Aleksandr Ometov

Social, Private, and Trusted Wearable Technology under Cloud-Aided Intermittent Wireless Connectivity

Julkaisu 1603 • Publication 1603

Tampere 2018
Aleksandr Ometov

Social, Private, and Trusted Wearable Technology under Cloud-Aided Intermittent Wireless Connectivity

Thesis for the degree of Doctor of Science in Technology to be presented with due permission for public examination and criticism in Tietotalo Building, Auditorium TB206, at Tampere University of Technology, on the 23rd of November 2018, at 12 noon.
Doctoral candidate: Aleksandr Ometov
Laboratory of Electronics and Communications Engineering
Faculty of Computing and Electrical Engineering
Tampere University of Technology
Tampere, Finland

Supervisor: Yevgeni Koucheryavy, Ph.D., Professor
Laboratory of Electronics and Communications Engineering
Faculty of Computing and Electrical Engineering
Tampere University of Technology
Tampere, Finland

Instructor: Sergey Andreev, Ph.D., Assistant Professor
Laboratory of Electronics and Communications Engineering
Faculty of Computing and Electrical Engineering
Tampere University of Technology
Tampere, Finland

Pre-examiners: Periklis Chatzimisios, Ph.D., Associate Professor
Department of Informatics
Alexander Technological Educational Institute of Thessaloniki
Thessaloniki, Greece

Tapani Ristaniemi, Ph.D., Professor
Department of Mathematical Information Technology
University of Jyväskylä
Jyväskylä, Finland

Opponent: Edmundo Monteiro, Ph.D., Professor
Department of Informatics Engineering
University of Coimbra
Coimbra, Portugal
Abstract

There has been an unprecedented increase in the use of smart devices globally, together with novel forms of communication, computing, and control technologies that have paved the way for a new category of devices, known as high-end wearables. While massive deployments of these objects may improve the lives of people, unauthorized access to the said private equipment and its connectivity is potentially dangerous. Hence, communication enablers together with highly-secure human authentication mechanisms have to be designed.

In addition, it is important to understand how human beings, as the primary users, interact with wearable devices on a day-to-day basis; usage should be comfortable, seamless, user-friendly, and mindful of urban dynamics. Usually the connectivity between wearables and the cloud is executed through the user’s more power independent gateway: this will usually be a smartphone, which may have potentially unreliable infrastructure connectivity. In response to these unique challenges, this thesis advocates for the adoption of direct, secure, proximity-based communication enablers enhanced with multi-factor authentication (hereafter refereed to MFA) that can integrate/interact with wearable technology. Their intelligent combination together with the connection establishment automation relying on the device/user social relations would allow to reliably grant or deny access in cases of both stable and intermittent connectivity to the trusted authority running in the cloud.

The introduction will list the main communication paradigms, applications, conventional network architectures, and any relevant wearable-specific challenges. Next, the work examines the improved architecture and security enablers for clusterization between wearable gateways with a proximity-based communication as a baseline. Relying on this architecture, the author then elaborates on the social ties potentially overlaying the direct connectivity management in cases of both reliable and unreliable connection to the trusted cloud. The author discusses that social-aware cooperation and trust relations between users and/or the devices themselves are beneficial for the architecture under proposal. Next, the author introduces a protocol suite that enables temporary delegation of personal device use dependent on different connectivity conditions to the cloud.

After these discussions, the wearable technology is analyzed as a biometric and behavior data provider for enabling MFA. The conventional approaches of the authentication factor combination strategies are compared with the ‘intelligent’ method proposed further. The assessment finds significant advantages to the developed solution over existing ones.

On the practical side, the performance evaluation of existing cryptographic primitives, as part of the experimental work, shows the possibility of developing the experimental methods further on modern wearable devices.

In summary, the set of enablers developed here for wearable technology connectivity is aimed at enriching people’s everyday lives in a secure and usable way, in cases when communication to the cloud is not consistently available.
Preface

This work was carried out at the Laboratory of Electronics and Communications Engineering of Tampere University of Technology (Finland) over the years 2016-2018.

First of all, I want to express my most profound appreciation to Prof. Yevgeni Koucheryavy, who helped and guided me through my work at the department. Also, I would like to emphasize the role of Asst. Prof. Sergey Andreev – as an instructor, he invested a lot of time and showed me the best side of academic life. Their strong collaboration played a crucial role in developing me as a person in both work and personal life. I would never forget their effort and trust in me.

Additionally, I wish to thank Dr. Alexander Pyattaev, Dr. Antonino Orsino, Dr. Mikhail Gerasimenko, and Vitaly Petrov for their contributions to this research and valuable advices. Taking the opportunity, I would like to acknowledge my industrial collaborators Dr. Gábor Fodor and Johan Torsner for providing valuable guidance and insight into future technological trends.

During this research, I have been closely collaborating with many professionals in the academic field. I would like to thank Prof. Sergey Bezzateev, who showed me an entirely new world of information security and research lifestyle. I would send my profound acknowledgments to Assoc. Prof. Jiří Hošek and Dr. Pavel Mašek who showed real engineering diligence and teamwork during my visits to Brno University of Technology, Czech Republic. I would like to take this opportunity and thank Dr. Dmitri Moltchanov for his valuable support and advices. My special thanks are to Prof. Tommi Mikkonen, Assoc. Prof. Thomas Olsson, and Dr. Niko Mäkitalo for their vision and support while delving into new software paradigms. I want to acknowledge my former supervisor Prof. Andrey Turlikov who made me believe that wireless telecommunications is the path worth taking.

I would like to acknowledge the reviewers of this thesis, Assoc. Prof. Periklis Chatzimisios (Greece) and Prof. Tapani Ristaniemi (Finland) for sharing their feedback on my research. This thesis was significantly improved based on their valuable suggestions. Additional gratitude goes to Prof. Edmundo Monteiro (Portugal) for agreeing to act as an opponent at my defense.

The support of this research was received from project “TAKE-5: the 5th Evolution Take of Wireless Communication Networks”, funded by Tekes (now Business Finland), from Nokia Foundation as a personal Nokia Scholarship grant, from Doctoral training network in ELectronics, Telecommunications and Automation (DELTA), as well as previously from the Academy of Finland, project “Empowering Secure, Private, and Trusted Network-Assisted Device-to-Device Communication”.

v
Remembering my time at TUT, I am profoundly grateful to my colleagues for making Finland warmer and brighter place in contrast to the weather. It was exciting being brothers-in-arms with Roman Florea, Adam Surák, Ekaterina Olshannikova, Roman Kovalchukov, Carlos Castillo Mateos, Anastasia Voropaeva, Margarita Gapeyenko, Olga Galina, Nikita Tafintsev, Yuliya Gural, Anastasia Yastrebova, Kryštof Zeman, Pavel Marek, Konstantin Zhidanov, Jani Urama, Andrey Samuylov, Aleksei Ponomarenko-Timofeev, and Dmitrii Solomitckii. It was my great pleasure to work and collaborate with such exceptional professionals.

I would also like to acknowledge Prof. Markku Renfors, Prof. Mikko Valkama, Elina Orava, and Sari Kinnari for their professional assistance and guidance throughout my work and studies at TUT. Additionally, I would like to thank Saara Benfield for providing the proofreading service.

Above all, my warmest thanks go to my family and friends. Their trust, care, and support helped me a lot during the time away from home.

Aleksandr Ometov. October 10, 2018, Tampere, Finland
Contents

Abstract iii

Preface v

List of Abbreviations ix

List of Publications xi

List of Figures xii

1 Introduction 1
  1.1 Research Motivation and Background 1
  1.2 Main Contributions and Scope 2
  1.3 Structure of the Thesis 3

2 Wearable Technology as a Part of the Advanced Internet of Things 5
  2.1 Historical Overview on Wearables 5
  2.2 Consumer Wearable Applications 9
  2.3 Enterprise Wearable Applications 10

3 Wearable Devices Communication Opportunities 13
  3.1 Wearable Communications State-of-the-Art 13
  3.2 Wearable-related Communication Paradigms 14
  3.3 Main Network Architectures 14
  3.4 Secure Group Formation for Wearables 16

4 Proximity Social Clustering 19
  4.1 Social Relationships for Human and Machine Connectivity 19
  4.2 Relation between Sociality and Proximity-based Communication 20
  4.3 Aspects of Proximity-based Social Interaction 21
  4.4 Selected Numerical Results 23

5 Privacy Preserving Strategies for Wearable Technology 27
  5.1 Authentication Methodology for Collective Smart Device Usage 27
  5.2 Broadcast Content Access for Mass Events 32
  5.3 Anonymized Content Dissemination Methodology 34
  5.4 Numerical Evaluation of Cryptographic Primitives on Small-Scale Devices 35

6 Wearable Technology as a Part of Multi-Factor Authentication 39
  6.1 Overview of Enabling Factors and Challenges 39
6.2 Proposed Multi-Factor Authentication Methodology Based on Reversed Lagrange Polynomial 42
6.3 Proposed Intelligent Factor Grouping for Multi-Factor Authentication 44
6.4 Future Perspectives 46

7 Conclusions and Future Perspective 47

8 Summary of Publications 49
  8.1 Publications Description 49
  8.2 Author’s Contribution 52

Bibliography 55

Publications 65
# List of Abbreviations

<table>
<thead>
<tr>
<th>Abbreviation</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>AP</td>
<td>Access Point</td>
</tr>
<tr>
<td>3GPP</td>
<td>3rd Generation Partnership Project</td>
</tr>
<tr>
<td>5G</td>
<td>Fifth generation</td>
</tr>
<tr>
<td>A-IoT</td>
<td>Advanced Internet of Things</td>
</tr>
<tr>
<td>AES</td>
<td>Advanced Encryption Standard</td>
</tr>
<tr>
<td>AR</td>
<td>Augmented Reality</td>
</tr>
<tr>
<td>BLE</td>
<td>Bluetooth Low Energy</td>
</tr>
<tr>
<td>BS</td>
<td>Base Station</td>
</tr>
<tr>
<td>C-LOR</td>
<td>Co-location object exchange</td>
</tr>
<tr>
<td>C-WOR</td>
<td>Co-work object relationship</td>
</tr>
<tr>
<td>CC</td>
<td>Cloud Computing</td>
</tr>
<tr>
<td>CER</td>
<td>Central Error Rate</td>
</tr>
<tr>
<td>D2D</td>
<td>Device-to-device</td>
</tr>
<tr>
<td>EER</td>
<td>Equal Error Rate</td>
</tr>
<tr>
<td>FAR</td>
<td>False Accept rate</td>
</tr>
<tr>
<td>FRR</td>
<td>False Reject rate</td>
</tr>
<tr>
<td>GPS</td>
<td>Global Positioning System</td>
</tr>
<tr>
<td>HSR</td>
<td>Human social relationship</td>
</tr>
<tr>
<td>IAM</td>
<td>Identity and access management</td>
</tr>
<tr>
<td>ID</td>
<td>Identifier</td>
</tr>
<tr>
<td>IEEE</td>
<td>Institute of Electrical and Electronics Engineers</td>
</tr>
<tr>
<td>IoT</td>
<td>Internet of Things</td>
</tr>
<tr>
<td>IoWT</td>
<td>Internet of Wearable Things</td>
</tr>
<tr>
<td>ITC</td>
<td>Information and Communications Technology</td>
</tr>
<tr>
<td>ITS</td>
<td>Intelligent Transportation Systems</td>
</tr>
<tr>
<td>LTE</td>
<td>Long Term Evolution</td>
</tr>
<tr>
<td>M2M</td>
<td>Machine-to-machine</td>
</tr>
<tr>
<td>MCC</td>
<td>Mobile Cloud Computing</td>
</tr>
<tr>
<td>MEC</td>
<td>Mobile Edge Computing</td>
</tr>
<tr>
<td>MFA</td>
<td>Multi-Factor Authentication</td>
</tr>
<tr>
<td>MPR</td>
<td>Market pricing relationship</td>
</tr>
<tr>
<td>OOR</td>
<td>Ownership object relationship</td>
</tr>
<tr>
<td>PKI</td>
<td>Public Key Infrastructure</td>
</tr>
<tr>
<td>QoS</td>
<td>Quality of Service</td>
</tr>
<tr>
<td>RAM</td>
<td>Random-access memory</td>
</tr>
<tr>
<td>RSA</td>
<td>Rivest-Shamir-Adleman cryptosystem</td>
</tr>
<tr>
<td>SHA</td>
<td>Secure Hash Algorithm</td>
</tr>
<tr>
<td>TA</td>
<td>Trusted authority</td>
</tr>
<tr>
<td>TETRA</td>
<td>Terrestrial Trunked Radio</td>
</tr>
<tr>
<td>Abbreviation</td>
<td>Description</td>
</tr>
<tr>
<td>--------------</td>
<td>------------------------------</td>
</tr>
<tr>
<td>VR</td>
<td>Virtual Reality</td>
</tr>
<tr>
<td>WiFi</td>
<td>Wireless Fidelity</td>
</tr>
<tr>
<td>WT</td>
<td>Wearable Technology</td>
</tr>
</tbody>
</table>
List of Publications


List of Figures

2.1 The evolution of the Internet paradigms [P3] ................................. 6
2.2 Human-centric Advanced IoT applications [P7] .............................. 7
2.3 Potential user-centric wearable applications [P4] ............................ 9

3.1 Personal cloud example [P3] ..................................................... 15
3.2 Urban wearable communications scenario [1] ................................. 16
3.3 Possible general operation states [1] .......................................... 17

4.1 Structure of proximate social networks [2] .................................... 22
4.2 Impact of social relationships on the system throughput [P1] .......... 25
4.3 Impact of LTE coverage on the degree of connectivity in the system [P1] . 25
4.4 Impact of social relationships on the user energy efficiency [P1] ...... 26

5.1 The lifecycle of a wearable during its delegation [P3] ....................... 28
5.2 Power consumption during different operation phases [P3] .............. 33
5.3 The scenario of wearable utilization during an ice hockey match [P4] . 34
5.4 Cryptographic primitives performance evaluation of small-scale devices [P2] . 36
5.5 Evaluation of RSA execution [P2] ............................................. 36
5.6 Evaluation of Hashing and AES execution [P2] .............................. 37

6.1 Main operational challenges of MFA [P6] .................................... 41
6.2 Classic Lagrange polynomial-based secret sharing methodology [P6] . 42
6.3 Proposed reversed methodology [P6] .......................................... 43
6.4 Possibility of the cloud assistance in the proposed methodology [P6] . 43
6.5 MFA system mode with selected threshold [P6] ............................. 44
6.6 Comparing alternative factor combining approaches [P7] ............... 46
1 Introduction

1.1 Research Motivation and Background

Today, the number of interconnected devices is growing rapidly worldwide. With this increase, conventional connectivity through wired technology can no longer meet the requirements of this highly dynamic and mobile communication ecosystem. Naturally, the spread of wireless technology has been driven by its ability to support communication between various types of electronic devices.

