrome $\mathbf{s}_A$. Privacy amplification produces a key of length strictly shorter than $|\mathbf{r}_A|$, at least by $|\mathbf{s}_A|$. At the same time, the goal is for the key to be uniform, i.e., to have maximum entropy. In brief, privacy amplification *reduces the overall output entropy* while at the same time *increases the entropy per bit* – compared to the input.

The privacy amplification is typically performed by applying either cryptographic hash functions such as those built using the Merkle-Damgard construction, or universal hash functions and has been proven to be secure, in an information theoretic sense, through the leftover hash lemma [79]. As an example, [40, 80] use a 2-universal hash family to achieve privacy amplification. Summarizing, the maximum key size after privacy amplification is:

$$|\mathbf{k}| \leq H(\mathbf{x}_A) - I(\mathbf{x}_A; \mathbf{x}_\mathbf{E}) - H(\mathbf{x}_A|\mathbf{x}_B) - r_0, \tag{2.6}$$

where $H(\mathbf{x}_A)$ represents the entropy of the measurement, $I(\mathbf{x}_A; \mathbf{x}_\mathbf{E})$ represents the mutual information between Alice's and Eve's observations, $H(\mathbf{x}_A|\mathbf{x}_B)$ represents the entropy revealed during information reconciliation and $r_0 > 0$ is an extra security parameter that ensures uncertainty on the key at Eve's side. For details and estimation of these parameters in a practical scenario please see [81].

As shown in this Section the SKG procedure requires only a few simple operations such as quantization, syndrome calculation and hashing. In future work we will examine the real possibilities of implementing such a mechanism in practical systems.

2.4.3 AE Using SKG

To develop a hybrid cryptosystem that can withstand tampering attacks, SKG can be introduced in standard AE schemes in conjunction with standard block ciphers in counter mode (to reduce latency), e.g., the advanced encryption standard (AES) in Galois counter mode (GCM). As a sketch of such a primitive, let us assume a system with three parties: Alice who wishes to transmit a secret message $\mathbf{m}$ with size $|\mathbf{m}|$, to Bob with confidentiality and integrity, and Eve, that can act as a passive and active attacker. The following algorithms are employed:

- The SKG scheme denoted by $\mathtt{G} : \mathbb{C} \to \mathcal{K} \times \mathcal{S}$, accepting as input the fading coefficients (modelled as complex numbers), and generating as outputs binary vectors $\mathbf{k}$ and $\mathbf{s}_A$ in the key and syndrome spaces, of sizes $|\mathbf{k}|$ and $|\mathbf{s}_A|$, respectively,

$$\mathtt{G}(\mathbf{h}) = (\mathbf{k}, \mathbf{s}_A)\,, \tag{2.7}$$

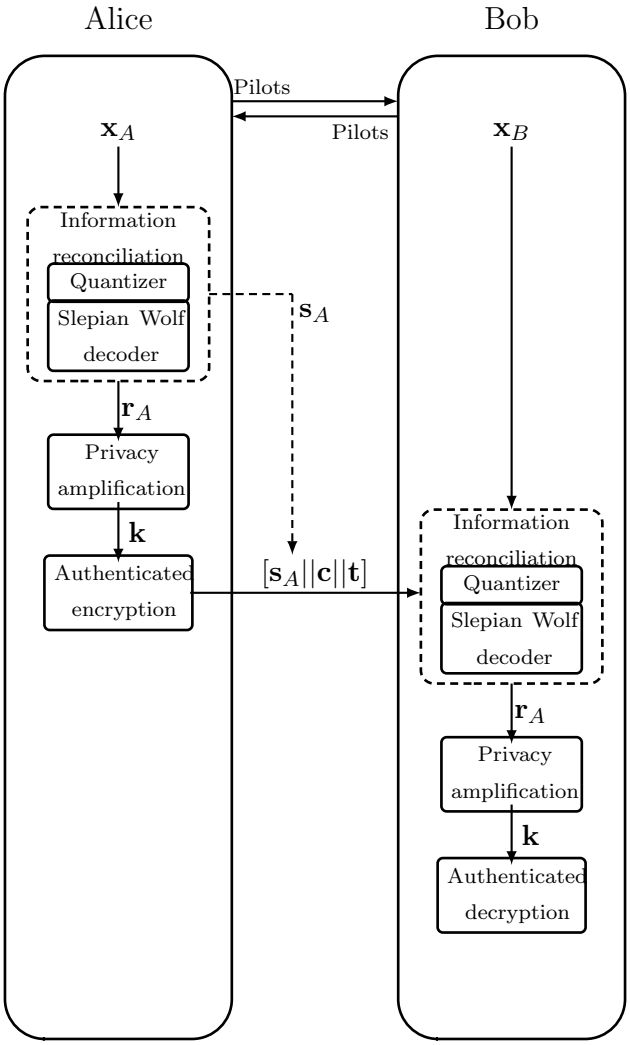

Figure 2.3: Pipelined SKG and encrypted data transfer between Alice and Bob.

where $\mathbf{k} \in \mathcal{K}$ denotes the key obtained from $\mathbf{h}$ after privacy amplification and $\mathbf{s}_A$ is Alice's syndrome.

- A symmetric encryption algorithm, e.g., AES GCM, denoted by $\texttt{Es} : \mathcal{K} \times \mathcal{M} \rightarrow \mathcal{C}_{\mathcal{T}}$ where $\mathcal{C}_{\mathcal{T}}$ denotes the ciphertext space with corresponding decryption $\texttt{Ds} : \mathcal{K} \times \mathcal{C}_{\mathcal{T}} \rightarrow \mathcal{M}$, such that

$$\texttt{Es}(\mathbf{k}, \mathbf{m}) = \mathbf{c}, \tag{2.8}$$
$$\texttt{Ds}(\mathbf{k}, \mathbf{c}) = \mathbf{m}, \tag{2.9}$$

  for $\mathbf{m} \in \mathcal{M}$, $\mathbf{c} \in \mathcal{C}_{\mathcal{T}}$.

- A pair of message authentication code (MAC) algorithms, e.g., in hashed-MAC (HMAC) mode, denoted by $\texttt{Sign} : \mathcal{K} \times \mathcal{M} \rightarrow \mathcal{T}$, with a corresponding verification algorithm $\texttt{Ver} : \mathcal{K} \times \mathcal{M} \times \mathcal{T} \rightarrow (yes, no)$, such that

$$\texttt{Sign}(\mathbf{k}, \mathbf{m}) = \mathbf{t}, \tag{2.10}$$
$$\texttt{Ver}(\mathbf{k}, \mathbf{m}, \mathbf{t}) = \begin{cases} yes, & \text{if integrity verified} \\ no, & \text{if integrity not verified} \end{cases} \tag{2.11}$$

A hybrid crypto-PLS system for AE SKG can be built as follows:

1. The SKG procedure is launched between Alice and Bob generating a key and a syndrome $\texttt{G}(\mathbf{h}) = (\mathbf{k}, \mathbf{s}_A)$.

2. Alice breaks her key into two parts $\mathbf{k} = \{\mathbf{k}_e, \mathbf{k}_i\}$ and uses the first to encrypt the message as $\mathbf{c} = \texttt{Es}(\mathbf{k}_e, \mathbf{m})$. Subsequently, using the second part of the key she signs the ciphertext using the signing algorithm $\mathbf{t} = \texttt{Sign}(\mathbf{k}_i, \mathbf{c})$ and transmits to Bob the extended ciphertext $[\mathbf{s}_A \| \mathbf{c} \| \mathbf{t}]$, as it is depicted in Fig. 2.3.

3. Bob checks first the integrity of the received ciphertext as follows: from $\mathbf{s}_A$ and his own observation he evaluates $\mathbf{k} = \{\mathbf{k}_e, \mathbf{k}_i\}$ and computes $\texttt{Ver}(\mathbf{k}_i, \mathbf{c}, \mathbf{t})$. The integrity test will fail if any part of the extended ciphertext was modified, including the syndrome (that is sent as plaintext); for example, if the syndrome was modified during the transmission, then Bob would not have evaluated the correct key and the integrity test would have failed.

4. If the integrity test is successful then Bob decrypts $\mathbf{m}=\texttt{Ds}(\mathbf{k}_e, \mathbf{c})$.

2.4.4 Resumption Protocol

In Section 2.4.1 we discussed that using PUF authentication can greatly reduce the computational overhead of a system. Authentication of new keys is required at the start of communication and at each key renegotiation. However, the number of challenges that can be applied to a single PUF is limited. Due to that we present a solution that is inspired by the 0-RTT authentication mode introduced in the 1.3 version of the TLS [23]. The use of 0-RTT obviates the need of performing a challenge for every re-authentication through the use of a resumption secret $\mathbf{r}_s$, thus reducing latency. Another strong motivation for using this mechanism is that it is forward secure in the scenario we are using here [24]. We first briefly describe the TLS 0-RTT mechanism before describing a similarly inspired 0-RTT mechanism applied to the information reconciliation phase of our SKG mechanism.

The TLS 1.3 0−RTT handshake works as follows: In the very first connection between client and server a regular TLS handshake is used. During this step the server sends to the client a look-up identifier $\mathbf{k}_l$ for a corresponding entry in session caches or it sends a session ticket. Then both parties derive a resumption secret $\mathbf{r}_s$ using their shared key and the parameters of the session. Finally, the client stores the resumption secret $\mathbf{r}_s$ and uses it when reconnecting to the same server which also retrieves it during the re-connection.

If session tickets are used the server encrypts the resumption secret using long-term symmetric encryption key, called a session ticket encryption key (STEK), resulting in a session ticket. The session ticket is then stored by the client and included in subsequent connections, allowing the server to retrieve the resumption secret. Using this approach the same STEK is used for many sessions and clients. On one hand, this property highly reduces the required storage of the server, however, on the other hand, it makes it vulnerable to replay attacks and not forward secure. Due to these vulnerabilities, in this work we focus on the session cache mechanism described next.

When using session caches the server stores all resumption secrets and issues a unique look-up identifier $\mathbf{k}_l$ for each client. When a client tries to reconnect to that server it includes its look-up identifier $\mathbf{k}_l$ in the 0-RTT message, which allows the server to retrieve the resumption secret $\mathbf{r}_s$. Storing a unique resumption secret $\mathbf{r}_s$ for each client requires server storage for each client but it provides forward security and resilience against replay attacks, when combined with a key generation mechanisms such as Diffie Hellman (or the SKG proposed in the present) which are important goals for security protocols [24]. In our physical layer 0-RTT, given that a node identifier state would be required for link-layer purposes, the session cache places little comparative load and thus is the mechanism proposed here for (re-)authentication.

The physical layer resumption protocol modifies the information reconciliation phase of Section 2.4.2 following initial authentication to provide a re-authentication mechanism between Alice and Bob. At the first establishment of communication we assume initial authentication is established, such as the mechanism shown in Section 2.4.1. During that Alice sends to Bob a look-up identifier $\mathbf{k}_l$. Then, both derive a resumption secret $\mathbf{r}_s$ that is identified by $\mathbf{k}_l$. Note, $\mathbf{r}_s$ and the session key have the same length $|\mathbf{k}|$. Then referring to the notation and steps in Section 4.1-4.3:

1. Advantage distillation phase is carried out as before (See section 2.4.2), where both parties obtain channel observations and obtain the vectors $\mathbf{r}_A$ and $\mathbf{r}_B$.

2. During the information reconciliation phase both Alice and Bob exclusive-or the resumption secret $\mathbf{r}_s$ with their observations $\mathbf{r}_A$ and $\mathbf{r}_B$, obtaining syndromes $\mathbf{s}'_A$ and $\mathbf{s}'_B$ with which both parties can carry out reconciliation to obtain the same shared value which is now $\mathbf{r}_A \oplus \mathbf{r}_s$.

3. The privacy amplification step in Section 4.2 is carried out as before, but now the hashing takes place on $\mathbf{r}_A \oplus \mathbf{r}_s$ to produce the final shared key $\mathbf{k}'$ that is a result of both the shared wireless randomness and the resumption secret.

Note that the key $\mathbf{k}'$ can only be obtained if both the physical layer generated key and the resumption key are valid and this method can be shown to be forward secure [24].

2.5 Pipelined SKG and Encrypted Data Transfer

As explained in the previous Section, if Alice and Bob follow the standard sequential SKG process they can exchange encrypted data only after both of them have distilled the key at the end of the privacy amplification step. In this Section, we propose a method to pipeline the SKG and encrypted data transfer. Alice can unilaterally extract the secret key from her observation and use it to encrypt data transmitted in the same "extended" ciphertext that contains the side information (see Fig. 2.3). Subsequently, using the side information, Bob can distill the same key $\mathbf{k}$ and decrypt the received data in one single step.

We have discussed in Section 2.4.2 how Alice and Bob can distill secret keys from estimates of the fading coefficients in their wireless link and in Section 2.4.3 how these can be used to develop an AE SKG primitive. At the same time CSI estimates are prerequisite in order to optimally allocate power across the subcarriers and achieve high data rates[4]. As a result, a question that naturally arises is

[4]As an example, despite the extra overhead, in URLLC systems advanced CSI estimation techniques are employed in order to be able to satisfy the strict reliability requirements.

whether the CSI estimates (obtained at the end of the pilot exchange phase), should be used towards the generation of secret keys or towards the reliable data transfer, and, furthermore, whether the SKG and the data transfer can be inter-woven using the AE SKG principle.

In this study, we are interested in answering this question and shed light into whether following the exchange of pilots, Alice should transmit reconciliation information on all subcarriers, so that she and Bob can generate (potentially) a long sequence of key bits, or, alternatively, perform information reconciliation only over a subset of the subcarriers and transmit encrypted data over the rest, exploiting the idea of the AE SKG primitive. Note here that the data can be already encrypted with the key generated at Alice, the sender of the side information, so that the proposed pipelining does not require storing keys for future use. We will call the former approach a *sequential* scheme, while we will refer to the latter as a *parallel* scheme. The two will be compared in terms of their efficiency with respect to the achievable data rates.

A simplified version of this problem, where the reconciliation rate is roughly approximated to the SKG rate, was investigated in [82]. In this study it was shown that in order to maximize the data rates in the *parallel* approach Alice and Bob should use the strongest subcarriers – in terms of SNR – for data transmission and the worst for SKG. Under this simplified formulation, the optimal power allocation for the data transfer has been shown to be a *modified* waterfilling solution.

Here, we explicitly account for the rate of transmitting reconciliation information and differentiate it from the SKG rate. We confirm whether the policy of using the strongest subcarriers for data transmission and not for reconciliation, is still optimal when the full optimization problem is considered, including the communication cost for reconciliation.

As discussed in Section 2.4.2, our physical layer system model assumes Alice and Bob exchange data over a Rayleigh BF-AWGN channel with N orthogonal subcarriers. Without loss of generality the variance of the AWGN in all links is assumed to be unity. During channel probing, constant pilots are sent across all subcarriers [16, 83] with power P. Using the observations (2.1), Alice estimates the channel coefficients as

$$\hat{h}_j = h_j + \tilde{h}_j, \tag{2.12}$$

for $j = 1, \ldots, N$ where $\tilde{h}_j$ denotes an estimation error that can be assumed to be Gaussian, $\tilde{h}_j \sim \mathcal{CN}(0, \sigma_e^2)$ [84]. Under this model, the following rate is achievable on the j-th subcarrier from Alice to Bob when the transmit power during data transmission is p_j [84]:

$$R_j = \log_2\left(1 + \frac{g_j p_j}{\sigma_e^2 P + 1}\right) = \log_2(1 + \hat{g}_j p_j), \tag{2.13}$$

where we use $\hat{g}_i = \frac{g_i}{\sigma_{i,e}^2 P + 1}$, to denote the estimated channel gains. As a result, the channel capacity $C = \sum_{j=1}^{N} R_j$ under the short term power constraint

$$\sum_{j=1}^{N} p_j \leq NP, \;\; p_j \geq 0, \; \forall j \in \{1, \ldots, N\}, \tag{2.14}$$

is achieved with the well known waterfilling power allocation policy $p_j = \left[\frac{1}{\lambda} - \frac{1}{\hat{g}_j}\right]^+$, where the water-level λ is estimated from the constraint (2.14). In the following, the estimated channel gains $\hat{g}_j$ are – without loss of generality – assumed ordered in descending order, so that:

$$\hat{g}_1 \geq \hat{g}_2 \geq \ldots \geq \hat{g}_N. \tag{2.15}$$

As mentioned above, the advantage distillation phase of the SKG process consists of the two-way exchange of pilot signals during the coherence time of the channel to obtain $\mathbf{r}_{A,j}, \mathbf{r}_{B,j}, j = 1, \ldots, N$.

On the other hand, the CSI estimation phase can be used to estimate the reciprocal channel gains in order to optimize data transmission using the waterfilling algorithm. In the former case, the shared parameter is used for generating symmetric keys, in the latter for deriving the optimal power allocation. In the parallel approach the idea is to inter-weave the two procedures and investigate whether a joint encrypted data transfer and key generation scheme as in the AE SKG in Section 4.3 could bear any advantages with respect to the system efficiency. While in the sequential approach the CSI across all subcarriers will be treated as a source of shared randomness between Alice and Bob, in the parallel approach it plays a dual role.

2.5.1 Parallel Approach

In the parallel approach, after the channel estimation phase, the legitimate users decide on which subcarrier to send the reconciliation information (e.g., the syndromes as discussed in Section 2.4.2) and on which data (*i.e.*, the SKG process here is not performed on all of the subcarriers). The total capacity has now to be distributed between data and reconciliation information bearing subcarriers. As a result, the overall set of orthogonal subcarriers comprises two subsets; a subset $\mathcal{D}$ that is used for encrypted data transmission with cardinality $|\mathcal{D}| = D$ and a subset $\breve{\mathcal{D}}$ with cardinality $|\breve{\mathcal{D}}| = N - D$ used for reconciliation such that, $\mathcal{D} \cup \breve{\mathcal{D}} = \{1, \ldots, N\}$. Over $\mathcal{D}$ the achievable sum data transfer rate, denoted by C_D is given by

$$C_D = \sum_{j \in \mathcal{D}} \log_2(1 + \hat{g}_j p_j), \tag{2.16}$$

while on the subset $\breve{\mathcal{D}}$, Alice and Bob exchange reconciliation information at rate

$$C_R = \sum_{j \in \breve{\mathcal{D}}} \log_2(1 + \hat{g}_j p_j). \tag{2.17}$$

As stated in Section 2.4.2 the fading coefficients are assumed to be zero-mean circularly-symmetric complex Gaussian random variables. Using the theory of order statistics, the distribution of the ordered channel gains of the SKG subcarriers, $j \in \breve{\mathcal{D}}$, can be expressed as [85]:

$$f(g_j) = \frac{N!}{\sigma^2 (N-j)!(j-1)!} \left(1 - e^{-\frac{\hat{g}_j}{\sigma^2}}\right)^{N-j} \left(e^{-\frac{\hat{g}_j}{\sigma^2}}\right)^{j}, \tag{2.18}$$

where σ^2 is the variance of the channel gains. As a result of ordering the subcarriers, the variance of each of the subcarriers, is now given by:

$$\sigma_j^2 = \sigma^2 \sum_{q=j}^{N} \frac{1}{q^2}, \quad j \in \{D+1, \ldots, N\}. \tag{2.19}$$

Thus, we can now write the SKG rate as (note that the noise variances are here normalized to unity for simplicity) [16, 83]:

$$C_{SKG} = \sum_{j \in \breve{\mathcal{D}}} \log_2 \left(1 + \frac{P\sigma_j^2}{2 + \frac{1}{P\sigma_j^2}}\right). \tag{2.20}$$

The minimum rate necessary for reconciliation was discussed in Section 4.2. Here, alternatively, we employ a practical design approach in which the rate of the encoder used is explicitly taken into account. Note that in a rate $\frac{k}{n}$ block encoder the side information is $n - k$ bits long, i.e., the rate of syndrome to output key bits after privacy amplification is $\frac{n-k}{k}$. However, in each key session a 0-RTT look-up identifier of length k is also sent. Therefore, we define the parameter $\kappa = \frac{n-k}{k} + 1 = \frac{n}{k}$, i.e., the inverse of the encoder rate, that reflects the ratio of the reconciliation and 0-RTT transmission

rate to the SKG rate. For example, for a rate $\frac{k}{n} = \frac{1}{2}$ encoder, $\kappa = 2$, etc. Based on this discussion, we capture the minimum requirement for the reconciliation rate through the following expression:

$$C_R \quad \geq \quad \kappa C_{SKG}. \tag{2.21}$$

Furthermore, to identify the necessary key rate, we note that depending on the exact choices of the cryptographic suites to be employed, it is possible to reuse the same key for the encryption of multiple blocks of data, e.g., as in the AES GCM, that is being considered for employment in the security protocols for URLLC systems [4]. In practical systems, a single key of length 128 to 256 bits can be used to encrypt up to gigabytes of data. As a result, we will assume that for a particular application it is possible to identify the ratio of key to data bits, which in the following we will denote by β. Specifically, we assume that the following security constraint should be met

$$C_{SKG} \geq \beta C_D, \;\; 0 < \beta \leq 1, \tag{2.22}$$

where, depending on the application, the necessary minimum value of β can be identified. We note in passing that the case $\beta = 1$ would correspond to a one-time-pad, *i.e.*, the generated keys could be simply x-ored with the data to achieve perfect secrecy without the need of any cryptographic suites.

Accounting for the reconciliation rate and security constraints in (2.21) and (2.22) we formulate the following maximization problem:

$$\max_{p_j, j\in\mathcal{D}} \sum_{j\in\mathcal{D}} R_j \tag{2.23}$$

$$\text{s.t. } (2.14), (2.21), (2.22),$$

$$\sum_{j\in\mathcal{D}} R_j + \sum_{j\in\check{\mathcal{D}}} R_j \leq C. \tag{2.24}$$

(2.22) can be integrated with (2.21) to the combined constraint

$$\sum_{j\in\mathcal{D}} R_j \quad \leq \quad \frac{\sum\limits_{j\in\check{\mathcal{D}}} R_j}{\kappa\beta}. \tag{2.25}$$

The optimization problem at hand is a mixed-integer convex optimization problem with unknowns both the sets $\mathcal{D}, \check{\mathcal{D}}$, as well as the power allocation policy $p_j, j \in \{1, \ldots, N\}$. These problems are typically NP hard and addressed with the use of branch and bound algorithms and heuristics.

In this work, we propose a simple heuristic to make the problem more tractable by reducing the number of free variables. In the proposed approach, we assume that the constraint (2.24) is satisfied with equality. The only power allocation that allows this is the waterfilling approach that uniquely determines the power allocation p_j and also requires that the constraint (2.14) is also satisfied with equality. Thus, if we follow that approach, we determine the power allocation vector uniquely and can combine the remanining constraints (2.24) and (2.25) into a single one as:

$$\sum_{j\in\mathcal{D}} R_j \leq \frac{C}{\kappa\beta + 1}. \tag{2.26}$$

Algorithm 1: Heuristic Greedy Algorithm for (2.27)-(2.28)

procedure HEURISTIC(start, end, R_j)
  $j \leftarrow 1, R_0 \leftarrow 0, R_{N+1} \leftarrow 0$
  while $j \leq N-1$ and $\sum_{j=1}^{N} R_j x_j \leq \frac{C}{1+\kappa\beta}$ **do**
    $\sum_{j=1}^{N} R_j x_j \leftarrow \sum_{j=1}^{N} R_{j-1} x_{j-1} + R_j x_j$
    if $\sum_{j=1}^{N} R_j x_j \leq \frac{C}{1+\kappa\beta}$ **then**
      $x_j \leftarrow 1; j \leftarrow j+1$
    else do $x_j \leftarrow 0; j \leftarrow j+1$
    end if
  end while
end procedure

The new optimization problem can be re-written as

$$\max_{x_j \in \{0,1\}} \sum_{j=1}^{N} R_j x_j \tag{2.27}$$

$$\text{s.t.} \sum_{j=1}^{N} R_j x_j \leq \frac{C}{1+\kappa\beta}. \tag{2.28}$$

The problem in (2.27)-(2.28) is a subset-sum problem from the family of $0-1$ knapsack problems, that is known to be NP hard [86]. However, these type of problems are solvable optimally using dynamic programming techniques in pseudo-polynomial time [86, 87]. Furthermore, it is known that greedy heuristic approaches are bounded away from the optimal solution by half [88].

We propose a simple greedy heuristic algorithm of *linear complexity,* as follows.[5] The data subcarriers are selected starting from the best – in terms of SNR – until (2.28) is not satisfied. Once this situation occurs the last subcarrier added to set $\mathcal{D}$ is removed and the next one is added. This continues either to the last index N or until (2.28) is satisfied with equality. The algorithm is described in *Algorithm 1.*

The efficiency of the proposed parallel method – measured as the ratio of the long-term data rate versus the average capacity – is evaluated as:

$$\eta_{\text{parallel}} = \frac{\mathbb{E}\left[\sum_{j \in \mathcal{D}} R_j\right]}{\mathbb{E}[C]}. \tag{2.29}$$

This efficiency quantifies the expected back-off in terms of data rates when part of the resources (power and frequency) are used to enable the generation of secret keys at the physical layer. In future work, we will compare the efficiency achieved to that of actual approaches currently used in 5G by accounting for the actual delays incurred due to the PKE key agreement operations [20].

2.5.2 Sequential Approach

In the sequential approach encrypted data transfer and secret key generation are two separate events; first, the secret keys are generated over the whole set of subcarriers, leading to a sum SKG rate given

[5] Without loss of generality, the algorithm assumes that the channel gains are ordered in decreasing order as in (2.15), and, consequently, the rates R_j are also ordered in descending order. The ordering is a $\mathcal{O}(N \log N)$ operation and required in common power allocation schemes such as the waterfilling, and, therefore does not come at any additional cost.

as

$$C_{SKG} = N \log_2 \left(1 + \frac{P\sigma^2}{2 + \frac{1}{P\sigma^2}} \right). \tag{2.30}$$

To estimate the efficiency of the scheme, we further need to identify the necessary resources for the exchange of the reconciliation information. We can obtain an estimate of the number of transmission frames that will be required for the transmission of the syndromes, as the expected value of the reconciliation rate (*i.e.*, it's long-term value) $\mathbb{E}[C_R]$. The average number of frames needed for reconciliation is then computed as:

$$M = \left\lceil \frac{\kappa C_{SKG}}{\mathbb{E}[C_R]} \right\rceil, \tag{2.31}$$

where $\lceil x \rceil$ denotes the smallest integer that is larger than x.

The average number of the frames that can be sent while respecting the secrecy constraint is:

$$L = \left\lfloor \frac{C_{SKG}}{\beta \mathbb{E}[C]} \right\rfloor, \tag{2.32}$$

where $\lfloor x \rfloor$ denotes the largest interger that is smaller than x. The efficiency of the sequential method is then calculated as:

$$\eta_{\text{sequential}} = \frac{L}{L + M}. \tag{2.33}$$

2.6 Effective Data Rate Taking into Account Statistical Delay QoS Requirements

In the previous section, we investigated the optimal power and subcarrier allocations strategy of Alice and Bob in order to maximize their long-term average data rate and proposed a greedy heuristic algorithm of linear complexity. Here, we extend our work from Section 2.5 by taking into account delay requirements. In detail, we investigate the optimal resource allocation for Alice and Bob, when their communication has to satisfy specific delay constraints. To this end, we use the theory of *effective capacity* [28] which gives a limit for the maximum arrival rate under delay-bounds with a specified violation probability.

We study the *effective data rate* for the proposed pipelined SKG and encrypted data transfer scheme; the effective rate is a data-link layer metric that captures the impact of statistical delay QoS constraints on the transmission rates. As background, we refer to [89] which showed that the probability of a steady-state queue length process $Q(t)$ exceeding a certain queue-overflow threshold x converges to a random variable $Q(\infty)$ as:

$$\lim_{x \to \infty} \frac{\ln(\Pr[Q(\infty) > x])}{x} = -\theta, \tag{2.34}$$

where θ indicates the asymptotic exponential decay-rate of the overflow probability. For a large threshold x, (2.34) can be represented as $Pr[Q(\infty) > x] \approx e^{-\theta x}$. Furthermore, the delay-outage probability can be approximated by [28] :

$$\Pr^{\text{out}}_{\text{delay}} = \Pr[\text{Delay} > D_{\max}] \approx \Pr[Q(\infty) > 0] e^{-\theta \zeta D_{\max}}, \tag{2.35}$$

where $D_{\max}$ is the maximum tolerable delay, $\Pr[Q(\infty) > 0]$ is the probability of a non-empty buffer, which can be estimated from the ratio of the constant arrival rate to the averaged service rate, ζ is the upper bound for the constant arrival rate when the statistical delay metrics are satisfied.

Using the delay exponent θ and the probability of non-empty buffer, the effective capacity, that

denotes the maximum arrival rate, can be formulated as [28]:

$$E_C(\theta) = -\lim_{t\to\infty} \frac{1}{\theta} \ln \mathbb{E}\left[e^{-\theta S[t]}\right] (\text{bits/s}), \tag{2.36}$$

where $S[t] = \sum_{i=1}^{t} s[i]$ denotes the time-accumulated service process, and $s[i], i = 1, 2, ...$ denotes the discrete-time stationary and ergodic stochastic service process. Therefore, the delay exponent θ indicates how strict the delay requirements are, *i.e.*, $\theta \to 0$ corresponds to looser delay requirements, while $\theta \to \infty$ implies exceptionally stringent delay constraints. Assuming a Rayleigh block fading system, with frame duration T_f and total bandwidth B, we have $s[i] = T_f B\tilde{R}_i$, with $\tilde{R}_i$ representing the instantaneous service rate achieved during the duration of the ith frame. In the context of the investigated data and reconciliation information transfer, $\tilde{R}_i$, is given by:

$$\tilde{R}_i = \frac{1}{F} \sum_{i\in\mathcal{D}} \log_2(1 + p_i\hat{g}_i), \tag{2.37}$$

where F is the equivalent frame duration, *i.e.*, the total number of subcarriers used for data transmission, so that for the parallel approach we have $F = |D|$ while for the sequential approach $F = N(L+M)L^{-1}$.

Under this formulation and assuming that Gärtner-Ellis theorem [90, 91] is satisfied, the *effective data rate*[6] $E_C(\theta)$ is given as:

$$E_{C,\mathcal{D}}(\theta) = -\frac{1}{\theta T_f B} \ln\left(\mathbb{E}\left[e^{-\theta T_f B\tilde{R}_i}\right]\right). \tag{2.38}$$

We set $\alpha = \frac{\theta T_f B}{\ln(2)}$. By inserting (2.37) into (2.38) we get:

$$\begin{aligned} E_{C,\mathcal{D}}(\theta) &= -\frac{1}{\ln(2)\alpha} \ln\left(\mathbb{E}\left[e^{-\ln(2)\alpha F^{-1}\sum_{i\in\mathcal{D}}\log_2(1+p_i\hat{g}_i)}\right]\right), \\ E_{C,\mathcal{D}}(\theta) &= -\frac{1}{\alpha}\log_2\left(\mathbb{E}\left[\prod_{i\in\mathcal{D}}(1+p_i\hat{g}_i)^{-\alpha F^{-1}}\right]\right). \end{aligned} \tag{2.39}$$

Assuming i.i.d. channel gains, by using the distributive property of the mathematical expectation, (2.39) becomes [92]:

$$E_{C,\mathcal{D}}(\theta) = -\frac{1}{\alpha}\log_2\left(\prod_{i\in\mathcal{D}}\mathbb{E}\left[(1+p_i\hat{g}_i)^{-\alpha F^{-1}}\right]\right). \tag{2.40}$$

We further manipulate by using the log-product rule to obtain:

$$E_{C,D}(\theta) = -\frac{1}{\alpha}\sum_{i\in\mathcal{D}}\log_2\left(\mathbb{E}\left[(1+p_i\hat{g}_i)^{-\alpha F^{-1}}\right]\right). \tag{2.41}$$

Similarly, the *effective syndrome rate* can be written as:

$$E_{C,\breve{D}}(\theta) = -\frac{1}{\alpha}\sum_{i\in\breve{\mathcal{D}}}\log_2\left(\mathbb{E}\left[(1+p_i\hat{g}_i)^{-\alpha\breve{F}^{-1}}\right]\right), \tag{2.42}$$

where the size of $\breve{F}$ here is $|N - D|$.

[6]Since part of the transmission rate is used for reconciliation information, and part for data transmission the terms "*effective syndrome rate*" and "*effective data rate*" are introduced instead of the term "effective capacity", for rigour. We note that we assume the information data and reconciliation information are accumulated in separate independent buffers within the transmitter.

Using that, we now reformulate the maximization problem given in (2.23) by adding a delay constraint. The reformulated problem can be expressed as follows:

$$\max_{p_j, j\in\mathcal{D}} E_{C,\mathcal{D}}(\theta), \tag{2.43}$$

$$\text{s.t. } (2.14), (2.25),$$

$$E_{C,\mathcal{D}}(\theta) + E_{C,\check{\mathcal{D}}}(\theta) \leq E_C^{\text{opt}}(\theta), \tag{2.44}$$

where $E_C^{\text{opt}}(\theta)$ represents the maximum achievable effective capacity for both key and data transmission for a given value of θ over N subcarriers:

$$E_C^{\text{opt}}(\theta) = \max_{p_i, i=1,2,\ldots N} \left\{ -\frac{1}{\alpha} \log_2 \left(\mathbb{E}\left[\prod_{i=1}^{N} (1 + p_i \hat{g}_i)^{-\alpha N^{-1}} \right] \right) \right\}. \tag{2.45}$$

In the proposed approach, we assume that the constraint (2.44) is satisfied with equality. Given that, the optimization problem in (2.43) can be evaluated as two sub-optimization problems: i) finding the optimal long term power allocation from (2.14) and (2.45); ii) finding the optimal subcarrier allocation that satisfies (2.25). We solve the first problem that gives the optimal power allocation using convex optimization tools. Next, as in Section 2.5 we use two methods to solve subcarrier allocation problem, i.e., by formulating a subset-sum 0 – 1 knapsack optimization problem or through a variation of *Algorithm 1.* The efficiency of both methods is compared numerically to the sequential method in Section 3.10.