Modern wireless technology is one of the most progressive fields of research in the telecommunications era. Wireless telecommunication has exponentially increased data volumes furthermore this trend is expected to continue, and mainly through mobile data usage in the years to come [3]. The lion’s share of this data traffic will be produced by the Internet of Things (IoT) devices, enabling cross-communication between humans and machines.

One of the most significant parts of IoT adoption is the emergence of personal wearable devices that are ‘worn’ by humans. These wearables have the ability to communicate with a chosen network, most commonly via short-range radio interface to a smartphone gateway, that in turn is equipped with long-range wireless technology delivering data to the cloud [4]. The wearable market has already developed extensively, and this growth trend is expected to continue [5].

There is a vast range of wearable devices on offer. From activity trackers, smartwatches, and AR glasses through to smart clothes, etc., wearable devices have a strong presence in the marketplace already in both consumer [6] and medical fields [7]. The possibility to customize and style them along with technological enhancements towards small-scale electronics and modern applications make wearables a strong contender in the IoT technological race. Almost one billion wearable devices are expected to join the IoT family by 2021 [3].

Currently, the strategy of making wearable technology (WT) communicate through a gateway device such as a smartphone leads to a number of security and privacy challenges. Conventionally, both groups were managed by a trusted centralized authority also controlling connectivity. However, one of the most significant tasks addressed in this work is related to operation under intermittent network connectivity constraints, i.e., when such centralized management is partially or entirely unavailable.

A reliable connection is needed between the user of a wearable device and the wearable device itself, requiring additional novel input. One option could be to build the trust relations on top of social connections between participants. Such relationships should be either active, preloaded in advance, or empirically based on shared interests. This tie
could be further used in many types of applications, including shared use of an electronic device, collaborative proximity-based gaming, and for access to services/devices.

Even when there is a pre-connection established and keeping in mind the fact that direct connectivity between two users can already be established manually on most smartphones, handling mobility and security may be challenging in high network dynamics [8]. Today, there are still no solutions available on the market that enable dynamic yet secure clustering when in proximity and in cases when a trusted authority is not present (tunnels, planes, distant resorts, or network overload due to a mass event). Despite that, social proximity would potentially enable varied wireless technology use cases whilst conventionally connecting to the cloud, i.e., through a known gateway of a socially-close, trusted user, or directly between devices in physical proximity [9].

Nonetheless, the fundamental basis of any secure and private communication is authentication. Checking the validity of a user/device has the potential to be significantly simplified by combining various biometric/behavior data from wearables together whilst also relying on a concept of MFA, i.e., the intellectual co-utilization of various data sensor points to enable usable and secure authentication.

This thesis focuses on the study of how WT could benefit from secure social clustering and, at the same time, may improve human lives by being a factor provider for MFA. Unlike most academic research and in addition to the design and evaluation of the solutions suggested, the data measurements taken from market-available wearable devices are provided.

1.2 Main Contributions and Scope

This thesis will focus on the challenging aspects of different wearables, including user sociality, security, privacy, trust, and any further related challenges to connectivity. This work looks at both theoretical and practical components, using simulations and prototyping as the primary drivers. The main contribution is made in enabling secure clustering and communications between wearable devices’ gateways in cases where connectivity to the trusted authority is not reliable.

This work continues the author’s M.Sc. research activity in the field of proximity clustering, showing how a framework based on secure sociality that had been developed previously may improve overall system-level metrics for both user- and machine-triggered communications.

Next, we developed a set of protocols that preserve privacy for the temporarily-delegated use of wearable devices in cases when there is only intermittent connection to the trusted authority. The complete set of protocols proposed allows for extension and modification of the delegation rules, whether or not the infrastructure network is accessible by any of the user gateways. Moreover, we propose an MFA framework enabling differentiated dissemination of content during a mass event in the broadcast communication link.

Additionally, a set of measurements is established on smartphones and smartwatches (currently available on the market) to assess possible implementation of these proposed frameworks and how users will be affected during the execution of cryptographic primitives.

Finally, there is an overview of the current state-of-the-art MFA solutions and the corresponding perspective brought by the broad adoption of WT and IoT in general. It is found that present connectivity strategies do not meet security or usability criteria. To overcome that challenge, we propose an intelligent factor combination approach.
1.3 Structure of the Thesis

There are seven chapters, followed by a compilation of seven publications [P1]–[P7]. There are also references to several co-authored publications, closely related to the topic of this thesis. Furthermore, proximity-based social clustering research considered in this work is partly used in the author’s master thesis [10].

Chapter 2 of the introductory section gives the motivation, objectives, and state-of-the-art in addition to a detailed overview of potential wearable applications for both consumer and enterprise markets.

Chapter 3 outlines current wearable communication architectures. This chapter also provides an overview of the main networking paradigms that involve communications and computing of wearable electronics. At the end of this chapter, an algorithm enabling secure group formation between the devices in proximity is described.

Chapter 4 details the applicability of social ties between users aiming to improve the communication metrics of interest such as throughput and energy efficiency. It starts with technology adoption, its relation to sociality and the corresponding aspects of proximity-based interaction. The chapter ends with the evaluation of the communication in one network cell where direct links can be established based on a pre-defined sociality/closeness between users or devices.

Chapter 5 details one of the most significant aspects of today’s information technology market – collective (or shared) use. The protocol suite enabling temporary delegation of device usage under intermittent connectivity to the cloud is provided and analyzed in terms of power consumption. Next, a methodology for differentiable access to the broadcast data during a mass event based on two or more authentication factors is described. The chapter ends with a comprehensive evaluation of cryptographic primitives on real portable devices.

Chapter 6 addresses how wearable electronics could benefit from the concept of MFA. Next, the comparison of conventional factor grouping strategies is given from a security vs. usability perspective, and an improved intelligent methodology is proposed. The chapter ends with a vision from the perspective of MFA application for the advanced IoT.

Chapter 7 concludes the introductory part, followed by a compilation of the publications.
2 Wearable Technology as a Part of the Advanced Internet of Things

This chapter provides a perspective and background to wearable market technology. First, there is a review of the global challenges from connectivity, security, and human perspectives. This is followed by current applications for consumer and enterprise wearables.

2.1 Historical Overview on Wearables

The global communications system of the Internet has faced incredible transformation in recent years, and the evolution of its main paradigms is shown in Figure 2.1. Just over 20 years ago, computers were interconnected by a fixed network, thus allowing those first millions of users to communicate over email, enabling the concept of ‘connected places’. Next, home and office Access Points (APs) took their position in supporting readily available connectivity. Simultaneously, users’ communication options also increased.

Since the beginning of the present century, there has been a dramatic transformation in connectivity. The proliferation of smartphones and tablets equipped with various wireless modules has allowed billions of users to connect to the Internet. Since this increased connectivity, global swathes of the population have become exposed to a richer style of media content and communications across the entertainment market. As this connectivity has grown and people expect more from their devices, it has been seen as time for machines to take the lead. Today, there is a shift towards supporting a high number of devices with entirely different traffic patterns, known as the phenomenon of ‘the Internet of Things’ (IoT) [11]. It aims to support tens of billions interconnected smart devices by the end of this decade [3].

The IoT era is depicted as an entirely new technological penetration into the everyday lives of modern humans – it covers a vast number of use cases, ranging from robotics to smart grids and cities [12]. Current IoT devices are already collecting, processing, storing, and combining enormous volumes of information that result in new areas of behavioral knowledge and can lead to efficient designs and decision-making. The primary task of all of the above is to improve society by automating critical and routine tasks in many areas of technology.

By simplifying people’s lives, IoT brings many challenges from a communications perspective. Interconnecting an unprecedentedly high number of devices with different technologies operating under different protocols and each with their own requirements concerning throughput, latency, reliability, etc. is an extremely challenging task for
system architects. Proper selection of all of the above requires not only careful planning but also vision and expertise while collaborating with other vendors and researchers.

As devices become more interconnected they rely more and more heavily on surrounding objects, not only for transmitting the data but also for its processing [13]. Conventional IoT systems usually involve a high number of low-cost devices such as actuators, sensors, and smart meters – these periodically transmit collected data to the chosen gateway responsible for pre-processing and aggregation. These massive IoT deployments are sometimes resource and power constrained, thus requiring the reconsideration of traditional security methods. In the last decade, IoT security and privacy have been one of the most significant topics of interest, especially in the areas of light-weight cryptography [14], and secure connection establishment [15, 16] conjointly to privacy-related problems [17].

In contrast to constrained devices, the trending topic of security today centers on the worldwide penetration of advanced IoT (A-IoT). This includes personal smart devices, smart vehicles, consumer drones, and other environmental devices (see Figure 2.2). These devices are defined by higher computational power and increased memory efficiency. They are more expensive and less battery dependent (e.g. we are already used to charging our smartphones every night). Therefore, delivering a higher level of security is a natural, progressive step to make – enabling guaranteed secure machine-to-machine (M2M) connections as an A-IoT feature.

Simultaneously, potentially unauthorized access to such powerful devices leads to new risks. These range from conventional theft to placing human life in danger. For example, by manipulating the data shown by AR glasses, one can change navigation-related information. Therefore, authentication and reliable assessment of who owns the device, directly projected into trust, become critical challenges in delivering easy-to-use and secure operation of IoT devices.

Utilization of a vast number of wireless access technologies [18] coming hand in hand

---

**Figure 2.1:** The evolution of the Internet paradigms [P3]
2.1. Historical Overview on Wearables

with a plethora of connected ‘things’ is only the beginning of a profound convergence between humans and machines that could, in turn, take humanity to an entirely new Internet epoch. Interconnected things carried by humans, namely, ‘wearables’ provide innumerable benefits for improving our life. They act not only as data collectors but as smart assistants when it comes to our health, safety, and many other aspects of life. Worldwide, the wearable market has already reached sales of approximately 25.1 million devices [19]. It takes us to an entirely new level of the evolution of the Internet: the Internet of Wearable Things (IoWT) [20].

Today’s consumer-wearable technology is beyond its first steps but is still very much in its infancy. Moreover, information gathered and used in context can deliver a truly personalized user experience for end users [21]. In addition to conventional sports trackers, smartwatches, on-body cameras, heart rate meters, and eyewear, the upcoming generation of wearables will also involve AR and VR devices, wearable smart clothes, and proprietary enterprise wearable equipment. As Ericsson envisions it, close to 70% early adopters are already interested in correlating their lives with next-generation wearables [22]. Even today, the process of improving health and obtaining sport-related information is made extremely simple by means of wearable technology that is tightly connected to the cloud services and delivered by Apple, Google, Garmin, Polar, TomTom, and others. Generally, modern IoWT devices are already providing a smartphone-like experience by employing voice and gesture control together with well-designed input and output interfaces.

Even though wearables can easily penetrate our lives and are small regarding form factor, they are more intuitively subject to power consumption limitations that other ‘portable’ devices [23]. Despite that, the swiftly growing number of wearables brings a problem of system scalability [24]. Current wearable connectivity architecture usually implies constant availability of the *gateway* device, e.g., a smartphone and any communications to the traditional Internet are expected to be organized through that ‘gateway’ node. By this means, a wearable device is not ready for communicating with another device in proximity (notably, made by different vendors). This is likely to be the case even if the alternative connection is a more efficient one than its gateway. To make this alternative
connectivity more feasible, new communication opportunities need to be explored to enhance user experience. However, current connectivity assumes the WD or user will face the challenge of not fulfilling the assumption of their constantly available cloud connection due to various factors – ranging from network failure, obstacles, or distances.

Therefore, we identify the primary wearable-specific constraints as *limited computational power and limited operation under intermittent network connectivity*. This leads to the primary goal of this research – the absolute need to rethink old data security, integrity, and reliability approaches for cases of intermittent connectivity to the operational cloud. Firstly, it must be considered that personal wearable devices usually hold extremely sensitive owner-related data and are naturally more exposed to the public than other portable devices. Some say that by using wearables, we are publicly ‘wearing’ a portion of the most significant and personal information about ourselves, including health, relationships, activity, and our favorite locations. Under this banner, wearables simultaneously become one of the most private and yet exposed sets of devices in our entire technological ecosystem. Thus, adequate protection of this private data is an immediate concern. Being already relied upon heavily for all kinds of medical data, information carried in wearables should adopt and improve the best practices of data privacy preservation as soon as developments allow.

Today, authentication is one of the most widely-used techniques to achieve data privacy [25]. The classical definition of authentication, that authors agree on, is when a “user identifies oneself by sending $x$ to the system; the system authenticates the identity by computing $F(x)$ and checking that it equals the stored value $y$” [26]. Over the years, the definition remains the same even though $x$ today has the potential not to be a password or key but a biometric factor such as a fingerprint. The main drawback of dynamic on-the-fly authentication is indeed the complexity of application in large-scale scenarios [27] and especially if there is a need to transmit and store data across heterogeneous environments that are not only in proximity but also remote [28]. Such a diversity regarding both devices and connectivity options also leads to increased complexity concerning access rights management. This further supports the need for beyond 5G-grade wireless communication opportunities inside the “personal cloud” or the user [29]. Since the wearable market is always being filled with new products and concepts available to an individual, new solutions and techniques are necessary to manage security. In particular, wearable devices in proximity to each other create opportunities to cluster together and establish trusted clouds between the services and devices themselves, unified by different users [1].

In a broader sense, data privacy is bounded to the control over user’s sensitive information which, in turn, implies the assured level of trust – especially during the transmissions when partners are expected to be whom they claim to be [30]. In the very beginning of the Internet era, trust was consistent with the physical terminal computers connected with wires where users authenticated themselves with local accounts. With systems development, the centralized remote servers became responsible for managing the access. More recently, it has become more expected for things to take advantage of connecting with their surrounding objects, the devices in their ecosystem, and also us. Once again, this is all aimed at improving usability and device security using a novel heterogeneous approach combined with a centrally distributed mechanism.
2.2 Consumer Wearable Applications

Smart consumer wearables devices have become part of many areas of our everyday activity [31, 32]. Emerging WT allows for several different types of multi-user interactions and user-centric contextual services that rely on dynamic local data sharing [33], such as proximity gaming or AR/VR services. Some consumer examples of wearable applications are given in Figure 2.3.

![Figure 2.3: Potential user-centric wearable applications [P4]](image)

Today, there is a large amount of wearable clothing and accessories available on the market that has been found to enhance human capabilities markedly. Many large conglomerates are currently in pursuit of promising wearable startups – often resulting in acquisitions. Examples include: Intel who acquired Recon Instruments and Replay Technologies, etc. or Fitbit who paid $23 million for Pebble. Some new releases have also come onto the market recently, such as Samsung brainBAND, utilized for body measurements during a match. Such deals have resulted in an extreme market rise, with expectations to exceed US$ 24.640 billion by 2022 [34].

In fact, the previous development and production of wearables were primarily focused on delivering a higher level of automation, and this was mainly in the field of healthcare. Most were pragmatically driven by the standard business model, which required a clear return-of-investment strategy [35]. Even by this time, the environmental and contextual features of such systems were not hard to model or monitor due to their predictable behavior – leading to solutions that operated flawlessly in a predefined context but failed in real life. One example of this: AR/VR applications [36].

One of the most significant niches of the wearable market is mass consumer applications. However, a recent survey claimed that more than a 50% of people who purchased a wearable device would most probably stop using it after only half of a year [37]. Consequently, some companies offer new wearables for a “try-before-buy” time interval, showcasing an
unconventional type’ of collective use application. The company Lumoid, for example, offers a way to pick the most suitable fitness band by trying it for only 20$. When the trial period ends, a person can either return the unwanted product or purchase it. In contrast, company ByeBuy uses a “pay-as-you-go” model for its advanced smart devices. Remarkably, when the device is returned to the shop, some sensitive data could still be stored in its memory even after a factory reset. The potential collective use of such personal devices is usually neglected when the device is sold or disposed of. In contrast, big companies have policies on how to properly recycle the hard drive of any smart device thus preventing sensitive data leakage.

In contrast to personal use, collective use or rental applications bring the attention back to questions of data privacy and trust. Further compelling use cases of this application are listed here. Despite conventional use cases, on-body cameras, Depth gauges, and spearfishing guns could be obtained directly from a scuba diving vessel. While talking about distant resorts, rental of smart skiing equipment with boot sensors, smart body armor, augmented reality glasses would be able to improve the entertainment of the visitor as well. The gigantic market of in-flight entertainment could be offered with a solution based on a VR headset for gaming or enjoying the movies. Access to properties or electronic devices could also be based on proximate connectivity. Here, the access could also be granted for special services in cases of emergency, i.e., for a medical staff, a firefighter, or a police officer.

Predicted within this thesis, the “pay-as-you-go” model is increasingly used in markets including car-sharing and apartment rental when granting physical access. Furthermore, companies such as Netflix or Spotify apply the same model to digital content. People are getting used to this convenient way of life, and the author believes that many services and objects will adopt the same level of flexibility and subscription-model shortly.

2.3 Enterprise Wearable Applications

In contrast to consumer electronics, enterprise or industrial wearable devices are usually less personalized and do not carry excessive personal information. Wearables can also be used for leisure and entertainment. Generally, this will be in areas such as proximity-based AR/VR gaming, non-confidential information sharing, and similar non-critical services that could be of significance from this perspective. These applications do not necessarily need evidence of explicit social relations between device owners, and the sociality of devices may instead drive trusted communications. Typical scenarios of interest in this category may focus on users distributed in a particular area with similar interests, e.g. information dissemination at a stadium, a university campus, or even a restaurant. In these scenarios people will be matched according to their interests, age, familiarity, etc. and will interact through their devices.

For urban scenarios, mobile wearable technology is already utilized for public safety and security [38]. Advanced wearables are being used across various branches of the special forces including. Those devices include smart-cameras, health monitors, and communication units [39]. One of the most notable projects has been developed by industrial giant Motorola and involves the so-called “Connected Law Enforcement Officer”. This project aims at improving response times within a mission-critical ecosystem by enabling real-time collaboration between different specialized devices. In other words, Motorola has found a solution which allows teams of people to communicate in an effective, secure, reliable, and cooperative way, in real time, regardless of their network or carrier
2.3. Enterprise Wearable Applications

with a centralized authority connection. The proposed system has applications that include data, exchange, voice and video streaming, ‘push-to-talk’, report writing, data capture and others. The proposed set of devices includes: (i) a body-mounted camera with a video speaker microphone and touchscreen interface. This will be utilized for intuitive capture and storage of the evidence; (ii) the important classical technology for special units communication is ‘push-to-talk’ radio, used for emergency two-way radio connectivity; (iii) a handheld smartphone that offers reliable broadband connectivity and enables users to enjoy the intelligent support of the entire team; (iv) AR glasses that are coming soon. They assist the user in a range of tasks and provide real-time analysis of the surrounding environment for faster response; and (v) a connected car can provide many opportunities – including TETRA communications, mobile computing features, mission-critical 3GPP and LTE broadband connectivity for the personal cloud. This vehicle is mainly of interest for our case study due to its superior computation and caching capabilities, as well as more stable connectivity to the network infrastructure.

Similar technology is adopted by firefighters [40, 41]. The main additional component is a custom in-mask thermal camera with the real-time display that offers a clear, preprocessed vision by continuously augmenting a thermal image in the helmet.

In contrast to the usability and reliability in the case of the special forces, professional sports WT was aimed, initially and primarily, at increasing the level of spectators’ entertainment and secondly at improving the player’s skills. WDs for sports players have been around for some time, and when worn during the training they make it possible to improve advice given by coaches and doctors. Note that such devices are still prohibited from being worn during actual competitions [42]. In football, for example, the only wearable allowed in the field is the referee’s watch – used to notify him/her when a goal is scored [43].

Outside of current rules, another reason why football players or some other active sports participants do not wear additional devices during a match is the weight. In contrast to football, hockey players carry a weighty set of protective pieces like pads, guards, and helmets. This theoretically allows for the installation of various lightweight sensors and small cameras around the body. Moreover, those wearables have been in development for more than 20 years already [44, 45, 46] and are very slowly coming to the market [47]. For example, the National Hockey League have already installed tracking devices in players’ clothes and the puck in 2015, thus allowing for more informed post-match discussion and overall performance increase.

One of the most promising entertainment trends is about providing spectators with an enhanced sense of immersion into a game. Wearable technology will hugely overtake the 20th century immersion solution that was the television broadcast, bringing with it a whole new range of opportunities. The data collected from players is currently only available to an extremely narrow circle of people. The author believes that new, emergent services will allow a spectator to have a first-person view of his/her favorite player whilst they are scoring a goal, for example, in just a few years.

Many companies are already integrating their devices into professional sports. Adidas offers a solution called miCoach that is a combination of wireless sensors capable of monitoring heart rate, speed, athletic performance, and other health metrics. The collected data is continually sent to the coach for better individual monitoring and training. Another device utilized in professional sports is called Viper Pod. It is mounted on the chest and collects data related to position, acceleration, compass, and heart rate. It is mainly used by rugby teams. Another notable device is the Catapult OptimEye G5,
designed explicitly for goalkeepers. It tracks their movement, together with a number of other statistical data [48]. The company Armour has created a smart t-shirt called the ‘E39 performance shirt’ with similar functions. Another exciting piece of technology is a ShotTracker – this is used in baseball and is a combination of a wristband and net monitor, allowing players to monitor, track, and then analyze their hits after a game.

This market-driven progress proves that wearable technology has many promising opportunities. This is not just about creating products for profit from consumers but also for application in mission-critical scenarios and professional sports. The examples listed here are about making the lives of many people more safe, reliable, and enjoyable.
3 Wearable Devices Communication Opportunities

This chapter will focus on the communication architectures analyzed during the research phase. It will start with an introduction to the Information and Communications Technology (ICT) global paradigms that may be applied to IoWT. Next, the central network architectural aspects required for WDs are discussed. Then, the information security framework developed for the dynamic clustering of the devices closes the chapter.

3.1 Wearable Communications State-of-the-Art

Overall, current wearable devices use the owner’s smartphone as a gateway to connect to the Internet. Operational efficiency can, therefore, be limited if the gateway’s battery is drained or the connectivity options are not adequate. Conversely, some wearable devices are already equipped with a long-range wireless radio, but continuous use is still not recommended due to the higher power consumption.

From a technological perspective, most of the ‘mobile’ wearables use short-range wireless technologies to access the gateway. These include, for example, IEEE 802.15.4 (BLE, ZigBee) or the more power hungry IEEE 802.11 (WiFi) [49]. Some vendors even suggest connection to conventional WiFi infrastructure from, for example, a smartwatch. In contrast, AR/VR wearable equipment requires a higher data rate and thus can only be connected through either 5G cellular or at least IEEE 802.11n links – this is due to the bulky data streams. Note, cellular connections are limited due to the licensed spectrum limitation, but there are researchers with a deep interest in licensed solutions for proximity-based communications, such as LTE-Direct [50]. However, many think that managing the interference and power control may be challenging in this scenario and thus practical device-to-device (D2D) applications should take their niche [51]. Note D2D is defined as proximity-based wireless communication which is maintained by the remote control center.

The technology for D2D is already available on many smartphones. WiFi-Direct allows us to use conventional WiFi chips without any additional modifications for D2D as well. Developers today are intensely researching technologies that operate on a higher, unlicensed frequency, e.g., IEEE 802.11ad and IEEE 802.11ac that can provide Gigabit data rate. Moreover, the utilization of this higher frequency has its propagation limitations, which are eventually a benefit for use in densely-packed scenarios, where human bodies act as a natural blocker and allow better spatial reuse [52].
3.2 Wearable-related Communication Paradigms

Clearly, proximity-based communication performance could be achieved using user-triggered links or conventional cloud solutions. However, when there is a failure in the gateway connectivity, this may be extremely challenging. One possible solution for improving the reliability of the wearable device’s operation is to use trusted nodes in close geographical proximity to relay essential data to the cloud or another personal device of the user. The following section will list the paradigms potentially involved in this type of communication.

Since computational power and battery commonly limit wearable devices, one of the related paradigms is Cloud Computing (CC) that was already adopted at the end of 00’s. This concept is about transferring the complex computations and data processing from the constrained device to large data centers [53]. The key driver in this cloud paradigm was initially the prevalence of wireless and broadband networking in a trade-off with storage and processing costs. Simply put, the network side is exceptionally overloaded, but the computations are complicated on the end node. Therefore there is a possibility of running wearable device services that cannot be executed in the remote cloud, in addition to simultaneously offloading cloud-enabled applications.

In 2012 Cisco Systems defined the Fog Computing as an “extension of the cloud computing paradigm (that) provides computation, storage, and networking services between end devices and traditional cloud servers” [54]. The computation is not necessarily performed in the actual cloud but at different nodes, side on, orchestrated by the main cloud [55]. Therefore, the first property of computational and communications heterogeneity is presented by the fog concept [56].

After just one year, IBM and Nokia Siemens Network introduced the concept of Mobile Edge Computing (MEC). The main idea behind this was to enable a flexible platform deployed on a different vendor network to the operators’ edge hardware – so that any applications could be executed immediately after the data was transmitted over the wireless channel [57]. The delay related to data traveling between the edge and the cloud could also be neglected, decreasing the load on intermediate routing hardware. One of the real examples of MEC is Mobile Edge Scheduler [58] which minimizes the LTE downlink traffic delay. From an application developer’s perspective, MEC functionality is expected to be open for third parties, i.e., not only kept for the cellular operator, and thus opening a road for entirely new services.

Interesting from the author’s perspective, one of the most significant technological aspects of wearables is Mobile Cloud Computing (MCC) [59]. The primary driver behind this technology is an attempt to redirect the computation and/or data storage to less constrained devices close by [60]. The main difference from MEC is the ability to run the execution on not only the infrastructure network edge nodes, but also on any neighboring devices and, potentially, the gateway.

3.3 Main Network Architectures

The paradigms mentioned above could not be discussed without also mapping out the supporting network architecture. As wearables mainly communicate through smartphone gateways, they naturally employ ‘direct’ communication that not only acts as a conventional data transfer technology but also enables many applications in M2M
3.3. Main Network Architectures

communication. Next, the communication flows to the edge devices and then to the cloud, as shown in Figure 3.1.

![Figure 3.1: Personal cloud example [P3]](image)

Below is a list of the main connectivity options that involve communication between wearable devices and/or gateways in MCC. This section focuses on providing an overview of architectures and potential solutions for securing the connectivity between nodes.

Standard network use and coverage today involves an implied central control node through which all communication and data exchange occurs. From an information security perspective, the infrastructure network is defined as a trusted authority (TA) that manages connections and distributes security- and trust-related information. Public Key Infrastructure (PKI) [61] solutions are commonly utilized for providing security in those cases.

In specific real-world scenarios, there will be times when connectivity to infrastructure is not available at all (see Figure 3.2). For example, this may happen on a cruise or a distant ski resort, etc. Here, users/wearers may still want to communicate, play AR-enabled games, and share memories from their head-mounted cameras. Today, there are no market-available solutions to these situations. However, Apple or Android users can use short-range communication to share data between smartphones through AirDrop or WiFi-Direct, though this is not always possible or straightforward. That only proves that the technology enabler is already integrated into most of the devices on the market, there is just a place for an additional level of extraction that allows the gateways to share the link with their wearable devices. From a security perspective, there needs to be an option to utilize a preloaded certificate from PKI, therefore providing some level of trust, through which ID-based cryptography or a one-time password could be used to make the final connection. Therefore, it would be possible to provide some level of service in cases where it was not previously offered at all. Moreover, the neighboring gateway may be utilized as a node relay without an appropriate connection to the infrastructure. However, challenges of data privacy, trust, and signaling overheads arise. New primitives need to be developed to enable such an operation.

Generally, wearable devices and related proximity-based communication are expected to generate a significant share of data traffic during the oncoming years. Thus, the main effect of D2D communication on future 5G systems will be to relieve conventional infrastructure
networks from the additional load. Even though the technology is only present in certain devices, this already has the potential to help. D2D links based on WiFi-Direct or BLE could be used to offload the traffic whenever used, especially for data distribution as generated by wearables between users [62]. Overall, D2D is slowly achieving Quality of Service (QoS) improvements from the user side and mitigating licensed spectrum congestion on the other [63]. Later on, this work will look at the security enablers making proximity-based communication more safe by integrating centralized and distributed systems together [64]. The corresponding clustering procedure is based on the majority rule and can be executed either automatically or by human choice based on predefined policies during when nodes are added/removed from the group.

Conventional solutions were not originally designed to provide reliable security and safety in cases where the quality of the infrastructure connectivity could not be guaranteed. Furthermore, they also require the group formation capabilities to be included in every edge-enabled service.

To unshackle the developers from adding the coalition forming as part of the application logic, an appropriate framework for operation needs to be developed along the network edge. The main idea behind the formation is to develop secure and private data delivery between proximate users in cases of both infrastructure and distributed operation. Nonetheless, some nodes may have a stable connection to the centralized authority, bringing an opportunity to create their own secure formations from both a logical and physical point of view. Today, user addition and removal is managed by the cloud while the proposed methodology extends such operation to scenarios without a reliable connection to the centralized entity [65].