Now, following the same steps as in (2.39)-(2.41) and using the fact that maximizing $E_C(\theta)$ is equivalent to minimizing $-E_C(\theta)$ (this is due to $\log(\cdot)$ being a monotonically increasing concave function for any $\theta > 0$) we formulate the following minimization problem:

$$\min_{p_i, i=1,2,\ldots N} \sum_{i=1}^{N} \left(\mathbb{E}\left[(1 + p_i \hat{g}_i)^{-\alpha N^{-1}} \right] \right), \tag{2.46}$$

$$\text{s.t. } (2.14).$$

where $F = N$ in this case as the full set of subcarriers is concerned. We form the Lagrangian function $\mathcal{L}$ as:

$$\mathcal{L} = \left(\mathbb{E}\left[(1 + p_i \hat{g}_i)^{-\alpha N^{-1}} \right] \right) + \lambda \left(\sum_{i=1}^{N} p_i - NP \right). \tag{2.47}$$

By differentiating (2.47) w.r.t. p_i and setting the derivative equal to zero [93] we get:

$$\frac{\partial \mathcal{L}}{\partial p_i} = \lambda - \frac{\alpha \hat{g}_i}{N} \left(\hat{g}_i p_i + 1 \right)^{-\frac{\alpha}{N} - 1} = 0. \tag{2.48}$$

Solving (2.48) gives the optimal power allocation policy:

$$p_i^* = \frac{1}{g_0^{\frac{N}{\alpha+N}} \hat{g}_i^{\frac{\alpha}{\alpha+N}}} - \frac{1}{\hat{g}_i}, \tag{2.49}$$

where $g_0 = \frac{N\lambda}{\alpha}$ is the cutoff value which can be found from the power constraint. By inserting p_i^* in

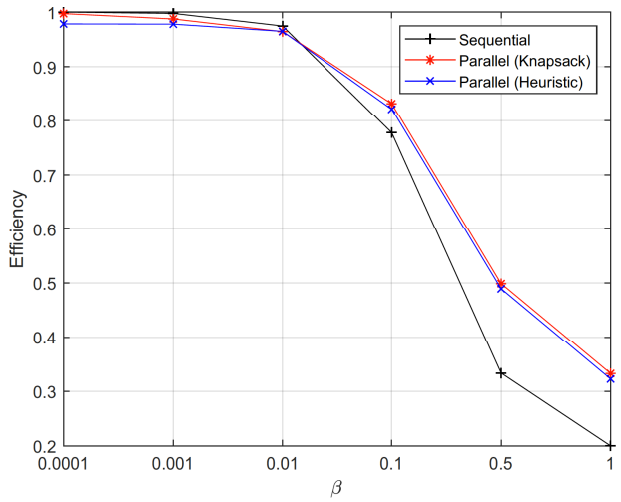

Figure 2.4: **a)** Efficiency comparison for $N = 12$, SNR=10 dB and $\kappa = 2$.

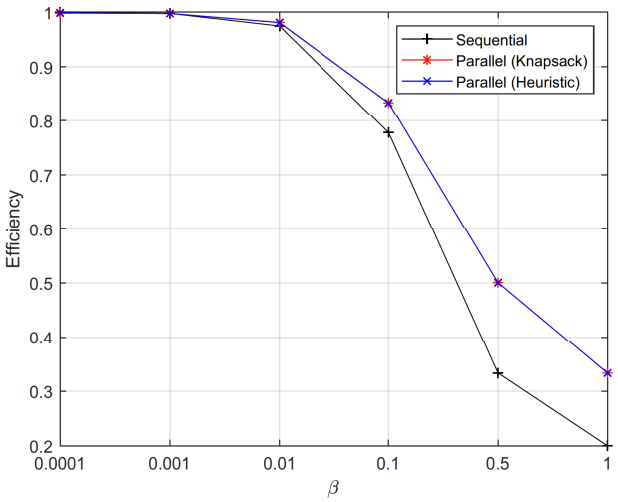

Figure 2.4: **b)** Efficiency comparison for $N = 64$, SNR=10 dB and $\kappa = 2$.

$E_C(\theta)$ we obtain the expression for $E_C^{\text{opt}}(\theta)$:

$$E_C^{\text{opt}}(\theta) = -\frac{1}{\alpha}\sum_{i=1}^{N}\log_2\left(\mathbb{E}\left[\left(\frac{\hat{g}_i}{g_0}\right)^{-\frac{\alpha}{\alpha+N}}\right]\right) \tag{2.50}$$

When $\theta \to 0$ the optimal power allocation is equivalent to water-filling and when $\theta \to \infty$ the optimal power allocation transforms to total channel inversion.

Now, fixing the power allocation as in (2.49) we can easily find the optimal subcarrier allocation that satisfies (2.25). As in Section 2.5 to do that we first formulate a subset-sum $0-1$ knapsack optimization problem that we solve using the standard dynamic programming approach. Furthermore we evaluate the performance of the heuristic algorithm presented in *Algorithm 1.*

2.7 Results and Discussion

In this Section we provide numerical evaluations of the efficiency that can be achieved with the presented methods (*i.e.*, sequential and parallel) for different values of the main parameters. With respect to the parallel approach, we provide numerical results of the optimal dynamic programming solution of the subset-sum $0-1$ knapsack problem, as well as of the greedy heuristic approach presented in *Algorithm 1.* For the case of the long term average data rate C_D (2.16), we compare the two methods through their efficiencies, *i.e.* $\eta_{\text{sequential}}$ and η_{parallel} given in (2.33) and (2.29), respectively. Next, to compare the two methods in the case of *effective data rate* we evaluate $E_{C,D}(\theta)$ given in (2.41). For better illustration of each case they are separated into different subsections.

2.7.1 Numerical results for the case long term average C_D

Figures 2.4a and 2.4b show the efficiency of the methods for $N = 12$, and $N = 64$, respectively, while $\kappa = 2$ and $P = 10$. We note that the proposed heuristic algorithm has a near-optimal performance (almost indistinguishable from the red curves achieved with dynamic programming). Due to this fact (which was tested across all scenarios that follow) only the heuristic approach is shown in subsequent figures for clarity in the graphs.

We see that when there are a small number of subcarriers (N=12, typical for NB-IoT) and small β the efficiency of both the parallel η_{parallel} and the sequential $\eta_{\text{sequential}}$ approaches are very close to

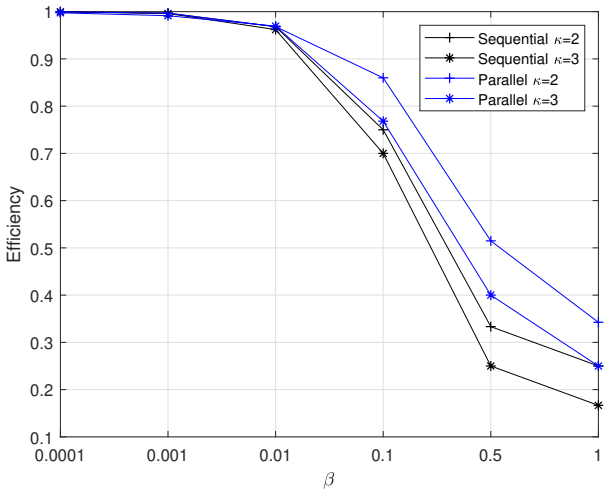

Figure 2.5: Efficiency vs κ, for $N = 24$, SNR=10 dB.

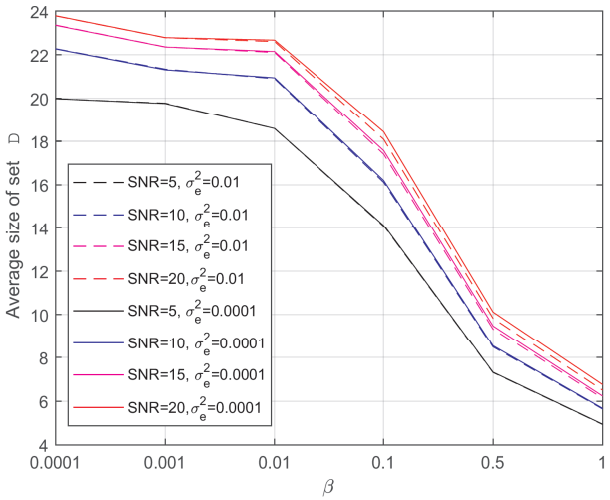

Figure 2.6: **a)** Size of set $\mathcal{D}$ for different SNR levels and σ_e^2 when $N = 24$.

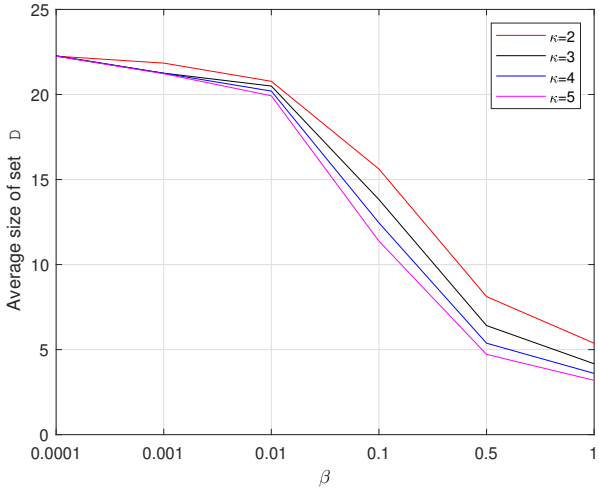

Figure 2.6: **b)** Size of set $\mathcal{D}$ for different values of κ when $N = 24$.

unity, a trend that holds for increasing N. With increasing β, due to the fact that more frames are needed for reconciliation in the sequential approach (*i.e.*, M increases), regardless of the total number of subcarriers, the parallel method proves more efficient than the sequential. While the efficiency of the sequential and parallel methods coincide almost until around $\beta = 0.01$ for $N = 12$, for $N = 64$ the crossing point of the curves moves to the left and the efficiency of the two methods coincide until around $\beta = 0.001$. This trend was found to be consistent across many values of N, only two of which are shown here for compactness of presentation.

Next, in Fig. 2.5 the efficiency of the parallel η_{parallel} and the sequential $\eta_{\text{sequential}}$ methods are shown for two different values of $\kappa \in \{2, 3\}$ for SNR = 10 dB and $N = 24$. It is straightforward to see that they both follow similar trends and when κ increases the efficiency decreases. On the other hand, regardless of the value of κ they both perform identically until around $\beta = 0.001$.

Finally, in Fig. 2.6, focusing on the parallel method, the average size of set $\mathcal{D}$ is shown for different values of σ_e^2 and SNR levels (Fig. 2.6a) and κ (Fig. 2.6b), for $N = 24$. As expected, in Fig. 2.6a we see when the SNR increases the size of the set increases, too. This is due to the fact that more power is used on any single subcarrier and consequently a higher reconcilliation rate can be sustained.

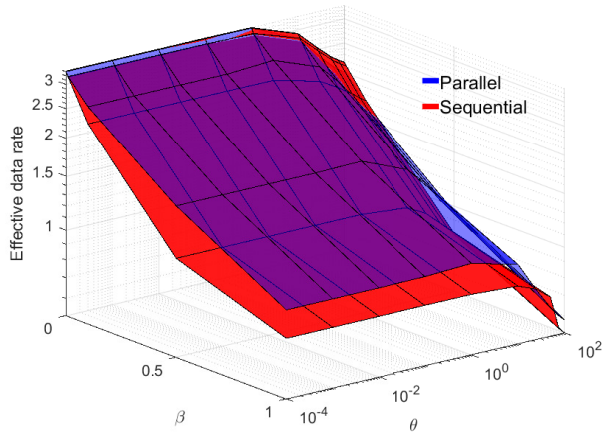

Figure 2.7: **a)** Effective data rate achieved by the parallel heuristic approach and the sequential approach when $N = 12$, SNR= 10 dB and $\kappa = 2$.

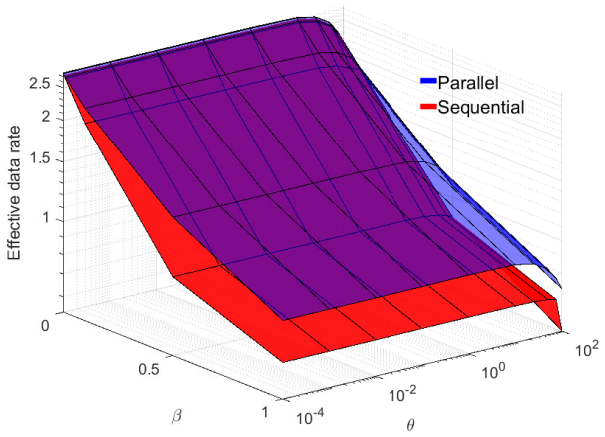

Figure 2.7: **b)** Effective data rate achieved by the parallel heuristic approach and the sequential approach when $N = 64$, SNR= 10 dB and $\kappa = 2$.

Regarding the estimation error σ_e^2 of the CSI, it only slightly affects the performance at high SNR levels. Hence more subcarriers have to be used for reconciliation, and fewer for data. The SNR level in Fig. 2.6b is set to 10 dB. The figure shows that when increasing κ the size of set $\mathcal{D}$ decreases. This result can be easily predicted from inequality (2.21), meaning, when κ increases more reconciliation data has to be sent, hence fewer subcarriers can be used for data. In both Fig. 2.6a and Fig. 2.6b when β increases the size of set $\mathcal{D}$ decreases; this effect is a consequence of constraint (2.28) as the data rate is decreasing with β.

2.7.2 Numerical results for the case of *effective data rate*

Inspired by the good performance of *Algorithm 1*, in the case where long-term average rate is the metric of interest, here, we continue our investigation with a variation of *Algorithm 1*, with the following differences: at lines 3 and 5 instead of (2.26) we use the constraint (2.25), the power allocation is fixed as in (2.49). The performance of our system is again compared with a sequential method and the metric of interest here is the *effective data rate*. The comparison is performed by taking into account the following parameters: signal to noise ration (SNR); number of subcarriers N; ratio of the reconciliation and $0-$RTT transmission rate to the SKG rate κ; delay exponent θ; and, the ratio of key bits to data bits β.

In Fig. 2.7 we give a three-dimensional plot showing the dependence of the achievable *effective data rate* $E_{C,D}(\theta)$ on β and θ. Figures 2.7a and 2.7b compare the parallel heuristic approach and the sequential approach for high SNR levels, whereas Fig. 2.8a and 2.8b compare their performance at low SNRs. In Fig. 2.7a and 2.8a we have $N = 12$ while in Fig. 2.7b and 2.8b the total number of subcarriers is $N = 64$. All graphs compare the performance of the heuristic parallel approach and the sequential approach for $\kappa = 2$.

As discussed in Section 2.6, when the delay exponent θ increases, the optimal power allocation transforms from waterfilling to total channel inversion. Consequently, the rate achieved on all subcarriers converges to the same value, hence when we a have small number of subcarriers (such as $N = 12$) and small values of β then using a single subcarrier for reconciliation data will use more capacity than needed and most of the rate on this subcarrier is wasted. Devoting a whole subcarrier for sending the reconciliation data for the case of $N = 12$ and $\beta = 0.0001$ is almost equivalent of losing $1/12$ of the achievable rate.

This can be seen in Fig. 2.7a and 2.8a where $N = 12$. When the SNR is high (See Fig. 2.7a),

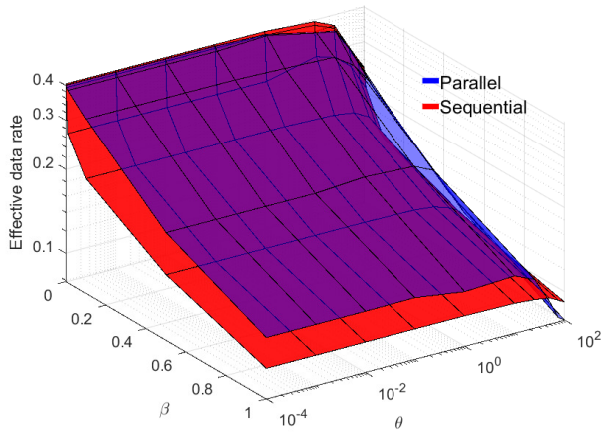

Figure 2.8: **a)** Effective data rate achieved by the parallel heuristic approach and the sequential approach when $N = 12$, SNR= 0.2 dB and $\kappa = 2$.

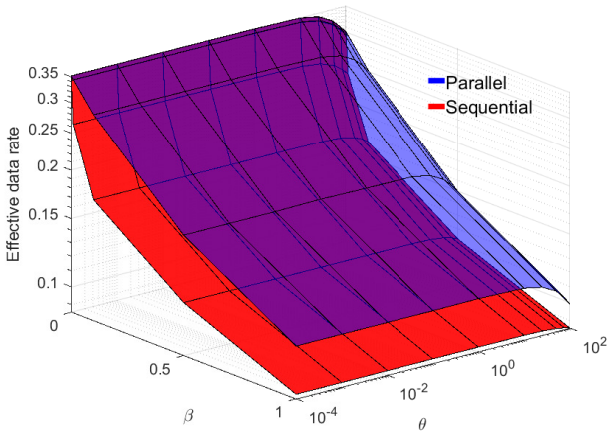

Figure 2.8: **b)** Effective data rate achieved by the parallel heuristic approach and the sequential approach when $N = 64$, SNR= 0.2 dB and $\kappa = 2$.

as discussed, this effect is mostly noticeable for large values of θ and small values of β[7], whereas for small values of β and θ both algorithms perform nearly identically. A similar trend can be seen at the low SNR regime in Fig. 2.8a. However, at a low SNR the sequential approach has a lower effective data rate. This happens because at high SNR levels each reconciliation frame will contain more information and hence more data frames will follow. Therefore, at the low SNR regime, the reconciliation information received will decrease, hence less data can be sent afterwards. This does not affect the parallel approach. However, in both scenarios high SNR Fig. 2.7a and low SNR Fig. 2.8a, when β increases regardless of the value of θ the parallel approach always achieves higher *effective data rate* $E_{C,D}(\theta)$.

In the next case, when the total number of subcarriers is $N = 64$, illustrated in Fig. 2.7b and 2.8b, we see that the penalty of devoting a high part of the achievable effective capacity $E_C^{\text{opt}}(\theta)$ to reconciliation disappears and the heuristic parallel approach always achieves higher or identical *effective data rate* $E_{C,D}(\theta)$ compared to the sequential approach. This trend repeats for high and low SNR levels as given in Fig. 2.7b and 2.8b, respectively.

Now, we take a closer look and transform some specific cases from the 3D plots to two-dimensional graphs. In Fig. 2.9 we see the achieved *effective data rate* $E_{C,D}(\theta)$ given in (2.41), for different values of N and θ while the SNR=5 dB and $\kappa = 2$. Fig. 2.9a gives the achieved effective rate on set $\mathcal{D}$ for $N = 12$ and $\theta = 0.0001$ (relaxed delay constraint). Similarly to the case of long term average value of C_D we see that for small values of β the sequential approach achieves slightly higher effective data rate. As before, the increase of β results in more reconciliation frames M required in the sequential case. This effect is not seen in the parallel case and for high values of β it performs better.

Fig. 2.9b illustrates the case when $N = 12$ and $\theta = 100$ (very stringent delay constraint). Similarly to before, we can see that for small values of β the sequential approach performs better than the parallel. As discussed, the efficiency loss is caused by the fact that the devoted part of the total achievable effective capacity $E_C^{\text{opt}}(\theta)$ to reconciliation (syndrome communication) is more than what is required. However, a higher β leads to an increase in the reconciliation information that needs to be sent, and the rate of the subcarriers in set $\breve{\mathcal{D}}$ will be fully or almost fully utilised and the parallel approach shows better performance for these values.

In the next two Fig.: 2.9c and 2.9d we show the performance of the two algorithms for higher value of $N = 64$. It is easy to see that regardless of the value of θ and β both algorithms perform identical

[7] *i.e.,* that the ratio of reconciliation information to data is small as seen from (2.25))

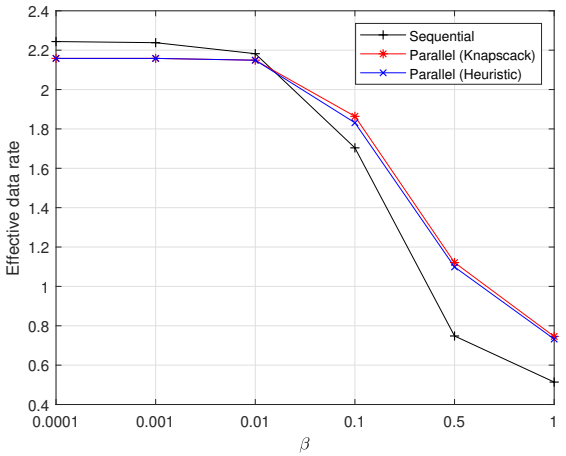

Figure 2.9: **a)** Effective data rate achieved by parallel and sequential approaches when $N = 12$, SNR= 5dB, $\theta = 0.0001, \kappa = 2$.

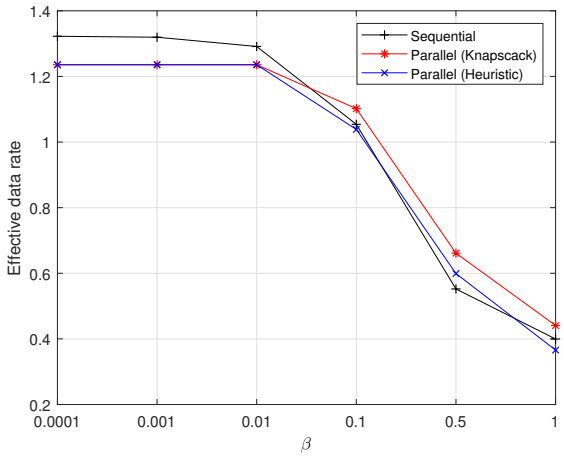

Figure 2.9: **b)** Effective data rate achieved by parallel and sequential approaches when $N = 12$, SNR= 5dB, $\theta = 100, \kappa = 2$.

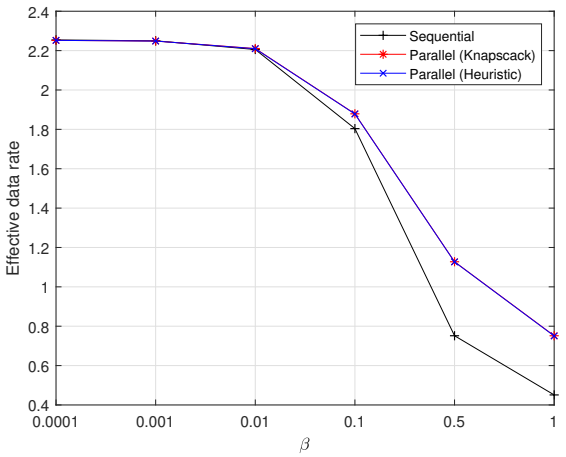

Figure 2.9: **c)** Effective data rate achieved by parallel and sequential approaches when $N = 64$, SNR= 5dB, $\theta = 0.0001, \kappa = 2$.

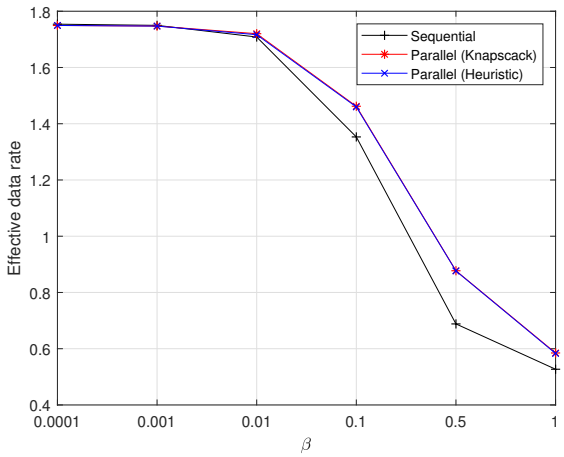

Figure 2.9: **d)** Effective data rate achieved by parallel and sequential approaches when $N = 64$, SNR= 5dB, $\theta = 100, \kappa = 2$.

or the parallel is better. In the previous case of $N = 12$ increasing θ might reduce the effectiveness of the parallel approach, however when $N = 64$ increasing θ does not incur such a penalty and the parallel is either identical to the sequential or outperforms it.

Another interesting fact from Fig. 2.9 is that looking at the parallel approach, it can easily be seen that in all cases the heuristic approach almost always performs as well as the optimal knapsack solution. The case of small values of θ is similar to the one when we work with long term average rate and choosing the best subcarriers for data transmission works as well as the optimal knapsack solution. Interestingly, *Algorithm 1* works well for high values of θ, too. This can be explained by the fact that when θ increases the rate on all of the subcarriers becomes similar and switching the subcarriers in set $\mathcal{D}$ does not incur high penalty.

2.8 Conclusions

In this work we discussed the possibility of using SKG in conjunction with PUF authentication protocols, illustrating this can greatly reduce the authentication and key generation latency compared to traditional mechanisms. Furthermore, we presented an AE scheme using SKG and a resumption protocol which further contribute to the system's security and latency reduction, respectively.

In addition, we explored the possibility of pipelining encrypted data transfer and SKG in a Rayleigh BF-AWGN environment. We investigated the maximization of the data transfer rate in parallel to performing SKG. We took into account imperfect CSI measurements and the effect of order statistics on the channel variance. Two scenarios were differentiated in our study: i) the optimal data transfer rate was found under power and security constraints, represented by the system parameters β and κ, which represent the minimum ratio of SKG rate to data rate and the maximum ratio of SKG rate to reconciliation rate; ii) by adding a delay constraint, represented by parameter θ, to the security and power constraint we found the optimal *effective data rate*.

To finalise our study we illustrated through numerical comparisons the efficiency of the proposed parallel method, in which SKG and data transfer are inter-weaved, to a sequential method where the two operations are done separately. The results of the two scenarios showed that in most of the cases the performance of both methods, parallel and sequential, is either equal or the parallel performs better. As the possible advantage of using the sequential is small and only applies in particular scenarios, we recommend the parallel scheme as a universal mechanism for general protocol design, when latency is an issue. Furthermore, a significant result is that although the optimal subcarrier scheduling is an NP hard $0-1$ knapsack problem, it can be solved in linear time using a simple heuristic algorithm with virtually no loss in performance.

References

[1] A. Mukherjee. Physical-layer security in the internet of things: Sensing and communication confidentiality under resource constraints. *Proceedings of the IEEE*, 103(10):1747–1761, Oct 2015.

[2] A. Yener and S. Ulukus. Wireless physical-layer security: Lessons learned from information theory. *Proceedings of the IEEE*, 103(10):1814–1825, Oct 2015.

[3] D. Karatzas, A. Chorti, N. M. White, and C. J. Harris. Teaching old sensors new tricks: Archetypes of intelligence. *IEEE Sensors Journal*, 7(5):868–881, May 2007.

[4] 3GPP TR 33.825 V0.3.0, Study on the Security for 5G URLLC (Release 16). 3rd Generation Partnership Project; Technical Specification Group Services and System Aspects, available online https://www.3gpp.org/ftp/Specs/archive/33_series/33.825/.

[5] Arsenia Chorti, Camilla Hollanti, Jean-Claude Belfiore, and H. Vincent Poor. Physical layer security: A paradigm shift in data confidentiality. *Lecture Notes in Electrical Engineering*, 358, 01 2016.

[6] A. Chorti, K. Papadaki, and H. V. Poor. Optimal power allocation in block fading channels with confidential messages. *IEEE Transactions on Wireless Communications*, 14(9):4708–4719, Sep. 2015.

[7] A. Chorti, S. M. Perlaza, Z. Han, and H. V. Poor. On the resilience of wireless multiuser networks to passive and active eavesdroppers. *IEEE Journal on Selected Areas in Communications*, 31(9):1850–1863, Sep. 2013.

[8] A. Chorti and H. V. Poor. Achievable secrecy rates in physical layer secure systems with a helping interferer. In *2012 International Conference on Computing, Networking and Communications (ICNC)*, pages 18–22, Jan 2012.

[9] M. Mitev, A. Chorti, and M. Reed. Subcarrier scheduling for joint data transfer and key generation schemes in multicarrier systems. In *2019 IEEE Global Communications Conference (GLOBECOM)*, pages 1–6, Dec 2019.

[10] M. Rodrigues I. Darwazeh Y. Kanaras, A. Chorti. An optimum detection for a spectrally efficient non orthogonal FDM system. In *Proc. 13th Int. OFDM WS*, pages 65–68, Aug 2008.

[11] A. Chorti and H. V. Poor. Faster than Nyquist interference assisted secret communication for OFDM systems. In *2011 Asilomar Conf. Signals, Systems and Computers (ASILOMAR)*, pages 183–187, Nov 2011.

[12] A. Chorti. Helping interferer physical layer security strategies for M-QAM and M-PSK systems. In *2012 46th Annual Conference on Information Sciences and Systems (CISS)*, pages 1–6, March 2012.

[13] Matti Latvaaho and Kari Leppänen. Key drivers and research challenges for 6G ubiquitous wireless intelligence. Published online by the University of Oulu, October 2019.

[14] U. M. Maurer. Secret key agreement by public discussion from common information. *IEEE Transactions on Information Theory*, 39(3):733–742, May 1993.

[15] R. Ahlswede and I. Csiszar. Common randomness in information theory and cryptography. i. secret sharing. *IEEE Transactions on Information Theory*, 39(4):1121–1132, July 1993.

[16] C. Ye, A. Reznik, and Y. Shah. Extracting secrecy from jointly gaussian random variables. In *2006 IEEE International Symposium on Information Theory*, pages 2593–2597, July 2006.

[17] Blaise Gassend, Dwaine Clarke, Marten van Dijk, and Srinivas Devadas. Silicon physical random functions. In *Proceedings of the 9th ACM Conference on Computer and Communications Security*, CCS '02, page 148–160, New York, NY, USA, 2002. Association for Computing Machinery.

[18] Ravikanth Pappu, Ben Recht, Jason Taylor, and Neil Gershenfeld. Physical one-way functions. *Science*, 297(5589):2026–2030, 2002.

[19] Roel Maes and Ingrid Verbauwhede. *Physically Unclonable Functions: A Study on the State of the Art and Future Research Directions*, pages 3–37. 10 2010.

[20] H. Schotten A. Weinand, M. Karrenbauer. Security solutions for local wireless networks in control applications based on physical layer security. 51:32–39, 2018.

[21] A. Mukherjee, S. A. A. Fakoorian, J. Huang, and A. L. Swindlehurst. Principles of physical layer security in multiuser wireless networks: A survey. *IEEE Communications Surveys Tutorials*, 16(3):1550–1573, Third 2014.

[22] Arsenia Chorti. A study of injection and jamming attacks in wireless secret sharing systems. In *in Proc. Workshop on Communication Security (WCS)*, 03 2017.

[23] The Transport Layer Security (TLS) Protocol Version 1.3. RFC 8446 (2018). Rescorla, E., available online https://rfc-editor.org/rfc/rfc8446.txt.

[24] Nimrod Aviram, Kai Gellert, and Tibor Jager. Session resumption protocols and efficient forward security for tls 1.3 0-rtt. Cryptology ePrint Archive, Report 2019/228, 2019. https://eprint.iacr.org/2019/228.

[25] Mihir Bellare and Chanathip Namprempre. Authenticated encryption: Relations among notions and analysis of the generic composition paradigm. *J. Cryptol.*, 21(4):469–491, September 2008.

[26] Ted Krovetz and Phillip Rogaway. The software performance of authenticated-encryption modes. In *FSE, Lecture Notes in Computer Science*, 2011.

[27] S. Koteshwara and A. Das. Comparative study of authenticated encryption targeting lightweight iot applications. *IEEE Design Test*, 34(4):26–33, Aug 2017.

[28] Dapeng Wu and R. Negi. Effective capacity: a wireless link model for support of quality of service. *IEEE Transactions on Wireless Communications*, 2(4):630–643, July 2003.

[29] Wenjie Che, Mitchell Martin, Goutham Pocklassery, Venkata K. Kajuluri, Fareena Saqib, and James F. Plusquellic. A privacy-preserving, mutual puf-based authentication protocol. *Cryptography*, 1:3, 2016.

[30] Blaise Gassend, Dwaine Clarke, Marten van Dijk, and Srinivas Devadas. Silicon physical random functions. In *Proceedings of the 9th ACM Conference on Computer and Communications Security*, CCS '02, page 148–160, New York, NY, USA, 2002. Association for Computing Machinery.

[31] C. Marchand, L. Bossuet, U. Mureddu, N. Bochard, A. Cherkaoui, and V. Fischer. Implementation and characterization of a physical unclonable function for IoT: A case study with the TERO-PUF. *IEEE Transactions on Computer-Aided Design of Integrated Circuits and Systems*, 37(1):97–109, Jan 2018.

[32] Jorge Guajardo, Sandeep S. Kumar, Geert-Jan Schrijen, and Pim Tuyls. FPGA intrinsic PUFs and their use for IP protection. In *Proceedings of the 9th International Workshop on Cryptographic Hardware and Embedded Systems*, CHES '07, page 63–80, Berlin, Heidelberg, 2007. Springer-Verlag.

[33] J. Aarestad, P. Ortiz, D. Acharyya, and J. Plusquellic. Help: A hardware-embedded delay puf. *IEEE Design Test*, 30(2):17–25, April 2013.

[34] Armin Babaei and Gregor Schiele. Physical unclonable functions in the internet of things: State of the art and open challenges. In *Sensors*, 2019.

[35] Pramod Maurya and Satya Bagchi. A secure PUF-based unilateral authentication scheme for RFID system. *Wireless Personal Communications*, 103, 05 2018.

[36] M. Yu, M. Hiller, J. Delvaux, R. Sowell, S. Devadas, and I. Verbauwhede. A lockdown technique to prevent machine learning on pufs for lightweight authentication. *IEEE Transactions on Multi-Scale Computing Systems*, 2(3):146–159, July 2016.

[37] Jeff Calhoun, Cyrus Minwalla, Charles Helmich, Fareena Saqib, Wenjie Che, and J. Plusquellic. Physical unclonable function (PUF)-based e-cash transaction protocol (PUF-Cash). *Cryptography*, 3:18, 07 2019.

[38] M. N. Aman, K. C. Chua, and B. Sikdar. Mutual authentication in iot systems using physical unclonable functions. *IEEE Internet of Things Journal*, 4(5):1327–1340, Oct 2017.

[39] Jeroen Delvaux, Roel Peeters, Dawu Gu, and Ingrid Verbauwhede. A survey on lightweight entity authentication with strong pufs. *ACM Comput. Surv.*, 48(2), October 2015.

[40] Suman Jana, Sriram Nandha Premnath, Mike Clark, Sneha K. Kasera, Neal Patwari, and Srikanth V. Krishnamurthy. On the effectiveness of secret key extraction from wireless signal strength in real environments. In *Proceedings of the 15th Annual International Conference on Mobile Computing and Networking*, MobiCom '09, page 321–332, New York, NY, USA, 2009. Association for Computing Machinery.

[41] Theodore Rappaport. *Wireless Communications: Principles and Practice*. Prentice Hall PTR, USA, 2nd edition, 2001.

[42] J. Wan, A. B. Lopez, and M. A. Al Faruque. Exploiting wireless channel randomness to generate keys for automotive cyber-physical system security. In *2016 ACM/IEEE 7th International Conference on Cyber-Physical Systems (ICCPS)*, pages 1–10, April 2016.

[43] B. Zan, M. Gruteser, and F. Hu. Key agreement algorithms for vehicular communication networks based on reciprocity and diversity theorems. *IEEE Transactions on Vehicular Technology*, 62(8):4020–4027, Oct 2013.