### 3.4 Secure Group Formation for Wearables

From an information security perspective, the establishment of a secure coalition without the connection to the trusted authority may be extremely complicated. Most of the existing networks use an IP to enable routing inside the network. This unique temporary address is usually provided by the network with dynamic behavior over time in most cases,
excluding Mobile IP. When the centralized network is unavailable, it becomes necessary to construct and use new policies and routing strategies on top of conventional means, mainly due to the limitations set by the LTE core itself \[66\] which in turn limits the potential of proximity-based communications.

Together with the Brno University of Technology, we have developed and tested, in the live 3GPP LTE core, a framework that allows for proximity-based communication between smartphones even when connection to the cloud was not 100% available. The framework is developed according to \[67\], where the centralized control node is deployed inside the cellular operator core. Users can establish secure coalitions connected to the cloud while user addition and removal was achieved without such connectivity. The server was mainly responsible for the management of security and connectivity \[68\].

Modern widely deployed cellular networks do not yet have efficient enablers for delivering cellular-assisted communication. User dynamics and continuously changing topology put limitations on the reliable operation of such systems. Particularly noteworthy in this work is users’ trust while being outside of the network coverage but still connected. The very moment a device reaches the infrastructure connectivity again – it would report all status updates to the cloud. This was the main reason why pure ad hoc solutions could not be applied in the D2D context, as our preliminary work \[64\] demonstrated.

Let us concentrate on an arbitrary coalition in the network. As it is shown in Figure 3.3, the options in cluster connectivity define some individual cases for our chosen scenario. The coalition may be located entirely in the cellar Base Station (BS) coverage. In this scenario, the information security procedures are performed over the infrastructure connections via the cloud. The first case in Figure 3.3 presumes that the actual data flow and security-related signaling is traveling through the cellular interfaces of the devices. The second case depicts a scenario when all users are connected to the network, and thus all the coalition establishment procedures are managed remotely. In the third case, the data and signaling are transmitted using direct links, providing a possibility for any users

![Figure 3.3: Possible general operation states \[1\]](image-url)
in proximity also to validate their communication when there is intermittent connectivity, i.e. when the connection to the cloud was not present for a certain period. Finally, the cluster could be entirely out of network coverage. Here, ad hoc strategies were used to establish adequate and timely direct connections for the users. In the following D2D paradigm, the presence of the centralized control is assumed, with initialization, execution, and cancellation being managed by the centralized network service.

The author’s framework developed and evaluated during his M.Sc. studies is built relying on the 3GPP model, allowing proximate communication when the device location is either handled by the cellular operator or by short-range wireless technology or, for example, BLE on the device side. Each user has a unique ID and certificate signed by the trusted authority. The certificate is either pre-installed with, for example, a SIM card, or is delivered during the registration of the user into the system. It is for secure group establishment and user validation. Based on this definition, the certificate retrieval requires some connectivity to the trusted authority.

Overall, the grouping procedure is the only one where connectivity to the cloud is required for future monitoring purposes and main group secret generation. After the devices have received their secret and public keys together with the corresponding certificate, any user can trigger the coalition initialization procedure with its neighbors. The willing user is then generating a request that contains the joining coalition node IDs to the cloud. Further, the users are polled to verify that they have agreed to enter the coalition. Note that it would be possible to automate this procedure. When users have agreed, a group certificate and group secret are generated by the cloud, refer to [64] for the details. Next, the server can be contacted by active group participants to add or remove users based on the majority rule, but the group secret remains the same.

Of course, users may come and leave the cellular connectivity alone or together as a group. If this happens, cloud connectivity cannot be guaranteed, so further techniques should be used to support communication between the users inside the coalition. The data, in this case, is sent directly between users. However, new coalitions cannot be created due to a lack of certificate authority for any future validation. On the other hand, users can still add and remove other ones based on community-oriented voting. Fundamentally, the Lagrange polynomial utilized in the clustering allows for the group secret to be recovered and creates a new secret share for a new user to join. Importantly, not all users in the coalition need to vote in favor for this to happen, but there must at least be a predefined amount of positive votes. The generation of the coalition is triggered by this number of positive votes.

In case a group of users is willing to exclude someone from the coalition, they may trigger a similar, but opposite, procedure. However, they would have to securely redistribute the shares to everyone in the active coalition excluding the unwanted user. The required number of users (defined during the coalition creation) would trigger another set of polling, and then group together and indirectly reconstruct the group secret, redistributing new shares between those. Management of this procedure would preferably happen on the most power-independent device, but it could also just be selected at random within the group.
4 Proximity Social Clustering

This chapter will center on the sociality aspect and the corresponding potential benefits brought by its utilization while interconnecting users in proximity. Starting with the adoption of this technology and possible applications, the chapter will focus more on how current technology may benefit from utilizing the over-the-top layer of social relations to improve overall communication. Next, the aspects of proximity-based social interactions, such as trust, ‘closeness’, ‘betweenness’, and others, are discussed. Numerical evaluation of the potential benefits in terms of conventional communication metrics is given at the end of this chapter.

4.1 Social Relationships for Human and Machine Connectivity

The rapid standardization and promised benefits of proximity-based communication under centralized network control brought the attention of the research community to the question of how those links could benefit from social user interactions that bridged the physical and virtual worlds?. Enabling such dynamic interactions would deliver a number of challenges related to establishing connections to both known nodes and unfamiliar devices. Historically, the field of identity and access management (IAM) aimed at offering solutions for flexible access to systems and services [69, 70]. The epoch of new and continually emerging applications supports the requirement for easy-to-use and straightforward IAM solution development. D2D communication in that mold is no different. The challenges range from data privacy preservation to operation in a mission-critical environment. An essential problem here is the fact that traditional IAM frameworks may fail under unreliable connectivity constraints.

In our previous works [64, 66], an information security framework was developed, allowing dynamic clustering in cases of intermittent connectivity to the trusted authority. This chapter provides a vision of the central vectors of trust and sociality in proximity-based communications. The numerical evaluation shows how devices can benefit from direct links even after the introduction of transitional and computational overheads of the clustering.

Despite straightforward benefits brought by using ‘simple’ D2D communications (higher system throughput, lower latency, etc.), the impact of efficient identification of human behavioral patterns and relationships may improve those even more. In addition to the fact that users in the same social group are likely to be interested in retrieving or sharing the same content, the trust in this group is intuitively considered to be higher when compared to a group of random strangers.

Interestingly, proximity-based communications may even affect external interactions between the gateway and the surrounding environment. One of the ways to improve
the establishment of the D2D trust relations from the user perspective is by monitoring
the common contacts (including friend-of-a-friend and other weak ties) by the trusted
authority and thus improving the overall decision-making process. Regrettably, neither
operator nor customers are ready for such steps yet. The main problem behind this is
a lack of value proposition for end users. In our model, a high level of trustworthiness
between users should be met to mitigate the risks of user distrust and rejection (see
Table 4.1).

Table 4.1: Social relationship factors between devices, possible applications, and the associated
trust value: UD – user-driven; DD – device driven.

<table>
<thead>
<tr>
<th>Relationship</th>
<th>Type</th>
<th>Description</th>
<th>Applications</th>
<th>Trust</th>
</tr>
</thead>
<tbody>
<tr>
<td>Human social relationship</td>
<td>UD</td>
<td>Familiarity degree with friends, relatives and colleagues</td>
<td>Leisure applications, confidential data, eHealth, mission-critical communications</td>
<td>[0–1]</td>
</tr>
<tr>
<td>Market pricing relationship</td>
<td>UD</td>
<td>Cooperative interactions with services triggered by the environment</td>
<td>Proximate marketing, proximity gaming, advertising</td>
<td>0.2</td>
</tr>
<tr>
<td>Ownership device relationship</td>
<td>DD</td>
<td>Relationship between devices with the same owner</td>
<td>Personal cloud, smart home</td>
<td>1</td>
</tr>
<tr>
<td>Co-location device relationship</td>
<td>DD</td>
<td>Devices that share personal experiences (e.g., cohabitation)</td>
<td>Information/data exchange at social aggregation points (concerts, sports events)</td>
<td>0.8</td>
</tr>
<tr>
<td>Co-work device relationship</td>
<td>DD</td>
<td>Devices that share public experiences (e.g., work)</td>
<td>Information/data exchange at work aggregation points (e.g., fairs, workshops)</td>
<td>0.6</td>
</tr>
</tbody>
</table>

4.2 Relation between Sociality and Proximity-based Communication

The utilization of social connections between people may significantly improve proximity-based communication in various applications and services [70, 71]. The utilization of only ‘social links’ without considering physical and technological limitations is not, however, an option. Researchers must examine the interactions between devices, keeping in mind any sociality impact. There has already been a study on trustworthiness for the simpler IoT device perspective, where all logic was already predefined and listed [72]. Therefore, we can classify the current interactions into two large groups.

The first group represents user-driven interactions. Here, the central activator for connectivity is a social human being. Human social relationship (HSR) factor is a metric that defines the degree of interest in data sharing between two individuals. It varies based on a multitude of connections that could be about family, friendship or between colleagues. HSR represents the trust and therefore likeliness to be connected with someone. From another perspective, the market pricing relational (MPR) model could be used to evaluate
trust between individuals but only between ones willing to achieve some mutual benefit. As an example, one may consider a flea market, a party, a gaming event, etc.

The second group is related to device-driven sociality, being of high interest in this thesis. Smart devices can interact with other ones automatically based on the set of predefined rules by either their owner or the manufacturer. Here, owners are not necessarily required to interact with their devices when the communication is triggered. To construct this sociality level, relevant context and mobility patterns can be used to effectively construct the appropriate forms of social relations [73], i.e., the wearable device could, for example, trigger the establishment of a link between its gateway and another smart device in proximity – or even temporarily change its gateway to a neighboring but trusted one.

Another significant challenge for proximity-based communications is the high dynamics concerning device mobility. Because all the nodes in such a network are highly mobile due to their portable nature, the dual mobility of any communication entity should be thus carefully taken into consideration, as to provide extended support for dynamic trust management that is directly related to the established communication links. Moreover, since the centralized control node does not know the mobility of the out-of-coverage node, the established connection should involve additional medium sensing. Further on, we focus on the following social scenarios.

Trust-based human applications correspond to interactions between humans who have a strong mutual trust. Here, each user wants to know whom its data is exchanged with. Primary examples are work-related scenarios including public safety, construction sites, or cargo handling stuff where safety requirements are stringent and dictated by this level of trust.

Leisure and entertainment applications correspond to gaming and non-confidential information exchange. The trust relation, as well as strong social relations between communicating users/devices, are not critical. Generally, users may be grouped based on location, interests, or participation in some social group.

The last group is critical M2M applications where human interaction is limited or non-existent at all. Thus, automatic trust management is utilized while establishing connections between the devices. The most exciting examples, despite conventional gateways’ connectivity, are related to industrial and hazardous robotic operation where devices communicate with mission-critical data. Here, trust management is defined by a specific set of policies leading to higher level of optimized communications performance.

### 4.3 Aspects of Proximity-based Social Interaction

The first aspect of proximity-based communications from a social perspective is indeed the idea of trust which in this technological context has challenged many minds for more than half of a century [74, 75]. The following definition has been given in [76], which we generally agree with: “Trust is the extent to which one party is willing to participate in a given action with a given partner, considering the risks and incentives involved”.

In the extremely dynamic context of proximity-based communications, developing a trusted relationship is complicated due to the requirement of anonymity in such scenarios [77]. Overall, one of the properties of wireless communication is a tendency to keep the nodes depersonalized, therefore achieving a lower level of potential identification by the malicious user. Keeping the presence of a centralized control node as a base, the author assumes that such a system can self-learn [78], and that by tracking the activities of the user it can
develop a high but relatively anonymized level of social trust. Initially, public information from social networks could be set as a foundation for any potential trustworthiness prediction [71, 79].

![Figure 4.1: Structure of proximate social networks [2]](image)

Strong social ties can also be evaluated by the categorization of the network elements, as shown in Figure 4.1. For example, *brokering* and *closure* [80] represent the idea of designing strong ties inside the group of people to gain trusted relationships in addition to their potential for clusterization. Therefore, we can conclude that *degree*, *betweenness*, and *closeness* are the essential characteristics of each selected network node. The *degree* corresponds to the number of connections per node retrieved from a social network data, for example. Hence, the higher the degree, the higher the probability of building a trusted cluster is. Such trust mapping allows for new potential links between less trusted nodes to be found, i.e., evaluation of second and higher degree level of connections. A ‘friend-of-friend’ concept lies at the basis of all connection opportunities to be used in future proximity-based networks.

The *closeness* factor represents how easy it is for the user to build a new connection to others. The straightforward way in both logical and physical worlds is to evaluate a potential acquaintance, i.e., the longer the distance, the more significant effort is needed
to start communication.

An exciting concept of betweenness deserves a mention, representing the potential of the node to become a ‘social relay’ between users. On the other hand, its presence also stands for the possibility of enabling or blocking the data flow between network sections in a worst-case scenario.

### 4.4 Selected Numerical Results

For validation of the proposed secure clustering framework, we have conducted a simulation campaign in custom-made WINTERsim environment\(^1\). In this assessment, we have neglected the effects of interference and propagation but instead introduced the full-scale user mobility. As metrics of interest, we select system throughput, energy efficiency, and degrees of connectivity, i.e., the proportion of users served by either cellular and proximity-based links.

For performance evaluation, we consider one 3GPP LTE BS placed in the center of 150 × 150 meters square area. The reliable cellular coverage is limited to 70% (we also have considered other values to understand the impact on connectivity better). The number of users is randomly deployed within the area of interest. The mobility model for each user follows the Levy flight pattern [81]. The remaining system parameters are given in Table 4.2.

---

\( ^1\)See “WINTERsim – An open simulation platform for the study of wireless systems”, 2018: [http://winter-group.net/downloads/](http://winter-group.net/downloads/)
In this work, we consider three different connectivity options:

- **Cellular (LTE) solution**, where the connectivity is only available through cellular links. There are no proximity-based communication possibilities;

- **Simple D2D solution**, where only devices with a reliable connection to the cellular network may establish direct links. The trusted authority is thus located in the cloud or on the edge and can serve any device with infrastructure connectivity;

- **Advanced (social-aware) D2D solution**, where the proposed secure clustering allows for establishing coalitions when a reliable connection to the cloud is not available. All direct links in areas both with low or no coverage are made trustworthy through the solution detailed in [64].