[44] Yicong Liu, Jiwu Jing, and Jun Yang. Secure underwater acoustic communication based on a robust key generation scheme. In *2008 9th International Conference on Signal Processing*, pages 1838–1841, Oct 2008.

[45] I. U. Zaman, A. B. Lopez, M. A. A. Faruque, and O. Boyraz. Physical layer cryptographic key generation by exploiting pmd of an optical fiber link. *Journal of Lightwave Technology*, 36(24):5903–5911, Dec 2018.

[46] D. Tian, W. Zhang, J. Sun, and C. Wang. Physical-layer security of visible light communications with jamming. In *2019 IEEE/CIC International Conference on Communications in China (ICCC)*, pages 512–517, Aug 2019.

[47] J. Zhang, T. Q. Duong, A. Marshall, and R. Woods. Key generation from wireless channels: A review. *IEEE Access*, 4:614–626, 2016.

[48] J. K. Tugnait, Lang Tong, and Zhi ding. Single-user channel estimation and equalization. *IEEE Signal Processing Magazine*, 17(3):17–28, May 2000.

[49] William C. Jakes and Donald C. Cox. *Microwave Mobile Communications*. Wiley-IEEE Press, 1994.

[50] H. Liu, Y. Wang, J. Yang, and Y. Chen. Fast and practical secret key extraction by exploiting channel response. In *2013 Proceedings IEEE INFOCOM*, pages 3048–3056, April 2013.

[51] Suhas Mathur, Wade Trappe, Narayan Mandayam, Chunxuan Ye, and Alex Reznik. Radio-telepathy: Extracting a secret key from an unauthenticated wireless channel. In *Proceedings of the 14th ACM International Conference on Mobile Computing and Networking*, MobiCom '08, page 128–139, New York, NY, USA, 2008. Association for Computing Machinery.

[52] S. T. Ali, V. Sivaraman, and D. Ostry. Eliminating reconciliation cost in secret key generation for body-worn health monitoring devices. *IEEE Transactions on Mobile Computing*, 13(12):2763–2776, Dec 2014.

[53] Suhas Mathur, Robert Miller, Alexander Varshavsky, Wade Trappe, and Narayan Mandayam. Proximate: Proximity-based secure pairing using ambient wireless signals. In *Proceedings of the 9th International Conference on Mobile Systems, Applications, and Services*, MobiSys '11, pages 211–224, New York, NY, USA, 2011. Association for Computing Machinery.

[54] M. Mitev, A. Chorti, E.V. Belmega, and M.J. Reed. Man-in-the-middle and denial of service attacks in wireless secret key generation. In *Proc. IEEE Global Commun. (GLOBECOM)*, Big Island, HI, 2019.

[55] Intrinsic-id company. https://www.intrinsic-id.com/sram-puf.

[56] ICTK holdings corporation. https://ictk-puf.com/puf-technology.

[57] A. Maiti, I. Kim, and P. Schaumont. A robust physical unclonable function with enhanced challenge-response set. *IEEE Transactions on Information Forensics and Security*, 7(1):333–345, Feb 2012.

[58] Monis Akhlaq, Baber Aslam, Muzammil A. Khan, and M. Noman Jafri. Comparative analysis of ieee 802.1x authentication methods. In *Proceedings of the 11th Conference on 11th WSEAS International Conference on Communications - Volume 11*, ICCOM'07, page 1–6, Stevens Point, Wisconsin, USA, 2007. World Scientific and Engineering Academy and Society (WSEAS).

[59] A. Chiornită, L. Gheorghe, and D. Rosner. A practical analysis of EAP authentication methods. In *9th RoEduNet IEEE International Conference*, pages 31–35, June 2010.

[60] Urbi Chatterjee, Rajat Chakraborty, and Debdeep Mukhopadhyay. A puf-based secure communication protocol for iot. *ACM Transactions on Embedded Computing Systems*, 16:1–25, 04 2017.

[61] M. N. Aman, M. H. Basheer, and B. Sikdar. Two-factor authentication for IoT with location information. *IEEE Internet of Things Journal*, 6(2):3335–3351, April 2019.

[62] M. H. Mahalat, S. Saha, A. Mondal, and B. Sen. A PUF based light weight protocol for secure WiFi authentication of IoT devices. In *2018 8th International Symposium on Embedded Computing and System Design (ISED)*, pages 183–187, Dec 2018.

[63] An Braeken. Puf based authentication protocol for iot. *Symmetry*, 10:352, 08 2018.

[64] Y. Yilmaz, S. R. Gunn, and B. Halak. Lightweight PUF-based authentication protocol for IoT devices. In *2018 IEEE 3rd International Verification and Security Workshop (IVSW)*, pages 38–43, July 2018.

[65] S. Ahmad, A. H. Mir, and G. R. Beigh. Latency evaluation of extensible authentication protocols in WLANs. In *2011 Fifth IEEE International Conference on Advanced Telecommunication Systems and Networks (ANTS)*, pages 1–5, Dec 2011.

[66] P. Gope and B. Sikdar. Lightweight and privacy-preserving two-factor authentication scheme for IoT devices. *IEEE Internet of Things Journal*, 6(1):580–589, Feb 2019.

[67] A. Ometov, P. Masek, L. Malina, R. Florea, J. Hosek, S. Andreev, J. Hajny, J. Niutanen, and Y. Koucheryavy. Feasibility characterization of cryptographic primitives for constrained (wearable) iot devices. In *2016 IEEE International Conference on Pervasive Computing and Communication Workshops (PerCom Workshops)*, pages 1–6, March 2016.

[68] J. Cho and W. Sung. Efficient software-based encoding and decoding of bch codes. *IEEE Transactions on Computers*, 58(7):878–889, July 2009.

[69] C. Chen and M. A. Jensen. Secret key establishment using temporally and spatially correlated wireless channel coefficients. *IEEE Transactions on Mobile Computing*, 10(2):205–215, Feb 2011.

[70] J. Zhang, A. Marshall, R. Woods, and T. Q. Duong. Efficient key generation by exploiting randomness from channel responses of individual ofdm subcarriers. *IEEE Transactions on Communications*, 64(6):2578–2588, June 2016.

[71] J. Zhang, B. He, T. Q. Duong, and R. Woods. On the key generation from correlated wireless channels. *IEEE Communications Letters*, 21(4):961–964, April 2017.

[72] C. Saiki and A. Chorti. A novel physical layer authenticated encryption protocol exploiting shared randomness. In *2015 IEEE Conference on Communications and Network Security (CNS)*, pages 113–118, Sep. 2015.

[73] Q. Wang, H. Su, K. Ren, and K. Kim. Fast and scalable secret key generation exploiting channel phase randomness in wireless networks. In *2011 Proceedings IEEE INFOCOM*, pages 1422–1430, April 2011.

[74] C. Ye, S. Mathur, A. Reznik, Y. Shah, W. Trappe, and N. B. Mandayam. Information-theoretically secret key generation for fading wireless channels. *IEEE Transactions on Information Forensics and Security*, 5(2):240–254, June 2010.

[75] Christopher Huth, Ren Guillaume, Thomas Strohm, Paul Duplys, Irin Ann Samuel, and Tim Gneysu. Information reconciliation schemes in physical-layer security. *Comput. Netw.*, 109(P1):84–104, November 2016.

[76] Li Guyue, Zheying Zhang, Yi Yu, and Aiqun Hu. A hybrid information reconciliation method for physical layer key generation. *Entropy*, 21:688, 07 2019.

[77] P. Treeviriyanupab, P. Sangwongngam, K. Sripimanwat, and O. Sangaroon. BCH-based Slepian-Wolf coding with feedback syndrome decoding for quantum key reconciliation. In *2012 9th International Conference on Electrical Engineering/Electronics, Computer, Telecommunications and Information Technology*, pages 1–4, May 2012.

[78] J. Etesami and W. Henkel. LDPC code construction for wireless physical-layer key reconciliation. In *2012 1st IEEE International Conference on Communications in China (ICCC)*, pages 208–213, Aug 2012.

[79] C. H. Bennett, G. Brassard, C. Crepeau, and U. M. Maurer. Generalized privacy amplification. *IEEE Transactions on Information Theory*, 41(6):1915–1923, Nov 1995.

[80] Furui Zhan and Nianmin Yao. On the using of discrete wavelet transform for physical layer key generation. *Ad Hoc Networks*, 64:22 – 31, 2017.

[81] M. Bloch, J. Barros, M. R. D. Rodrigues, and S. W. McLaughlin. Wireless information-theoretic security. *IEEE Transactions on Information Theory*, 54(6):2515–2534, June 2008.

[82] M. Mitev, A. Chorti, and M. Reed. Optimal resource allocation in joint secret key generation and data transfer schemes. In *2019 15th International Wireless Communications Mobile Computing Conference (IWCMC)*, pages 360–365, June 2019.

[83] E. V. Belmega and A. Chorti. Protecting secret key generation systems against jamming: Energy harvesting and channel hopping approaches. *IEEE Transactions on Information Forensics and Security*, 12(11):2611–2626, Nov 2017.

[84] M. Medard. The effect upon channel capacity in wireless communications of perfect and imperfect knowledge of the channel. *IEEE Transactions on Information Theory*, 46(3):933–946, May 2000.

[85] Hong-Chuan Yang and Mohamed-Slim Alouini. *Order Statistics in Wireless Communications: Diversity, Adaptation, and Scheduling in MIMO and OFDM Systems*. Cambridge University Press, USA, 1st edition, 2011.

[86] Silvano Martello and Paolo Toth. *Knapsack Problems: Algorithms and Computer Implementations*. John Wiley and Sons, Inc., USA, 1990.

[87] H. Kellerer, U. Pferschy, and D. Pisinger. *Knapsack Problems*. Springer, Berlin, Germany, 2004.

[88] Vijay V. Vazirani. *Approximation Algorithms*. Springer-Verlag, Berlin, Heidelberg, 2001.

[89] Cheng-Shang Chang. Stability, queue length, and delay of deterministic and stochastic queueing networks. *IEEE Transactions on Automatic Control*, 39(5):913–931, May 1994.

[90] J. Gärtner. On large deviation from invariant measure. *Theory Prob. Appl.*, 22:24–39, 1977.

[91] Richard Ellis. Large deviations for a general class of random vectors. *The Annals of Probability*, 12, 02 1984.

[92] T. Abrao, S. Yang, L. D. H. Sampaio, P. J. E. Jeszensky, and L. Hanzo. Achieving maximum effective capacity in ofdma networks operating under statistical delay guarantee. *IEEE Access*, 5:14333–14346, 2017.

[93] Stephen Boyd and Lieven Vandenberghe. *Convex Optimization*. Cambridge University Press, USA, 2004.

Chapter 3

Application of Change Point Analysis in Edge Resource Allocation and Intrusion Detection

3.1 Introduction

Edge computing emerges as a critical element in future networks, provisioning storage and computation resources in the proximity of end devices to provide low latency services. The joint allocation and management of communication, computing and storage resources will improve the quality of service (QoS) and user experience, especially for various delay-sensitive applications. At the same time, software-defined networking (SDN) is a technology that can help bridge the gap when combining Edge computing and traditional Clouds. For example, the SDN controller can make decisions on whether tasks should be uploaded and processed in the Cloud or at the Edge. The advancement of Edge computing poses many challenges, e.g., in the deployment and management of distributed resources; in parallel, SDNs are prone to new security threats due to the separation of the control and data planes. In this Chapter, we first focus on content distribution at the Edge using change point (CP) analysis. Next, motivated by the good performance of the developed algorithms, we investigate their application in intrusion detection in software defined wireless sensor networks (SDWSNs), showcasing that the wide range of applications that can be covered.

Beginning with resource allocation at the Edge servers, we propose a novel and flexible approach exploiting popular virtualization technologies, such as unikernels [1] or containers; as a use case we consider video content distribution, which accounts for more than 70% of the global IP traffic. The core idea in our proposal is that virtual servers could hold individual video content and could be "live" for as long as there is corresponding demand; in case of an increase in demand, more replicas of the virtual servers could be put up, or alternatively put down if the demand dies off. We note that bringing up or down a unikernel is typically very fast, with reported numbers for the boot time as little as 20 milliseconds [2].

In this context, due to high volatility in the respective demand, it is important for video content delivery infrastructures to rapidly detect and respond to changes in "content popularity" dynamics. We explore the employment of on-line CP analysis to implement real-time, autonomous and low-complexity video content popularity detection. Our proposal, denoted as *real-time change point detector (RCPD)*, estimates the existence, the number and the direction of changes on the average number of video visits by combining: (i) off-line and on-line CP detection algorithms; (ii) an improved time-series segmentation heuristic for the reliable detection of multiple CPs; and (iii) two algorithms for the identification of the direction of changes. The proposed detector is validated against synthetic data, as well as a large database of real YouTube video visits.

Finally, we note that customarily CP analysis is employed in the detection of anomalies in times

series. Therefore, as a natural extension of this work, we further consider the application of the RCPD for anomaly detection in SDWSNs. SDN is a promising technology to overcome many challenges in wireless sensor networks (WSN), particularly with respect to flexibility and reuse. Notably, it is now argued that SDN and related technologies should be integrated to facilitate the management and operations of Edge servers and various IoT devices [3]. Conversely, the centralization and the planes' separation turn SDNs vulnerable to new security threats in the general context of distributed denial of service (DDoS) attacks, which carry over to SDWSNs. State-of-the-art approaches to identify DDoS do not always take into consideration restrictions in typical WSNs, e.g., computational complexity and power constraints, while further performance improvement is always a target. Our objective in this study is to propose a lightweight but very efficient DDoS attack detection approach using the RCPD.

3.2 Contributions and Chapter Organization

3.2.1 CP Analysis in Resource Allocation

Video content is projected to account for 82% of the global Internet traffic by 2020, significantly increased from 72% in 2016 [4]. In parallel, novel emerging networking, Cloud and Edge computing paradigms with significant elasticity capabilities appeared recently, e.g., SDNs [5], Cloud orchestration proposals [6] and content distribution networks (CDNs) [7]. These advances offer the means to respond quickly to changes in content popularity dynamics with appropriate adaptations, e.g., in terms of efficient server resource allocation schemes, load balancing or content caching. As a result, the early detection of changes in content popularity [8], [9] is proving a highly important topic and can have a significant impact on the network traffic and the utilization of servers.

So far, the vast majority of research efforts have focused on the *prediction* of content popularity dynamics, as opposed to their *real time detection*, which is the focus of this study. There is a multitude of reasons as to why the precision of even state-of-the-art prediction algorithms can be impaired. A variety of factors – both from the digital and the physical world – can influence the users' Internet surfing behavior, e.g., [8]: (i) the quality, type (e.g., commercial or user-provided) and life-time of content; (ii) its relevance to users and physical events; (iii) the social interactions between users; and (iv) the content promotion strategies involved. Importantly, mid-term and long-term content popularity prediction [10] – and corresponding adaptations in the network or cloud environment – can prove highly inaccurate [11] and thus result in sub-optimal service planning, provisioning, and utilization of resources or violation of service level agreements.

In this work, to address the aforementioned shortcomings of the commonly employed prediction algorithms, we propose a corresponding detector, referred to as the "real-time change point detector" (RCPD). The RCPD is compatible with modern, flexible networking and Cloud approaches, that are highly adaptive and can respond to short-term network dynamics. With accurate, on-line content popularity detection, discrepancies between inaccurate predictions and actual changes can be alleviated. The RCPD is real-time, lightweight, accurate and is parameterized autonomously by analyzing historical data.

In the RCPD, we employ the CP detection theory and algorithms; their suitability is confirmed against a large number of synthetic as well as real YouTube video datasets. In this contribution, the early detection of changes in the average content popularity is addressed with a novel CP detection methodology, consisting of a training phase, using historical data, and, an on-line phase. In the training phase, we employ a modified off-line CP detection scheme to configure the on-line (sequential) algorithm's parameters. This approach is shown to greatly improve the accuracy of the on-line detector, as in essence, the algorithm parameterization is not arbitrary but rather extracted from corresponding historical data. To the best of our knowledge, this was the first proposal in the literature on combining retrospective (off-line) and sequential (on-line) CP detection schemes in a single algorithm operating autonomously (i.e., without manual configuration of parameters).

Besides that, our approach complements the off-line scheme with an improved time-series segmen-

tation heuristic for the detection of multiple CPs. Furthermore, we propose two possible variations for the on-line CP algorithm, the first based on the standard cumulative sum (CUSUM) procedure [12] and the second on the ratio-type CUSUM procedure [13][1]. Additionally, we introduce two alternative indicators to detect the direction of changes: the first one is directly derived from the statistical test of the on-line CP procedure, while the second is based on a modified exponential moving average filter, extensively used in econometrics. As discussed in Sections 3.4 and 3.5, the RCPD combines all the above mentioned algorithmic elements, and is based on sufficiently general and convenient assumptions. Moreover, unlike other approaches e.g., [14], we employ methods that allow dependence between observations (in the form of $t-$dependence), leading to more realistic assumptions for the statistical structure of the content visits.

We evaluate the proposed detector and its individual algorithmic components (i.e., the off-line / on-line test statistics, the time-series segmentation algorithm and the trend indicator), over synthetic and real YouTube content views data. Our experiments using synthetic data, generated by an autoregressive moving average (ARMA) filter, demonstrate:

- The superior performance of the proposed time-series segmentation heuristic over the standard approach, improving the true alarm rates by up to 43%.
- The ability of the two proposed trend indicators to identify the direction of estimated changes, with successful identification rates exceeding 99%, in all cases.
- The RCPD performance; the true alarm rates surpass 94% for medium / large changes in the mean number of content views, while the corresponding CP identification lag ranges between 10 to 20 instances, confirming the real-time operation of the detector. On the other hand, the RCPD achieves very small false alarm rates, well within the limits of the statistical error specified by the chosen significance level of the CP algorithms.

Furthermore, our tests on real YouTube content views datasets show that:

- YouTube video views match the underlying assumptions of the RCPD, i.e., the content popularity time-series datasets can be modeled as t-dependent.
- The RCPD can detect CPs in more than 70% of the videos in our dataset, implying a sufficiently high number of content popularity changes and the suitability of the CP theory framework for content popularity detection.
- The successful CP direction identifications exceed 91%, i.e., the proposed trend indicators work for real data.
- The average dynamic time warping (DTW) distance [15], [16] between the identified CPs and a benchmark off-line algorithm was estimated to be 52 time instances on average, showcasing the rapid responsiveness of the RCPD.
- The overall processing cost of the RCPD is very low; notably, it took less than one second to process 882 videos on a typical personal computer (PC).

As a proof-of-concept, we demonstrate the applicability of the proposed algorithm in a real load balancing scenario. We provide a set of measurements showcasing improvements in terms of the clients' connectivity time to download specific content, without a significant impact on the utilization of the content servers. This is achieved due to the deployment of additional content caches, an event triggered by the output of the proposed RCPD detector.

[1]The advantage of ratio-type CUSUM is that it does not require the estimation of long-run covariance (variance) matrices, which is the case for the standard CUSUM method.

3.2.2 CP Analysis for Anomaly Detection in SDWSNs

Next, we explore the application of the RCPD for anomaly (intrusion) detection in SDWSNs. The SDN paradigm was devised to simplify network management, avoid configuration errors and automate infrastructure sharing in wired networks [17]. The aforementioned benefits motivated the discussion of combining SDN and WSNs as a solution to many WSN challenges, in particular concerning flexibility and resource reuse [18]. This combination is referred to as SDWSN. The SDWSN approach decouples the control plane from the data plane and centralizes the control decisions; its main characteristic is the ability to program the network operation dynamically [19]. Recent results show that SDWSNs can perform as well as the IPv6 routing protocol for low-power and lossy networks (RPL) [20].

On the other hand, the SDN centralization and the planes' separation turn the network vulnerable to new security threats (explained in Section 3.9.1), a property that is inadvertently passed on to SDWSNs. Shielding SDNs from these vulnerabilities has already attracted a lot of attention in the literature with proposals to implement attack detection in IoT networks using SDN. Overall, in the case of SDWSNs, due to the resource constraints of the nodes, most of the security mechanisms designed for non-resource constrained SDNs have to be adapted or redesigned. This is one of the major challenges for SDWSN security.

Considering the limitations of previous works, our main objective is to propose a mechanism for DDoS detection with, i) a high detection rate, and, ii) low complexity, so that it would be suitable for "restricted" networks. To this end, we propose the employment of the RCPD [21] [22]. We study two DDoS attacks: a false data flow forwarding (FDFF) attack, and a false neighbor information (FNI) attack, chosen to illustrate the proposed algorithm's capabilities in the case of specific SDWSN vulnerabilities that exhibit largely different behavior. Both attacks are explained in Section 3.9.1. We have tested our approach on the IT-SDN framework[2] [20] and our results show that we can detect these attacks with a detection rate close to 100%, improving the state of the art; importantly, it is further possible to gain insight regarding the *type of the attack*, based on the metric that provides the quickest detection, a feature, that to the best of our knowledge, breaks new ground in the domain of DDoS analysis for SDWSNs.

3.2.3 Chapter Organization

The rest of the Chapter is organized as follows. In Section 3.3 we provide a comprehensive literature review of related topics. In Section 3.4, we present the off-line (training) phase of the RCPD algorithm, while the on-line phase is discussed in Section 3.5. In Section 3.6, we present four experiments over synthetic video content data, providing an extensive validation of the RCPD and its subroutines, while in Section 3.7, we discuss corresponding experiments using a database of real YouTube video views. In Section 3.8, we demonstrate the load balancing gains achieved through the use of the RCPD, in a realistic content provisioning scenario.

Moving to intrusion detection in SDWSNs, Section 3.9.1 illustrates the FDFF and FNI attacks and their impact on the network performance. Experimental methods are presented in Section 3.9.3 and results on intrusion detection using the RCPD are presented in Section 3.10.

Finally, Section 3.11 concludes the Chapter.

3.3 Related Works

In this Section, we discuss how this work relates to the literature of video content popularity prediction, and, anomaly detection in general and in SDNs and SDWSNs in particular.

The topic of content popularity attracted a lot of attention in recent years, because of its importance in a number of applications, such as network dimensioning (e.g., capacity planning or scaling of resources), on-line marketing (e.g., advertising, recommendation systems) or real-world outcome

[2] http://www.larc.usp.br/users/cbmargi/www/it-sdn/

prediction (e.g., analysis of economical trends) [8]. The main approaches used for content popularity estimation can be categorized as: (i) cumulative growth studies, estimating the "amount of attention" from the publication instance to the prediction moment [9]; (ii) temporal analysis approaches, i.e., how content visits evolve over time [23]; and (iii) clustering methods of content with similar popularity trends [10]. We note that many content popularity studies consider the aggregate behavior of a particular content, e.g., [9], [23], whereas we study the real-time behavior of video views time-series. In addition, studies using clustering methods [10] are based on content popularity prediction and adopt parametric models, unlike the RCPD algorithm that is non-parametric.

To the best of our knowledge, our conference paper [24] is the first in the literature proposing CP techniques [25] for content popularity detection. The RCPD algorithm falls into the general category of anomaly detection [26]; in essence, we assume that no changes in popularity constitutes the normal behavior of video content and search for deviations from this behavior. Non-parametric anomaly detection has typically been considered for the detection of abnormalities in the network traffic. As an example, in [27] an algorithm was proposed based on the Shiryaev-Roberts procedure for anomaly detection in computer network traffic. In [28] and [29], CUSUM based approaches were introduced for the detection of SYN attacks.

As opposed to previous content popularity prediction works, in this Chapter we introduce a novel CP detection methodology that provides accurate, lightweight, autonomous and on-line CP detection of content popularity. We formulate the detection of a change in the average content popularity as a statistical hypothesis test and employ non-parametric procedures to avoid a particular distribution assumption (such as a specific copula model). This context ensures low convergence time since it avoids estimating a large number of model parameters and restrictive assumptions that may not match the structure of the time-series. Furthermore, we avoid problems of parametric models that require parameters' fitting and selection, which become challenging as new data become available. In the proposed RCPD algorithm, an off-line phase specifies important parameters for the on-line phase; these parameters are re-evaluated dynamically after a detected CP. Our load-balancing experiments, elaborated in [7], demonstrate the RCPD's behavior in a real test-bed deployment.

Up to now there are only a handful of proposals addressing the challenges of new flexible networking and Cloud architectures accounting for content popularity. Exceptions include [30] in which a machine learning approach to content popularity prediction is applied for a Fog radio access network (RAN) environment, and, our recent papers [7] and [24]. In [7], the algorithm – outlined in [24] and presented extensively here – is integrated into an elastic content distribution network (CDN) framework based on lightweight Cloud capabilities using Unikernels. [7] focuses on the platform details rather than on the CP algorithm; it confirms experimentally the suitability of the latter for relevant flexible network and cloud architectures. A detailed description of the proposed CP detection algorithm is presented in the following Sections, along with a rich set of validation results.

Further examples of parametric anomaly detection methods include [31], in which a bivariate sequential generalized likelihood ratio test (LRT) was proposed, accounting for the packet rate – assumed to follow a Poisson distribution – and the packet size – assumed to follow a normal distribution. Other parametric anomaly detection approaches assume a particular underlying process for the normal behavior and search for anomalies on the residuals of the process. For example, in [32], Kalman filtering is combined with several CP methods, such as CUSUM and LRT, to detect anomalies in origin-destination flows. In [33], traffic flows (in the form of TCP's finite state machine), are modeled using Markov chains and an anomaly detection mechanism based on the generalized LRT algorithm is developed.

On the other hand, looking at existing literature in SDN anomaly detection, the authors in [34] proposed *softhings*, an SDN-based IoT framework with security support. The framework was developed for OpenFlow [19], which, however, can be a limiting factor for its use in networks composed of low-end nodes. The use of support vector machines (SVM) was proposed to detect control plane attacks; it was shown that a detection rate of around 96% and 98% could be achieved. The algorithm was tested in Mininet, simulating scenarios with only five nodes and considering one node as attacker. Furthermore

Yin *et al.* [35] developed the framework SD-IoT, which included a security system for DDoS attacks detection, based on the difference of packets received by the controller. The difference was calculated using the *cosine similarity* method. This mechanism was devised for networks where all the nodes had periodic communication with the controller, which could be not optimal for very "restricted" networks with low-end nodes. The authors tested their proposal through simulations using Mininet. The network size was not explicitly specified, but can be inferred to be around 50 to 60 nodes.

Furthermore, Wang *et al.* [36] proposed an SDWSN trust management and routing mechanism. They compared their proposal to SDN-WISE when both networks were under attack. The focus of the work was on the selective forwarding attacks and new flow requests. The first attack applied to any type of WSNs, while the second was specific to SDNs. The mechanism was tested in simulations with 100 nodes, varying the number of attackers between 5 and 20. Their results showed an attack detection rate between 90% and 96% when 5 nodes were attackers, and between 60% and 79% when 20 nodes were attackers. Compared to these previous works, our proposal for the employment of the RCPD is SDWSN anomaly detection has the advantages of being i) lightweight, ii) fast and iii) highly accurate as will be demonstrated in Sections 3.9 to 3.10.

We begin the description of the RCPD by first elaborating on the off-line and on-line phases in Sections 3.4 and 3.5 respectively, where we also provide the corresponding pseudo-code.

3.4 Training (Off-line) Phase

In this Section, the training phase of the algorithm is discussed and the fundamental components of the off-line scheme are presented. We note that standard off-line CP schemes can only detect a single CP. To address the issue of detection of multiple CPs, we modify the basic algorithm with a novel time-series segmentation heuristic, that belongs to the family of binary segmentation algorithms.

3.4.1 Basic Off-line Approach

Let $\{X_n : n \in \mathbb{N}\}$ be a sequence of r- dimensional random vectors (r.v.). The first dimension represents the number of views for a specific video content within a time period $n \in \{1, \ldots, N\}$, while the other dimensions could be optionally used to represent other content popularity features, such as likes, comments, etc. We assume that $X_1, ..., X_N$ can be written as,

$$X_n = \mu_n + Y_n, \quad 1 \leqslant n \leqslant N \tag{3.1}$$

where $\{\mu_n : n \in \mathbb{N}\}$ is the mean value of video visits, $\{Y_n : n \in \mathbb{N}\}$ a random component with zero mean $\mathbb{E}[Y_n] = 0$ and positive definite covariance matrix, $\mathbb{E}\left[Y_n Y_n^T\right] = \Sigma$, while $\mathbb{E}[\cdot]$ denotes expectation. We further assume that the time-series is t-dependent, implying that for $t_1, t_2, t \in \mathbb{N}$, Y_{t_1} is independent of Y_{t_2} if $|t_1 - t_2| > t$.

The model in (3.34) and the underlying assumption of $t-$dependence are in agreement with statistical characterizations of the distribution of visits, which have been shown in numerous analyses to follow either a Zipf [37] or a Zipf-Mandelbrot [38] distribution for both commercial and user-generated content. Furthermore, it is confirmed in the real YouTube datasets used in the present work through the evaluation of the time-series's Hurst exponents, as will be discussed in Section 3.7.1.

The off-line analysis tests the constancy (or not) of the mean values up to the current time N. Hence, we define the following null hypothesis of constant mean,

$$H_0 : \quad \mu_1 = \ldots = \mu_N,$$

against the alternative,

$$H_1 : \quad \mu_1 = \ldots = \mu_{k^*_{off}} \neq \mu_{k^*_{off}+1} = \ldots = \mu_N,$$

indicating that the mean value changed at the unknown (time) point $k^*_{off} \in \{1, \ldots, N\}$.

Considering (3.34) and the corresponding assumptions for the stochastic process X_n, we develop a non-parametric CUSUM test statistic following [39]. The test statistic TS_{off}, can be viewed as a max-type procedure,

$$TS_{off} = \max_{1 \leqslant n \leqslant N} C_n^T \widehat{\Omega}_N^{-1} C_n, \tag{3.2}$$

where the parameter C_n is the retrospective CUSUM detector,

$$C_n = \frac{1}{\sqrt{N}} \left(\sum_{i=1}^{n} X_i - n\overline{X}_{1,N} \right), \tag{3.3}$$

while $\overline{X}_{1,N} = \frac{1}{N} \sum_{i=1}^{N} X_i$ denotes the sample mean. $\widehat{\Omega}_N$ represents a suitable estimator of the long-run covariance Ω, where

$$\Omega = \sum_{i=-\infty}^{\infty} \mathbf{Cov} \left(X_n X_{n-i} \right). \tag{3.4}$$

The estimator should satisfy,

$$\widehat{\Omega}_N \xrightarrow{P} \Omega \tag{3.5}$$

where $\xrightarrow{P}$ denotes convergence in probability.

Several estimators have been proposed in the literature that satisfy (3.5), including kernel-based [40], bootstrap-based [41], etc. Considering our requirement for real-time detection (low computational time), a kernel-based estimator is more suitable; in this context, we employ the Bartlett estimator, so that

$$\widehat{\Omega}_N = \widehat{\Sigma}_0 + \sum_{w=1}^{W} k_{BT} \left(\frac{w}{W+1} \right) \left(\widehat{\Sigma}_w + \widehat{\Sigma}_w^T \right), \tag{3.6}$$

which satisfies (3.5), while the function $k_{BT}(.)$ corresponds to the Bartlett weight,

$$k_{BT}(x) = \begin{cases} 1 - |x|, & \text{for } |x| \leqslant 1 \\ 0, & \text{otherwise} \end{cases}, \tag{3.7}$$

and $\widehat{\Sigma}_w$ denotes the empirical auto-covariance matrix for lag w,

$$\widehat{\Sigma}_w = \frac{1}{N} \sum_{n=w+1}^{N} \left(X_n - \overline{X} \right) \left(X_{n-w} - \overline{X} \right)^T. \tag{3.8}$$

Finally, we chose $W = \log_{10}(N)$ as in [40].

The long-run covariance is involved in the test statistic to incorporate the dependence structure of the r.v. into the statistical analysis, through the integration of second order statistical properties. This approach is suitable for the targeted context since we avoid a restrictive assumption for the dependence structure of the observations.

Going back to the basic question of rejecting or not H_0, we need to obtain critical values, denoted by cv_{off}, for the test statistic. We approach this issue by considering the asymptotic distribution of the test statistic under H_0,

$$TS_{off} \xrightarrow{D} cv_{off} = \sup_{0 \leqslant t \leqslant 1} \sum_{j=1}^{r} B_j^2(t) \quad (N \to \infty), \tag{3.9}$$

where $\xrightarrow{D}$ denotes convergence in distribution, $(B_j(t) : t \in [0,1])$, $1 \leqslant j \leqslant r$, are independent standard Brownian bridges $B(t) = W(t) - tW(1)$, and $W(t)$ denotes the standard Brownian motion with mean

0 and variance t. The critical values for several significance levels α can be computed using Monte Carlo simulations that approximate the paths of the Brownian bridge on a fine grid. The last step is to estimate the unknown CP, defined previously as k^*_{off}, under H_1, given by:

$$\hat{k}^*_{off} = \frac{1}{N} \underset{1 \leqslant n \leqslant N}{\operatorname{argmax}} TS_{off}. \tag{3.10}$$

3.4.2 Extended Off-line Approach

The above hypothesis test identifies the existence of at most one CP and does not ensure that the sample remains statistically stationary in either direction of the detection. In particular, by construction (see (3.2)), the off-line test statistic detects the CP with the highest magnitude. Therefore, for the detection of multiple CPs we need to rephrase the hypothesis test H_1, as follows:

$$H_1: \quad \mu_1 = \ldots = \mu_{k_1} \neq \mu_{k_1+1} = \ldots = \mu_{k_2} \neq \ldots \cdots \neq \mu_{k_{\tau-1}+1} = \ldots = \mu_{k_\tau} \neq \mu_{k_\tau+1} = \ldots = \mu_N.$$

A greedy technique to identify multiple CPs is the binary segmentation (BS) algorithm. The standard BS algorithm relies on the general concept of binary segmentation and is an extension of the single CP estimator. First, a single CP is searched for in the time-series. In case of no change, the procedure stops and H_0 is accepted. Otherwise, the detected CP is used to divide the time-series into two segments in which new searches are performed. The procedure is iterated until no more CPs are detected. The BS algorithm is lightweight (computational time $O(N\log N)$), while its conceptual simplicity leads to efficient implementations. On the other hand, it has been shown in the literature [42], [43], that the standard BS algorithm tends to overestimate the number of CPs, as it does not cross-validate them after their detection.

In the extended off-line approach, we propose the modification of the standard BS with a cross-validation step of the estimated CPs. The cross-validation step is similar to that used in the iterative cumulative sum of squares (ICSS) segmentation algorithm [44], which is used to search for CPs on the marginal variance of independent and identically distributed (i.i.d.) r.vs. In the extended off-line algorithm we consider the CPs estimated from the standard BS in pairs and check if H_0 is rejected in the segment delimited by each pair. If H_0 is not rejected in a particular segment, then no change can be detected in it; as a result, all CPs that fall in the respective segment are eliminated. The improvement, in terms of accuracy, is shown through simulation results in Section IV.

3.5 On-line Phase

In this Section, we describe the on-line scheme that includes: (i) two alternative CUSUM-type approaches for the detection of a change in the mean; and (ii) two alternative approaches to estimate the direction of a change.

3.5.1 On-line Analysis

We rewrite equation (1) in the form,

$$X_n = \begin{cases} \mu + Y_n, & n = 1, \ldots, m + k^* - 1 \\ \mu + Y_n + I, & n = m + k^*, \ldots \end{cases} \tag{3.11}$$

where μ, $I \in \mathbb{R}^r$ represents the mean parameters before and after the unknown time of possible change $k^* \in \mathbb{N}^*$ respectively. As a reminder, the first dimension of the time-series represents the video views; the rest could be likes, comments, etc., and $\{Y_n : n \in \mathbb{N}\}$ is a random component. The term $m \in \mathbb{N}$ denotes the length of the training period, i.e., an interval of length m over the historical period during

which the mean is assumed to remain unchanged, so that,

$$\mu_1 = \cdots = \mu_m. \tag{3.12}$$

To satisfy this assumption, the modified off-line CP test previously presented is run in order to identify a suitable m. With m determined, the on-line procedure can be used to check whether (3.12) holds as new data become available. In the form of a statistical hypothesis test, the on-line problem becomes,

$$\begin{aligned} H_0 &: I = 0, \\ H_1 &: I \neq 0. \end{aligned} \tag{3.13}$$

The on-line sequential analysis belongs to the category of stopping time stochastic processes. In general, a chosen on-line test statistic $TS_{on}(m,l)$ and a given threshold $F(m,l)$ define the stopping time $\tau(m)$:

$$\tau(m) = \begin{cases} \min\{l \in \mathbb{N} : TS_{on}(m,l) \geqslant F(m,l)\}, \\ \infty, \text{ if } TS_{on}(m,l) < F(m,l)\ \forall l \in \mathbb{N}, \end{cases} \tag{3.14}$$

implying that $TS_{on}(m,l)$ is calculated on-line for every l in the monitoring period. The procedure stops if the test statistic exceeds the value of the threshold function $F(m,l)$. As soon as this happens, the null hypothesis is rejected and a CP is detected. The following properties should hold for $\tau(m)$,

$$\lim_{m\to\infty} Pr\{\tau(m) < \infty | H_0\} = \alpha,$$

ensuring that the probability of false alarm is asymptotically bounded by $\alpha \in (0,1)$, and,

$$\lim_{m\to\infty} Pr\{\tau(m) < \infty | H_1\} = 1,$$

ensuring that under H_1 the asymptotic power of the statistical test is unity. The threshold $F(m,l)$ is given by,

$$F(m,l) = cv_{on,a} g(m,l), \tag{3.15}$$

where: (i) the critical value $cv_{on,a}$ is determined from the asymptotic behavior of the stopping time procedure under H_0 by letting $m \to \infty$; and (ii) the weight function,

$$g(m,l) = \sqrt{m}\left(1 + \frac{l}{m}\right)\left(\frac{l}{l+m}\right)^{\gamma} \tag{3.16}$$

depends on the sensitivity parameter $\gamma \in [0, 1/2)$.

We use two different CUSUM approaches; the standard [12], with test statistic denoted by TS_{on}^{ct}, and, the ratio-type [13], with test statistic denoted by TS_{on}^{rt}. Their corresponding critical values are denoted by $cv_{on,a}^{ct}$ and $cv_{on,a}^{rt}$, respectively, and their stopping rules by $\tau_{ct}(m)$ and $\tau_{rt}(m)$, correspondingly. Both tests are based on the sequential CUSUM detector, $E(m,l)$,

$$E(m,l) = \left(\overline{X}_{m+1,m+l} - \overline{X}_{1,m}\right) \tag{3.17}$$

The standard CUSUM test is expressed as:

$$TS_{on}^{ct}(m,l) = l\widehat{\Omega}_m^{-\frac{1}{2}} E(m,l), \tag{3.18}$$

where $\widehat{\Omega}_m$ is the estimated long-run covariance, defined as in (4), that captures the dependence between observations. Then, the stopping rule $\tau_{ct}(m)$, is defined as:

$$\tau_{ct}(m) = \min\{l \in \mathbb{N} : \|TS_{on}^{ct}(m,l)\|_1 \geq cv_{on,a}^{ct} g(m,l)\}, \tag{3.19}$$

where the ℓ_1 norm is involved to modify TS_{on}^{ct} so that it can be compared to a one dimensional threshold function. The critical value, $cv_{on,a}^{ct}$, is derived from the asymptotic behavior of the stopping rule under H_0:

$$\lim_{m\to\infty} Pr\{\tau(m) < \infty\} = \lim_{m\to\infty} Pr\left\{ \sup_{1\leqslant l\leqslant\infty} \frac{\|TS_{on}^{ct}(m,l)\|_1}{g(m,l)} > cv_{on,\alpha}^{ct} \right\}$$
$$= Pr\left\{ \sup_{t\in[0,1]} \frac{\|W(t)\|_1}{t^\gamma} > cv_{on,\alpha}^{ct} \right\} = \alpha. \tag{3.20}$$

Unlike standard CUSUM tests, ratio type statistics do not require to estimate the long-run covariance and are also considered for this reason in this analysis. The precise form of the chosen statistic is given in the following quadratic form,

$$TS_{on}^{rt}(m,l) = \frac{l^2}{m} E^T(m,l) \left\{ \frac{1}{m^2} \sum_{j=1}^{m} j^2 \left(\overline{X}_{1,j} - \overline{X}_{1,m}\right)\left(\overline{X}_{1,j} - \overline{X}_{1,m}\right)^T \right\}^{-1} E(m,l), \tag{3.21}$$

with its equivalent stopping rule,

$$\tau_{rt}(m) = \min\{l \in \mathbb{N} : TS_{on}^{rt} \geq cv_{on,a}^{rt} g^2(m,l)\}. \tag{3.22}$$

Similarly to the standard CUSUM, the critical value, $cv_{on,a}^{rt}$, is estimated by,

$$\lim_{m\to\infty} Pr\{\tau(m) < \infty\} = Pr\left\{ \sup_{t\in[0,\infty)} \Delta_\gamma(t) > cv_{on,\alpha}^{rt} \right\} = \alpha, \tag{3.23}$$

where,

$$\Delta_\gamma(t) = \frac{1}{\eta_\gamma^2(t)} B^T(1+t) \left(\int_0^1 B(r) B^T(r) dr \right)^{-1} B(1+t), \eta_\gamma^2(t) = (1+t)\left(\frac{t}{1+t}\right)^\gamma,$$

and $B(t)$ is a standard Brownian bridge, $t \in [0,\infty)$.

Similarly to the off-line case, the on-line critical values for both test statistics can be computed using Monte Carlo simulations, considering that,

$$cv_{on,\alpha}^{ct} = \sup_{t\in[0,1]} \frac{W(t)}{t^\gamma}, \tag{3.24}$$

$$cv_{on,\alpha}^{rt} = \sup_{t\in[0,\infty)} \Delta_\gamma(t). \tag{3.25}$$

The estimated on-line CP, $\hat{k}_{on}^*$, is derived directly from the value of the stopping time $\tau(m)$, as,

$$\hat{k}_{on}^* = m + \{\tau(m) | \tau(m) < \infty\}. \tag{3.26}$$

3.5.2 Trend Indicator

Considering the on-line procedure, the hypothesis H_1 is two-tailed because the test statistics TS_{on}^{rt} and TS_{on}^{ct} are formulated in a quadratic form and a ℓ_1 norm, respectively. This means that the stopping time rule $\tau_{ct}(m)$ (or $\tau_{rt}(m)$) cannot be an indicator of the direction of a detected change. Thus, to estimate the direction of a change we introduce two indicators: i) based on the CUSUM detector in (3.17), denoted by TI_{ts}; and ii) based on the moving average convergence divergence (MACD) filter [45], denoted by TI_f.

Focusing on TI_{ts}, the indicator is directly derived from the form of the sequential CUSUM detector $E(m,l)$. The detector compares the mean value of the observations that are collected on-line for a chosen monitoring period l, with the mean value of a subsample of the historical data over the predetermined training sample. Hence, for a detected CP, we have that,

$$\begin{cases} E(m,l) > 0, \text{ denotes an upward change} \\ E(m,l) < 0, \text{ denotes a downward change} \end{cases} . \tag{3.27}$$

However, in certain cases, limiting the window over which the direction of a change is estimated to the immediate neighbourhood of a detected CP can be unreliable due to the continuous variability of the time-series. In such cases, we have to estimate the direction of a change by incorporating more elaborate filters; in this context, we estimate the direction of detected changes by applying the MACD indicator. The MACD is based on an exponential moving average (EMA) filter, of the form,

$$EMA_p(n) = \frac{2}{p+1} X_n + \frac{p-1}{p+1} EMA_p(n-1), \tag{3.28}$$

with p denoting the lag parameter. The MACD series can be derived from the subtraction from a short p_2 lag EMA (sensitive filter) of a longer p_3 lag EMA (blunt filter), as described below:

$$MACD(n) = EMA_{p_2} - EMA_{p_3}. \tag{3.29}$$

The trend indicator TI_f is then obtained by the subtraction of a short p_1 lag EMA filter of a MACD series from the raw MACD series, as described below

$$TI_f(n) = MACD(n) - EMA_{p_1}(MACD(n)),\ p_1 < p_2 < p_3. \tag{3.30}$$

In the evaluation of TI_f three exponential filters are involved. In essence, TI_f is an estimation of the second derivative over an interval around the change (considering that the subtraction of a filtered variable from the variable generates an estimate of its time derivative). In contrast to other works [45], we only adopt TI_f to characterize the direction from the specific value of TI_f at the estimated time of change. We announce an upward change if $TI_f(\hat{k}^*_{on}) > 0$, otherwise, if $TI_f(\hat{k}^*_{on}) < 0$, a downward change.

Finally, we propose a modification of the trend indicator TI_f, converting it from a point estimator to an interval estimator; instead of evaluating $TI_f(\hat{k}^*_{on})$, we propose to evaluate the trend indicator at a time interval $(\hat{k}^*_{on}, \hat{k}^*_{on} + h)$, where h is a threshold parameter:

$$TI_f(\hat{k}^*_{on}, h) = \sum_{l=\hat{k}^*_{on}}^{\hat{k}^*_{on}+h} TI_f(l). \tag{3.31}$$

The proposed $TI_f(\hat{k}^*_{on}, h)$ modification improves the estimator's accuracy; the calculation of the sum of a multitude of observations, after a CP, can smooth out a potential false one-point estimation, especially in the case of small changes.

3.5.3 Overall Algorithm

We outline in *Algorithm 1* the RCPD algorithm, as a combination of the off-line and the on-line phase, in the form of pseudo-code. Beginning from the initial value set for the monitoring starting period, denoted by m_s, the modified off-line algorithm is applied over the whole historical period; the training period m is then defined as the interval elapsed from the last detected off-line CP (if one exists) to m_s. As a second step, the on-line test statistic, $TS_{on}(m,l)$ in (14), is applied for a specified monitoring time frame l. If a content popularity change is detected at time instance $\hat{k}^*_{on}$, the trend indicator subroutine

Algorithm 1: The Real-time CP Detector (RCPD)

procedure RCPD(X_n,m_s,k)
; X_n: time-series of video views
; m_s: running end of training period
; m: training period
; l: monitoring time frame
; d: period assuming no change
; TS_{on}: on-line test statistic (eq. 3.18 or 3.21)
; cv_{on}: critical value (eq. 3.24 or 3.25)
; $\hat{k}^*_{on}$: the estimated on-line CP (eq. 3.26)
; TI: trend indicator (TI_{ts} or TI_f)
for n in X_n **do**
if $n = m_s$ **then**
s=MBS(1,m_s,1) ; calculate off-line CPs
if `array_length`(s) > 0 **then**
m=$\{\max(s), m_s\}$; $max(s)$ is the latest CP
else
m=$\{max(1, m_s - u), m_s\}$; u a large value
end if
else if $m_s < n < m_s + l$ **then**
calculate $\texttt{TS}_{on}$`(m,l)`
if $\texttt{TS}_{on}$`(m,l)`$>\text{cv}_{on}$ **then**
calculate `TI`
signal CP and estimated direction
$m_s = \widehat{cp}_{on} + d$; keep a distance from $\widehat{cp}_{on}$
end if
else if $n = m_s + l$ **then**
$m_s = m_s + l$; start a new training period
end if
end for
end procedure

is called to reveal the direction of change.[3] At this point the procedure stops and a new starting point for the monitoring window is defined as $m_s = \hat{k}^*_{on} + d$, where d is a constant value specifying a period assuming no change. Otherwise, if no change is detected after a maximum of l instances, the procedure restarts from the last time point, $m_s = m_s + l$.

3.6 Validation of the RCPD Using Synthetic Data

In this Section, we validate the performance of the overall algorithm by performing a series of four different experiments on synthetic data. The use of synthetic data allows us to regulate the parameters of the time-series in terms of mean changes and thus obtain quantitative metrics for the performance of the proposed algorithms.

The choice of the time-series model for the generation of the synthetic data is based on the fact that several studies have shown that ARMA models capture very well content popularity evolution. For example, in [10] it has been concluded that an ARMA model can efficiently describe the daily access patterns of YouTube content, based on an extensive analysis of 100,000 videos. Similarly, in [46]

[3] In the load balancing scenario discussed in Section VII, in the case of an increase in the content popularity a new content cache is being deployed, while conversely a decrease leads to the removal of an existing cache.

Table 3.1: Percentage of the successful CP detections for the standard and modified BS algorithm

	Test 1: two CPs		Test 2: four CPs	
μ	BS	modified BS	BS	modified BS
	True (false) alarm rate		True (false) alarm rate	
μ_1=1	0.94 (0.06)	0.95 (0.05)	0.5 (0.258)	0.7 (0.05)
μ_2=1.5	0.95 (0.05)	0.95 (0.05)	0.5 (0.258)	0.9 (0.08)
μ_3=2	0.95 (0.05)	0.95 (0.05)	0.47 (0.53)	0.9 (0.1)

Table 3.2: Success rates of trend indicators

	Test 1: two CPs		Test 2: four CPs	
μ	TI_{ts}	TI_f	TI_{ts}	TI_f
	Success rate		Success rate	
μ_1=1	0.99	0.99	0.99	0.99
μ_2=1.5	1	1	1	1
μ_3=2	1	1	1	1

an ARMA model has been proposed for the estimation of the popularity of video content. Motivated by these findings, for the validation of the proposed algorithm we use an ARMA$(1,1)$ time-series. We generate $1,000$ time-series of length $N = 600$ samples. Without loss of generality, we assume an initial mean value $\mu_0 = 0$, noting that the performance of the RCPD is independent of the initial mean value and only depends on the magnitude of the variation of the mean value before and after a CP.

In the first experiment, we begin with a comparison of the standard BS to the proposed modified BS algorithms described in Section 3.4. We perform two tests; in the first test we introduce two CPs at the instances $k_i^* = (iN)/3,\ i = 1, 2$, while in second test, we introduce four CPs at $k_i^* = (iN)/5,\ i = 1, \ldots, 4$. The two tests are repeated for three different values of the magnitude of a change $\mu_1 = 1,\ \mu_2 = 1.5, \mu_3 = 2$, i.e., we randomly increase or decrease the mean value by $\mu_j,\ j = 1, \ldots, 3$ at the time of change. Table 3.1 summarizes our findings regarding the true and false alarm rates of the two algorithms.

Both the standard and the modified BS algorithms provide similar true alarm rates, exceeding 94%, in the first test. On the contrary, in the more challenging second test, the superiority of the modified BS over the standard BS algorithm is clear. The modified BS algorithm achieves true alarm rates in excess of 70%, even in the demanding scenario of a relatively small change in the mean $\mu_1 = 1$. On the other hand, the standard BS algorithm has in all cases a true alarm rate of less than 50%, rendering any CP detection highly questionable. The second test confirms that the standard BS algorithm is prone to an overestimation of the number of CPs as shown by the high false alarm rates (in excess of 25% in all cases), an issue that can be effectively addressed by the modified BS algorithm which scores false alarm rates below 10%.

Next, in the second experiment, using the same test sets as above, we measure the success rates achieved by the proposed trend indicators TI_{ts} and TI_f for $h = 0$ (larger thresholds provided the same true identification rates). The results are summarized in Table 3.2. The two trend indicators successfully identify the direction of a change in more than 99% of the cases, which shows that they can be interchangeably employed. In the assessment of the performance using real datasets in Sections 3.6 and 3.7, we solely employ the TI_f trend indicator.

Table 3.3: Results of the RCPDs' algorithm CPs detection for one change in the mean value.

		ARMA(1,1)							
μ	l	standard CUSUM				ratio-type CUSUM			
		Number of detected CPs			$\hat{k}^*$	Number of detected CPs			$\hat{k}^*$
		0	1	> 1	med	0	1	> 1	med
	25	0.95	0.05	0	-	0.95	0.05	0	-
$\mu = 0$	50	0.95	0.05	0	-	0.95	0.05	0	-
	100	0.94	0.06	0	-	0.95	0.05	0	-
	25	0.7	0.29	0.01	-	0.8	0.19	0.01	-
$\mu = 0.5$	50	0.16	0.8	0.04	343	0.55	0.43	0.02	-
	100	0	0.93	0.07	341	0.2	0.76	0.04	348
	25	0.26	0.73	0.01	332	0.69	0.3	0.01	-
$\mu = 0.7$	50	0	0.96	0.04	326	0.3	0.65	0.05	328
	100	0.01	0.91	0.08	331	0.05	0.89	0.06	335
	25	0.01	0.97	0.02	327	0.52	0.46	0.02	-
$\mu = 1$	50	0	0.96	0.04	316	0.08	0.86	0.06	321
	100	0	0.92	0.08	321	0	0.95	0.05	323
	25	0.01	0.97	0.02	323	0.43	0.54	0.03	331
$\mu = 1.2$	50	0	0.95	0.05	316	0.02	0.93	0.05	317
	100	0	0.93	0.07	318	0	0.93	0.07	318
	25	0	0.97	0.03	320	0.36	0.6	0.04	329
$\mu = 1.5$	50	0	0.95	0.05	310	0	0.94	0.06	313
	100	0	0.93	0.07	314	0	0.94	0.06	318
	25	0	0.97	0.03	310	0.26	0.71	0.03	317
$\mu = 2$	50	0	0.95	0.05	307	0	0.93	0.07	310
	100	0	0.94	0.06	310	0	0.94	0.06	313

We proceed by assessing the proposed RCPD algorithm using both the standard and the ratio type CUSUM. In this third experiment, we measure the average number of CPs detected, averaged over $1,000$ simulations when a single CP is introduced in the ARMA time-series at the time instance $\frac{N}{2} = 300$. We consider different values for the magnitude of change $\mu \in \{0, 0.5, 0.7, 1, 1.2, 1.5, 2\}$ and the monitoring window length $l \in \{25, 50, 100\}$. We note that we included the case $\mu = 0$ – which corresponds to the absence of a change – to evaluate the false alarm rate of the overall algorithm. We omit results with true alarm rates lower than 50% as they are statistically unreliable. In terms of the remaining algorithmic parameters, we have set the minimum distance between two successive CPs to $d = 50$,[4] the sensitivity parameter to $\gamma = 0.25$ [47] (we choose a neutral value as the behaviour of γ is well studied), and, the significance level to $\alpha = 0.05$. In each test of the third experiment we measure the exact number of CPs detected, tabulated as one the following three values: i) 0 when (falsely[5]) no

[4]This choice is justified by our observations of the minimum distance between successive CPs in real data sets, presented in Section VI.

[5]Except for the $\mu = 0$ case.

Table 3.4: Results of the RCPDs' algorithm CPs detection for two mean changes.

		ARMA(1,1)									
μ	l	standard CUSUM					ratio-type CUSUM				
		Number of detected CPs			$\hat{k}_1^*$	$\hat{k}_2^*$	Number of detected CPs			$\hat{k}_1^*$	$\hat{k}_2^*$
		< 2	2	> 2	med		< 2	2	> 2	med	
$\mu_1 = 0.5$	25	0.88	0.12	0	-	-	0.95	0.05	0	-	-
	50	0.38	0.60	0.02	251	440	0.79	0.2	0.01	-	-
	100	0.1	0.87	0.03	242	443	0.54	0.44	0.02	-	-
$\mu_1 = 0.7$	25	0.41	0.58	0.01	230	427	0.9	0.1	0	-	-
	50	0.06	0.91	0.03	223	427	0.58	0.41	0.01	-	-
	100	0.01	0.93	0.06	227	428	0.25	0.72	0.03	231	439
$\mu_1 = 1$	25	0.04	0.93	0.03	219	420	0.74	0.25	0.01	-	-
	50	0.03	0.93	0.04	215	419	0.26	0.71	0.03	221	423
	100	0	0.94	0.06	217	420	0.05	0.9	0.05	220	424
$\mu_1 = 1.2$	25	0.01	0.96	0.03	214	414	0.56	0.42	0.02	-	-
	50	0	0.95	0.05	212	416	0.17	0.79	0.04	215	428
	100	0	0.94	0.06	217	420	0.02	0.93	0.05	216	421
$\mu_1 = 1.5$	25	0	0.98	0.02	211	411	0.33	0.63	0.04	213	417
	50	0	0.94	0.06	209	413	0.1	0.85	0.05	213	415
	100	0	0.94	0.06	211	415	0	0.96	0.04	216	419
$\mu_1 = 2$	25	0	0.98	0.02	208	407	0.12	0.85	0.03	210	412
	50	0	0.95	0.05	207	410	0.3	0.91	0.06	209	413
	100	0	0.94	0.06	209	411	0	0.96	0.04	211	414

CP is detected; ii) 1 when (correctly) a single CP is detected; and iii) > 1 when (falsely) multiple CPs are detected. Finally, we measure the median of the time instance of the single CP detection, denoted by $\hat{k}^*$.[6] The results of this experiment are presented in Table 3.3 and are discussed below.

Firstly, we observe that both the standard and the ratio type CUSUM achieve very small false alarm rates, inferior to 6% when no CP is inserted, irrespective of the choice of l. On the contrary, the choice of l readily affects the algorithm's success rate for $\mu > 0$; for small changes in the mean value, $\mu = 0.5, 0.7$, a larger monitoring window l increases the algorithm's true alarm rates in identifying correctly the existence of the CP. For medium and high changes in the magnitude of change $\mu = 1, 1.2, 1.5, 2$, it is observed that a high true alarm rate – in excess of 93% for the standard CUSUM – is achieved, while choosing a smaller l can slightly increase the true alarm rates. As a result, depending on the application, a choice of a larger l can be appropriate if the algorithm is to be employed as a universal CP detector. Alternatively, a smaller l can be chosen when the focus is on the identification of large changes in the mean value, i.e., we are interested primarily in detecting CPs of larger magnitude.

Secondly, we observe that overall, the ratio type CUSUM is outperformed by the standard CUSUM in all tests. Consequently, the standard CUSUM based detector can be considered as an efficient universal choice. Finally, we observe that the lag between $\hat{k}^*$ and the actual instance of change at

[6] We omit the results with true detection rate lower than 50%.

the point 300 decreases with increasing μ, ranging from 343 to 307, while it appears less sensitive to changes in l. This demonstrates that, intuitively, larger magnitude changes can be detected faster. This result is important for load balancing applications as it provides us with the means to quickly respond to significant changes in the network traffic.

Subsequently, in Table 3.4 in the previous page, we present the outputs of the fourth experiment in which we assess the performance, averaged over $1,000$ simulations, of the RCPD algorithm when two CPs are inserted in the ARMA time-series. We introduce a change at the time instance $k_1^* = \frac{N}{3} = 200$ and a second CP at the time instance $k_2^* = \frac{2N}{3} = 400$. We investigate the true and false alarm rates for $\mu \in \{0.5, 0.7, 1, 1.2, 1.5, 2\}$ and $l \in \{25, 50, 100\}$, while the rest of the parameters retain the values of the third experiment. In each test of the fourth experiment we measure the exact number of CPs detected, tabulated as one the following three values: i) < 2 when (falsely) less than two CPs are detected, ii) 2 when (correctly) two CPs are detected, and iii) > 2 when (falsely) more than two CPs are detected. Finally, we measure the median of the detection instances of the two CPs, denoted by $\hat{k}_1^*$ and $\hat{k}_2^*$, respectively (we omit the results with true detection rate lower than 50%).

Similarly to the third experiment, we observe that increasing l increases the true alarm rates for small magnitudes in the mean changes $\mu = 0.5,\ 0.7$, while this trend is reversed in high magnitudes $\mu = 1.5,\ 2$. For medium values $\mu = 1,\ 1.2$ the effect of l on the true alarm rates is less than 2%. Furthermore, in agreement with the outputs of the third experiment, with increasing μ the algorithms achieve increasingly high success rates, over 93% for the standard CUSUM when $\mu \geq 1$.

In addition, the superior performance of the standard CUSUM is re-confirmed in all the tests of the fourth experiment. Finally, with respect to the lag in the estimation of the time instances of the CPs, we observe that, as in experiment three, larger magnitude changes can be detected faster, e.g., for $\mu = 2$ a lag inferior to 11 instances is observed for both CPs with the standard CUSUM, irrespective of l.

Concluding this Section, we have presented an extensive set of experiments that provide strong evidence for the efficiency of the proposed algorithms. We have explicitly demonstrated the superiority of the modified BS over the standard BS algorithm and confirmed the validity of the proposed trend indicators. Subsequently, we evaluated the performance of the overall algorithm for various values of μ and l. We have shown that the RCPD algorithm achieves extremely high true alarm rates for larger values of μ, while increasing the length of the monitoring window l can significantly impact the performance for small values of μ. Finally, overall, the standard type CUSUM outperforms the ratio type CUSUM and should be preferred.

3.7 Performance Evaluation Using Real Data

In this Section we investigate the performance of the proposed algorithms using a real dataset provided within the framework of the CONGAS project [48]; the dataset consists of the number of views of 882 YouTube videos, observed over $N = 1,000$ instances.

3.7.1 Statistical Properties of the Real Dataset

First, we evaluate the validity of the most important underlying assumption of this analysis, that the content popularity can be modelled as the sum of a constant mean and a weak-dependent (t-dependent) stochastic process, as given in (3.34). A first intuitive method to test whether the time-series is short-range dependent (SRD) is through its autocorrelation function (ACF). The ACF for a weakly-stationary process $\{X_t : t \in \mathbb{N}$ with mean value μ is given by,

$$\rho(k) = \frac{(X_t - \mu)(X_{t+k} - \mu)}{\sigma^2}.$$

Note that if $\sum_{k=-\infty}^{\infty} \rho(k) \to \infty$ the process has long-range dependence (LRD), while if $\sum_{k=-\infty}^{\infty} |\rho(k)| < \infty$ it exhibits SRD. To distinguish between these two phenomena, we use the following functional form

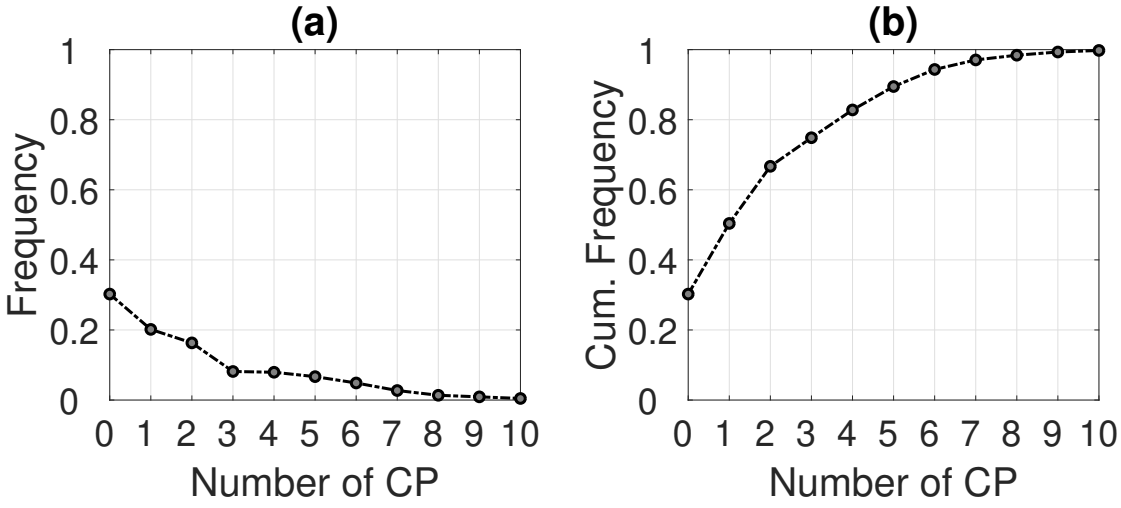

Figure 3.1: Estimated a) frequency and b) cumulative frequency of the number of CPs per time-series.

Table 3.5: Success rates of TI_f trend indicator

h	0	3	5	7	10
Video Set 1	0.69	0.91	0.95	0.97	0.98
Video Set 2	0.90	0.99	0.99	0.99	0.99

of the ACF,

$$\rho(k) \sim C_i^{2H-2}, \text{ as } i \to \infty,$$

where $C_i > 0$ and $H \in (0, 1)$ is the Hurst exponent characterizing the LRD,i.e., $H \in (1/2, 1)$ indicates the presence of LRD. It is challenging to accurately estimate the Hurst exponent out of real data [49] and several methods have been proposed in the literature [50]. In this work, we apply two semi-parametric tests, identified as accurate options among others presented in the survey paper [50]. The first method uses the discrete second order derivative in the time domain while the second uses the discrete second order derivative in the wavelet domain. Both methods estimate an $H \leq 0.5$ for 95% of the YouTube time-series, indicating the validity of our assumptions related to the equation (3.34).

3.7.2 Performance of the Off-line Training Phase

First, we test the hypothesis H_0 of no change in the mean structure on our dataset. H_0 is rejected in approximately 70% of the cases, for a significance level of $a = 0.05$. This outcome indicates that CP algorithms can identify changing content dynamics in real times series. Next, we estimate the number of CPs, by applying the extended off-line algorithm. The corresponding results are illustrated in Fig. 3.1 and indicate a sufficiently high number of content popularity anomalies (i.e., mean changes). Hence, a CP analysis is indeed a suitable tool for content popularity detection.

To evaluate the performance of the proposed trend indicator TI_f, we need a baseline independent assessment of the direction of change. We declare that a real increase in the mean value of content visit exists if

$$\mathsf{E}[X(\hat{k}^*_{i-1,off}) : X(\hat{k}^*_{i,off})] < \mathsf{E}[X(\hat{k}^*_{i,off}) : X(\hat{k}^*_{i+1,off})], \tag{3.32}$$

or, that a real decrease in the number of visits exists if

$$\mathsf{E}[X(\hat{k}^*_{i-1,off}) : X(\hat{k}^*_{i,off})] > \mathsf{E}[X(\hat{k}^*_{i,off}) : X(\hat{k}^*_{i+1,off})], \tag{3.33}$$

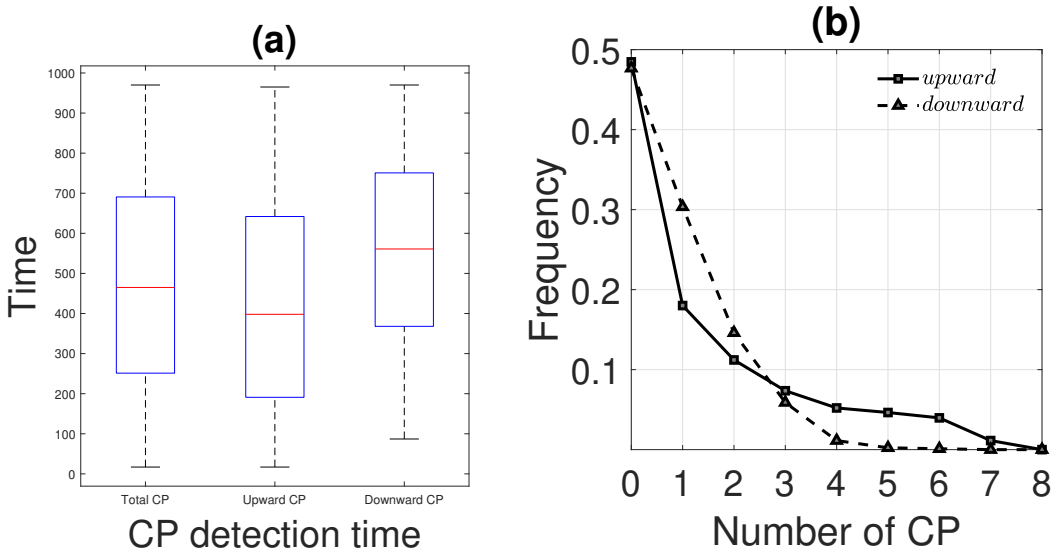

Figure 3.2: Frequency values of the number of upward and downward CPs, per time-series.

where $i = 2, \ldots, N-1$ and $\mathsf{E}[\cdot]$ denotes the numerical average. We test the modified MACD TI_f on two sets of videos. The first set, Video Set 1, comprises the whole dataset, while the second set, Video Set 2, comprises only the videos with a considerable average number of visits (> 10), i.e., for which, $\mathsf{E}[X(1) : X(1000)] > 10$.

The percentage of successful TI_f identifications are tabulated in Table 3.5 for five values of the parameter h, namely $h = 0$, 3, 5, 7 and 10, where h denotes the TI_f's calculation threshold. Commenting on the results for Video Set 1, the TI_f trend indicator works well, except for $h = 0$, providing at least 90% correct direction identifications. As expected, as h increases the procedure works better. More specifically, an $h \geq 5$ parameter choice yields a success rate of 95%, while if a more agile estimation is needed then an $h \geq 3$ still maintains a 91% accuracy. Considering the interim time between consecutive changes, we deduce that an $h \leq 7$ is preferable. Regarding Video Set 2, we see that the results are highly improved, indicating that the procedure works even better for the most popular videos. In practice, this represents the more interesting scenario as it will have a greater impact in terms of the applied load balancing mechanism.

Furthermore, in Fig. 3.2, the time instances of upward and downward changes are shown in the form of a boxplot. It is intuitive that upward changes occur earlier than downward changes. Moreover, Fig. 3.2 demonstrates that the multitude of upward changes is greater than the respective of downward changes, indicating that decreases in popularity are sharper than increases. In particular, we estimated that out of the total number of changes, 67% are upward.

Finally, we analyze the interim time between consecutive CPs. The results presented in Fig. 3.3 illustrate the existence of a sufficiently large gap between consecutive potential changes. 90% of the intervals corresponding to consecutive CPs exceed 70 time instances and only 5% of them are shorter than 50 time instances, ensuring that a sufficiently large training window can be applied. The results depicted in Fig. 3.3 allow adjusting parameters of the on-line phase, in particular the minimum time interval between consecutive changes, denoted by the parameter d.