Parameter $\alpha$ is the representation of human-only or device-only sociality scenarios. For the applications, where both human- and device-driven types of sociality are considered, we set $\alpha = 0.5$. In particular, the value of $D_{i,j}$ is determined by the relationships between the humans/devices as reported in Table 4.1. In those cases, where two nodes are connected by two or more types of social factors, we should consider the strongest tie with the highest degree.

Further, the weighting term $\alpha \in [0, 1]$ is needed to adjust the impact of two contributions described above, according to a specific application and scenario. Hence, the role played by $\alpha$ in our view is to augment the model with a weighting factor that may be adjusted according to the scenario of interest. More specifically, we assign the value of 1 to $\alpha$ when considering the first case and the value of 0 for the third case.

The rest of this chapter will collate the aggregated data that was transferred under reliable cellular for ‘simple D2D’ and ‘baseline LTE’ for all users as **aggregated system throughput**. In ‘advanced D2D’ case, we also count the data transferred outside the cellular network coverage.

The aggregate system throughput as a function of the number of deployed devices is given in Figure 4.2. The proposed methodology of secure communication provided better results than both baseline LTE and simple D2D cases. In particular, for device-driven sociality where $\alpha = 0$ the best results were achieved. It is followed by human-driven sociality ($\alpha = 0.5$ and $\alpha = 1$). Based on the above, the level of interdevice sociality may introduce significant benefits to the operation if the trust relations are predefined in advance.

Next, the overall proportion of users that could be served in a given area by accounting for both the simple D2D and the social-aware D2D is given in the left subplot of Figure 4.3. We observed a higher percentage of served users were reached when we considered a social layer of awareness among devices and humans. Further, we learn that this positive effect is higher for lower degrees of LTE coverage. The right subplot of Figure 4.3 reports the proportion of users with a simple D2D solution and the proposed social-aware D2D solution. The benefits brought by social D2D are achieved when users established a direct link instead of downloading content over the LTE infrastructure and also through users with D2D connections in locations where there was no coverage. Clearly, with network coverage degradation, the benefits of the utilization of the proposed framework increase.

In summary, the LTE coverage and user connectivity were tightly connected by varying the area where the LTE infrastructure was present. Different numbers of users were found to have been served over a cellular link with the LTE BS. In fact, as can be seen
from Figure 4.3, when the LTE coverage is particularly low (i.e., only 50% of the area of interest) using a legacy LTE approach led to only 40% of users being served.

One of the most significant metrics for mobile devices is energy efficiency. For consistency with the aggregated system throughput, we measured the aggregated energy efficiency as well, and the results are displayed in Figure 4.4, based on the transmit power values per technology stated in Table 4.2. Here, the social-aware D2D approach outperformed both the considered baseline LTE and the simple D2D options. For $\alpha = 0$, the proposed solution reached its highest gain in contrast to the benchmark LTE operation. This is due to a lower transmit power of small-scale devices (i.e., connected machines) as compared...
to more power-hungry handheld UEs.

Overall, the social ties between humans and their connected devices had a significant impact on the proposed social-aware methodology, according to the analysis above. Notably, a higher degree of social relationships delivered higher system throughput without energy efficiency degradation. The most significant drawback of the proposed system is the potential increase in latency due to the time it took to establish proximity-based communication – this should be considered when making the application. In particular, handheld devices need additional time to complete the security methods from [64], leading to slightly higher latencies as the number of the communicating entities grows. However, the implementation efficiency of these security mechanisms can be optimized further to reduce the computation time, which can be left for subsequent study.
5 Privacy Preserving Strategies for Wearable Technology

This chapter will elaborate on the future potential and associated challenges in the collective use of private wearable devices and the possibility of secure content dissemination for AR/VR equipment during mass events. First, the properties and methodology of the electronic devices temporary delegation of use are discussed followed by a short description of protocols that have been developed and that enable delegation when cloud connectivity is intermittent. These were established with some numerical evaluation. Next, the broadcast procedure and corresponding authentication scheme for the content dissemination are discussed. The end of this chapter provides a summary of the cryptographic primitives and possibilities in their execution for modern smartwatches and smartphones, stating that modern security requirements could be fulfilled by both groups.

5.1 Authentication Methodology for Collective Smart Device Usage

As it was noted previously, one of the most promising trends for electronic devices is collective use. We proposed a set of protocols to enable temporary delegation of device usage in scenarios when cellular connection, i.e., link to the cloud, may or may not be available.

In the scenario when a person willing to borrow the device for some time is trusted or at least known, the solution is rather straightforward and conventional. However, the legitimate return according to the agreement could not be guaranteed, even in this case. The protocols developed were based on a Trusted Authority (TA) that is indirectly involved in lending the device and could provide some level of confidence during the process. For example, the authority can define a set of basic rules applied for any delegation procedure and set the process of landing, retrieving, and modifying the lease even if the connection to the infrastructure network is not continuously available. Being more specific, we assume that ‘initialization’ of the device after purchase requires connection to the TA. During the delegation phase, however, this connection is unnecessary, which is exceptionally beneficial for distant locations or when roaming. The designed set of protocols explicitly accommodates those requirements. The assumptions about the network architecture in this chapter are the same as in previous chapters, i.e., the smartphone assumes the role of a gateway for the wearable under examination.

In Figure 5.1, we show the primary stages of the system operation. Main notations and constructions are given in Table 5.1. The description and protocols themselves are presented in paper [P3].
Chapter 5. Privacy Preserving Strategies for Wearable Technology

We place the TA for our system inside the operator’s cloud. The gateway of each user has its $SK_A$ and certificate $sign_{cloud}(ID_A, PK_A)$ with a public key from TA. $sign_{cloud}$ is delivered from TA according to the corresponding signature construct with the secret key $SK_{CA}$ of the TA. Note, each owner’s main device has $cert_{cloud} = PK_{CA}$ (for device and message verification) and each wearable $w_i$ has a unique hardware-locked identifier $ID_i$ and factory-set PIN stored separately.

Each $w_i$ has pre-installed necessary libraries so as to support our functionality. At any time, the device could be rolled back to its trusted factory state [82]. We imply safety against person-in-the-middle attacks. Next, the user can initially setup a timer $t_f$ to obtain additional resistance against misuse in cases when the device is stolen or lost $sign_A(t_f)$. In the following, we show the main protocol components required for full-scale implementation of the proposed methodology.

- **Association phase** represents the case when the future owner buys a new ‘factory-clean’ wearable device and is willing to connect it to its smartphone/personal cloud. Here, new $w_i$ device is associated with Alice’s unique identifier (for example, email address or account) through the cloud, as it is summarized in Algorithm 1.
### Table 5.1: Main delegation protocol suite notations [P3]

<table>
<thead>
<tr>
<th>Construct</th>
<th>Container</th>
<th>Explanation</th>
</tr>
</thead>
<tbody>
<tr>
<td>$A, B, cloud$</td>
<td>–</td>
<td>Alice (owner), Bob (temporary user), and Cloud names.</td>
</tr>
<tr>
<td>$w_i$</td>
<td>–</td>
<td>$i^{th}$ wearable device.</td>
</tr>
<tr>
<td>$PK_A, SK_A$</td>
<td>–</td>
<td>Owner’s public and private keys.</td>
</tr>
<tr>
<td>$\text{sign}_\text{cloud}(PK_A)$</td>
<td>–</td>
<td>Definition that $PK_A$ is signed by cloud certificate.</td>
</tr>
<tr>
<td>$t_d, t_f$</td>
<td>–</td>
<td>Secure timers for delegation and reset periods.</td>
</tr>
<tr>
<td>$S_A$</td>
<td>–</td>
<td>Owner’s device secret key utilized for communication with a wearable.</td>
</tr>
<tr>
<td>$\text{hash}(SW_i)$</td>
<td>–</td>
<td>Wearable software hash.</td>
</tr>
<tr>
<td>$\text{cert}_\text{cloud}$</td>
<td>$\text{sign}_\text{cloud}(w_i, PK_A, ID_A, \text{hash}(SW_i))$</td>
<td>Certificate used for data integrity retrieved from the cloud.</td>
</tr>
<tr>
<td>$\text{cert}_A$</td>
<td>$\text{sign}<em>A(\text{cert}</em>\text{cloud})$</td>
<td>The owner’s envelope for storing on the wearable side.</td>
</tr>
<tr>
<td>$m[D]_\text{cloud}$</td>
<td>$\text{sign}_\text{cloud}(m[D]_A)$</td>
<td>TA verified message envelope.</td>
</tr>
<tr>
<td>$m[D]_A$</td>
<td>$\text{sign}_A(w_i, t_d, ID_A, ID_B, {\text{rules for delegation}})$</td>
<td>Initialization message.</td>
</tr>
<tr>
<td>$m[R]_B$</td>
<td>$\text{sign}_B(w_i, R)$</td>
<td>Request for the device retrieval.</td>
</tr>
<tr>
<td>$m[C(S_A)]_A$</td>
<td>$\text{sign}_A(w_i, C[S_A])$</td>
<td>Message for removal the user secret key from wearable device.</td>
</tr>
</tbody>
</table>

**Algorithm 1** Association while connected to TA

1: $A$ generates $S_A$ for $w_i$ and transmits it to $w_i$;
2: $w_i$ transmits ($\text{hash}(SW_i)$) to TA via $A$;
3: $A$ also transmits $PK_A$ and $ID_A$ to TA;
4: TA generates $\text{cert}_\text{cloud} = \text{sign}_\text{cloud}(w_i, PK_A, ID_A, \text{hash}(SW_i))$;
5: TA transmits $\text{cert}_\text{cloud}$ to $A$;
6: $A$ signs $\text{cert}_\text{cloud}$ and generates $\text{cert}_A = \text{sign}_A(\text{cert}_\text{cloud})$;
7: $A$ transmits $\text{cert}_A$ to $w_i$.

- Here, the most significant phase in our suite is *Delegation*, which can take place under both reliable and unreliable coverage to the TA. Note, the owner aims to temporarily delegate the use of its device (i.e., lend) to its friend/colleague/customer with some limitations. Overall, the delegation process may also be run remotely if both the owner’s and user’s gateways are connected to the cloud. *Delegation under network coverage* requires that both users’ gateways have a stable connection to the TA, which will handle all the verification processes (see Algorithm 2). *Delegation out of network coverage*, i.e., runs without a stable link to the TA. The steps are summarized in Algorithm 3.
Algorithm 2 Delegation while connected to TA

1: A transmits $m[D]_A = \text{sign}_A(w_i, t_d, ID_A, ID_B, \{\text{rules for delegation}\})$ to $w_i$ thus setting the delegation time;
2: A transmits $m[D]_A$ to TA;
3: TA verifies $m[D]_A$ based on $PK_A$. Process is terminated if the verification fails;
4: TA signs $m[D]_{\text{cloud}} = \text{sign}_{\text{cloud}}(m[D]_A)$;
5: TA transmits $m[D]_{\text{cloud}}, \text{cert}_A$ to $B$;
6: A removes $S_A$ using $m[C(S_A)]_A$ at $w_i$ side;
7: if $B$ does not trust $A$ then
8: $w_i$ is reset to factory defaults;
9: $B$ verifiers if $\text{hash}(SW_i)$ in $\text{cert}_A$ equals current $\text{hash}(SW_i)$. Process is terminated if the verification fails.
10: else
11: $B$ obtains a right to use $A$'s application according to the delegation rules.
12: end if
13: $B$ generates new $S_B$ for $w_i$;
14: $B$ transmits $S_B$ to $w_i$;
15: $B$ signs $\text{sign}_B(w_i, SW_i)$ for integrity reasons;
16: if Delegation timer $t_d$ expires then
17: $w_i$ is reset to factory defaults. $t_d$ could be updated upon request while $w_i$ has the connection to TA and $A$ transmits $m[D]_A = \text{sign}_A(w_i, t_d, ID_A, ID_B, \{\text{rules for delegation}\})$ to it via $B$ or directly.
18: end if

Algorithm 3 Delegation while not connected to TA

1: A transmits $m[D]_A = \text{sign}_A(w_i, t_d, ID_A, ID_B, \{\text{rules for delegation}\})$ to $w_i$ thus setting the delegation time;
2: A transmits $\text{cert}_A, m[D]_A$ to $B$.
3: $B$ verifies if $\text{cert}_A$ and $m[D]_A$ are valid by $\text{cert}_{\text{cloud}}$. Process is terminated if the verification fails;
4: A removes $S_A$ using $m[C(S_A)]_A$ at $w_i$ side;
5: if $B$ does not trust $A$ then
6: $w_i$ is reset to factory defaults;
7: $B$ verifiers if $\text{hash}(SW_i)$ in $\text{cert}_A$ equals current $\text{hash}(SW_i)$. Process is terminated if the verification fails.
8: else
9: $B$ obtains a right to use $A$’s application according to the delegation rules.
10: end if
11: $B$ generates new $S_B$ for $w_i$;
12: $B$ transmits $S_B$ to $w_i$;
13: $B$ signs $\text{sign}_B(w_i, SW_i)$ for integrity reasons;
14: if Delegation timer $t_d$ expires then
15: $w_i$ is reset to factory defaults. $t_d$ could be updated upon request while $w_i$ has the connection to TA and $A$ transmits $m[D]_A = \text{sign}_A(w_i, t_d, ID_A, ID_B, \{\text{rules for delegation}\})$ to it via $B$ or directly.
16: end if
Next, we outline protocols handling the process of device reclaiming by its owner. Similarly to the delegation, the operation may be under both reliable and unreliable connectivity constraints.

- **Reclaiming under network coverage** is executed similarly to delegation (see Algorithm 4).

**Algorithm 4** Reclaiming while connected to TA

1: $B$ transmits $m[R]_B = \text{sign}_B(w_i, R)$ to TA.
2: TA verifies $m[R]_B$ based on $PK_B$. Process is terminated if the verification fails;
3: TA signs $m_{\text{cloud}} = \text{sign}_{\text{cloud}}(m[R]_B)$;
4: TA transmits $m_{\text{cloud}}$ to $A$;
5: $B$ removes $S_B$ using $m[C(S_B)]_B$ at $w_i$ side;
6: if $A$ does not trust $B$ then
7: $w_i$ is reset to factory defaults;
8: $B$ verifiers if $\text{hash}(SW_i)$ in $\text{cert}_A$ equals current $\text{hash}(SW_i)$. Process is terminated if the verification fails and $A$ should reinitialize the device with the PIN preset by the factory.
9: else
10: The software data is unchanging and $A$ can continue using local applications after returning from $B$.
11: end if
12: $A$ transmits stored $S_A$ to $w_i$;
13: $A$ signs $\text{sign}_A(w_i, SW_i)$ for data integrity reasons. Thus, $w_i$ has to store $\text{cert}_A, \text{cert}_{\text{cloud}}$ now.