3.7.3 Evaluation of the RCPD Algorithm

In the previous subsection we have evaluated the performance of the off-line algorithm and demonstrated its efficiency as well as how it is employed in determining parameters of the on-line phase, such as the interval assuming no change d and the threshold parameter of TI_f h.

We further employ the off-line algorithm as a benchmark against which the performance of the

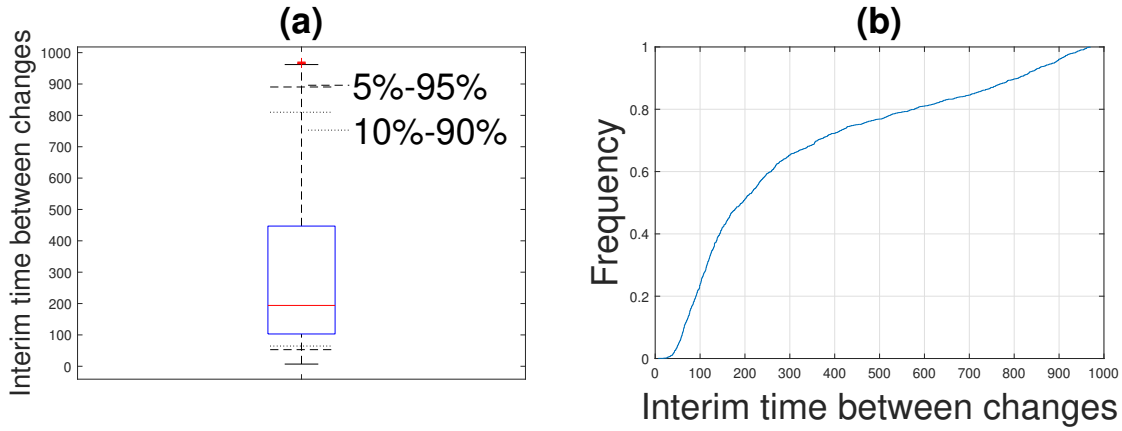

Figure 3.3: a) Boxplot including the interval $(5\% - 95\%)$ (dashed line) and $(10\% - 90\%)$ interval (dotted line), b) Cumulative frequency for the interim time of consecutive CPs.

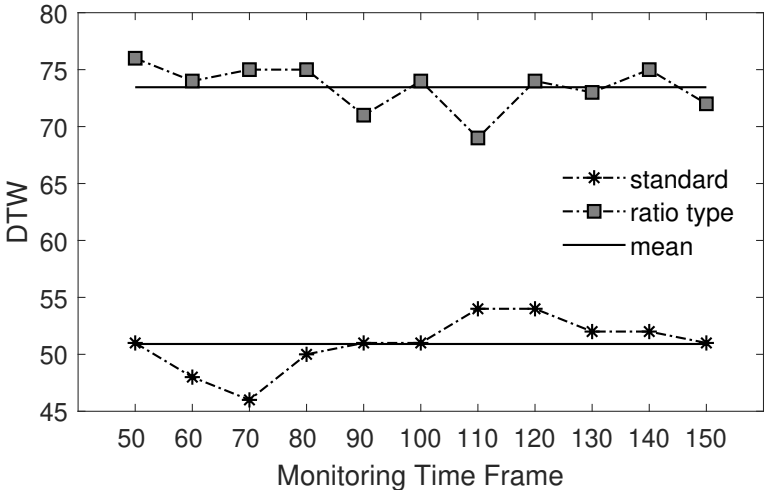

Figure 3.4: DTW distances for the two on-line detection schemes.

RCPD algorithm will be evaluated. We note that the off-line analysis provides the *best possible statistical detection* of the actual mean changes, as off-line algorithms operate retrospectively over the entirety of each of the time-series. Thus, in absence of a priori knowledge of the actual CPs in the real data (as opposed to the synthetic data in which the CPs were controlled), we evaluate the performance of the RCPD procedure by measuring the "similarity" of its outputs (detected CPs, instances of detection and trends) to the corresponding outputs of the off-line version.

As the number of detected CPs and / or their exact positions are likely to differ at the output of the retrospective (off-line) and of the RCPD algorithm, in order to obtain a measure of their similarity, we estimate their dynamic time warping (DTW) distance. The DTW is a dynamic programming tool that measures distances between asynchronous sequences and is widely used by the speech processing community [15].

The results are presented in Fig. 3.4, where the estimated DTW distances are depicted for several values of the monitoring window length $l \in [40, 150]$, to investigate the consistency of parameter l over different values. In the RCPD algorithm we use $d = 50$ (minimum distance between two changes) and have set the sensitivity parameter to $\gamma = 0.25$. The estimated mean DTW distance for the standard CUSUM is 52 and for the ratio-type CUSUM is 73. For comparison purposes, we note that the corresponding DTW distance over the synthetic data is 20 for medium / large changes, while the true CP detections are around 95%. As a result, we can infer, that the outputs of the on-line algorithm,

Table 3.6: Empirical percentiles of mean values change rate.

	Percentiles Threshold			
	10%	15%	25%	50%
Standard	9%	13.1%	20.8%	42.21%
Ratio type	9.5%	14.82%	28.22%	67.40%

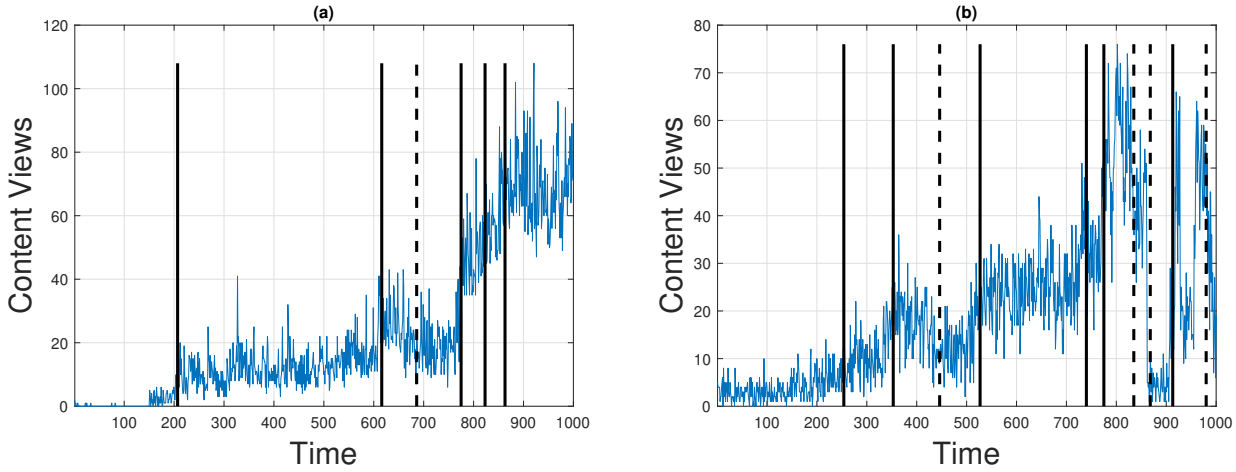

Figure 3.5: Outputs of the RCPD algorithm using standard CUSUM for different time-series. Solid and dashed lines depict an upward and a downward change, respectively.

using the standard CUSUM, are "very close" to the outputs of the benchmark off-line algorithm. In agreement with our observations over the synthetic data, the DTW distance using the ratio-type CUSUM is clearly larger.

We also study the magnitude of the detected CPs. We define as the CP magnitude the percentage-wise change in the mean values before and after the CP. We group the measured magnitudes for all change points using the four percentile threshold values 10%, 15%, 25% and 50%, i.e., reflecting the frequency of magnitudes exceeding the respective thresholds. The results are summarized in Table 3.6. According to our results, both the standard and ratio type CUSUM algorithms detect the most significant changes in the content popularity. Moreover, ratio-type CUSUM detects, in general, CPs with the largest magnitude of change, in agreement with synthetic data results.

Additionally, for illustration purposes, we depict the RCPD algorithm's outputs for four different time-series. We set the beginning of the monitoring period at $m_s = 200$ and monitoring horizon $l = 50$, the on-line parameter $g = 0.25$ and the significance level to $a = 0.05$. The corresponding results are depicted in Fig. 3.5 and 3.6, showing the estimated CPs by applying the standard CUSUM and the ratio type CUSUM procedures, respectively. In both cases, the estimated changes correspond to the real content popularity changes; visual inspection suggests that the performance of the standard CUSUM is more reasonable (e.g., Fig. 3.6d). The RCPD, as it is illustrated in Fig. 3.5b seems to be adaptable to "fast" changes; without getting "confused" by random peaks in the time-series, such as those in Fig. 3.5a or in Fig. 3.6c.

3.7.4 Time Dependencies of Piecewise time-series

We also measure the autocorrelation function of the piecewise - divided by the detected CPs - time-series. Results are tabulated in Table 3.7 and verify the short dependence structure of the dataset; significant lags in time dependencies higher than 30 instances can be found in less than 5% of the time-series. Furthermore, the fact that the ACF of the piecewise time-series drops to zero quickly indicates that

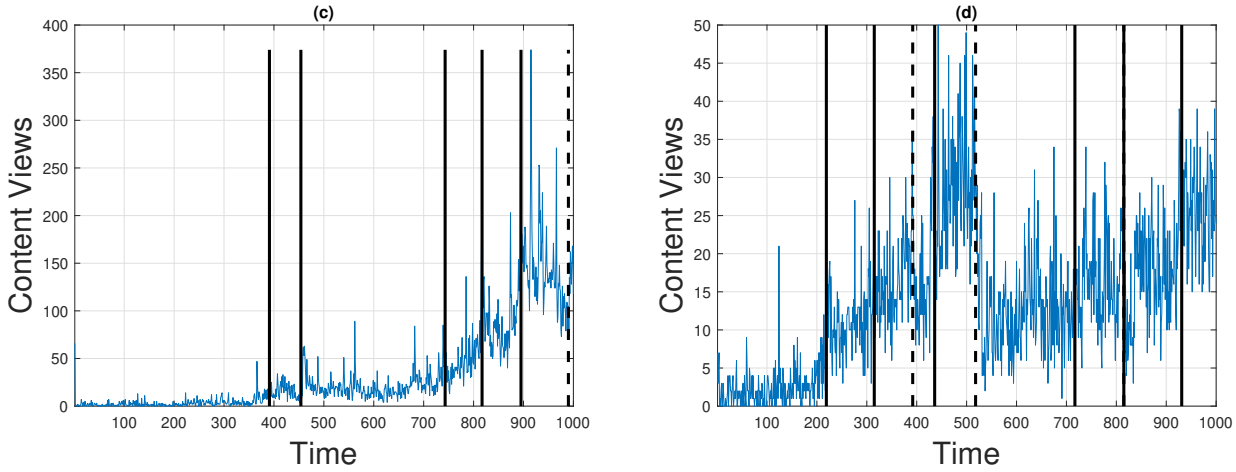

Figure 3.6: Outputs of the RCPD algorithm using standard type CUSUM for different time-series. Solid and dashed lines depict an upward and a downward change, respectively.

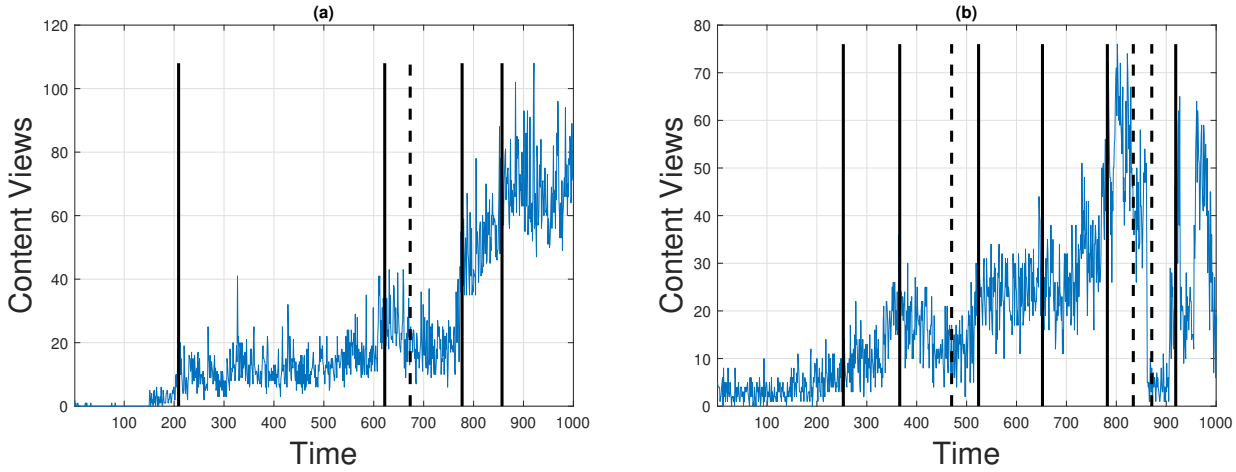

Figure 3.7: Outputs of the RCPD algorithm using ratio type CUSUM for different time-series. Solid and dashed lines depict an upward and a downward change, respectively.

the detected CPs split the time-series into stationary segments, which, additionally, confirms indirectly the accuracy of the off-line CP estimations over the changes in the real data.

3.7.5 Computational Complexity and Scalability

Finally, we present a MATLAB ® implementation of the overall algorithm with a large number of time-series (882 in this experiment) to quantify its performance in terms of processing cost. The computational time is measured on a Lenovo IdeaPad 510-15IKB laptop, with an Intel Core i7-7500U @ 2.70 GHz processor and 12 GB RAM. In Fig. 3.9, we show the aggregate processing cost per time instance for the two on-line methods and the total number of time-series. For the first 100 time instances, the algorithm collects the initial data, since it bootstraps. The peaks indicate the off-line part of the algorithm, which is more processing demanding mainly due to the segmentation algorithms running in parallel. The on-line part in the standard on-line algorithm indicates a linear complexity, since it is based on (3.18), while the equivalent quantity in (3.21) of the ratio-type is more CPU intensive, justifying the comparatively higher processing cost of the latter algorithm. In both cases, the aggregate processing cost is typically much less than a second, which demonstrates the lightweight nature of the proposed scheme. Such results could be further improved with a distributed deployment

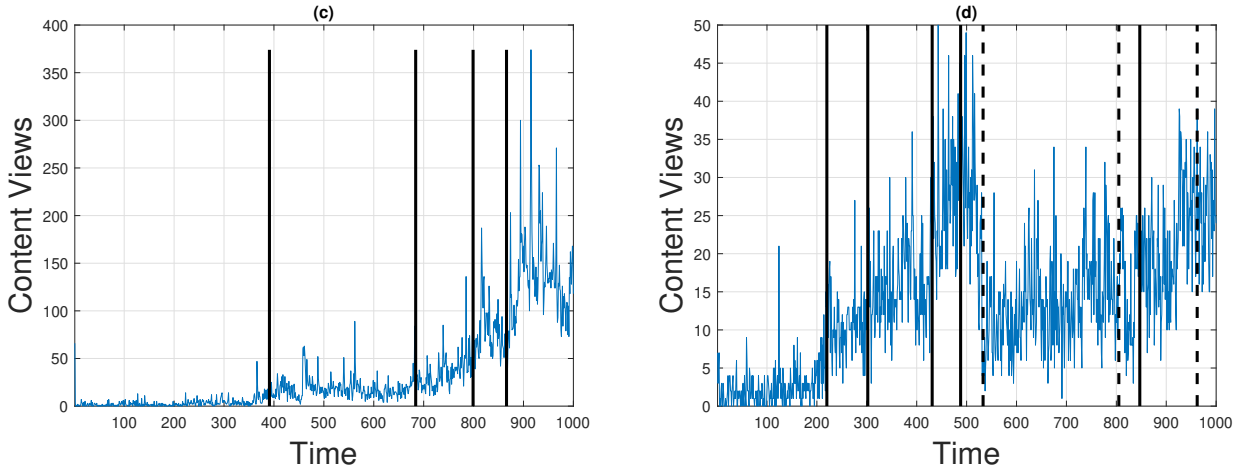

Figure 3.8: Outputs of the RCPD algorithm; using ratio type CUSUM for different time-series. Solid and dashed lines depict an upward and a downward change, respectively.

Table 3.7: Percentages of time-series with Time Dependencies Exceeding t Samples

t	≥ 1	≥ 5	≥ 15	≥ 30	≥ 50
piecewise	0.93	0.57	0.23	0.05	0.04

of scheme replicas since each of the time-series could be processed independently.

3.8 The RCPD Algorithm in a Load Balancing Scenario

In this Section, we demonstrate our proposal in a real content distribution scenario, balancing the traffic between web clients and content caches with a bespoke DNS-based load-balancer. We implement the RCPD algorithm as a client-server MATLAB ® application. The RCPD engine receives periodic content popularity measurements; if a CP is detected, the corresponding upward or downward changes are signalled to the load balancer. The load balancer: (i) distributes the load between the deployed content caches, in a round-robin fashion; (ii) tracks content visits and communicates them to the RCPD engine; and (iii) deploys or removes content caches based on the RCPD outputs.

We implement the web clients using with the httpperf tool (https://github.com/httperf/httperf). The number of clients at each time instance is based on a real time-series of YouTube content views, illustrated in Fig. 3.10a. In practice, an experimental run without the RCPD mechanisms uses three content caches constantly and a run with the RCPD mechanism enabled uses initially two and then three, four and five content caches, after each of the three detected change points, respectively. As we show in Fig. 3.10b, the web clients improve their connectivity times to download the content, while as demonstrated in Fig. 3.10c the CPU utilization in the servers hosting the content remains almost the same. A relevant experimental platform is presented in [7].

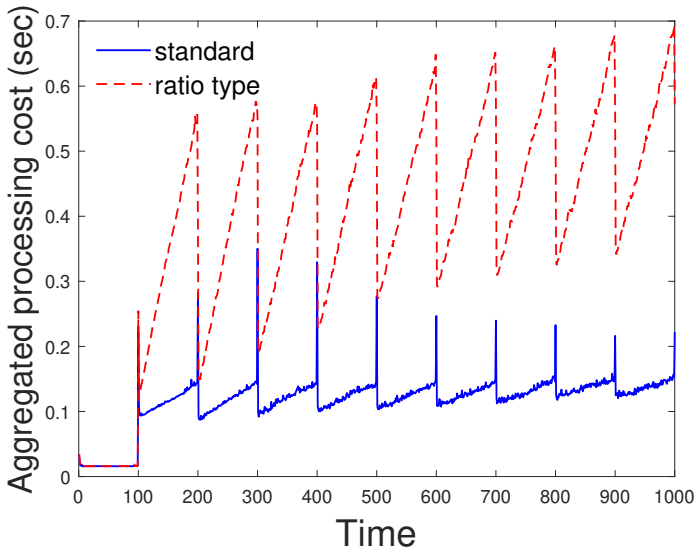

Figure 3.9: The aggregated overall processing cost, per time-instance, of the RCPD algorithm over 882 time-series.

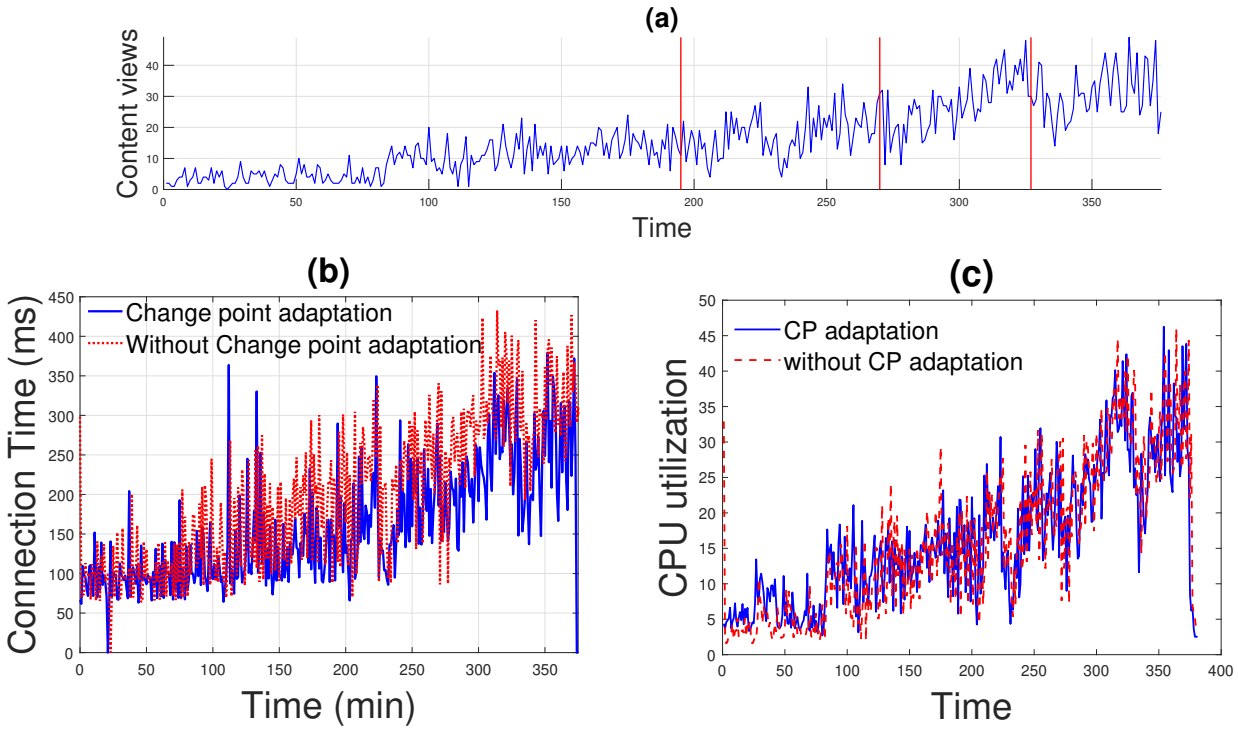

Figure 3.10: a) time-series of video content views, red lines depict the detected CPs, b) the connection time with and without RCPD adaptation and c) the equivalent servers' CPU utilization.

3.9 Application of the RCPD for Intrusion Detection in SDWSNs

Considering the limitations of previous works in SDWSN anomaly detection, outlined in Section 3.3, our main objective is to propose in the remainder of this Chapter a mechanism for DDoS detection with, i) a high detection rate, and, ii) low complexity, so that it would be suitable for "restricted" networks. To this end, we propose the employment of the RCPD. As will be explained in detail next, we study two different DDoS attacks: a false data flow forwarding (FDFF) attack, and a false neighbor

information (FNI) attack, chosen to illustrate the proposed algorithm's capabilities in the case of specific SDWSN vulnerabilities that exhibit largely different behavior. Both attacks are explained in Section 3.9.1, next.

3.9.1 SDWSN security analysis

The SDN networks security threats are grouped in three sets [51]: application plane attacks, control plane attacks, and data plane attacks. Among the three, the control plane attacks are pointed out as the most high impact and attractive [51] [52], as the control plane is responsible for the overall management of the network [53]. This characteristic turns the control plane prone to distributed denial of service (DDoS) attacks. For example, an intruder may flood the network with flow rule requests, which could lead to an exhaustion of the controller's resources. This attack can be intensified using multiple intruders.

The threats and vulnerabilities explained before also apply to SDWSNs. Moreover, there are specific attacks that can attain SDWSNs due to resources constraints, for example: in SDWSN the forwarding devices have low storage capacity, which limits the memory assigned for flow tables and buffers. These constraints make the forwarding devices prone to saturation attacks. Also, SDWSN networks are characterized for having a limited bandwidth and low processing power. This means that a saturation attack can also result in a DoS attack.

Another vulnerability concerns the gateway between the SDN controller and the WSN. The gateway has a radio module of limited bandwidth, rendering it a weak link even when the controller has enough resources to overcome an attack.

For the reasons outlined above, most of the security mechanisms designed for standard SDN networks have to be adapted or redesigned. This is one of the major challenges for SDWSN security.

3.9.2 Impact of DDoS Attacks on Network Performance

Based on SDWSN specific security vulnerabilities, in a previous work, we studied the impact of three DDoS attacks on SDWSN performance [54]. The attacks investigated were: false flow request (FFR), false data flow forwarding (FDFF), and false neighbor information (FNI).

The FFR attack aimed at increasing the SDWSN controller's processing overhead, as well as the packets' traffic, thus, increasing the number of collisions. Each attacker sent multiple flow rule requests to the controller, while the latter calculated the rule and replied to the request. The impact of the attack was observed to be negligible. The FDFF attack followed the FFR attack main idea of sending false flow rule requests to the controller, however, the execution was based on using each attacker's neighbors (benign nodes). Each attacker sent one data packet to its neighbors tagged with an unknown flow identifier; as the neighbors did not have a rule to apply to the packet, they sent a flow request to the controller asking a rule for the unknown flow identifier. Thus, compared to the FFR, the intensity of the attack was multiplied by the number of neighbors. The FDFF attack tripled the number of control packets in the whole network, but had a minor impact on the delivery rate. For both control and data packets, the delivery rate decreased only between 2% and 4%.

In the FNI attack, each attacker intercepted packets containing neighbor information, modified them with false neighbor information and forwarded them to the controller. The controller updated the network topology graph using the false information, and then reconfigured the network with wrong forwarding rules. Our main results [54] showed that the FNI attack could double the number of control packets in the whole network and had a significant impact on the delivery rate. In the case of the control packets, the delivery rate decreased between 35% and 50%. In the case of the data packets, the delivery rate decreased between 20% and 70%.

3.9.3 RCPD for Intrusion Detection

We employed the RCPD algorithm in SDWSNs under FDFF and FNI attacks. We simulated grid topologies with 36 and 100 nodes, varying the number of attackers in the network (5% and 20%). Each simulation run during 10 hours and each scenario was replicated 30 times. During the first 8 hours the network operated normally, then the attack was triggered. The choice of 8 hours was made because empirically it was seen that we needed at least 250 samples for the training period and we obtained one sample every 2 minutes. The simulations were performed using the COOJA simulator [55] and sky motes. The MAC layer was the IEEE 802.15.4, configured to work without radio duty cycle (`nullrdc_driver`). The data sink received the application data, while the management sink received performance metrics information. Notice that the SDN controller is a different node from the sink. Table 3.8 depicts the simulation parameters.

Table 3.8: Simulation Parameters

Simulation parameters	
Topology	Square grid
Number of nodes	36 and 100
Simulation duration	36000 s
Node boot interval	$[0, 1]$ s
Number of sinks	2
Sinks position	Middle of the grid edge
Data traffic rate	1 packet every 30 seconds
Management traffic rate	1 packet every two minutes
Data payload size	10 bytes
Management payload size	10 bytes
Data traffic start time	$[2, 3]$ min
Radio module power	0 dB
Distance between neighbors	50 m
Attacks begins after	28800 s

IT-SDN parameters	
Controller position	center
ND protocol	Collect-based
Link metric	ETX
CD protocol	none
Flow setup	source routed
Route calculation algorithm	Dijkstra
Route recalculation threshold	10%
Flow setup types	regular or source routed
Flow table size	10 entries

We analyzed the data packets delivery rate and the control packets overhead. The delivery rate

was calculated by dividing the total number of packets successfully received by the total number of packets sent. The control packets overhead was quantified as the total amount of control packets sent. Those metrics were updated every two minutes.

The metrics measuring the performance of the intrusion detection algorithm were the following: i) the detection rate (DR); ii) the false positive rate (FPR); iii) the false negative rate (FNR); iv) the detection time median (DTM), indicating the median of the time instances elapsed from the launch of the attack to the instance it was identified; and v) the median absolute deviation (MAD). The detection rate is defined as the ratio between the correctly detected attacks and the total number of attacks. The false positive rate is defined as the ratio between the number of attack events classified as attack and the total number of attack events. The false negative rate is defined as the ratio between attack events classified as non-attack event and the number of attack events. The detection time median is defined as the median of the number of samples required to detect the attack. The median absolute deviation measures the variability of the detection times and is calculated as shown in (3.34), where X_i is the detection time for replication i, and $\tilde{X}$ is the median of all the detection times,

$$\text{MAD} = \text{median}(|X_i - \tilde{X}|). \tag{3.34}$$

The delivery rate and control overhead time series were analyzed for three monitoring windows and three critical values. We used monitoring periods $K \in \{50, 100, 150\}$ samples. This means that the test statistic was run over K samples to extract changes in the mean value. As critical values we used $\alpha \in \{90\%, 95\%, 99\%\}$. Finally, in this analysis, we discarded the first 15 samples because during this time the network was bootstrapping.

3.10 Results and Analysis

In this Section we present and analyze the simulation results. In Section 3.10.1 we compare the FDFF attack detection performance when monitoring the data packets delivery rate and the control overhead. In Section 3.10.2 we repeat this analysis for the FNI attack.

3.10.1 FDFF attack detection

Tables 3.9 and 3.10 summarize the FDFF attack detection results when 5% of nodes are attackers. The results show that when monitoring the data packets delivery rate, the DR is between 57% and 73% for 36 nodes, and between 60% and 83% for 100 nodes. The results when monitoring the control packets overhead show two main points: (i) the algorithm has the same detection performance if configured with a monitoring period K of 50 or 150 samples, and (ii) when the monitoring period is configured as $K = 100$ samples we obtained a DR between 97% and 100%.

Comparing the FPR and the FNR metrics, we observed that the number of cases classified as false negative is higher than the number of cases classified as false positive. This means, it is more common for the algorithm not to detect a change in the metrics when the network is under attack than to detect a suspicious change in a network without attackers. For example, looking at the results when monitoring the control overhead in Table 3.9, only in one out of nine cases the FPR was different than zero. Conversely, the FNR was different than zero in six of nine cases.

The DTM (detection time median) results show that when monitoring the control packets overhead, the attack detection is faster than when monitoring the delivery rate in all the cases. When monitoring the data packets delivery rate, the DTM is between 31 and 37 samples for 36 nodes, and between 20 and 31 samples for 100 nodes. When monitoring the control packets overhead, the DTM is between 9 and 19 samples for 36 nodes, and between 10 and 19 samples for 100 nodes. The fastest detection is obtained monitoring the control packets overhead using a monitoring period of 100 samples, highlighted in red color.

Table 3.9: FDFF Attack Detection, 36 Nodes, 5% Attackers

Data packets delivery rate

K	**50**			**100**			**150**		
α	**90**	**95**	**99**	**90**	**95**	**99**	**90**	**95**	**99**
DTM	31	33	31	31	37	33	31	31	31
MAD	4	6	4	8	9	10	4	4	4
DR	63	67	67	57	70	63	67	73	70
FPR	7	10	7	0	0	0	0	0	0
FNR	30	23	27	43	30	37	33	27	30

Control overhead

K	**50**			**100**			**150**		
α	**90**	**95**	**99**	**90**	**95**	**99**	**90**	**95**	**99**
DTM	19	16	18	12	9	11	19	16	18
MAD	3	3	3	3	2	2	3	3	3
DR	67	73	67	100	97	100	67	73	67
FPR	0	0	0	0	3	0	0	0	0
FNR	33	27	33	0	0	0	33	27	33

Table 3.10: FDFF Attack Detection, 100 nodes, 5% Attackers

Data packets delivery rate

K	**50**			**100**			**150**		
α	**90**	**95**	**99**	**90**	**95**	**99**	**90**	**95**	**99**
DTM	24	26	27	22	20	21	29	31	31
MAD	7	6	13	9	10	11	13	9	15
DR	60	67	67	77	83	73	63	67	63
FPR	23	20	20	10	7	13	0	3	7
FNR	17	13	13	13	1	13	37	30	30

Control overhead

K	**50**			**100**			**150**		
α	**90**	**95**	**99**	**90**	**95**	**99**	**90**	**95**	**99**
DTM	19	17	19	13	10	12	19	17	19
MAD	3	3	3	3	2	3	3	3	3
DR	60	73	63	100	100	100	60	73	63
FPR	0	0	0	0	0	0	0	0	0
FNR	40	27	37	0	0	0	40	27	37

Tables 3.11 and 3.12 summarize the FDFF attack detection results when 20% of nodes are attackers. In the case of 36 nodes, the DR was between 73% and 83% when monitoring the data packets delivery

Table 3.11: FDFF Attack Detection, 36 nodes, 20% Attackers

Data packets delivery rate

K	**50**			**100**			**150**		
α	**90**	**95**	**99**	**90**	**95**	**99**	**90**	**95**	**99**
DTM	28	28	28	30	24	28	29	28	28
MAD	5	8	6	11	7	8	6	5	8
DR	77	80	73	73	83	73	77	80	77
FPR	3	07	7	0	3	0	0	3	0
FNR	20	13	20	27	13	27	23	17	23

Control overhead

K	**50**			**100**			**150**		
α	**90**	**95**	**99**	**90**	**95**	**99**	**90**	**95**	**99**
M	8	7	7	5	5	5	8	7	7
MAD	2	2	2	1	1	1	2	2	2
DR	100	100	100	97	87	97	100	100	100
FPR	0	0	0	3	13	3	0	0	0
FNR	0	0	0	0	0	0	0	0	0

Table 3.12: FDFF Attack Detection, 100 nodes, 20% Attackers

Data packets delivery rate

K	**50**			**100**			**150**		
α	**90**	**95**	**99**	**90**	**95**	**99**	**90**	**95**	**99**
DTM	15	13	14	8	7	7	15	14	14
MAD	5	6	5	6	5	5	5	5	5
DR	100	93	100	97	93	97	100	97	97
FPR	0	7	0	3	7	3	0	3	3
FNR	0	0	0	0	0	0	0	0	0

Control overhead

K	**50**			**100**			**150**		
α	**90**	**95**	**99**	**90**	**95**	**99**	**90**	**95**	**99**
DTM	4	4	4	3	3	3	4	4	4
MAD	0	0	0	0	0	0	0	0	0
DR	100	97	100	97	90	97	100	97	100
FPR	0	3	0	3	10	3	0	3	0
FNR	0	0	0	0	0	0	0	0	0

rate, and between 87% and 100% when monitoring the control packets overhead. In terms of detection time, the best DTM when monitoring the data packets delivery rate was 24 samples and the DTM

Table 3.13: FNI Attack Detection, 36 nodes, 5% Attackers

Data packets delivery rate

K	**50**			**100**			**150**		
α	**90**	**95**	**99**	**90**	**95**	**99**	**90**	**95**	**99**
DTM	7	6	7	8	7	6	7	6	6
MAD	3	4	3	4	3	3	2	4	4
DR	93	83	93	93	80	93	93	83	87
FPR	0	10	0	0	13	0	0	10	7
FNR	7	7	7	7	7	7	7	7	6

Control overhead

K	**50**			**100**			**150**		
α	**90**	**95**	**99**	**90**	**95**	**99**	**90**	**95**	**99**
DTM	28	25	27	35	26	33	28	25	27
MAD	6	7	9	4	3	5	6	7	9
DR	27	33	27	20	27	23	27	33	27
FPR	3	3	3	0	0	0	0	0	0
FNR	70	63	70	80	73	77	73	67	73

Table 3.14: FNI Attack Detection, 100 nodes, 5% Attackers

Data packets delivery rate

K	**50**			**100**			**150**		
α	**90**	**95**	**99**	**90**	**95**	**99**	**90**	**95**	**99**
DTM	6	6	6	6	6	6	6	6	6
MAD	4	4	3	3	3	2	4	4	4
DR	87	93	83	83	83	83	83	90	87
FPR	13	7	17	17	17	17	13	10	13
FNR	0	0	0	0	0	0	3	0	0

Control overhead

K	**50**			**100**			**150**		
α	**90**	**95**	**99**	**90**	**95**	**99**	**90**	**95**	**99**
DTM	34	29	33	35	37	37	34	29	33
MAD	7	7	7	10	7	8	7	8	8
DR	63	70	67	30	47	37	63	70	67
FPR	0	0	0	0	0	0	0	0	0
FNR	37	30	33	70	53	63	37	30	33

when monitoring the control packets overhead was 5 samples. Configuring the monitoring period in 100 we obtain the best DTM, but there was a drop in the DR if compared with the cases when using

Table 3.15: FNI Attack Detection, 36 nodes, 20% Attackers

Data packets delivery rate

K	**50**			**100**			**150**		
α	**90**	**95**	**99**	**90**	**95**	**99**	**90**	**95**	**99**
DTM	7	7	7	7	7	7	8	7	7
MAD	2	2	2	3	4	3	2	2	2
DR	100	100	100	100	100	100	100	100	100
FPR	0	0	0	0	0	0	0	0	0
FNR	0	0	0	0	0	0	0	0	0

Control overhead

K	**50**			**100**			**150**		
α	**90**	**95**	**99**	**90**	**95**	**99**	**90**	**95**	**99**
DTM	26	24	26	26	24	27	26	24	26
MAD	8	7	7	17	11	13	8	7	7
DR	57	70	60	43	63	57	57	70	60
FPR	0	0	0	0	0	0	0	0	0
FNR	43	30	40	57	37	43	43	30	40

Table 3.16: FNI Attack Detection, 100 nodes, 20% Attackers

Data packets delivery rate

K	**50**			**100**			**150**		
α	**90**	**95**	**99**	**90**	**95**	**99**	**90**	**95**	**99**
DTM	9	10	10	8	9	8	10	12	11
MAD	5	8	7	4	6	4	5	9	8
DR	100	100	100	100	100	100	100	100	97
FPR	0	0	0	0	0	0	0	0	3
FNR	0	0	0	0	0	0	0	0	0

Control overhead

K	**50**			**100**			**150**		
α	**90**	**95**	**99**	**90**	**95**	**99**	**90**	**95**	**99**
DTM	27	24	26	26	25	25	27	24	26
MAD	6	3	6	6	6	6	6	3	6
DR	93	97	97	93	97	93	93	97	97
FPR	0	0	0	0	0	0	0	0	0
FNR	7	3	3	7	3	7	7	3	3

monitoring periods of 50 and 150 samples.

The results for 100 nodes showed it is possible to obtain a DR of 100% monitoring any of the

metrics, but there were significant differences in the detection time. The DTM when monitoring the control overhead is between 3 and 4 samples, while when monitoring the data packets delivery rate the DTM was between 7 and 15 samples. Considering the earliest detection with the highest DR for both monitoring metrics, it occurred when using a monitoring period of 100 samples. For both cases the DR obtained was 97%. In terms of FPR and FNR, the best performance was obtained when monitoring the control overhead and using a monitoring period of 50 and 150 samples. Monitoring the control overhead using a monitoring window of 100 samples provided a FPR between 3% and 10%.

Summarizing, the algorithm was able to detect the FDFF attack using either the data packet packets delivery rate or the control packets overhead as inputs. Notably, the algorithm obtained a DR of 100% with both metrics when 20% of nodes behave as attackers. However, aiming for the quickest detection captured through the detection time median, the algorithm achieved far better results when monitoring the control packets overhead in all scenarios. This is a direct consequence of the type of the attack; the attacker creates multiple flow rule request packets to increase the packet traffic and the controller processing overhead. After some time, the flow table of the nodes around the attacker start to saturate, affecting the data packets delivery rate. This means that the change in the delivery will be detected only after the tables saturation; on the contrary, the number of control packets start to change immediately after the attack is triggered.

3.10.2 FNI attack detection

Tables 3.13 and 3.14 summarize the FNI attack detection results when 5% of nodes were attackers. Opposite to the FDFF attack results, the algorithm obtained a better performance detecting the FNI attack when monitoring the data packets delivery rate. In the case of 36 nodes, the DR when monitoring the data packets delivery rate was between 80% and 93%, and the DR when monitoring the control packets overhead was between 23% and 33%. In the case of 100 nodes, the DR when monitoring the data packets delivery rate was between 83% and 93%, and the DR when monitoring the control packets overhead was between 30% and 70%. This means, even the best DR when monitoring the control packets overhead was under the worse DR when monitoring the data packets delivery rate. Also, the results showed that using a critical value of 90%, we obtained a negligible FPR (in our simulation calculated zero). With respect to the DTM, the best result was obtained by monitoring the data packets delivery rate and the control packets overhead were 6 and 25 samples, respectively. This means the algorithm detected the attack four times faster when monitoring the data packets delivery rate. For 100 nodes, the best DTM when monitoring the data packets delivery rate remained in 6 samples, but when monitoring the control packets overhead it was 29 samples.

Lastly, Tables 3.15 and 3.16 summarize the FNI attack detection results when 20% of nodes were attackers. For 36 nodes, the results remained similar to the case of 5% of nodes are attackers. In the case of 100 nodes, the DR when monitoring the data packets delivery rate was between 97% and 100%, and the DR when monitoring the control packets delivery rate was between 93% and 97%. About the DTM, the results for the scenarios when monitoring the data packets delivery rate were between 4 and 9 samples. The results for this same metric when monitoring the control packets overhead were between 24 and 26 samples. This means, for grid topologies with 100 nodes where 20% of nodes were attackers, we obtained similar DRs regardless of the monitoring metric, but when monitoring the delivery rate the detection was at least 3 times faster.

Summarizing our findings, the algorithm was able to detect the FNI attack monitoring either the data packet packets delivery rate or the control packets overhead. Then, comparing the detection performance based on the detection rate and the detection time median, the algorithm obtained a far better performance when monitoring the data packets delivery rate in all scenarios. This effect was directly related to the type of the attack; in the FNI attack, the attackers intercept the control packets that contained neighbor information, modify them, and then forward them to the controller. This means this attack could lead to a network misconfiguration using few control packets.

3.11 Conclusion

In this Chapter, we proposed the RCPD, a novel algorithm for the real-time detection of changes in the mean value of content popularity. Approaching the problem statistically, we efficiently combined off-line and on-line non-parametric CUSUM procedures to avoid restrictive assumptions for content popularity behavior and to reduce the overall computational cost. We divided the algorithm in two phases. The first phase was an extended retrospective (off-line) procedure with a modified BS algorithm and was used to adjust on-line parameters, based on historical data of the particular video. The second phase integrated one of two alternative trend indicators to the sequential (on-line) procedure, to reveal the direction of a detected change. We provided extensive simulations, using synthetic and real data, that demonstrated the performance of the proposed algorithm for the successful identification of content popularity changes in real-time. We also demonstrated through experimental measurements that the RCPD's processing cost is almost imperceptible. Finally we provided proof-of-concept by applying the algorithm in a load balancing application, highlighting its efficiency in a realistic setting.

Furthermore, we have used the RCPD for intrusion detection in SDWSNs. We performed experiments for two SDWSN DDoS attacks, in topologies of 36 and 100 nodes, and with varying number of attackers. Our results showed that it is feasible to detect different types of attacks by monitoring either the data packets delivery rate or control packets metrics. As the detector's algorithmic complexity is linear to the size of the network and the number of metrics monitored, the proposed approach could scale to include other metrics.

References

[1] Tom Goethals, Merlijn Sebrechts, Ankita Atrey, Bruno Volckaert, and Filip De Turck. Unikernels vs containers: An in-depth benchmarking study in the context of microservice applications. In *IEEE Int. Symp. Cloud Service Comput. (SC2)*, Nov. 2018.

[2] Joao Martins, Ahmed Mohamed, Costin Raiciu, and Felipe Huici. Enabling fast, dynamic networking processing with ClickOS. In *Second ACM SIGCOMM Workshop on Hot Topics in Software Defined Networking*, pages 67–72, 2013.

[3] A. Wang, Z. Zha, Y. Guo, and S. Chen. Software-defined networking enhanced edge computing: A network-centric survey. *Proc. IEEE*, 107:1500–1519, Aug. 2019.

[4] CISCO Visual Networking. Cisco global cloud index: forecast and methodology, 2015-2020. San Jose, CA, USA, CISCO, Tech. Rep., 2017.

[5] Nick McKeown, Tom Anderson, Hari Balakrishnan, Guru Parulkar, Larry Peterson, Jennifer Rexford, et al. Openflow: enabling innovation in campus networks. *ACM SIGCOMM Comput. Commun. Rev.*, 38(2):69–74, Mar. 2008.

[6] Necos project: Towards lightweight slicing of cloud federated infrastructures. https://intrig.dca.fee.unicamp.br/2017/09/05/necos-2-year-eu-brazil-collaborative-project-starting-in-nov2017/.

[7] Polychronis Valsamas, Sotiris Skaperas, and Lefteris Mamatas. Elastic content distribution based on unikernels and change-point analysis. In *Proc. 24th Eur. Wireless Conf. (EW)*, pages 1–7, Catania, Italy, May 2018.

[8] Alexandru Tatar, Marcelo Dias De Amorim, Serge Fdida, and Panayotis Antoniadis. A survey on predicting the popularity of web content. *J. Internet Services Appl.*, 5(1):8, Dec. 2014.

[9] Gabor Szabo and Bernardo A Huberman. Predicting the popularity of online content. *Commun. ACM*, 53(8):80–88, Aug. 2010.

[10] Gonca Gürsun, Mark Crovella, and Ibrahim Matta. Describing and forecasting video access patterns. In *Proc. IEEE Int. Conf. Comput. Commun. (IEEE INFOCOM)*, pages 16–20, Shanghai, China, Apr. 2011.

[11] Justin Cheng, Lada Adamic, P Alex Dow, Jon Michael Kleinberg, and Jure Leskovec. Can cascades be predicted? In *Proc. 23rd Int. Conf. World Wide Web (WWW)*, pages 925–936, Seoul, Republic of Korea, Apr. 2014.

[12] Stefan Fremdt. Asymptotic distribution of the delay time in page's sequential procedure. *J. Statist. Planning Inference*, 145:74–91, Feb. 2014.

[13] Yannick Hoga. Monitoring multivariate time series. *J. Multivariate Anal.*, 155:105–121, Mar. 2017.

[14] E Brodsky and Boris S Darkhovsky. *Nonparametric methods in change point problems*. Dordrecht, The Netherlands: Kluwer, 2013.

[15] Donald J Berndt and James Clifford. Using dynamic time warping to find patterns in time series. In *Proc. AAAI Workshop Knowl. Disc. Databases (KDD)*, volume 10, pages 359–370, Seattle, USA, Aug. 1994.

[16] Rohit J Kate. Using dynamic time warping distances as features for improved time series classification. *Data Mining Knowledge Discovery*, 30:283–312, Mar. 2016.

[17] D. Kreutz, F. M. V. Ramos, P. E. Veríssimo, C. E. Rothenberg, S. Azodolmolky, and S. Uhlig. Software-Defined Networking: A Comprehensive Survey. *Proc. IEEE Proc.*, 103(1):14–76, Jan 2015.

[18] H. I. Kobo, A. M. Abu-Mahfouz, and G. P. Hancke. A Survey on Software-Defined Wireless Sensor Networks: Challenges and Design Requirements. *IEEE Access*, 5:1872–1899, 2017.

[19] Nick McKeown, Tom Anderson, Hari Balakrishnan, Guru Parulkar, Larry Peterson, Jennifer Rexford, Scott Shenker, and Jonathan Turner. OpenFlow: Enabling Innovation in Campus Networks. *SIGCOMM Comput. Commun. Rev.*, 38(2):69–74, March 2008.

[20] R. C. A. Alves, D. A. G. Oliveira, G. A. Nunez Segura, and C. B. Margi. The Cost of Software-Defining Things: A Scalability Study of Software-Defined Sensor Networks. *IEEE Access*, 7:115093–115108, Aug 2019.

[21] Sotiris Skaperas, Lefteris Mamatas, and Arsenia Chorti. Early Video Content Popularity Detection with Change Point Analysis. In *IEEE Global Commun. Conf. (GLOBECOM)*, Abhu-Dhabi, United Arab Emirates, December 2018.

[22] S. Skaperas, L. Mamatas, and A. Chorti. Real-Time Video Content Popularity Detection Based on Mean Change Point Analysis. *IEEE Access*, 7:142246–142260, 2019.

[23] Henrique Pinto, Jussara M Almeida, and Marcos A Gonçalves. Using early view patterns to predict the popularity of youtube videos. In *Proc. 6th ACM Int. Conf. Web Search and Data Mining (WSDM)*, pages 365–374, Rome, Italy, Feb. 2013.

[24] Sotiris Skaperas, Lefteris Mamatas, and Arsenia Chorti. Early video content popularity detection with change point analysis. In *Proc. IEEE Global Commun. Conf. (IEEE GLOBECOM)*, pages 1–7, Abu Dhabi, UAE, Dec. 2018.

[25] Michèle Basseville, Igor V Nikiforov, et al. *Detection of abrupt changes: theory and application*, volume 104. Prentice Hall Englewood Cliffs, 1993.

[26] Varun Chandola, Arindam Banerjee, and Vipin Kumar. Anomaly detection: A survey. *ACM Comput. Surveys (CSUR)*, 41(3):1–58, Sept. 2009.

[27] Alexander G Tartakovsky, Aleksey S Polunchenko, and Grigory Sokolov. Efficient computer network anomaly detection by changepoint detection methods. *IEEE J. Sel. Topics Signal Process.*, 7(1):4–11, Feb. 2013.

[28] Alexander G Tartakovsky, Boris L Rozovskii, Rudolf B Blazek, and Hongjoong Kim. A novel approach to detection of intrusions in computer networks via adaptive sequential and batch-sequential change-point detection methods. *IEEE Trans. Signal Process.*, 54(9):3372–3382, Sept. 2006.

[29] Haining Wang, Danlu Zhang, and Kang G Shin. Change-point monitoring for the detection of dos attacks. *IEEE Trans. Depend. Sec. Comput.*, 1(4):193–208, Oct.-Dec. 2004.

[30] Yanxiang Jiang, Miaoli Ma, Mehdi Bennis, Fuchun Zheng, and Xiaohu You. A novel caching policy with content popularity prediction and user preference learning in fog-ran. In *Proc. IEEE Global Commun. Conf. (IEEE GLOBECOM) Workshops*, pages 1–6, 2017.

[31] Gautam Thatte, Urbashi Mitra, and John Heidemann. Parametric methods for anomaly detection in aggregate traffic. *IEEE/ACM Trans. Netw. (TON)*, 19(2):512–525, Apr. 2011.

[32] Augustin Soule, Kavé Salamatian, and Nina Taft. Combining filtering and statistical methods for anomaly detection. In *Proc. 5th ACM SIGCOMM Conf. Internet Measurement*, pages 1–14, New York, NY, USA, Oct. 2005.

[33] Ido Nevat, Dinil Mon Divakaran, Sai Ganesh Nagarajan, Pengfei Zhang, Le Su, Li Ling Ko, and Vrizlynn LL Thing. Anomaly detection and attribution in networks with temporally correlated traffic. *IEEE/ACM Trans. Netw. (TON)*, 26(1):131–144, Feb. 2018.

[34] S. S. Bhunia and M. Gurusamy. Dynamic attack detection and mitigation in IoT using SDN. In *27th Int. Telecommun. Netw. and Appl. Conf. (ITNAC)*, pages 1–6, Nov 2017.

[35] D. Yin, L. Zhang, and K. Yang. A DDoS Attack Detection and Mitigation With Software-Defined Internet of Things Framework. *IEEE Access*, 6:24694–24705, 2018.

[36] Rui Wang, Zhiyong Zhang, Zhiwei Zhang, and Zhiping Jia. ETMRM: An Energy-efficient Trust Management and Routing Mechanism for SDWSNs. *Computer Networks*, 139:119 – 135, 2018.

[37] Xiaobo Zhou and Cheng-Zhong Xu. Optimal video replication and placement on a cluster of video-on-demand servers. In *in Proc. Int. Conf. Parallel Process. (ICPP)*, pages 547–555, Vancouver, Canada, Aug. 2002.

[38] Wenting Tang, Yun Fu, Ludmila Cherkasova, and Amin Vahdat. Modeling and generating realistic streaming media server workloads. *Comput. Netw.*, 51(1):336–356, Jan. 2007.

[39] Alexander Aue and Lajos Horváth. Structural breaks in time series. *J. Time Series Anal.*, 34(1):1–16, Jan. 2013.

[40] Donald WK Andrews. Heteroskedasticity and autocorrelation consistent covariance matrix estimation. *Econometrica: J. Econometric Soc.*, 59:817–858, May 1991.

[41] Dominik Wied. A nonparametric test for a constant correlation matrix. *Econometric Rev.*, 36(10):1157–1172, Apr. 2017.

[42] Marc Lavielle and Gilles Teyssiere. Adaptive detection of multiple change-points in asset price volatility. In *Long Memory in Economics*, pages 129–156. Springer, G. Teyssiere and A. Kirkman, Eds. Berlin, Germany: Springer–Verlag, 2007.

[43] Daniele Angelosante and Georgios B Giannakis. Sparse graphical modeling of piecewise-stationary time series. In *Proc. IEEE Int. Conf. Acoust., Speech and Signal Process (IEEE ICASSP)*, pages 1960–1963, Prague, Czech Republic, May 2011.

[44] Carla Inclan and George C Tiao. Use of cumulative sums of squares for retrospective detection of changes of variance. *J. Amer. Statist. Assoc.*, 89(427):913–923, Sept. 1994.

[45] Huang Kai, Qi Zhengwei, and Liu Bo. Network anomaly detection based on statistical approach and time series analysis. In *Proc. Int. Conf. Advanced Inform. Netw. Appl. (WAINA) Workshops*, pages 205–211, Bradford, UK, May 2009.

[46] Nesrine Ben Hassine, Ruben Milocco, and Pascale Minet. Arma based popularity prediction for caching in content delivery networks. In *Proc. Wireless Days*, pages 113–120, Porto, Portugal, Mar. 2017.

[47] Dominik Wied and Pedro Galeano. Monitoring correlation change in a sequence of random variables. *J Statist. Planning Inference*, 143(1):186–196, Jan. 2013.

[48] Mattia Zeni, Daniele Miorandi, and Francesco De Pellegrini. Youstatanalyzer: a tool for analysing the dynamics of youtube content popularity. In *Proc. 7th Int. Conf. Perform. Eval. Methodol. Tools*, pages 286–289, Torino, Italy, Dec. 2013.

[49] Richard G Clegg. A practical guide to measuring the hurst parameter. *Int. J. Simul. Syst. Sci. Technol.*, 7(2):3–14, Nov. 2006.

[50] Jean-Marc Bardet, Gabriel Lang, Georges Oppenheim, Anne Philippe, Stilian Stoev, and Murad S Taqqu. Semi-parametric estimation of the long-range dependence parameter: a survey. *Theory and applications of long-range dependence*, pages 557–577, 2003.

[51] I. Ahmad, S. Namal, M. Ylianttila, and A. Gurtov. Security in Software Defined Networks: A Survey. *IEEE Commun. Surveys Tuts.*, 17(4):2317–2346, Fourthquarter 2015.

[52] Zhaogang Shu, Jiafu Wan, Di Li, Jiaxiang Lin, Athanasios V. Vasilakos, and Muhammad Imran. Security in Software-Defined Networking: Threats and Countermeasures. *Mobile Netw. and Appl.*, 21(5):764–776, Oct 2016.

[53] A. Akhunzada, E. Ahmed, A. Gani, M. K. Khan, M. Imran, and S. Guizani. Securing software defined networks: taxonomy, requirements, and open issues. *IEEE Commun. Mag.*, 53(4):36–44, April 2015.

[54] Gustavo A. Nunez Segura, Cintia B. Margi, and Arsenia Chorti. Understanding the Performance of Software Defined Wireless Sensor Networks Under Denial of Service Attack. *Open Journal of Internet Of Things (OJIOT)*, 2019. Special Issue: Proc. Int. Workshop Very Large Internet of Things (VLIoT 2019) in conjunction with the VLDB 2019 Conf. Los Angeles, United States.

[55] F. Osterlind, A. Dunkels, J. Eriksson, N. Finne, and T. Voigt. Cross-Level Sensor Network Simulation with COOJA. In *Proc. IEEE Conf. Local Comput. Netw. (LCN)*, pages 641–648, Nov 2006.

Chapter 4

Uplink Non-Orthogonal Multiple Access (NOMA) Under Statistical QoS Delay Constraints

4.1 Introduction

Various verticals in 5G and beyond (B5G) networks require very stringent latency guarantees, while at the same time envisioning massive connectivity. As a result, choosing the optimal multiple access (MA) technique to achieve low latency is a key enabler of B5G. In particular, this issue is more acute in uplink transmissions due to the potentially high number of collisions. On this premise, in the present contribution we discuss the issue of delay-sensitive uplink connectivity; to this end, we perform a comparative analysis of various MA approaches with respect to the achievable effective capacity (EC). As opposed to standard rate (PHY) or throughput (MAC) analyses, we propose the concept of the effective capacity as a suitable metric for characterizing jointly PHY-MAC layer delays. The palette of investigated MA approaches includes standard orthogonal MA (OMA) and power domain non-orthogonal MA (NOMA) in uplink scenarios.

For two-user networks, we propose novel closed-form expressions for the EC of the NOMA users and show that in the high signal to noise ratio (SNR) region, the "strong" NOMA user has a limited EC, assuming the same delay constraint as the "weak" user. We demonstrate that for the weak user, OMA achieves higher EC than NOMA at small values of the transmit SNR, while NOMA outperforms OMA in terms of EC at high SNRs. On the other hand, for the strong user the opposite is true, i.e., NOMA achieves higher EC than OMA at small SNRs, while OMA becomes more beneficial at high SNRs. This result raises the question of introducing "adaptive" OMA / NOMA policies, based jointly on the users' delay constraints as well as on the available transmit power.

4.2 Contributions and Chapter organization

Non-orthogonal multiple access (NOMA) schemes have attracted a lot of attention recently, allowing multiple users to be served simultaneously with enhanced spectral efficiency; it is known that the boundary of achievable rate pairs (in the case of two users) using NOMA is outside the capacity region achievable with orthogonal multiple access (OMA) techniques [1] or other schemes [2]. Superior achievable rates are attainable through the use of superposition coding at the transmitter and of successive interference cancellation (SIC) at the receiver [3, 4]. The SIC receiver decodes multi-user signals with descending received signal power and subtracts the decoded signal(s) from the received superimposed signal, so as to improve the signal-to-interference ratio. The process is repeated until the signal of interest is decoded. In uplink NOMA networks, the strongest user's signal is decoded first (as opposed to downlink NOMA networks in which the inverse order is applied).

Besides, in a number of emerging applications, delay QoS requirements become increasingly important, e.g., for URLLC systems. Furthermore, in future wireless networks, users are expected to necessitate flexible delay guarantees for achieving different service requirements. In order to satisfy diverse delay requirements, a simple and flexible delay QoS model is imperative to be applied and investigated. In this respect, the EC theory can be employed [5], [6] [7], with EC denoting the maximum constant arrival rate which can be served by a given service process, while guaranteeing the required statistical delay provisioning. We studied delay-constrained downlink NOMA networks in [4] and with secrecy constraints [8] in [9]. The present analysis complements [4], focusing on uplink transmissions. In the following Sections, we derive novel closed-form expressions for the ECs of a two user network; we then provide four Lemmas for the asymptotic performance of the network with NOMA and OMA. The conclusions drawn are supported by an extensive set of simulations.

The rest of the Chapter is organized as follows. In Section 4.3 we investigate the EC of a two user uplink NOMA system under statistical delay QoS constraints. Simulation results are given in Section 4.4, followed by conclusions in Section 4.5.

4.3 Effective Capacity of Two-user NOMA Uplink Network

Assume a two-user NOMA uplink network with users U_1 and U_2 in a Rayleigh block fading propagation channel, with respective channel gains during a transmission block denoted by $|h_1|^2 < |h_2|^2$. The users transmit corresponding symbols S_1, S_2 respectively, with power $\mathbb{E}[|S_i|^2] = P_i, i = 1, 2$ and total power $P_T = \sum_{i=1}^{2} P_i = 1$. Here, P_i is the power coefficient for the user i [10]. The received superimposed signal can be expressed as [11]

$$Z = \sum_{i=1}^{2} \sqrt{P_i} h_i S_i + w, \tag{4.1}$$

where w denotes a zero mean circularly symmetric complex Gaussian random variable with variance σ^2. The receiver will first decode the symbol of the strong user treating the transmission of the weak as interference. After decoding it, the receiver will suppress it from Z and decode the signal of the weak user. Following the SIC principle and denoting by $\rho = \frac{1}{\sigma^2}$ the transmit SNR, the achievable rates, in b/s/Hz, for user $\mathrm{U}_i, i = 1, 2$, is expressed as: [12]

$$R_i = \log_2 \left(1 + \frac{\rho P_i |h_i|^2}{1 + \rho \sum_{l=1}^{i-1} P_l |h_l|^2} \right). \tag{4.2}$$

Introducing statistical delay QoS constraints, let θ_i be the statistical delay QoS exponent of the i-th user, and assume that the service process satisfies the Gärtner-Ellis theorem [6]. The delay exponent θ_i captures how strict the delay constraint is [6]. A slower decay rate can be represented by a smaller θ_i, which indicates that the system is more delay tolerant, while a larger θ_i corresponds to a system with more stringent QoS requirements. Applying the EC theory in a uplink NOMA with two users, the i-th user's EC over a block-fading channel, is defined as:

$$E_c^i = -\frac{1}{\theta_i T_f B} \ln \left(\mathbb{E} \left[e^{-\theta_i T_\mathrm{f} B R_i} \right] \right) \quad \text{(in b/s/Hz)}, \tag{4.3}$$

where T_f is the fading-block length, B is the bandwidth and $\mathbb{E}[\cdot]$ denotes expectation over the channel gains. By inserting R_i into (4.3), we obtain the following expression for the EC of the i-th user

$$E_c^i = \frac{1}{\beta_i} \log_2 \left(\mathbb{E} \left[(1 + \frac{\rho P_i |h_i|^2}{1 + \rho \sum_{l=1}^{i-1} P_l |h_l|^2})^{\beta_i} \right] \right), \tag{4.4}$$

where $\beta_i = -\frac{\theta_i T_f B}{\ln 2}$, $i = 1, 2$, is the normalized (negative) QoS exponent.

4.3.1 ECs in a Two-user NOMA Uplink Network

For the ordering of the channel gains we make use of the theory of order statistics in the following analysis [13]. Assuming a Rayleigh wireless environment, the channel gains, denoted by $x_i = |h_i|^2, i = 1, 2$, are exponentially distributed with probability density function (PDF) and cumulative density function (CDF) respectively given by $f(x_i) = e^{-x_i}$, $F(x_i) = 1 - e^{-x_i}$. Then, according to order statistics [13], the ordered channel gains have respective PDFs $f_{i:2}(x_i), i = 1, 2$, and joint PDF $f(x_1, x_2)$ that are expressed as

$$f_{1:2}(x_1) = 2e^{-2x_1}, \tag{4.5}$$
$$f_{2:2}(x_2) = 2e^{-x_2}\left(1 - e^{-x_2}\right), \tag{4.6}$$
$$f(x_1, x_2) = 2e^{-x_1}e^{-x_2}. \tag{4.7}$$

As a result, the EC of User 1, denoted by E_c^1 is expressed as

$$\begin{aligned} E_c^1 &= \frac{1}{\beta_1}\log_2(\mathbb{E}[(1+\rho P_1 x_1)^{\beta_1}]) = \frac{1}{\beta_1}\log_2\left(\int_0^\infty (1+\rho P_1 x_1)^{\beta_1} f_{1:2}(x_1)dx_1\right) \\ &= \frac{1}{\beta_1}\log_2\left(\frac{2}{P_1\rho}\times U\left(1, 2+\beta_1, \frac{2}{\rho P_1}\right)\right). \end{aligned} \tag{4.8}$$

where $U(\cdot,\cdot,\cdot)$ denotes the confluent hypergeometric function [4]. On the other hand, the EC of the User 2 is evaluated as

$$\begin{aligned} E_c^2 &= \frac{1}{\beta_2}\log_2\left(\mathbb{E}\left[\left(1+\frac{\rho P_2 x_2}{1+\rho P_1 x_1}\right)^{\beta_2}\right]\right) = \frac{1}{\beta_2}\log_2\left(\int_0^\infty\int_{x_1}^\infty\left(1+\frac{\rho P_2 x_2}{1+\rho P_1 x_1}\right)^{\beta_2} f(x_1,x_2)dx_2dx_1\right) \\ &= \frac{1}{\beta_2}\log_2\left(2P_2^{1-\beta_2}(\rho P_2)^{\beta_2}e^{\frac{1}{\rho P_2}}e^{-\frac{(P_1-P_2)}{\rho P_2}}\right) + \frac{1}{\beta_2}\log_2\left(\sum_{j=0}^{-\beta_2}\binom{-\beta_2}{j}(\rho P_1)^j \times \sum_{k=0}^{\infty}\frac{(-1)^k(P_2-P_1)^k}{k!(1+j+k)}\right. \\ &\left.\times\left[\Gamma[2+\beta_2+j+k, \frac{1}{\rho P_2}] - (\rho P_2)^{-1-j-k}\Gamma[1+\beta_2, \frac{1}{\rho P_2}]\right]\right), \end{aligned} \tag{4.9}$$

with $\Gamma(\cdot,\cdot)$ denoting the incomplete Gamma function [4]. The proof for deriving E_c^1 is omitted as it can be verified with standard software (MAPLE or Mathematica) while for E_c^2 is provided in Appendix I.

In order to perform a comparative performance analysis, here we provide the achievable data rates for a two-user OMA network, denoted by $\widetilde{R}_i, i = 1, 2$, given as

$$\widetilde{R}_i = \frac{1}{2}\log_2\left(1+\rho P_T|h_i|^2\right), i = 1, 2 \tag{4.10}$$

Note that the coefficient $\frac{1}{2}$ is due to the equal allocation of resources to both users. The corresponding expressions are obtained for the ECs of both users in a OMA network, denoted by $\widetilde{E}_c^i$, given as

$$\tilde{E}_c^i = \frac{1}{\beta_i}\log_2\left(\mathbb{E}\left[(1+\rho P_T|h_i|^2)^{\frac{\beta_i}{2}}\right]\right)$$

so that,

$$\begin{aligned} \tilde{E}_c^1 &= \frac{1}{\beta_1}\log_2\left(\frac{2}{\rho}\times U\left(1, 2+\frac{\beta_1}{2}, \frac{2}{\rho}\right)\right) \\ \widetilde{E}_c^2 &= \frac{1}{\beta_2}\log_2\left(\frac{2}{\rho}\sum_{k=0}^{1}\binom{1}{k}(-1)^k \times U\left(1, 2+\frac{\beta_2}{2}, \frac{1+k}{\rho}\right)\right) \end{aligned} \tag{4.11}$$

The proof is omitted as it can be verified with software (MAPLE or Mathematica).

4.3.2 Asymptotic Analysis

We first perform an asymptotic analysis with respect to the SNR. Our results are summarized in Lemma 1.

Lemma 1: In the low and high SNR regimes, respectively, the following conclusions hold:

1. *When* $\rho \to 0$, *then,* $E_c^1 \to 0$, $E_c^2 \to 0$, $\widetilde{E}_c^1 \to 0$, $\widetilde{E}_c^2 \to 0$, $E_c^1 - \widetilde{E}_c^1 \to 0$, $E_c^2 - \widetilde{E}_c^2 \to 0$;
2. *When* $\rho \to +\infty$, *then* $E_c^1 \to +\infty$, $E_c^2 \to \frac{1}{\beta_2}\log_2\left(\mathbb{E}\left[\left(1+\frac{P_2|h_2|^2}{P_1|h_1|^2}\right)^{\beta_2}\right]\right)$, $\widetilde{E}_c^1 \to +\infty$, $\widetilde{E}_c^2 \to +\infty$, $E_c^1 - \widetilde{E}_c^1 \to +\infty$, $E_c^2 - \widetilde{E}_c^2 \to -\infty$.

Proof: The proof is provided in Appendix II.

Lemma 1 indicates that the ECs of both users are vanishingly small at low values of ρ, irrespective of employing NOMA or OMA. On the other hand, at high SNRs, we notice that the EC of the strong user with NOMA is limited to a finite value. On the contrary, for the weaker user, when $\rho >> 1$, its achievable EC in the NOMA uplink increases without bound. This is the exact opposite of the downlink scenario, where it is the weaker user which is limited in terms of EC, when $\rho >> 1$ [4].

Now, the question is how the ECs evolve with ρ between the two asymptotic regimes. To answer this question and to further analyze the impact of ρ on the individual EC, we look at the derivatives with respect to ρ [4] in Lemma 2.

Lemma 2: For the EC of User 1, in a two-user uplink network the following hold:

1. $\frac{\partial E_c^1}{\partial \rho} \geq 0$ *and* $\frac{\partial \widetilde{E}_c^1}{\partial \rho} \geq 0$, $\forall \rho$;
2. *When* $\rho \to 0$, *then* $\lim_{\rho\to 0}\left(\frac{\partial(E_c^1 - \widetilde{E}_c^1)}{\partial \rho}\right) = \frac{P_1 - \frac{1}{2}}{\ln 2}\mathbb{E}[|h_1|^2]$;
3. *When* $\rho >> 1$, *then* $\frac{\partial(E_c^1 - \widetilde{E}_c^1)}{\partial \rho} \approx \frac{1}{2\rho \ln 2} \geq 0$ *and it approaches* 0 *when* $\rho \to \infty$.

Proof: The proof is provided in Appendix III.

Lemma 2 indicates that for User 1, when the transmit SNR ρ is very small, the EC with OMA increases faster than the EC with NOMA. On the other hand, Lemma 2 shows that when the transmit SNR is very large, the EC with NOMA increases faster than with OMA.

Combining Lemma 2 and Lemma 1, we can conclude that, $E_c^1 - \widetilde{E}_c^1$ starts at vanishingly small value, first decreases, and subsequently increases to ∞ at a gradually reducing speed. This means that for the weaker user, OMA achieves higher EC than NOMA at small values of the transmit SNR ρ. At high values of ρ, NOMA becomes more beneficial for the weak user. Finally, when $\rho \to \infty$ the performance gain of NOMA over OMA reaches a constant value in the case of User 1.