- **Reclaiming out of network coverage** is also similar to delegation (see Algorithm 5).

**Algorithm 5** Reclaiming while not connected to TA

1: $B$ transmits $m[R]_B = \text{sign}_B(w_i, R)$ to $A$ over a direct link;
2: $B$ removes $S_B$ using $m[C(S_B)]_B$ at $w_i$ side;
3: $A$ verifies $m[R]_B$ by $\text{cert}_{\text{cloud}}$.
4: if $A$ does not trust $B$ then
5: $w_i$ is reset to factory defaults;
6: $B$ verifiers if $\text{hash}(SW_i)$ in $\text{cert}_A$ equals current $\text{hash}(SW_i)$. The process is terminated if the verification fails and $A$ should reinitialize the device with the PIN preset by the factory.
7: else
8: The software data is unchanging and $A$ can continue using local applications after returning from $B$.
9: end if
10: $A$ transmits stored $S_A$ to $w_i$;
11: $A$ signs $\text{sign}_A(w_i, SW_i)$ for data integrity reasons. Thus, $w_i$ has to store $\text{cert}_A, \text{cert}_{\text{cloud}}$ now.

- Another significant phase of the protocol suite is de-association of the wearable device from the personal cloud if it is sold or disposed of. Therefore, it is necessary to clean personal data, keys, and certificates from the device that were obtained previously from the TA. Generally, there are two options for de-association: (i) manual (in case of 'triggered
by the owner’ process, see Algorithm 6), and (ii) automatic (if the rental regulations were not met, the device was lost, damaged, or stolen. Thus, the data should be erased, see Algorithm 7).

**Algorithm 6 Manual de-association**

1. \(A\) transmits \(m[F]_A\) to \(w_i\);
2. \(w_i\) is reset to the factory state with total removal of all the data, the certificate storage is erased.
3. \(w_i\) could thus be reinitialized only by the factory PIN and when the new owner has a connection to the cloud.

**Algorithm 7 Automatic de-association**

1. The \(t_f\) is initialized when the device leaves a personal cloud of the user after preset threshold time interval.
2. **if** \(t_f\) expires **then**
3. \(w_i\) is reset to the factory state with total removal of all the data, the certificate storage is erased.
4. \(w_i\) could thus be reinitialized only by the factory PIN and when the new owner has a connection to the cloud.
5. **end if**

Since we found no option to implement the protocol in the wearable devices available on the market, we numerically assessed the energy consumption on the gateway (in case of potential implementation) as it would become the most significant limiting factor.

Here, the transmitting conditions are kept static for all network interfaces, and the results are given in absolute numbers due to heterogeneity concerning potential devices. In particular, the power consumption data for the cellular interface are reproduced from [83]. We have obtained the results related to power consumption from works [84, 85, 86]. Here, power consumption for WiFi is set as 720mW, for BLE as 147mW, and for ZigBee as 71.402mW. These values were used as a baseline for our numerical evaluation regarding transmission overheads while keeping the packet payloads equal.

As the main explanatory example, we summarize relative power consumption for cases of reliable and intermittent connectivity in Figure 5.2. Indeed, the delegation and association phases are more power hungry compared to others, and this is due to a higher number of signaling messages transmitted over the wireless medium. We also show that the utilization of WiFi for direct connectivity is more energy consuming in comparison with other studied short-range wireless technologies. Still, power consumption is relatively low for any technology.

### 5.2 Broadcast Content Access for Mass Events

So far, we have shown the breadth of what must be considered with wearable technology and devices, and yet there are even more complex contexts to be found in the field of entertainment. This area is also profoundly affected by such requirements as privacy, security, and connectivity. Unfortunately, the market still faces a lack of appropriate information security enablers.
In this and the next section, there is a discussion on flexible authentication that has the potential to be used while accessing broadcast content during the physical attendance to a sports event, meaning directly during the event (with micro-transactions). A broader list of applications is as follows: (i) video content retrieval from new sources (players, gates, drones, etc.); (ii) obtaining team-related information (statistics, history, future games, etc.); (iii) service and critical information (evacuation plans, alerts, etc.); (iv) advertisement (taxi, fast food, souvenirs, etc.). Despite listed monetization opportunities, the stadium owner can also acquire more significant statistical data, such as the proportion and distribution of the occupied seats; spectators interests and requests, and many others. This statistical data can be used to improve the experience of fans through better planning of frequent consumables purchases, improving the advertisement content, etc. All this can positively influence any future event planning.

Generally, Figure 5.3 shows a given scenario depicting the deep penetration of wearable technology in professional sport, e.g., an ice hockey match with different categories of users. The numerous wearable devices are expected to be carried by content consumers – mass spectators. This group of people have purchased their ticket and obtained access to a personalized AR kit. The next big group is support personnel that include referees, security, medics, technicians, advertisers, and other specialists who have access to highly specific and sensitive data. Then, the most significant component is the competing teams and their coaches. The special requirement for this class is long-term protection against misuse of context-oriented information regarding their abilities, health, and behavior during the game – since it could be easily analyzed.

The main pragmatic outcome of the proposed framework is a secure, wearable-aware data streaming system that opens out new avenues for entertainment and advertisement efficiency. Since players would wear a multitude of monitoring devices, including those for heart rate, lung capacity, metabolism, and location [87], the game-oriented data may be presented to the public in an entirely new way. At the same time, the information
should be carefully classified as sensitive and nonsensitive, and thus divided between authorized users and spectators. Nonetheless, sensitive data could be further masked and even delivered to promote third-party services for fans’ AR glasses or home screen [88].

5.3 Anonymized Content Dissemination Methodology

This section is focused on the proposed authentication technique that could be used for the AR/VR broadcast data, with differentiated access based on a specific subscription. Overall, current authentication systems mainly rely on symmetric, asymmetric, or hybrid cryptosystems [89]. Many of these use a hashing process as part of their basis, allowing for the reconstruction of “shares” into one secret key. Asymmetric PKI from this perspective allows for serving a high number of users through flexible and an easy to use authentication procedure.

In the case of access control during a mass event, we combine a set of available secret shares obtained from a number of unique digital sources, i.e., ticket number, seat number, etc. Moreover, a user of such a system could be uniquely verified based on their combination, and thus diversified access could be provided to different users even in the same seat. One of the solutions to enable such functionality is the so-called hybrid Yoking-Proof protocol [90].

Further, we discuss a solution that involves interaction with stadium equipment (see [P4] for details). The central assumptions are: (i) Radio-frequency identification (RFID) tags are integrated in seats and tickets [91]; (ii) every spectator has an NFC-enabled
smartphone. In order to verify the user, the system simultaneously checks the ticket and seat via the stadium wireless communication technology. After that, the actual authentication code is delivered either via the application’s push notification mechanism or SMS.

Therefore, the event organizer or owner of the stadium could obtain the seat identifier, the ticket identifier, and, optionally, the subscription identifiers. Based on this data, the level of service for the selected user may vary if using the hashing function and the authentication code together. Note, the user is assumed to have an anonymous ticket for any third-party.

In other words, the unique composition of the seat number, the provided level of service, and public key-based signature (stored in the cloud for previous fields verification) could be considered as the digital subscription ticket. Here, the trusted authority is responsible for any cloud management. The main scenarios of interest are listed further:

- The service is provided by the event organizer anonymously. We need to achieve the following in addition to any authentication:

  - ID-based scheme should be used instead of conventional PKI-based ones [92, 93];
  - Private key generator (PKG) should be used on the trusted authority side;
  - The secret key should not have any direct relation to the user. It must be connected with the ticket, subscription, seat, and additionally connected to specific event properties such as event name, its date, and time. Therefore, ID-based key generation would produce $SK_i$ with PKG.

- The event organizer would require just ticket and event-related data during the signature verification process, in turn providing an adequate level of anonymization of the user.

- This private data could still be obtained, if necessary, upon specific requests in case of, for example, mass riots.

5.4 Numerical Evaluation of Cryptographic Primitives on Small-Scale Devices

During his research visit to the Brno University of Technology and after the development of secure proximity-based clustering framework, the author decided to evaluate the overall ‘executability’ of cryptographic primitives on modern IoWT devices, and the primary results are provided in [P2]. The focus was given to calculations using a big integer as multiplication, division, power functions, and classical elliptic curves’ algorithms utilized for constructing the Rivest-Shamir-Adleman cryptosystem (RSA) signature, hash functions (Secure Hash Algorithm (SHA1), SHA-256), and block cipher (Advanced Encryption Standard, AES). This list appeared mainly due to some interoperability challenges on constrained devices and could be extended as the selected primitives became available for evaluation. Paper [P2] provides a more in-depth explanation of those also taking into consideration more complicated cases, such as bilinear pairing and IoT boards, while this thesis focuses specifically on wearables and widely used functionality.

Note that modern wearable devices can now rival the computational performance of handheld devices that are a few years old (see [P2] for details). This computational
functionality and performance have consistently grown across several hardware parameters and contemporary wearable devices have become able to perform similar computing tasks to handheld devices or even laptop computers. Thus, we selected a number of those for our evaluation (see Figure 5.4). Specifically, we used Apple iPhone 6 with iOS 9.1, Samsung Galaxy S4, and Jiayu S3 Advanced, with Android 4.4.2. In terms of wearables, we selected Apple Watch with Watch OS 2.0 and Sony Smart Watch 3 with Android Wear 5.1.1.

![Figure 5.4: Cryptographic primitives performance evaluation of small-scale devices][1]

For the next phase, the security primitives described above were implemented for all listed platforms, without delegating any computations to more powerful devices. The applications were developed in Java for Android-based devices and in Objective-C for those by Apple. All background processes were terminated to make the executions fairer. The evaluation was executed at each device at least 1000 times, in order to achieve statistically reliable data.

![Figure 5.5: Evaluation of RSA execution][2]

Figure 5.5 shows the overheads brought about by utilizing RSA encryption and decryption procedures. OpenSSL was used to generate private and public keys for 1024 and 2048 bits [94] and default public exponent (3 bytes). We found that a typical smartphone

---

[1]: #Figure 5.4: Cryptographic primitives performance evaluation of small-scale devices
[2]: #Figure 5.5: Evaluation of RSA execution
spends around 1 ms for RSA encryption while a smartwatch uses almost three times as much time to perform the same task. At the same time, decryption time provided a much worse result; Android required 25 ms on average, while the iPhone needed up to 100 ms. Smartwatches demonstrated similar behavior, the values are 35 ms for Android and 200 ms for Apple.

The hashing function is another fundamental primitive evaluated in this work. The measurements of SHA1 and SHA-256 are shown in Figure 5.6. Here, it was concluded that on most of the devices the execution is optimized and mainly varies because of equipment differences. The results for AES encryption are also present in Figure 5.6 following the pattern of PKI operation.

In summary, even smartwatches available on the market today can execute the current necessary cryptographic primitives. At the same time, the main wearable gateway – a smartphone – already has the computational power of a three year-old stationary computer, meaning security requirements need to be met whilst developing secure communications between wearables, their gateways, and cloud.
6 Wearable Technology as a Part of Multi-Factor Authentication

This chapter examines multi-factor authentication around modern wearable electronics. Firstly, the main factor providers (sensors) and corresponding challenges are given. The author then proposes the threshold authentication methodology with cloud assistance for situations when some of the factors could not be retrieved. Further, some numerical evaluation of factor grouping is given from the security vs. usability perspective. Future perspectives of MFA for wearables are given at the end of this chapter.

Historically, the authentication was based on not just a single factor. Less than a decade ago, it was realized that single-factor authentication could not guarantee adequate protection due to the ever-growing number of threats [95]. Thus, two-factor authentication [96] was first proposed by grouping any two factors together through a combination of the conventional username/password with an extra request to present a physical token, i.e., a smart card or phone [97, 98].

Generally, the factors currently present on the market can be classified into three main groups [99]: (i) knowledge; (ii) ownership; and (iii) biometrics/behavior.

Next, MFA was introduced to the public to ensure an even higher level of safety, simultaneously employing more than two factors [100]. The key technology behind MFA is biometrics, which is defined by its continuous recognition of human behavior [101] and biological characteristics [102]. The use of actual person-related data provided an entirely new level of security by using someone’s identity as evidence [103].

Today, MFA systems can be classified into the following market groups: Forensic segment – corpse investigation, missing person search, criminal investigation, etc. [104]; Government segment – border control, driver license, government ID, etc. [105]; Commercial segment – account access, e-commerce, physical access control, etc. [106].

6.1 Overview of Enabling Factors and Challenges

Although wearables provide biometric and behavior data, the communication between devices and smart sensors is also a significant question since the authentication between and by machines may also take place during the MFA process [107]. Utilization of ‘active’ biometric input also has its drawbacks concerning usability, since humans need to be educated on how to operate the system [108]. Transparent retrieval of biometric data from wearables may play a tremendous role in improving system usability in the most transparent way for any user. As an example, we may consider a fingerprint scanner that recently achieved a high level of adoption through its usability, mainly due to its
deep functional integration by smartphone vendors [109]. At the same time, using this technique as a standalone method is not recommended [110].

One of the biggest and well-studied challenges in biometric authentication systems is the binary behavior of the decision given from the sensor. There is a broad range of solutions enabling slight mismatch and control of the sensed data with stored samples. The most widely adopted strategies involve measuring the False Reject Rate (FRR) – representing the security, and False Accept Rate (FAR) – representing usability [111]. In a broader sense, MFA systems highly depend on the appropriate selection of FAR and FRR parameters. Another important metric is Equal Error Rate (EER), which in literature is referred to as Central Error Rate (CER). It defines the ‘equilibrium’ between FAR and FRR, i.e., the point where they are equal.

From the perspective of wearable devices, one of the widely utilized methods for MFA could be accelerometer fingerprinting [112, 113]. The user can be continuously verified based on their gait or gesture patterns, being almost impossible to replicate by another person, in a similar way to one’s heart rate. One more option is the analysis of the user’s lifestyle and behavior [114], i.e., places often visited based on the GPS location, wireless APs in proximity, or even purchases made [115].

Generally, any factor provider could be analyzed based on a set of specific parameters [116]. The first one is universality or the property of the factor ‘presence’ in each person. Keeping the fingerprint in mind, an MFA system architect should consider that a person may lose a limb during an accident and the system should still be useful for him/her. The second one is uniqueness – to offer differentiation between two persons. The third one is collectability, representing the ease of data acquisition. The fourth one is performance, being the accuracy, maximum detection speed, and environmental robustness. After this comes acceptability, i.e., the degree of potential community adoption. The final parameter stands for potential spoofing, representing the ease of capturing and spoofing of the data sample. The list of main factor providers with the corresponding classification is given in Table 6.1.