Lemma 3: For the EC of User 2, in a two-user uplink network the following hold:

1. $\frac{\partial E_c^2}{\partial \rho} \geq 0$ *and* $\frac{\partial \widetilde{E}_c^2}{\partial \rho} \geq 0$, $\forall \rho$;
2. *When* $\rho \to 0$, *then* $\lim_{\rho\to 0}\left(\frac{\partial(E_c^2 - \widetilde{E}_c^2)}{\partial \rho}\right) = \frac{P_2}{2\ln 2}\mathbb{E}[|h_2|^2]$
3. *When* $\rho >> 1$, *then* $\frac{\partial(E_c^2 - \widetilde{E}_c^2)}{\partial \rho} \approx -\frac{1}{2\ln 2}\frac{1}{\rho} < 0$ *and it approaches 0 when* $\rho \to \infty$.

Proof: The proof is provided in Appendix IV.

Lemma 3 indicates that, for User 2, when the transmit SNR ρ is very small, the uplink EC with NOMA increases faster than that with OMA. On the other hand, when the transmit SNR is very large, the uplink EC with OMA increases faster than that with NOMA. Combining Lemma 3 and Lemma 1, we can conclude that, $E_c^2 - \widetilde{E}_c^2$ starts at an initial vanishingly small value, first increases, and subsequently decreases to $-\infty$ with a gradually diminishing rate. This means that for the stronger user, NOMA achieves higher EC than OMA at small values of the transmit SNR ρ. At high values of

ρ, OMA becomes more beneficial for the strong user. Finally, when $\rho \to \infty$ the performance gain of OMA over NOMA reaches a constant value, for the stronger user.

Finally, we investigate the sum ECs when using OMA and NOMA, denoted by V_N and V_O,

$$V_N = E_c^1 + E_c^2, \tag{4.12}$$
$$V_O = \tilde{E}_c^1 + \tilde{E}_c^2. \tag{4.13}$$

Our conclusions are drawn in Lemma 4.

Lemma 4: For the sum EC with NOMA, denoted by V_N, and with OMA, denoted by V_O, in a two-user uplink network, the following hold:

1. $\frac{\partial V_N}{\partial \rho} \geq 0$ *and* $\frac{\partial V_O}{\partial \rho} \geq 0$, $\forall \rho$;
2. *When* $\rho \to 0$, $V_N \to 0$, $\lim_{\rho \to 0}(\frac{\partial V_N}{\partial \rho}) = \frac{P_1}{\ln 2}\mathbb{E}[|h_1|^2] + \frac{P_2}{\ln 2}\mathbb{E}[|h_2|^2] \geq 0$, *and* $V_O \to 0$, $\lim_{\rho \to 0}(\frac{\partial V_O}{\partial \rho}) = \frac{P_1}{2\ln 2}\mathbb{E}[|h_1|^2] + \frac{P_2}{2\ln 2}\mathbb{E}[|h_2|^2] \geq 0$;
3. *When* $\rho >> 1$, $V_N \to \infty$, $\lim_{\rho \to \infty}(\frac{\partial V_N}{\partial \rho}) = 0$, *and* $V_O \to \infty$, $\lim_{\rho \to \infty}(\frac{\partial V_O}{\partial \rho}) = 0$.

The proof is provided in Appendix V.

Lemma 4 indicates that when NOMA is applied, the sum EC has a constant increasing rate at small value of the transmit SNR ρ that depends on the average of the channel power gains and the allocated power coefficients. A similar conclusion is reached when using OMA. On the other hand, when $\rho >> 1$, Lemma 4 indicates that the rate at which the sum ECs increase reaches a plateau, both in the case of NOMA and OMA.

4.4 Numerical Results

In this Section, the Lemmas presented in Section 4.3 are validated through Monte Carlo simulations. We consider a two user uplink NOMA system, with the following settings: normalized transmission power levels for both users, $P_1 = 0.2$, $P_2 = 0.8$, normalized delay exponent $\beta_1 = \beta_2 = -1$ for both users, unless otherwise stated.

In Fig. 4.1 the ECs of the two-user uplink NOMA and OMA networks are depicted versus the transmit SNR. We note that for the weak user, OMA is more advantageous than NOMA for low transmit SNRs, and NOMA is more advantageous than OMA at high transmit SNRs. Reverse conclusions can be drawn for the strong user. We notice also that the EC of the strong user converges at high SNRs. This provides numerical validation for Lemma 1.

Figs. 4.2 and 4.3, show respectively the EC of User 1 and User 2, versus the transmit SNR, for different values of $\beta_1 = \beta_2 = \beta$. When the delay constraints become more stringent, i.e., β decreases (equivalently, θ increases), the individual link-layer rates in NOMA decrease, for both users.

In Fig. 4.4, the ECs of the strong and weak users are depicted across different SNR values, $\rho \in \{1, 10, 30, 40, 50\}$ dB, as functions of the (negative) normalized delay exponent, for NOMA and OMA scenarios. We notice that the EC of each user is identical for NOMA and OMA, for small and large values of the normalized delay exponent. And with incresing transmit SNR ρ, the EC increases for both users.

Fig. 4.5 shows the difference of the EC in NOMA and the EC in OMA of the weak user. This curve starts initially at zero, then decreases to a certain minimum and starts increasing at the high values of transmit SNR. This confirms Lemma 2. When the delay is equal to -1, we see that for $\rho \in [0, 30]$ dB, the difference values are negative, indicating that OMA outperforms NOMA in this range. But when $\rho > 30$ dB, the values are positive, i.e., NOMA offers better link-layer rates. However, the particular ranges depend not only on the delay exponents but also on the power allocation coefficients. By increasing the transmission power of the weak user and reducing the transmission power of the

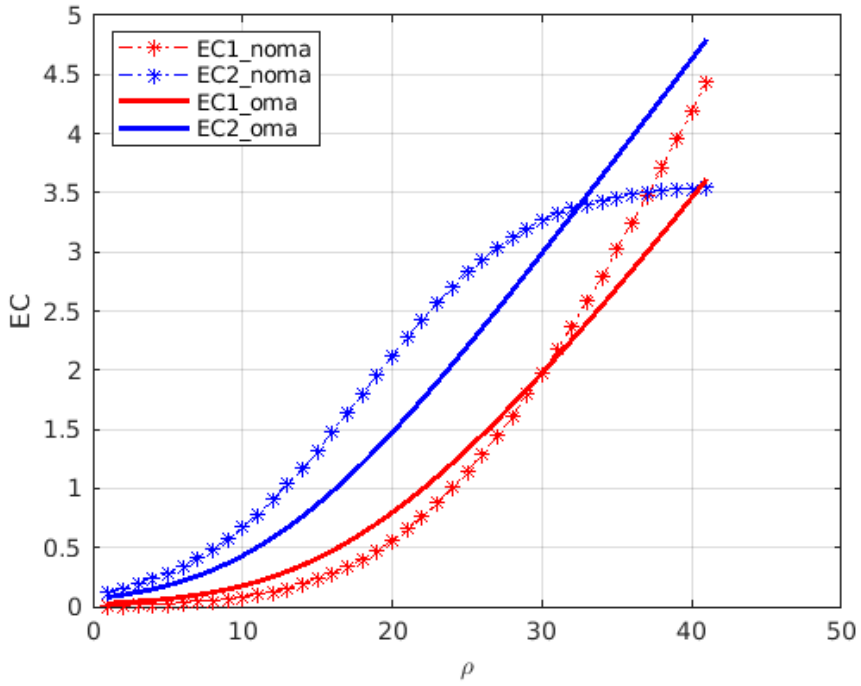

Figure 4.1: E_c^1, E_c^2in a two-user NOMA uplink network compared to Ecs of two users OMA, versus ρ

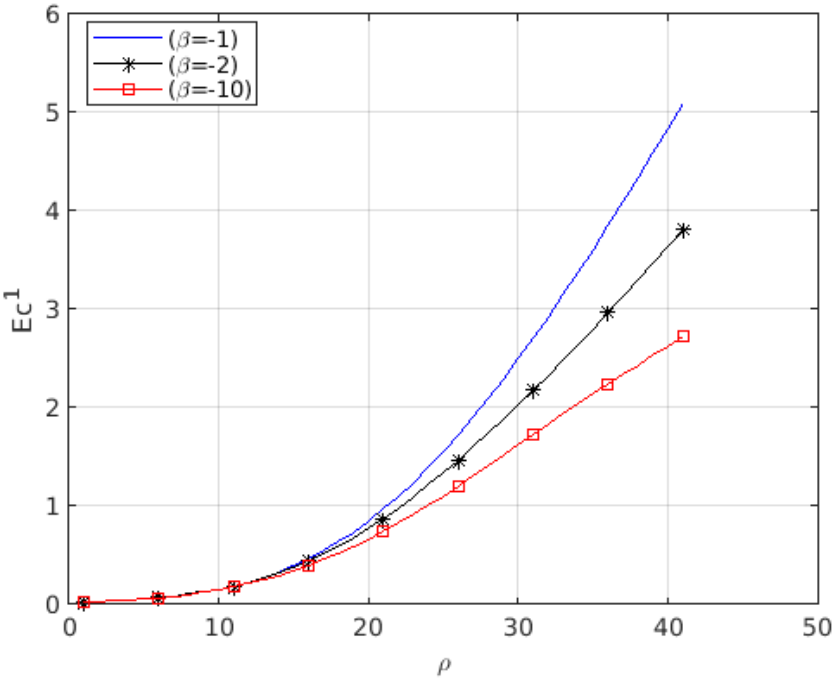

Figure 4.2: E_c^1 versus the transmit SNR, for several delays.

strong user, we notice that the range is reduced. That range expands when we do the inverse. Also, when the delay becomes more stringent, e.g., β_1=β_2=-2, the zero crossing moves from 30 to 36 dB.

Figure 4.6 shows the difference of the EC in NOMA and the EC in OMA for the strong user. This

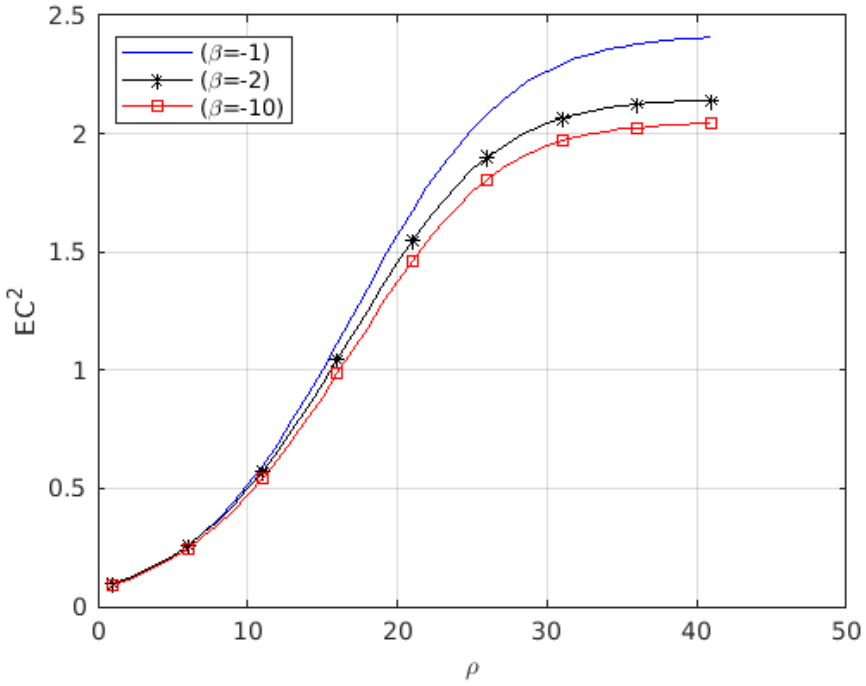

Figure 4.3: E_c^2 versus the transmit SNR ρ for several delays.

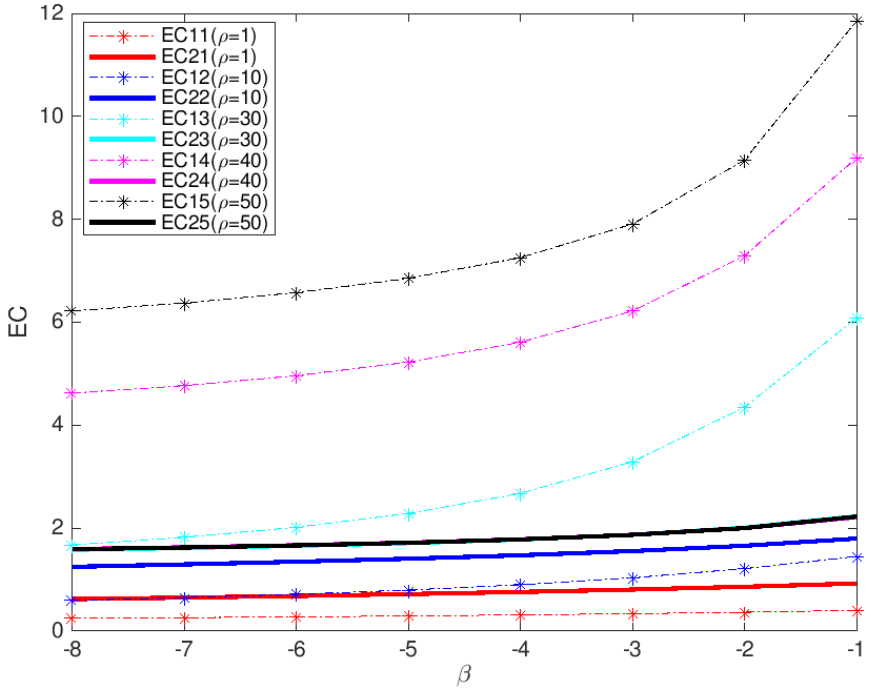

Figure 4.4: E_c^1 and E_c^2 in a two-user NOMA compared to ECs of two users OMA, versus normalized delay β, for different values of ρ.

curve starts initially at zero, then increases to a certain maximum and starts decreasing without bound at high values of the transmit SNR. This confirms Lemma 3. We note that the maximum of these curves decreases when the delay becomes more stringent.

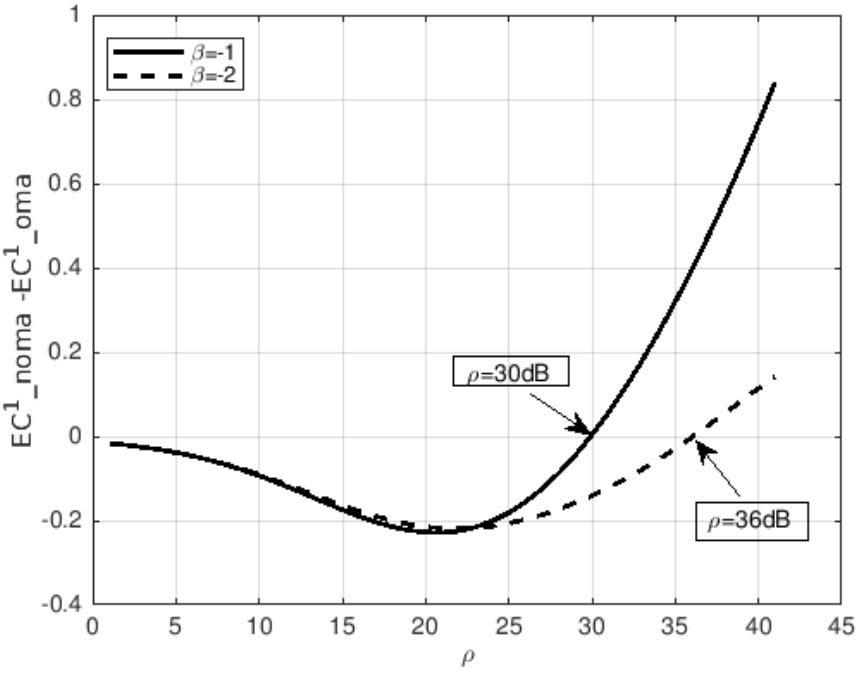

Figure 4.5: $E_c^1 - \tilde{E}_c^1$ versus ρ, for several values of the normalized delay exponent.

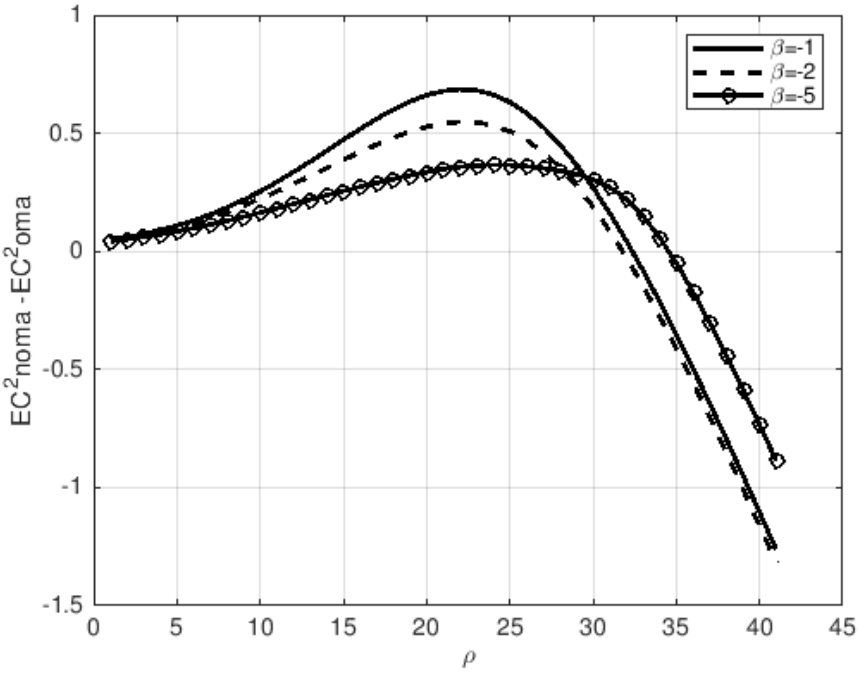

Figure 4.6: $E_c^2 - \tilde{E}_c^2$ versus ρ, for various normalized delay exponent.

To investigate the impact of ρ on the performance of the total link-layer rate for the two-user system, in Fig. 4.7 the plots for V_N in NOMA and V_O in OMA, versus the transmit SNR are depicted for various delay exponents. The curves demonstrate that for both NOMA and OMA, the total EC for

the two users starts at the initial value of 0 and then increases with the transmit SNR, as outlined in Lemma 4. When ρ is very small, the total link-layer rate for the two user in NOMA, V_N, increases faster than V_O in OMA. On the contrary, with the increase of the transmit SNR, V_O becomes gradually higher than V_N. At very high values of the transmit SNR, the gap between the sum EC with NOMA and OMA increases further. Finally, when the delay becomes more stringent, the sum EC of both NOMA and OMA decreases.

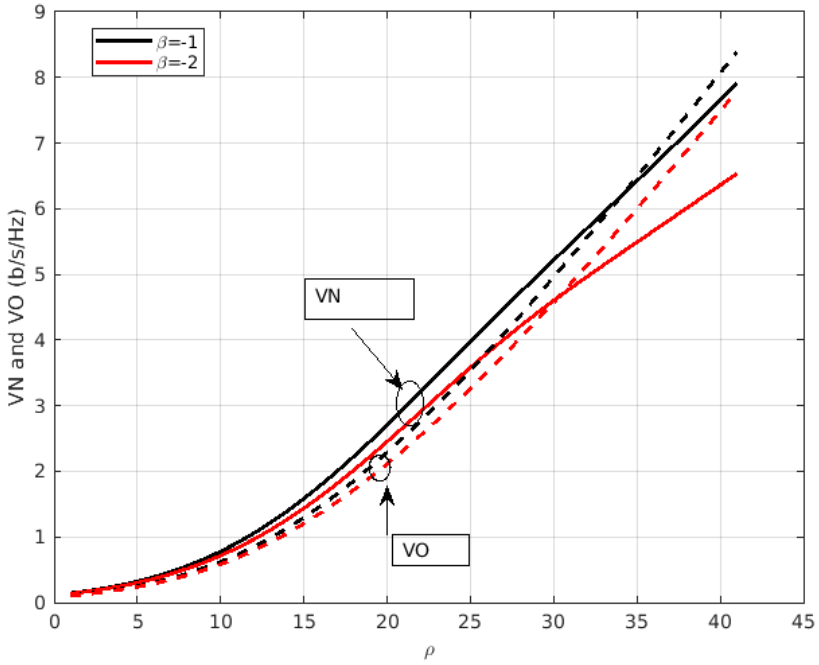

Figure 4.7: V_N and V_O versus ρ, for several values of normalized delay exponent.

Finally, Figs 4.8 and 4.9 depict the sum ECs versus ρ, for several values of the (negative) normalized delay exponent. In Fig. 4.8, the delay of the strong user is fixed, while the delay exponent of the weak user varies. It is shown that in this case, the highest delay QoS (i.e., the smallest negative normalized delay exponent) of the weak user corresponds to the highest gap between the sum ECs $V_N - V_O$. On the other hand, when the delay of the weak user is fixed, Fig. 4.9 shows that the smallest delay Qos (i.e., the highest negative normalized delay exponent) for the strong user corresponds to the largest gap in $V_N - V_O$.

The curve of $V_N - V_O$ starts at zero, increases to a maximum, and returns to negative values. The transition to zero is at $\rho = 31$, and ρ =36 respectively for the figures 4.8 and 4.9. That means from 0 to 31dB (36dB in the Figure 4.9), the total link-layer rate of NOMA is higher than the OMA one. And when ρ becomes larger than this transition point, the total link-layer rate of OMA outperforms the NOMA one.

4.5 Conclusions

The concept of the EC enabled us to study the achievable data-link layer rates when statistical delay QoS guarantees are in place, expressed in the form of delay exponents. We investigated the EC for the uplink of a two-user NOMA network, assuming a Rayleigh block fading channel. We derived novel closed-form expressions for the ECs of the two users and provided a comparison between NOMA and OMA. In NOMA networks, we showed that the ECs of both users decrease as the delay constraints

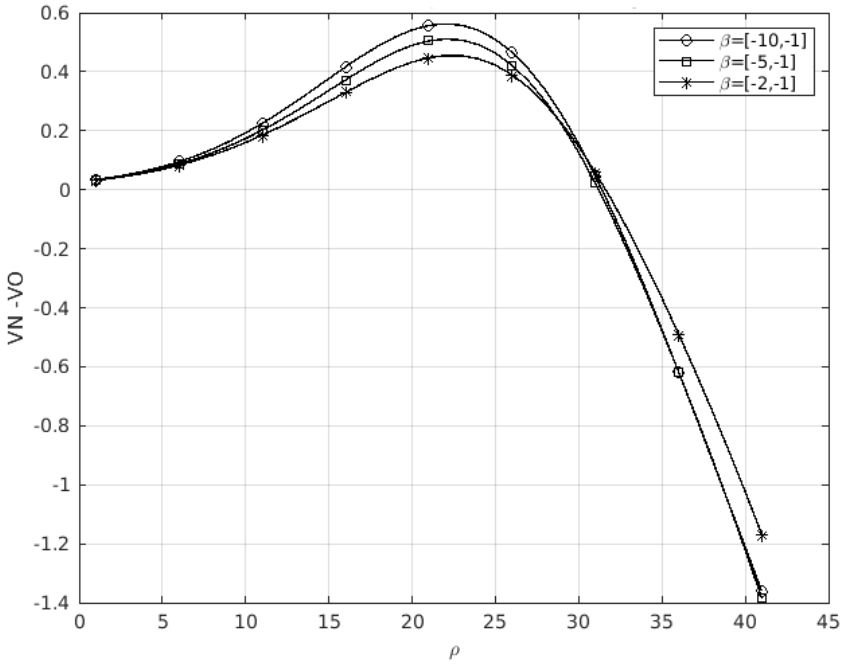

Figure 4.8: V_N - V_O versus ρ for various normalized delay.

become stringent. On the other hand, at high transmit SNRs, the EC of the weak user can surpass the EC of the strong user, as the latter is limited due to interference. This provides the possibility of switching between NOMA and OMA according to the individual users' delay constraints and transmit power levels.

Appendix I

For the second user, we have:

$$E_c^2 = \frac{1}{\beta_2}\log_2\left(2\int_0^\infty \left(\frac{\rho P_2}{1+\rho P_1 x_1}\right)^{\beta_2} e^{-x_1}\int_{x_1}^\infty \left(\frac{1+\rho P_1 x_1}{\rho P_2}+x_2\right)^{\beta_2} e^{-x_2} dx_2 dx_1\right).$$

Set $z = \frac{1+\rho P_1 x_1}{\rho P_2} + x_2$, which means we have $x_2 = z - \frac{1+\rho P_1 x_1}{\rho P_2}$ and $dx_2 = dz$. Then,

$$\begin{aligned}
E_c^2 &= \frac{1}{\beta_2}\log_2\left(2e^{\frac{1}{\rho P_2}}\int_0^\infty \left(\frac{\rho P_2}{1+\rho P_1 x_1}\right)^{\beta_2} e^{-x_1} e^{\frac{P_1 x_1}{P_2}}\int_{\frac{1+\rho x_1}{\rho P_2}}^\infty z^{\beta_2} e^{-z} dz dx_1\right) \\
&= \frac{1}{\beta_2}\log_2\left(2(\rho P_2)^{\frac{\beta_2}{2}} e^{\frac{1}{2\rho P_2}}\int_0^\infty (1+\rho P_1 x_1)^{-\beta_2}(1+\rho x_1)^{\frac{\beta_2}{2}} e^{\frac{(2P_1-2P_2-1)x_1}{2P_2}}\left[\mathbf{W}_{\frac{\beta_2}{2},\frac{1+\beta_2}{2}}\left(\frac{1+\rho x_1}{\rho P_2}\right)\right] dx_1\right) \\
&= \frac{1}{\beta_2}\log_2\left(2P_2(\rho P_2)^{\beta_2} e^{\frac{1}{\rho P_2}} e^{-\frac{(P_1-P_2)}{\rho P_2}}\int_{\frac{1}{\rho P_2}}^\infty P_2^{-\beta_2}(1+\rho P_1 y)^{-\beta_2} e^{(P_1-P_2)y}\Gamma(1+\beta_2, y) dy\right),
\end{aligned}$$

where $\mathbf{W}$ is the Whittaker W function.

Using the binomial expansion, we have $(1+\rho P_1 y)^{-\beta_2} = \sum_{j=0}^{-\beta_2}\binom{-\beta_2}{j}(\rho P_1 y)^j$ and we get the expression given in (4.9).

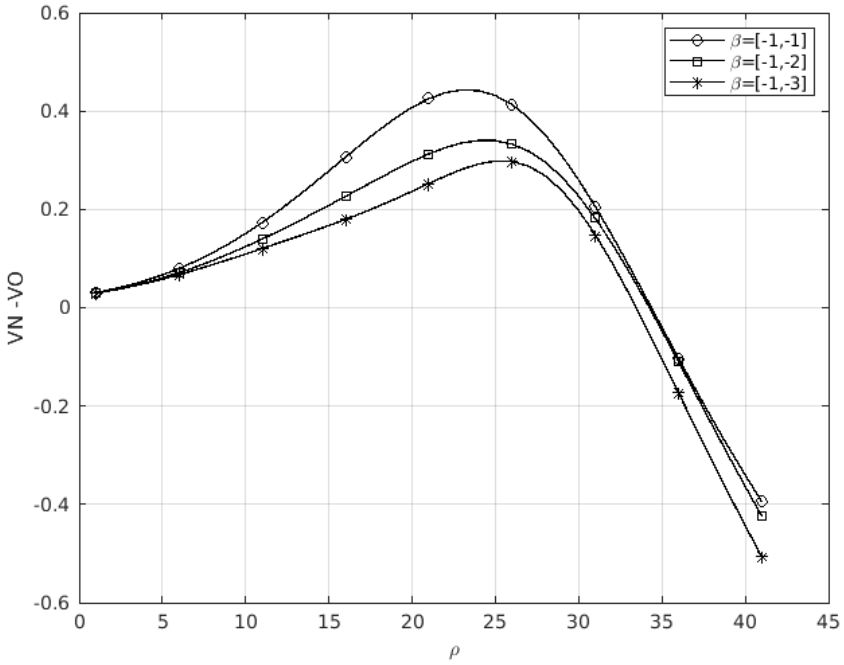

Figure 4.9: V_N - V_O versus ρ for various normalized delay.

The closed-form expression for the EC OMA, of the m-th user with M total users, is determined in [4] as follow.

$$\begin{aligned}\tilde{E}_c^m &= \frac{1}{\beta_m}\log_2\left(E\left[(1+\rho|h_m|^2)^{\frac{\beta_m}{2}}\right]\right) \\ &= \frac{1}{\beta_m}\log_2\left(\frac{\psi_m}{\rho}\int_0^\infty (1+\gamma_m)^{\beta_m} e^{-\frac{(M-m+1)\gamma_m}{\rho}}(1-e^{\frac{-\gamma_m}{\rho}})^{m-1}d\gamma_m\right) \\ &= \frac{1}{\beta_m}\log_2\left(\frac{\psi_m}{\rho}\sum_{k=0}^{m-1}\binom{m-1}{k}(-1)^k\int_0^\infty (1+\gamma_m)^{\beta_m} e^{-\frac{(M-m+1+k)\gamma_m}{\rho}}d\gamma_m\right).\end{aligned} \tag{4.14}$$

$$\tilde{E}_c^m = \frac{1}{\beta_m}\log_2\left(\frac{\psi_m}{\rho}\sum_{k=0}^{m-1}\binom{m-1}{k}(-1)^k U\left(1,2+\frac{2}{M}\beta_m,\frac{M-m+1+k}{\rho}\right)\right) \tag{4.15}$$

For the two users case, $M=2$, we have:

$$\widetilde{E}_c^1 = \frac{1}{\beta_1}\log_2\left(\frac{2}{\rho}\times U\left(1,2+\beta_1,\frac{2}{\rho}\right)\right) \tag{4.16}$$

and,

$$\widetilde{E}_c^2 = \frac{1}{\beta_2}\log_2\left(\frac{2}{\rho}\sum_{k=0}^{1}\binom{1}{k}(-1)^k\times U\left(1,2+\beta_2,\frac{1+k}{\rho}\right)\right) \tag{4.17}$$

where $U(\cdot,\cdot,\cdot)$ is the confluent hypergeometric function of the second kind, defined as

$$U(a,b,z) = \frac{1}{\Gamma(a)}\int_0^\infty e^{-zt}t^{a-1}(1+t)^{b-a-1}dt \tag{4.18}$$

Appendix II

By inserting $\rho \to 0$ into (4.8) and (4.9), we get 1) of Lemma 1, i.e.,

$$\lim_{\rho\to 0}(E_c^1 - \widetilde{E}_c^1) = \frac{1}{\beta_1}\log_2\left(\frac{\mathbb{E}[(1+\rho P_1|h_1|^2)^{\beta_2}]}{\mathbb{E}[(1+\rho|h_1|^2)^{\frac{\beta_2}{2}}]}\right) = 0,$$

$$\lim_{\rho\to 0}(E_c^2 - \widetilde{E}_c^2) = \frac{1}{\beta_2}\log_2\left(\frac{\mathbb{E}[(1+\frac{\rho P_2|h_2|^2}{1+\rho P_1|h_1|^2})^{\beta_2}]}{\mathbb{E}[(1+\rho|h_2|^2)^{\frac{\beta_2}{2}}]}\right) = 0.$$

In the same way, by inserting $\rho \to \infty$ into (4.8) and (4.9), we get 2) in Lemma 1, given below.

$$\lim_{\rho\to\infty} E_c^2 \to \frac{1}{\beta_2}\log_2(\mathbb{E}[(1+\frac{P_2|h_2|^2}{P_1|h_1|^2})^{\beta_2}]),$$

$$\lim_{\rho\to\infty}(E_c^1 - \widetilde{E}_c^1) = \frac{1}{\beta_1}\log_2\left(\rho^{\frac{\beta_1}{2}}\frac{\mathbb{E}[(\frac{1}{\rho}+P_1|h_1|^2)^{\beta_2}]}{\mathbb{E}[(\frac{1}{\rho}+|h_1|^2)^{\frac{\beta_2}{2}}]}\right) = \infty,$$

$$\lim_{\rho\to\infty}(E_c^2 - \widetilde{E}_c^2) = \frac{1}{\beta_2}\log_2\left(\frac{\mathbb{E}[(\frac{\frac{1}{\rho}+P_1|h_1|^2+P_2|h_2|^2}{\frac{1}{\rho}+P_1|h_1|^2})^{\beta_2}]}{\rho^{\frac{\beta_2}{2}}\mathbb{E}[(\frac{1}{\rho}+|h_2|^2)^{\frac{\beta_2}{2}}]}\right) = -\infty.$$

Appendix III

To analyze the trends of E_c^1 and $\widetilde{E}_c^1$ with respect to ρ, we start with

$$\frac{\partial E_c^1}{\partial\rho} = \frac{1}{\beta_1\ln 2}\frac{\left(\mathbb{E}[(1+\rho P_1|h_1|^2)^{\beta_1}]\right)'}{\mathbb{E}[(1+\rho P_1|h_1|^2)^{\beta_1}]} = \frac{P_1}{\ln 2}\frac{\mathbb{E}[|h_1|^2(1+\rho P_1|h_1|^2)^{\beta_1-1}]}{\mathbb{E}[(1+\rho P_1|h_1|^2)^{\beta_1}]} \geq 0.$$

Similarly, for user 1 in OMA we have

$$\frac{\partial \widetilde{E}_c^1}{\partial\rho} = \frac{1}{\beta_1\ln 2}\frac{\left(\mathbb{E}[(1+\rho|h_1|^2)^{\frac{\beta_1}{2}}]\right)'}{\mathbb{E}[(1+\rho|h_1|^2)^{\frac{\beta_1}{2}}]} = \frac{1}{2\ln 2}\frac{\mathbb{E}[|h_1|^2(1+\rho|h_1|^2)^{\frac{\beta_1}{2}-1}]}{\mathbb{E}[(1+\rho|h_1|^2)^{\frac{\beta_1}{2}}]} \geq 0.$$

Then, we get that

$$\frac{\partial(E_c^1 - \widetilde{E}_c^1)}{\partial\rho} = \frac{P_1}{\ln 2}\frac{\mathbb{E}[|h_1|^2(1+\rho P_1|h_1|^2)^{\beta_1-1}]}{\mathbb{E}[(1+\rho P_1|h_1|^2)^{\beta_1}]} - \frac{1}{2\ln 2}\frac{\mathbb{E}[|h_1|^2(1+\rho|h_1|^2)^{\frac{\beta_1}{2}-1}]}{\mathbb{E}[(1+\rho|h_1|^2)^{\frac{\beta_1}{2}}]}, \tag{4.19}$$

and $\lim_{\rho\to 0}(\frac{\partial(E_c^1-\widetilde{E}_c^1)}{\partial\rho}) = \frac{(P_1-\frac{1}{2})}{\ln 2}\mathbb{E}[|h_1|^2] \leq 0$. When $\rho >> 1$, we have

$$\frac{\partial(E_c^1 - \widetilde{E}_c^1)}{\partial\rho}) = \frac{P_1}{\ln 2}\frac{\mathbb{E}[|h_1|^2(\rho P_1|h_1|^2)^{\beta_1-1}]}{\mathbb{E}[(\rho P_1|h_1|^2)^{\beta_1}]} - \frac{1}{2\ln 2}\frac{\mathbb{E}[|h_1|^2(\rho|h_1|^2)^{\frac{\beta_1}{2}-1}]}{\mathbb{E}[(\rho|h_1|^2)^{\frac{\beta_1}{2}}]} = \frac{1}{2\rho\ln 2} \geq 0. \tag{4.20}$$

When $\rho \to \infty$, this term approaches 0.

Appendix IV

$E_c^2 = \frac{1}{\beta_2}\log_2(\mathbb{E}[(1+\frac{\rho P_2|h_2|^2}{1+\rho P_1|h_1|^2})^{\beta_2}])$, and

$$\frac{\partial E_c^2}{\partial \rho} = \frac{1}{\beta_2 \ln 2}\frac{\left(\mathbb{E}[(1+\frac{\rho P_2|h_2|^2}{1+\rho P_1|h_1|^2})^{\beta_2}]\right)'}{\mathbb{E}[(1+\frac{\rho P_2|h_2|^2}{1+\rho P_1|h_1|^2})^{\beta_2}]} = \frac{1}{\ln 2}\frac{\mathbb{E}[\frac{P_2|h_2|^2}{(1+\rho P_1|h_1|^2)^2}(1+\frac{\rho P_2|h_2|^2}{1+\rho P_1|h_1|^2})^{\beta_2-1}]}{\mathbb{E}[(1+\frac{\rho P_2|h_2|^2}{1+\rho P_1|h_1|^2})^{\beta_2}]} \geq 0. \tag{4.21}$$

In the same way, for the user 2 in OMA, we have

$$\frac{\partial \widetilde{E}_c^2}{\partial \rho} = \frac{1}{\beta_2 \ln 2}\frac{\left(\mathbb{E}[(1+\rho|h_2|^2)^{\frac{\beta_2}{2}}]\right)'}{\mathbb{E}[(1+\rho|h_2|^2)^{\frac{\beta_2}{2}}]} = \frac{1}{2\ln 2}\frac{\mathbb{E}[|h_2|^2(1+\rho|h_2|^2)^{\frac{\beta_2}{2}-1}]}{\mathbb{E}[(1+\rho|h_2|^2)^{\frac{\beta_2}{2}}]} \geq 0, \tag{4.22}$$

and

$$\frac{\partial (E_c^2-\widetilde{E}_c^2)}{\partial \rho} = \frac{1}{\ln 2}\frac{\mathbb{E}[\frac{P_2|h_2|^2}{(1+\rho P_1|h_1|^2)^2}(1+\frac{\rho P_2|h_2|^2}{1+\rho P_1|h_1|^2})^{\beta_2-1}]}{\mathbb{E}[(1+\frac{\rho P_2|h_2|^2}{1+\rho P_1|h_1|^2})^{\beta_2}]} - \frac{1}{2\ln 2}\frac{\mathbb{E}[|h_2|^2(1+\rho|h_2|^2)^{\frac{\beta_2}{2}-1}]}{\mathbb{E}[(1+\rho|h_2|^2)^{\frac{\beta_2}{2}}]}. \tag{4.23}$$

When $\rho \to 0$, we have that $\lim_{\rho\to 0}(\frac{\partial(E_c^2-\widetilde{E}_c^2)}{\partial\rho}) = \frac{(P_2-\frac{1}{2})}{\ln 2}\mathbb{E}[|h_2|^2]$. When ρ is very large,

$$\begin{aligned}
\frac{\partial (E_c^2-\widetilde{E}_c^2)}{\partial \rho} &= \frac{\mathbb{E}[\frac{P_2|h_2|^2}{\rho^2(\frac{1}{\rho}+P_1|h_1|^2)^2}(1+\frac{\rho}{\rho}\frac{(P_2|h_2|^2)}{(\frac{1}{\rho}+P_1|h_1|^2)})^{\beta_2-1}]}{\ln 2\mathbb{E}[(1+\frac{\rho}{\rho}\frac{P_2|h_2|^2}{(\frac{1}{\rho}+P_1|h_1|^2)})^{\beta_2}]} - \frac{1}{2\ln 2}\frac{1}{\rho}\frac{\mathbb{E}[|h_2|^2(\frac{1}{\rho}+|h_2|^2)^{\frac{\beta_2}{2}-1}]}{\mathbb{E}[(\frac{1}{\rho}+|h_2|^2)^{\frac{\beta_2}{2}}]} \\
&= \frac{\mathbb{E}[\frac{P_2|h_2|^2}{\rho^2(P_1|h_1|^2)^2}(1+\frac{P_2|h_2|^2}{P_1|h_1|^2})^{\beta_2-1}]}{\ln 2\mathbb{E}[(1+\frac{P_2|h_2|^2}{P_1|h_1|^2})^{\beta_2}]} - \frac{1}{2\ln 2}\frac{1}{\rho}\frac{\mathbb{E}[(|h_2|^2)^{\frac{\beta_2}{2}}]}{\mathbb{E}[(|h_2|^2)^{\frac{\beta_2}{2}}]} \\
&= \frac{P_2}{\rho^2 P_1^2}\frac{\mathbb{E}[\frac{|h_2|^2}{(|h_1|^2)^2}(1+\frac{P_2|h_2|^2}{P_1|h_1|^2})^{\beta_2-1}]}{\ln 2\mathbb{E}[(1+\frac{P_2|h_2|^2}{P_1|h_1|^2})^{\beta_2}]} - \frac{1}{2\ln 2}\frac{1}{\rho} = \frac{\frac{P_2}{P_1^2\ln 2}A - \frac{1}{2\ln 2}\rho}{\rho^2},
\end{aligned} \tag{4.24}$$

where $A = \frac{\mathbb{E}[\frac{|h_2|^2}{(|h_1|^2)^2}(1+\frac{P_2|h_2|^2}{P_1|h_1|^2})^{\beta_2-1}]}{\mathbb{E}[(1+\frac{P_2|h_2|^2}{P_1|h_1|^2})^{\beta_2}]}$, unrelated to ρ. Hence, when ρ is very large, $\frac{\partial(E_c^2-\widetilde{E}_c^2)}{\partial\rho}$ can be approximated by $-\frac{1}{2\ln 2}\frac{1}{\rho}$, and it gradually approaches 0 when $\rho \to \infty$.

Appendix V

Note that $V_N = E_c^1 + E_c^2$. By using Lemma 1, we have $\lim_{\rho\to 0}(V_N) = 0$ and $\lim_{\rho\to\infty}(V_N) = \infty$. Then, we get that

$$\frac{\partial V_N}{\partial \rho} = \frac{\partial(E_c^1+E_c^2)}{\partial\rho} = \frac{P_1}{\ln 2}\frac{\mathbb{E}[|h_1|^2(1+\rho P_1|h_1|^2)^{\beta_1-1}]}{\mathbb{E}[(1+\rho P_1|h_1|^2)^{\beta_1}]} + \frac{1}{\ln 2}\frac{\mathbb{E}[\frac{P_2|h_2|^2}{(1+\rho P_1|h_1|^2)^2}(1+\frac{\rho P_2|h_2|^2}{1+\rho P_1|h_1|^2})^{\beta_2-1}]}{\mathbb{E}[(1+\frac{\rho P_2|h_2|^2}{1+\rho P_1|h_1|^2})^{\beta_2}]} \geq 0. \tag{4.25}$$

When $\rho \to 0$, we have $\lim_{\rho\to 0}(\frac{\partial V_N}{\partial \rho}) = \frac{P_1}{\ln 2}\mathbb{E}[|h_1|^2] + \frac{P_2}{\ln 2}\mathbb{E}[|h_2|^2]$. When $\rho \to \infty$, we get that

$$\lim_{\rho\to\infty} \frac{\partial V_N}{\partial \rho} = \frac{1}{\rho \ln 2} + \frac{\mathbb{E}[\frac{P_2|h_2|^2}{(P_1|h_1|^2)^2}(1+\frac{P_2|h_2|^2}{P_1|h_1|^2})^{\beta_2-1}]}{\rho^2 \ln 2 \mathbb{E}[(1+\frac{P_2|h_2|^2}{P_1|h_1|^2})^{\beta_2}]} = 0.$$

For V_O in the case of OMA, we note that $V_O = \widetilde{E}_c^1 + \widetilde{E}_c^2$. By using Lemma 1, we have $\lim_{\rho\to 0}(V_0) = 0$ and $\lim_{\rho\to\infty}(V_0) = \infty$. Then,

$$\frac{\partial V_0}{\partial \rho} = \frac{\partial(\widetilde{E}_c^1 + \widetilde{E}_c^2)}{\partial \rho} = \frac{1}{2\ln 2}\frac{\mathbb{E}[|h_1|^2(1+\rho|h_1|^2)^{\frac{\beta_1}{2}-1}]}{\mathbb{E}[(1+\rho|h_1|^2)^{\frac{\beta_1}{2}}]} + \frac{1}{2\ln 2}\frac{\mathbb{E}[|h_2|^2(1+\rho|h_2|^2)^{\frac{\beta_2}{2}-1}]}{\mathbb{E}[(1+\rho|h_2|^2)^{\frac{\beta_2}{2}}]} \geq 0.$$

When $\rho \to 0$, we have $\lim_{\rho\to 0}(\frac{\partial V_O}{\partial \rho}) = \frac{1}{2\ln 2}\mathbb{E}[|h_1|^2] + \frac{1}{2\ln 2}\mathbb{E}[|h_2|^2]$. When $\rho \to \infty$, we have that $\lim_{\rho\to\infty}(\frac{\partial V_O}{\partial \rho}) = \lim_{\rho\to\infty}(\frac{1}{2\rho\ln 2} + \frac{1}{2\rho\ln 2}) = \lim_{\rho\to\infty}(\frac{1}{\rho\ln 2})$, which equals to 0.

References

[1] SM Riazul Islam, , Nurilla Avazov, Octavia A Dobre, and Kyung-Sup Kwak. Power-domain non-orthogonal multiple access (NOMA) in 5G systems: Potentials and challenges. *IEEE Commun. Surv, & Tut.*, 19(2):721–742, 2016.

[2] Y. Kanaras and I. Darwazeh A. Chorti, M. Rodrigues. An optimum detection for a spectrally efficient non orthogonal FDM system. In *Proc. 13th Int. OFDM WS*, pages 65–68, Aug 2008.

[3] Yuya Saito, Yoshihisa Kishiyama, Anass Benjebbour, Takehiro Nakamura, Anxin Li, and Kenichi Higuchi. Non-orthogonal multiple access (NOMA) for cellular future radio access. In *Proc. IEEE VTC Spring*, pages 1–5, 2013.

[4] Wenjuan Yu, Leila Musavian, and Qiang Ni. Link-layer capacity of NOMA under statistical delay QoS guarantees. *IEEE Trans. Commun.*, 66(10):4907–4922, 2018.

[5] Wenjuan Yu, Leila Musavian, and Qiang Ni. Tradeoff analysis and joint optimization of link-layer energy efficiency and effective capacity toward green communications. *IEEE Trans. Wireless Commun.*, 15(5):3339–3353, 2016.

[6] Dapeng Wu and Rohit Negi. Effective capacity: a wireless link model for support of quality of service. *IEEE Trans. Wireless Commun.*, 2(4):630–643, 2003.

[7] Jia Tang and Xi Zhang. Cross-layer modeling for quality of service guarantees over wireless links. *IEEE Trans. Wireless Commun.*, 6(12):4504–4512, 2007.

[8] A. Chorti and H. V. Poor. Achievable secrecy rates in physical layer secure systems with a helping interferer. In *2012 Int Conf. Computing, Networking Commun. (ICNC)*, pages 18–22, Maui, HI, 2012.

[9] W. Yu, A. Chorti, L. Musavian, H. Vincent Poor, and Q. Ni. Effective secrecy rate for a downlink NOMA network. *IEEE Trans. Wireless Commun.*, pages 5673–5690, Dec. 2019.

[10] Zheng Yang, Zhiguo Ding, Pingzhi Fan, and Naofal Al-Dhahir. A general power allocation scheme to guarantee quality of service in downlink and uplink noma systems. *IEEE Transactions on Wireless Communications*, 15(11):7244–7257, 2016.

[11] Ningbo Zhang, Jing Wang, Guixia Kang, and Yang Liu. Uplink nonorthogonal multiple access in 5G systems. *IEEE Commun. L.*, 20(3):458–461, 2016.

[12] Li Fan, Shi Jin, Chao-Kai Wen, and Haixia Zhang. Uplink achievable rate for massive MIMO systems with low-resolution ADC. *IEEE Commun. L.*, 19(12):2186–2189, 2015.

[13] Hong-Chuan Yang and Mohamed-Slim Alouini. *Order statistics in wireless communications: diversity, adaptation, and scheduling in MIMO and OFDM systems*. Cambridge University Press, 2011.

Chapter 5

Perspectives

ToDo!

5.1 Introduction

While security protocols predominantly focus on the core network, the enhancement of the security of the B5G access network becomes of critical importance. Despite the strengthening of 5G security protocols with respect to LTE, there are still open issues that have not been fully addressed. The current tutorial is articulated around the premise that rethinking the security design bottom up, starting at the physical layer, is not only viable in 6G but importantly, arises as an efficient way to overcome security hurdles in novel use cases, notably mMTC and URLLC. In this tutorial, we begin with a review of fundamental concepts in security overall and physical layer security in particular. We then move to provide a comprehensive review of the state-of the-art in i) secret key generation from shared randomness, ii) the wiretap channel in the mMIMO era, iii) authentication of devices using physical unclonable functions (PUFs) and localization based authentication, protocols using multi-factor authentication, iv) jamming attacks and intrusion detection at PHY. We finally conclude with the proposers' aspirations for the 6G security landscape, in the hyper-connectivity and semantic communications era.

5.2 6G Research Topics

Lorem ipsum dolor sit amet, consectetuer adipiscing elit. Ut purus elit, vestibulum ut, placerat ac, adipiscing vitae, felis. Curabitur dictum gravida mauris. Nam arcu libero, nonummy eget, consectetuer id, vulputate a, magna. Donec vehicula augue eu neque. Pellentesque habitant morbi tristique senectus et netus et malesuada fames ac turpis egestas. Mauris ut leo. Cras viverra metus rhoncus sem. Nulla et lectus vestibulum urna fringilla ultrices. Phasellus eu tellus sit amet tortor gravida placerat. Integer sapien est, iaculis in, pretium quis, viverra ac, nunc. Praesent eget sem vel leo ultrices bibendum. Aenean faucibus. Morbi dolor nulla, malesuada eu, pulvinar at, mollis ac, nulla. Curabitur auctor semper nulla. Donec varius orci eget risus. Duis nibh mi, congue eu, accumsan eleifend, sagittis quis, diam. Duis eget orci sit amet orci dignissim rutrum.

Nam dui ligula, fringilla a, euismod sodales, sollicitudin vel, wisi. Morbi auctor lorem non justo. Nam lacus libero, pretium at, lobortis vitae, ultricies et, tellus. Donec aliquet, tortor sed accumsan bibendum, erat ligula aliquet magna, vitae ornare odio metus a mi. Morbi ac orci et nisl hendrerit mollis. Suspendisse ut massa. Cras nec ante. Pellentesque a nulla. Cum sociis natoque penatibus et magnis dis parturient montes, nascetur ridiculus mus. Aliquam tincidunt urna. Nulla ullamcorper vestibulum turpis. Pellentesque cursus luctus mauris.

Nulla malesuada porttitor diam. Donec felis erat, congue non, volutpat at, tincidunt tristique, libero. Vivamus viverra fermentum felis. Donec nonummy pellentesque ante. Phasellus adipiscing

semper elit. Proin fermentum massa ac quam. Sed diam turpis, molestie vitae, placerat a, molestie nec, leo. Maecenas lacinia. Nam ipsum ligula, eleifend at, accumsan nec, suscipit a, ipsum. Morbi blandit ligula feugiat magna. Nunc eleifend consequat lorem. Sed lacinia nulla vitae enim. Pellentesque tincidunt purus vel magna. Integer non enim. Praesent euismod nunc eu purus. Donec bibendum quam in tellus. Nullam cursus pulvinar lectus. Donec et mi. Nam vulputate metus eu enim. Vestibulum pellentesque felis eu massa.

Quisque ullamcorper placerat ipsum. Cras nibh. Morbi vel justo vitae lacus tincidunt ultrices. Lorem ipsum dolor sit amet, consectetuer adipiscing elit. In hac habitasse platea dictumst. Integer tempus convallis augue. Etiam facilisis. Nunc elementum fermentum wisi. Aenean placerat. Ut imperdiet, enim sed gravida sollicitudin, felis odio placerat quam, ac pulvinar elit purus eget enim. Nunc vitae tortor. Proin tempus nibh sit amet nisl. Vivamus quis tortor vitae risus porta vehicula.

Fusce mauris. Vestibulum luctus nibh at lectus. Sed bibendum, nulla a faucibus semper, leo velit ultricies tellus, ac venenatis arcu wisi vel nisl. Vestibulum diam. Aliquam pellentesque, augue quis sagittis posuere, turpis lacus congue quam, in hendrerit risus eros eget felis. Maecenas eget erat in sapien mattis porttitor. Vestibulum porttitor. Nulla facilisi. Sed a turpis eu lacus commodo facilisis. Morbi fringilla, wisi in dignissim interdum, justo lectus sagittis dui, et vehicula libero dui cursus dui. Mauris tempor ligula sed lacus. Duis cursus enim ut augue. Cras ac magna. Cras nulla. Nulla egestas. Curabitur a leo. Quisque egestas wisi eget nunc. Nam feugiat lacus vel est. Curabitur consectetuer.

Suspendisse vel felis. Ut lorem lorem, interdum eu, tincidunt sit amet, laoreet vitae, arcu. Aenean faucibus pede eu ante. Praesent enim elit, rutrum at, molestie non, nonummy vel, nisl. Ut lectus eros, malesuada sit amet, fermentum eu, sodales cursus, magna. Donec eu purus. Quisque vehicula, urna sed ultricies auctor, pede lorem egestas dui, et convallis elit erat sed nulla. Donec luctus. Curabitur et nunc. Aliquam dolor odio, commodo pretium, ultricies non, pharetra in, velit. Integer arcu est, nonummy in, fermentum faucibus, egestas vel, odio.

Sed commodo posuere pede. Mauris ut est. Ut quis purus. Sed ac odio. Sed vehicula hendrerit sem. Duis non odio. Morbi ut dui. Sed accumsan risus eget odio. In hac habitasse platea dictumst. Pellentesque non elit. Fusce sed justo eu urna porta tincidunt. Mauris felis odio, sollicitudin sed, volutpat a, ornare ac, erat. Morbi quis dolor. Donec pellentesque, erat ac sagittis semper, nunc dui lobortis purus, quis congue purus metus ultricies tellus. Proin et quam. Class aptent taciti sociosqu ad litora torquent per conubia nostra, per inceptos hymenaeos. Praesent sapien turpis, fermentum vel, eleifend faucibus, vehicula eu, lacus.

5.3 The Role of of PLS in 6G

The rollout of fifth-generation (5G) mobile networks and the forthcoming 6G will bring about fundamental changes in the way we communicate, access services and entertainment. In the context of security, inarguably, 5G security enhancements present a big improvement with respect to LTE. However, as the complexity of the application scenarios increases with the introduction of novel use cases, notably ultra-reliable low latency (URLLC) and massive machine type communications (mMTC), novel security challenges arise that might be difficult to address using the standard paradigm of complexity based classical crypto solutions. Specific use cases with open security issues are described in detail in a number of 3GPP technical reports, e.g., on the false base station attack scenario [1] and on the security issues in URLLC [2]. Indeed, for beyond 5G (B5G) systems, there exist security aspects that can be further enhanced by exploiting different approaches, as classical mechanisms either fall short in guaranteeing all the security and privacy relevant aspects, or, can be strengthened with mechanisms that could provide a second layer of protection.

In the past years, physical layer security (PLS) [3, 4] has been studied and indicated as a possible way to emancipate networks from classical, complexity based, security approaches. Notably, it is explicitly mentioned in the first white paper on 6G: “The strongest security protection may be achieved at the physical layer.” Furthermore, it is stated as an enabling technology in the IEEE International

Network Generations Roadmap (INGR) 1st Edition 2019 in the Chapters on "Security" (Section 1.1 pp. 1-2) and on "Massive MIMO" (Section 4.3 pp. 8-9) and is expected to be more closely monitored in the 2nd Edition (currently under writing). Based on this, the objective of this tutorial is to investigate how it could be possible in B5G to move some of the security core functions down to the physical layer, exploiting both the communication radio channel and the hardware, as unique entropy sources.

Since the wireless channel is reciprocal, time-varying and random in nature, it offers a valid, inherently secure source for key agreement (KA) protocols between two communicating parties. This is pertinent to many forthcoming B5G applications that will require strong, but nevertheless, lightweight KA mechanisms; in this direction, PLS may offer such solutions, or complement existing algorithms, with minimal changes in the control plane. With respect to authentication, physical unclonable functions (PUFs), wireless fingerprinting / localization, combined with more classical approaches, could also enhance authentication and key agreement (AKA) in demanding scenarios, including (but not limited to) device to device (D2D) and Industry 4.0. In parallel, mmWave in the Terahertz range will rely upon setting up wireless "wires"; although on their own they cannot ensure confidentiality, they will provide a concrete scenario for the wiretap channel. It is therefore pertinent to discuss advancements in wiretap secrecy encoders. In a nutshell, several advantages can be envisioned by rethinking the security design bottom-up, and in particular: 1) PLS can provide information-theoretic security guarantees with lightweight mechanisms (e.g., using LDPC encoders); 2) hybrid crypto-PLS protocols can provide alternatives in scenarios where classical mechanisms fall short such as in [1] and [2], and 3) PLS can act as an extra security layer, complementing other approaches.

Our motivation in this tutorial on PLS stems from the fact that in B5G PLS emerges as a complementary means to enhance the security in demanding low latency and massive connectivity scenarios. A few supporting examples include: 1) the security vulnerabilities identified in [1] arise during the establishment of the radio link; in this aspect, standard security protocols that build on the premise that the communication link has already been established, cannot offer solutions when this is not the case, whereas, PLS schemes can be seamlessly incorporated (e.g., can be interwoven with channel estimation); 2) in the realm of massive IoT in which standard authentication and key distribution /management becomes challenging (it is unrealistic to use digital certificates for billions of devices), PLS can offer complementary, device oriented solutions; 3) in 6G we will move away from the standard client/server networking paradigm on which the most successful security protocols build on and incorporate D2D and D2Edge at massive scales; and 4) the standard "rigid" on-or-off security approach of current protocols might not be the best fit to future generations of "semantic" communications between smart devices. A further benefit comes from the fact that PLS techniques – if implemented correctly – can offer quantum resistance. In this sense, PLS could pave the way out of the low latency impasse introducing novel lightweight mechanisms to post-quantum security.

The tutorial's two core objectives are: 1) to inform the audience on how PLS schemes can work either as stand-alone or as complementary schemes to address open security issues in 5G, and 2) to discuss how PLS can fit in the palette of 6G security solutions.

5.3.1 Information theoretic security

Fundamental results of information theory, notably in terms of the channel capacity of various classes of wireless communication channels, have to a large extent materialized in working communication systems, bringing wireless technology to the current fifth generation (5G). However, confidentiality in communication exchange took a different route to practice than the one prescribed by Shannon, whose negative result on perfect secrecy is referred to as the one-time pad scheme. The requirement of perfect secrecy was abandoned for decades and security studies focuses on semantic security, i.e., indistinguishability results in polynomial time, giving rise to the domain of computational security. Yet, recent work on information theoretic secrecy, such as using channel noise to assist in secret communication or using shared aléa, e.g., manifested as quantum entanglement to exchange secret keys, removes some of the barriers for perfect secrecy. This advancement become increasingly important as

quantum computers seem finally to be tangible in the not so distant future.

However, quantum computing is not the only precarity for public key encryption based authentication and key agreement (AKA). The rollout of 5G mobile networks and the forthcoming sixth generation (6G) will bring about fundamental changes in the way we communicate, access services and entertainment. In the context of security, inarguably, 5G security enhancements present a big improvement with respect to LTE. However, as the complexity of the application scenarios increases with the introduction of novel use cases, notably ultra-reliable low latency (URLLC) and massive machine type communications (mMTC), commonly referred to as the Internet of things IoT, novel security challenges arise that might be difficult to address using the standard paradigm of complexity based crypto solutions. Blockchain technologies could offer a viable alternative for the registration and normal activity tracking of massive IoT; however, these are also computationally intensive and might not be the best approach in the domain of fast authentication. Security under latency constraints is still considered as a high priority open issue.

In the past years, physical layer security (PLS) has been studied and indicated as a possible way to emancipate networks from classic, complexity based, security approaches. Notably, it is explicitly mentioned as a 6G enabling technology in the first white paper on 6G and in two IEEE INGR (International Network Generations Roadmap) chapters. Importantly, several advantages can be envisaged by rethinking the security design bottom-up, and in particular:

1) PLS can provide information-theoretic security guarantees and can offer a lightweight mechanism towards quantum resistance; 2) hybrid crypto-PLS protocols can provide alternatives in scenarios where classic mechanism falls short such as in low latency scenarios; 3) PLS can act as an extra security layer, complementing other approaches.

A primary direction of my proposed research project is to move some of the security core functions down to the physical layer, exploiting both the communication radio channel and the hardware, as unique entropy sources. My proposed research on PLS encompasses research on authentication, data confidentiality and anomaly detection, described below.

5.3.2 Authentication

With respect to authentication, physical unclonable functions (PUFs), wireless fingerprinting and localization could enhance AKA in demanding scenarios, including (but not limited to) device to device (D2D) and ultra-low latency applications such as autonomous vehicles or smart factories, enhanced reality and tactile Internet. Related research questions in this direction would include:

- The design of novel information reconciliation code designs from the families of Slepian Wolf or Wyner Ziv distributed source encoders, with a particular focus on the short block-length; furthermore, benchmarking their performance against the best known second order approximation results for the respective achievable rates.

- Characterization of the wireless channel from a security (as opposed to the standard point-to-point communication) point of view. The baseline idea boils down to the fact that the predictable element of the wireless coefficient is useful for authentication (localization), while the purely random for extracting secret keys. Developing the mathematical and machine learning tools to resolve the two is an unchartered area of research so far. Together with the systematic study of the short block-length reconciliation, it can give important answers on the required amount of privacy amplification in real systems.

- The evaluation of the security level achieved with the proposed PLS methods will be sought, scrutinizing the hypothesis that security level 5 (post-quantum) is attainable with PLS in the finite block-length regime.

- The proposal of hybrid crypto-PLS systems and the resilience of related systems to active attacks. Unlike in the network security paradigm, active attacks at the physical layer (PHY) can be alleviated by PHY remedies, e.g., narrow beamforming, pilot randomization, energy harvesting and frequency hopping.

5.3.3 Data Confidentiality

Additionally, with respect to data confidentiality achieved in wiretap channels with the use of secrecy encoders, practical designs have so far been presented only for the wiretap-II channel (i.e., noiseless main channel) and the erasure channel. Building secrecy encoders for the standard wiretap-I channel is timely, especially as degradedness of the eavesdropping channel can be substantiated in mmWave technologies enabled by narrow beamforming using multiple antennas. In my proposed research direction in this domain, I intend to seek input from the design of core building blocks of symmetric block ciphers, in particular of reversible S-boxes. While linear encoders purposed for reliability have proved instrumental to reach the Shannon limit for reliable communication in (linear) channel wireless settings, they might not be the optimal choice for secrecy. The study of bent functions from the crypto community can constitute the starting point of the design of non-linear secrecy encoders, purposed to guarantee reliability in the communication and secrecy with respect to an eavesdropper.

5.3.4 Anomaly Detection

In terms of anomaly detection, my focus is on proposing novel approaches to identify hacking and distributed denial of service (DDoS) attacks in IoT networks, starting at the device (PHY) level and accounting for the wireless edge. The primary aim is to break away from incremental earlier approaches that rely on traffic monitoring and (deep) packet inspection, performed at the upper layers of the network stack. In the proposed research direction, IoT resilience to hacking and DDoS will be investigated leveraging my recent results in change point analysis, on one hand, and, deep learning techniques, on the other. The primary innovation is on rethinking the overall design bottom up, noting that robust, early detection tools should optimally aggregate behavioural information both from the network side (e.g., IP addresses, number of TCP segments, type of messages, etc.) as well as the device's physical side channels (e.g., power consumed, duty cycle, temperature variations, variations in the number of memory read /write operations, etc.).

With respect to this latter dimension, side channels have customarily been used for negative security proofs, e.g., showcasing it is possible to compromise symmetric block ciphers such as the DES with smart power monitoring. The idea of using side channels to identify intrusions or anomalous events breaks completely new ground and can offer a straightforward solution to issues related to monitoring a huge number of network layer parameters (an issue typically tackled with data thinning approaches). On this premise, the early detection of hacking / DDoS can be developed using state-of-the-art, real-time, lightweight anomaly detection algorithms, suitable for the constraints of IoT devices, such as on-line change point detection or deep learning tools. A second important aspect is the development of distributed anomaly detection algorithms. Understanding the trade-off between cluster size and speed of detection would be the primary initial goal in this setting.

5.4 Longer Term Perspectives

Lorem ipsum dolor sit amet, consectetuer adipiscing elit. Ut purus elit, vestibulum ut, placerat ac, adipiscing vitae, felis. Curabitur dictum gravida mauris. Nam arcu libero, nonummy eget, consectetuer id, vulputate a, magna. Donec vehicula augue eu neque. Pellentesque habitant morbi tristique senectus et netus et malesuada fames ac turpis egestas. Mauris ut leo. Cras viverra metus rhoncus sem. Nulla et lectus vestibulum urna fringilla ultrices. Phasellus eu tellus sit amet tortor gravida placerat. Integer sapien est, iaculis in, pretium quis, viverra ac, nunc. Praesent eget sem vel leo ultrices bibendum. Aenean faucibus. Morbi dolor nulla, malesuada eu, pulvinar at, mollis ac, nulla. Curabitur auctor semper nulla. Donec varius orci eget risus. Duis nibh mi, congue eu, accumsan eleifend, sagittis quis, diam. Duis eget orci sit amet orci dignissim rutrum.

Nam dui ligula, fringilla a, euismod sodales, sollicitudin vel, wisi. Morbi auctor lorem non justo. Nam lacus libero, pretium at, lobortis vitae, ultricies et, tellus. Donec aliquet, tortor sed accumsan

bibendum, erat ligula aliquet magna, vitae ornare odio metus a mi. Morbi ac orci et nisl hendrerit mollis. Suspendisse ut massa. Cras nec ante. Pellentesque a nulla. Cum sociis natoque penatibus et magnis dis parturient montes, nascetur ridiculus mus. Aliquam tincidunt urna. Nulla ullamcorper vestibulum turpis. Pellentesque cursus luctus mauris.

Nulla malesuada porttitor diam. Donec felis erat, congue non, volutpat at, tincidunt tristique, libero. Vivamus viverra fermentum felis. Donec nonummy pellentesque ante. Phasellus adipiscing semper elit. Proin fermentum massa ac quam. Sed diam turpis, molestie vitae, placerat a, molestie nec, leo. Maecenas lacinia. Nam ipsum ligula, eleifend at, accumsan nec, suscipit a, ipsum. Morbi blandit ligula feugiat magna. Nunc eleifend consequat lorem. Sed lacinia nulla vitae enim. Pellentesque tincidunt purus vel magna. Integer non enim. Praesent euismod nunc eu purus. Donec bibendum quam in tellus. Nullam cursus pulvinar lectus. Donec et mi. Nam vulputate metus eu enim. Vestibulum pellentesque felis eu massa.

Quisque ullamcorper placerat ipsum. Cras nibh. Morbi vel justo vitae lacus tincidunt ultrices. Lorem ipsum dolor sit amet, consectetuer adipiscing elit. In hac habitasse platea dictumst. Integer tempus convallis augue. Etiam facilisis. Nunc elementum fermentum wisi. Aenean placerat. Ut imperdiet, enim sed gravida sollicitudin, felis odio placerat quam, ac pulvinar elit purus eget enim. Nunc vitae tortor. Proin tempus nibh sit amet nisl. Vivamus quis tortor vitae risus porta vehicula.

Fusce mauris. Vestibulum luctus nibh at lectus. Sed bibendum, nulla a faucibus semper, leo velit ultricies tellus, ac venenatis arcu wisi vel nisl. Vestibulum diam. Aliquam pellentesque, augue quis sagittis posuere, turpis lacus congue quam, in hendrerit risus eros eget felis. Maecenas eget erat in sapien mattis porttitor. Vestibulum porttitor. Nulla facilisi. Sed a turpis eu lacus commodo facilisis. Morbi fringilla, wisi in dignissim interdum, justo lectus sagittis dui, et vehicula libero dui cursus dui. Mauris tempor ligula sed lacus. Duis cursus enim ut augue. Cras ac magna. Cras nulla. Nulla egestas. Curabitur a leo. Quisque egestas wisi eget nunc. Nam feugiat lacus vel est. Curabitur consectetuer.

Suspendisse vel felis. Ut lorem lorem, interdum eu, tincidunt sit amet, laoreet vitae, arcu. Aenean faucibus pede eu ante. Praesent enim elit, rutrum at, molestie non, nonummy vel, nisl. Ut lectus eros, malesuada sit amet, fermentum eu, sodales cursus, magna. Donec eu purus. Quisque vehicula, urna sed ultricies auctor, pede lorem egestas dui, et convallis elit erat sed nulla. Donec luctus. Curabitur et nunc. Aliquam dolor odio, commodo pretium, ultricies non, pharetra in, velit. Integer arcu est, nonummy in, fermentum faucibus, egestas vel, odio.

Sed commodo posuere pede. Mauris ut est. Ut quis purus. Sed ac odio. Sed vehicula hendrerit sem. Duis non odio. Morbi ut dui. Sed accumsan risus eget odio. In hac habitasse platea dictumst. Pellentesque non elit. Fusce sed justo eu urna porta tincidunt. Mauris felis odio, sollicitudin sed, volutpat a, ornare ac, erat. Morbi quis dolor. Donec pellentesque, erat ac sagittis semper, nunc dui lobortis purus, quis congue purus metus ultricies tellus. Proin et quam. Class aptent taciti sociosqu ad litora torquent per conubia nostra, per inceptos hymenaeos. Praesent sapien turpis, fermentum vel, eleifend faucibus, vehicula eu, lacus.