From the author’s perspective, acceptability and integration are always the most complicated steps during the technology adoption phase. There exist many other challenges that could be faced during MFA integration, with the most critical ones summarized in Figure 6.1 and detailed in [P6].
Table 6.1: Comparison of suitable factors for MFA [P6]: H–high; M–medium; L–low; n/a–unavailable.

<table>
<thead>
<tr>
<th>Factor</th>
<th>Collectability</th>
<th>Uniqueness</th>
<th>Performance</th>
<th>Spoofing</th>
<th>Universality</th>
<th>Acceptability</th>
</tr>
</thead>
<tbody>
<tr>
<td>Behavior</td>
<td>L</td>
<td>H</td>
<td>L</td>
<td>L</td>
<td>H</td>
<td>L</td>
</tr>
<tr>
<td>Facial</td>
<td>M</td>
<td>L</td>
<td>L</td>
<td>M</td>
<td>H</td>
<td>H</td>
</tr>
<tr>
<td>Fingerprint</td>
<td>M</td>
<td>H</td>
<td>H</td>
<td>H</td>
<td>M</td>
<td>M</td>
</tr>
<tr>
<td>Hand geometry</td>
<td>M</td>
<td>M</td>
<td>M</td>
<td>M</td>
<td>M</td>
<td>M</td>
</tr>
<tr>
<td>Location</td>
<td>M</td>
<td>L</td>
<td>H</td>
<td>H</td>
<td>n/a</td>
<td>M</td>
</tr>
<tr>
<td>Ocular-based</td>
<td>M</td>
<td>H</td>
<td>M</td>
<td>H</td>
<td>H</td>
<td>L</td>
</tr>
<tr>
<td>Password</td>
<td>H</td>
<td>L</td>
<td>H</td>
<td>H</td>
<td>n/a</td>
<td>H</td>
</tr>
<tr>
<td>Thermal image</td>
<td>L</td>
<td>H</td>
<td>M</td>
<td>H</td>
<td>H</td>
<td>H</td>
</tr>
<tr>
<td>Token</td>
<td>H</td>
<td>M</td>
<td>H</td>
<td>H</td>
<td>n/a</td>
<td>H</td>
</tr>
<tr>
<td>Vein</td>
<td>M</td>
<td>M</td>
<td>M</td>
<td>M</td>
<td>M</td>
<td>M</td>
</tr>
<tr>
<td>Voice</td>
<td>M</td>
<td>L</td>
<td>L</td>
<td>H</td>
<td>M</td>
<td>H</td>
</tr>
</tbody>
</table>

Figure 6.1: Main operational challenges of MFA [P6]
6.2 Proposed Multi-Factor Authentication Methodology Based on Reversed Lagrange Polynomial

Users (i.e., humans) have a tendency, however, to forget their keys, passwords, etc. that may potentially make MFA access impossible. To overcome this issue, the proposed methodology was based on a well-known Lagrange polynomial, which was required to collect any $l$ previously distributed shares $\{S_{ID_1}, S_{ID_2}, \ldots, S_{ID_l}\}$ for access. The corresponding polynomial of the degree $n > l$ is shown in Figure 6.2. Here, we aim to define the curve $S$ and utilize random coefficients $a_i$ for generation of the secret share $S_i$ [117, 118].

That approach, however, is not suitable for modern authentication systems involving biometrics, since most of the biometric factors do not change over time, i.e., we can neither modify the existing share nor assign a new $S_i$ to a factor. Generally, the user can constantly change its knowledge-based factors, but this would make the system use complicated. Here, we need to ‘reverse’ the approach and set $S_{ID_i}$ as the already known factor values $S_i$, i.e., $S_i$ values are static per factor per user as $\{S_1, S_2, \ldots, S_l\}$. Therefore, curve $S$ should be not the baseline for generating shares, but vice versa.

![Figure 6.2: Classic Lagrange polynomial-based secret sharing methodology [P6]](Image)

To improve the overall usability and keep the level of security relatively high, we base our MFA system on the reversed Lagrange polynomial with $l$ shares from each individual factor $F_i$. In the following lines, we expand on a tagged legitimate user. Each $F_i$ has its own secret $S_i$ obtained from the sensor. Importantly, the probability of this value changing over time (for biometric data) is extremely low. Thus, we can write these combinations as: $F_1 : S_1; F_2 : S_2; \ldots; F_l : S_L; F_{l+1} : T$, where $F_{l+1}$ stands for a timestamp at time moment $T$.

To keep user-sensitive biometric samples even more well-hidden from the TA, we define that $S_i$ are directly retrieved from the sensors and $S$ is actually the polynomial main secret (see Figure 6.2). The proposed system, in contrast, generates the main secret $S$ based
6.2. Proposed Multi-Factor Authentication Methodology Based on Reversed Lagrange Polynomial

on results collected from the sensors $S_i$ instead of assigning ones to factors. The main secret required to access the service or device $S$ may be reconstructed from previously distributed $l$ shares. The improvements must satisfy the uniqueness requirements of the collected data, together with the timestamp, as in Figure 6.3, as these measures give robustness to counteract incidences when the collected shares do not change over time.

Figure 6.3: Proposed reversed methodology [P6]

With the system proposed here, one of the main benefits is that when the cloud connection is available, the user can be supplied with additional shares by the cloud. This may
happen if secret components have been forgotten, lost or mismatched. The cloud-provided shares could be delivered when, for example, the second channel is used for identity verification. Here, the TA may be requested to supplement the necessary number of temporary shares (see Figure 6.4).

Consider a case when the user could not provide a few factors to the system for verification, such as $F_1$ and $F_2$. The cloud could generate two temporary shares in order for the main secret to be reconstructed as $S_{\Phi_1} = f(\Phi_1)$ and $S_{\Phi_2} = f(\Phi_2)$. The shares are sent to the user via a secure channel and as $F_1 : S_1; F_2 : S_2; \ldots; F_L : S_L; F_{L+1} : T; \Phi_1 : S_{\Phi_1}; \Phi_2 : S_{\Phi_2}$ [119]. By this means, the access to the device or service would be granted.

6.3 Proposed Intelligent Factor Grouping for Multi-Factor Authentication

In conventional authentication systems based on knowledge and ownership factors, the decisions always either pass or fail based on how accurate the input data is. At the very moment biometric data is captured, any sensor is subject to mistakes during the sample capture process. Therefore, the interoperability of binary decision factors and probabilistic ones is an exciting task to focus on.

Conventionally, statistically collected FAR/FRR properties of the sensors are provided by the vendor. To evaluate the MFA framework, we assume two typical decisions that could take place during the authentication phase (see Figure 6.5). The first one represents the illegitimate user authentication ($H_0$) and the second stands for legitimate ($H_1$). Therefore, those two decisions form a sample space of $P(H_0) + P(H_1) = 1$. The MFA system architect then sets the distributions of $P(H_0)$ and $P(H_1)$ and also tunes the system according to chosen parameters.

![Figure 6.5: MFA system mode with selected threshold [P6]](image)

Here, the user input data could be collected from $n$ biometric sensors. Each sensor sample from the set $Z = \{z_1, \ldots, z_n\}$ is within $[0, 1]$. The corresponding capture could be analyzed further based on two different strategies.

The first strategy is called the **Strict Decision Method**. Here, all sensors (including biometric ones) return either accept or reject by preprocessing the data on the sensor side. The results are later combined and provide a group decision based on the resulting vector.
or results. Therefore, the threshold could be used for the final MFA decision. Basically, each sensor returns value $z_i$, $z_i = [0; 1]$. Next, conditional probabilities $P(z_i | H_0)$ and $P(z_i | H_1)$ are defined by $FAR_i$ and $FRR_i$ values.

The second strategy is called the Probabilistic Decision Method. Here, the sensor does not provide a binary decision but uses a probabilistic feature instead. Next, the data is processed in the MFA system. The results are complemented with the sample comparison as a match score of $z_i (0 \leq z_i \leq 1)$. Thus, the conditional probability $P(z_i | H_0)$ is estimated based on the $FAR_i$ values for each $z_i$. We base the calculation of the conditional probability $P(z_i | H_1)$ on $FRR_i$ values at $z_i$. Generally, the first strategy could be considered as a simplified version of the second one if $FAR_i$ and $FRR_i$ are given only at one point.

Consider an example of knowledge-based authentication. Here, FAR represents the probability of guessing the secret and FRR stands for the possibility of entering the data wrong. At the same time, FAR and FRR also define cases of stealing the secret knowledge and correspondingly forgetting it. This also applies in the case of ownership factors, and therefore most of the binary decision factors could be represented in a unified output format and supplemented with proper FAR/FRR values.

Despite that, the use of several individual factors together does not offer any immediate advantage since it is still unclear how to combine them effectively. As a straightforward solution, one can end up with two different strategies: a user should successfully pass all the checks to get access (All) and a user should successfully pass any of the checks to get access (Any). However, those strategies have their own advantages and weaknesses. To prove that, we present a numerical example comparing the corresponding weaknesses in addition to showing the importance of intelligent factor combination. For this evaluation, we assume a set of factors with the corresponding FAR and FRR values. Following the self-explanatory plot, we assume all the FARs be equal to 0.03% whereas all the FRRs are equal to 2%. The Law of Total Probability then derives the resultant values for FAR/FRR.

The numerical example results are summarized in Figure 6.6. Here, the All approach has the lowest FAR for any number of factors combined thus offering the highest security level. At the same time, this approach shows the highest FRR, proving that such a system is extremely uncomfortable to use since the user has to input all the requested data successfully. Based on the above, we conclude that its application for the daily basis, i.e., while accessing the services on-the-fly, is not an effective option.

In contrast, FAR is increasing at the expense of much better FRR for the Any approach. Therefore, the system stays insecure (since there is a need to hijack the weakest factor to obtain access) but very easy to use, which is also not acceptable for everyday use. Consequently, none of the trivial MFA combinations are directly usable in the challenging A-IoT scenarios.

Therefore, we propose a novel Balanced approach. Here, a user should successfully pass some of the checks to obtain access to the system thus representing a compromise between usability and security for the modern MFA system by decreasing both FAR and FRR values. The quantitative gains highly depend on the input parameters and reach $10^4$ vs. $10^8$ when 7 factors are combined. This example also highlights the importance of threshold value selection, since incorrect combining may often result in rapid system performance degradation [P6]. The same holds true for any other values of FAR/FRR, even though they may actually vary for different factors.
Chapter 6. Wearable Technology as a Part of Multi-Factor Authentication

Figure 6.6: Comparing alternative factor combining approaches [P7]

6.4 Future Perspectives

Most of the conventional ICT systems typically utilize a binary authentication strategy based on a set of explicitly defined steps that must be followed in order to gain access to the system. Considering the highly dynamic world of tomorrow, the main authentication phases should be complemented with additional verification taking place before or after the actual authentication of the user. The MFA system may vastly benefit from both the surrounding environment of the user, i.e., Smart City, Intelligent Transportation Systems (ITS), Surveillance Cameras, and Smart devices around, as well as from the user’s personal cloud, in this way seamlessly monitoring user biometrics and behavior. Therefore, the entire process could be divided into three main phases:

- **Preauthentication** could be described as the most dynamic and unpredictable phase when the user is approaching the target, for example, a vehicle. This case is one of the most critical ones, and the use of surrounding ecosystem may significantly improve the authentication procedure by ‘observing’ the user biometrics/behavior that could be delivered by wearables worn by the subject, devices they are carrying, and other vehicles/infrastructure.

- **Active authentication** is the authentication we know today, i.e., the user is forced to execute some actions with the system directly. This phase is the most conventional one and can use any type of authentication factor but commonly relies on prior knowledge and ownership.

- **Continuous authentication** of the user, meaning they remain legitimate to operate the system even after the previous phases are completed successfully. From the author’s perspective, it becomes possible to monitor and analyze the user by the personal cloud, the smart vehicle, surrounding infrastructure, and other cars. Consider a case where the driver has provided all of the tokens, passed all of the biometric tests but faces a seizure during a highway trip. In this case, the vehicle may automatically overtake the control, connect with neighboring cars, and safely stop by the wayside. As an example, recent works confirm that it is necessary to monitor the driver for just under 2.5 minutes in order to validate the behavior with 95% accuracy [120].
7 Conclusions and Future Perspective

The conclusion will be a summary of the work presented in this thesis, beginning with outlining the overall research goals achieved:

• The evolution of the wearable electronics concept was analyzed, together with existing consumer and enterprise applications, as well as communication paradigms and architectures.

• The impact of social user connections on establishing secure direct links between wearable gateways for intermittent connectivity to the cloud was analyzed. The solution to enable functionality was proposed and evaluated, representing the corresponding communication benefits. In the future, the developed solution may be used together with an underlying social networks layer to enable efficient content dissemination between users/devices in proximity.

• A protocol suite that enabled privacy to be preserved with a temporary delegation of wearable usage, when there was both reliable and unreliable connectivity to the cloud, was proposed and evaluated. Perspectively, the developed set of protocols may be integrated as part of modern smart wearable ecosystem thus creating a flexible environment for enabling shared or temporary use of the devices.

• The main strategy of combination factors for MFA system from the biometric data perspective was evaluated, taking into consideration both security and usability. An intelligent methodology for this combination was proposed showing its benefits compared to conventional solutions. In prospect, the use of the proposed methodology together with the wearable devices data may facilitate the daily access procedure.

In the course of this work, our research group also developed several frameworks for assessing information security primitives on modern wearable devices, as well as modules for system-level modeling in our custom simulation tool.

Generally, the development of wearable technology and its mass adoption would have a substantial impact on modern communication paradigms. The discussed in this work enablers may be used as a solid foundation for the development of enabling technology components, including actual direct communications both under and outside of cellular network coverage and corresponding security-related tasks. The utilization of the developed components would allow enabling distributed secure data exchange among groups
of humans (and/or associated wearable devices) communicating over short-range and infrastructure links, possibly without a connection to the centralized trusted authority.
8 Summary of Publications

8.1 Publications Description

The main publications used in this thesis are referred to as [P1]-[P7]. The publications include four works ([P1], [P3-P6]) published or accepted with minor revision ([P7]) in scientific journals. Publication [P2] is the conference paper. None of the publications are used in any other dissertation, to the best of the author’s knowledge. This section clarifies the main contents and results of the contributions.


**Description**

In [P1], we argue on the widespread adoption of the direct communications paradigm being very unlikely without embracing the concepts of trust and social-aware cooperation concerning both aspects of user-operators and user-devices. We first elaborate on the state-of-the-art market and the corresponding effects. Next, we focus on the user adoption, trust-related challenges, and potential incentives for its improvement. We provide a vision that sociality has the potential to become a core incentive across proximate users. We show that sociality could be considered from different perspectives, mainly: user-driven and device-driven. Further, we propose a social-aware framework aimed at enabling trusted D2D-centric data delivery for proximate users in mobile environments to free the licensed operator spectrum. We evaluate the proposed framework by comparing it to conventional direct communication and show the corresponding benefits concerning throughput, number of served users, and energy efficiency. Finally, we show a standardization perspective.

This paper is a collaborative work of the author and his supervisor with Dr. Dmitri Moltchanov and Asst. Prof. Sergey Andreev from the same research group in Tampere University of Technology (Finland), Ekaterina Olshannikova and Assoc. Prof. Thomas Olsson were also with Tampere University of Technology (Finland), Dr. Antonino Orsino, Dr. Leonardo Militano, Asst. Prof. Giuseppe Araniti and Prof. Antonio Iera were with University Mediterranea of Reggio Calabria (Italy), Dr. Gabor Fodor was with Ericsson Research (Sweden), and Johan Torsner was with Ericsson Research (Finland).
Chapter 8. Summary of Publications


Description

In [P2], we elaborate on the information security primitives and cryptosystems currently utilized for delivering the level of data privacy and security required for modern constrained and wearable devices. Next, we develop a testbed framework allowing us to evaluate some of the devices on the market from different vendors. We analyze the real-life execution time of symmetric/asymmetric cryptography, hashing functions, and elliptic curve-based cryptography. The main conclusions of this paper are so that market-available wearable and constrained devices already have enough computational power to execute modern crypto primitives.

This paper is a collaborative work of the author and his supervisor with Roman Florea, and Asst. Prof. Sergey Andreev from the same research group in Tampere University of Technology (Finland), Dr. Pavel Masek, Dr. Lukas Malina, Assoc. Prof. Jiri Hosek, and Dr. Jan Hajny were with Brno University of Technology (Czech Republic), and Jussi Niutanen was with Intel Finland.


Description

In [P3], we continue the discussion related to the penetration of wearable devices in the everyday lives of the humanity. We discuss a paradigm shift from the Internet of Things to the Internet of Wearable Things, which in turn brings along a truly personalized user experience by capitalizing on the rich contextual information followed by new data privacy challenges. Next, we develop a set of protocols enabling the authentication of wearable devices when the presence of a reliable infrastructure link is missing. By this means, we develop a solution for temporary device delegation in cases of intermittent cellular connectivity that allow for collective use fulfilling data privacy. The paper also provides potential attacks on the developed protocol and corresponding countermeasures. We also show the power consumption based on the utilized short-range technology for communication between the wearable device and the corresponding gateway.

This paper is a collaborative work of the author and his supervisor with Asst. Prof. Sergey Andreev from the same research group in Tampere University of Technology (Finland), Joona Kannisto and Prof. Jarmo Harju were also with Tampere University of Technology, and Prof. Sergey Bezzateev was with Saint-Petersburg University of Aerospace Instrumentation (Russia).

Description
In [P4], we first provide a historical and then a future vision of wearables usage on a mass event by both content providers (players, coaches, etc.) and consumers (fans, operators, medics, etc.). We provide the main characteristics of the scenario and the corresponding limitations. From the communications perspective, we evaluate the data delivery broadcast for the target scenario based on a real hockey stadium case, considering different connectivity solutions (IEEE 802.11n at 2.4 GHz, IEEE 802.11ac at 5 GHz, and IEEE 802.11ad at 60 GHz) with the help of custom developed by our team Ray-Launcher software. Finally, we propose a multi-factor authentication methodology suitable for utilization at the stadium by a ticket holder and the corresponding set of attacks.

This paper is a collaborative work of the author and his supervisor with Asst. Prof. Sergey Andreev from the same research group in Tampere University of Technology (Finland), Assoc. Prof. Thomas Olsson and Prof. Jarmo Harju were also with Tampere University of Technology, and Anna Shchesniak and Prof. Sergey Bezzateev were with ITMO University (Russia).


Description
In [P5], we propose a programming model and an associated secure-connectivity framework for leveraging safe, coordinated device proximity as an additional degree of freedom between the remote cloud and the safety-critical network edge, especially under uncertain environment constraints. We first elaborate on conventional scenarios such as pure centralized communications or distributed ones showing the importance of cases when the connectivity becomes more heterogeneous. We propose the use of the so-called ‘liquid software’ paradigm allowing more flexible executions, and provide an overview of our dynamic clustering mechanism allowing for efficient communication at the network peripheral.

This paper is a collaborative work of the author and his supervisor with Asst. Prof. Sergey Andreev from the same research group in Tampere University of Technology (Finland), Joona Kannisto was also with Tampere University of Technology, and Dr. Niko Mäkitalo and Prof. Tommi Mikkonen were with University of Helsinki (Finland).


Description
In [P6], we provide a survey of the multi-factor authentication concept for modern devices. We first overview factor types and the corresponding evolution from single- to multi-factor authentication methodology. Next, we focus on the currently available factor provider techniques/sensors and compare the most suitable for MFA factors. Next, we list the MFA integration, security, privacy, and usability challenges and perspective on the flexible MFA operation, based on ‘reversed’ Lagrange secret sharing methodology. Finally, we develop a methodology for the MFA systems’ evaluation allowing to combine knowledge and biometric factors.
This paper is a collaborative work of the author and his supervisor with Asst. Prof. Sergey Andreev from the same research group in Tampere University of Technology (Finland), Prof. Sergey Bezzateev was with ITMO University (Russia), and Dr. Niko Mäkitalo and Prof. Tommi Mikkonen were with University of Helsinki (Finland).


Description

Work [P7] is a logical continuation of [P6]. First, we analyze the main challenges of ownership determination in advanced IoT. Next, we compare existing factors from a usability perspective keeping in mind that usability always goes as a trade-off to security. Next, we propose to intelligently use ‘weighted’ factors for MFA instead of binary ones and compare those from both usability and security perspectives. Finally, we lay out a future vision on how to dynamically improve the authentication experience by switching from active authentication, i.e., when a user is asked for some input, to preliminary and continuous authentication where environmental sensory devices directly communicate with the target user devices.

8.2 Author’s Contribution

The core listed publications were completed at the Laboratory of Electronics and Communications Engineering, Tampere University of Technology (TUT), Finland.

In addition to this, the author was on a visit to Brno University of Technology (Czech Republic), where he completed part of the simulation and prototype development-related assignments. The author of this thesis is the first author and primary contributor to [P1]–[P4] and [P6]–[P7]. The author’s thesis-related research was guided by his supervisor, Prof. Evgeny Kucheryavy, and instructor, Asst. Prof. Sergey Andreev. Historically, the author is in strong collaboration with Prof. Sergey Bezzateev who guided him while working on information security-related tasks. The main results summarized in this thesis were obtained collaborating with colleagues at TUT and international co-authors from BUT and UNIRIC. The following text provides the author’s contribution in all included publications, [P1]–[P7].

In [P1], the author was responsible for providing state-of-the-art on all aspects except for sociality and incentives. He was also involved in designing the scenarios and numerical performance evaluation. In [P2], the author was responsible for executing the literature review, determination of the requirements, selection of the most significant primitives, main development of the evaluation software-based framework, and partially for the actual testing. In [P3], the author was responsible for proposing the most important scenarios, main development of the protocol, numerical evaluation, and partially for determination of attacks together with corresponding recommendations. In [P4], the author was
responsible for the formulation of the research hypothesis, principally for the development and execution of the experiment. He was the main developer of the authentication methodology. In [P5], the author’s contribution was dedicated to developing the concept of secure coalition formation methodology and classification of the communication paradigms. In [P6], the author is a primary contributor responsible for providing an extensive literature review, analyzing suitable factors, distinguishing potential ones, and developing the MFA methodology based on ‘reversed’ Lagrange polynomial. Finally, in [P7], the author is a primary contributor developing the evaluation methodology, the vision of the utilization of dynamic MFA in the world of tomorrow, and recommendations on how to group the obtained factors.
Bibliography


[67] 3GPP, “TS 33.303 Universal Mobile Telecommunications System (UMTS); LTE; Proximity-based Services (ProSe); Security aspects (version 12.1.0 Release 12),” Tech. Rep., 2014.


Publication I

© 2016 IEEE. Reprinted, with permission, from


In reference to IEEE copyrighted material which is used with permission in this thesis, the IEEE does not endorse any of Tampere University of Technology’s products or services. Internal or personal use of this material is permitted. If interested in reprinting/republishing IEEE copyrighted material for advertising or promotional purposes or for creating new collective works for resale or redistribution, please go to http://www.ieee.org/publications_standards/publications/rights/rights_link.html to learn how to obtain a License from RightsLink.
Toward Trusted, Social-Aware D2D Connectivity: Bridging Across the Technology and Sociality Realms

Aleksandr Ometov†, Antonino Orsino, Leonardo Militano, Dmitri Moltchanov, Giuseppe Araniti, Ekaterina Olshannikova, Gabor Fodor, Sergey Andreev, Thomas Olsson, Antonio Iera, Johan Torsner, Yevgeni Koucheryavy, and Tommi Mikkonen

Abstract—Driven by the unprecedented increase of mobile data traffic, device-to-device (D2D) communications technology is rapidly moving into the mainstream of fifth-generation (5G) networking landscape. While D2D connectivity has originally emerged as a technology enabler for public safety services, it is likely to remain in the heart of the 5G ecosystem by spawning a wide diversity of proximate applications and services. In this work, we argue that the widespread adoption of the direct communications paradigm is unlikely without embracing the concepts of trust and social-aware cooperation between end users and network operators. However, such adoption remains conditional on identifying adequate incentives that engage humans and their connected devices into a plethora of collective activities. To this end, the mission of our research is to advance the vision of social-aware and trusted D2D connectivity, as well as to facilitate its further adoption. We begin by reviewing the various types of underlying incentives with the emphasis on sociality and trust, discuss these factors specifically for humans and for networked devices (machines), as well as propose a novel framework allowing to construct the much needed incentive-aware D2D applications. Our supportive system-level performance evaluations suggest that trusted and social-aware direct connectivity has the potential to decisively augment the network performance. We conclude by outlining the future perspectives of its development across research and standardization sectors.

I. INTRODUCTION AND RATIONALE

In recent years, we have been witnessing an increased proliferation of bandwidth-hungry user applications, which are becoming ubiquitous in the form of multimedia services, interactive games, and social networking solutions. To effectively cope with the resulting avalanche of mobile traffic, fifth generation (5G) networks demand innovative technologies capable of supporting the ambitious system requirements. To this end, unprecedentedly high targets were set for the 5G system design, such as seamless wide-area coverage (with 100 Mbps user rate) and extremely high-capacity hot-spot access (1 to around 10 Gbps user rate). Among the candidate 5G technologies, direct device-to-device (D2D) communications attracts an increased research attention [1] as it promises to deliver improved throughputs, provide more efficient spatial reuse, lead to extended network coverage, and enhance user energy efficiency. Broadly, D2D communications refers to a radio technology that enables devices to communicate directly with each other, that is, without routing the data paths through a network infrastructure.

With the widespread adoption of D2D communications, we expect the user devices to take a more active part in 5G service provisioning and, in some cases (e.g., in partial coverage situations), even assume some of the roles of the network infrastructure. In particular, they can aid in providing wireless connectivity such as offering D2D-based data relaying, proximity gaming, content distribution and caching, as well as other forms of cooperative communications. This paradigm shift from the conventional cellular model is driven by the natural progress in communications technologies: the user devices are decisively augmenting their capabilities, whereas the base stations (BSs) are becoming smaller as a result of the ongoing network densification [2]. Consequently, the original functional disparity between these key components of the maturing 5G ecosystem – the user equipment (UE) and the BS infrastructure – is gradually becoming blurred.

However, there remains a fundamental difference between the UE and the BS, which is rooted in the ownership rights of the corresponding equipment. Hence, cellular operators may become interested in employing user devices as an important asset in their networks, to benefit from their improved computational power, storage and caching capacity, wireless access and sensing capability, as well as efficient support for proximity services. Accordingly, adequate sources of motivation that facilitate the end-user decisions to lend their personal devices for the collective tasks need to be involved. In return, to compensate for the corresponding reduction in the networking
and computation power actually available to the individual user, more capable network assistance protocols will have to be developed – guiding the UE toward the best opportunities to receive its desired service (e.g., user-in-the-loop [3] and similar concepts). This rationale brings into focus the role that social relations and interactions between an individual human user and its proximate neighbors may play in supporting the maturing D2D communications paradigm.

In the past, community-centric incentives were exploited frequently, which means agreeing to engage into direct connectivity to cooperate with other like-minded individuals in certain well-defined scenarios (such as a conference, concert, sports event, etc.). However, in order for this solution to scale to network-wide applications, operator-driven incentive mechanisms are strongly demanded, such as dynamic pricing technique in [4]. Indeed, recent D2D-centric studies are already exploring benefits from the integration between social and communications domains [5], but most existing work implicitly assumes that all the users are equally likely to cooperate and share data. However, this is not the case in practice as users acquire and own digital content based on their individual interests and may not be willing to expose it unless trust is established with a potential D2D partner. As a result, our main motivation behind this research is a possibility to construct a trustworthy 5G-grade D2D connectivity environment (see Fig. 1) featuring both the offline human interactions (i.e., driven by the user encounter patterns) as well as the online human interactions (i.e., driven by social applications similar to Facebook, Twitter, and LinkedIn).

Fig. 1. Urban network-assisted D2D applications.

In this work, we concentrate on introducing a novel layer of social awareness, which empowers the communicating devices to become the autonomously deciding entities. Our main objective is thus to explore how the two domains – the human social awareness and the D2D-enabled proximate connectivity – may interplay to improve the resulting communications performance (in terms of better system throughput) as well as achieve higher levels of service quality (in terms of better connectivity). These attractive improvements, together with the resulting growth in the UE energy efficiency, may therefore constitute the much needed incentives for the eventual user adoption of the promising D2D paradigm.

II. D2D Market and User Adoption

Presently, the 5G market is still at its developing stage indicating the projected compound annual growth rate (CAGR) of 58.2% during 2013-2020 [6]. In particular, future cellular networks are expected to be employed across a variety of market segments, including the proximity-based applications and multimedia services, along the avenues of public safety, social networking, and Internet of Things (IoT). As network operators have near-exclusive opportunities to handle the transmitted data, device location (proximity), and other user ecosystem information, we may expect the advent of a new generation of mobile services based on such context and proximity knowledge. However, direct connectivity is inherently constrained by certain real-life factors (such as contact time, location, duration of connectivity, user preferences, etc.), which makes it challenging to reach the critical mass of D2D users in today’s networks [7].

With appropriate user adoption mechanisms, we envision the rapid proliferation of D2D communications scenarios, which would include not only public safety and emergency situations together with vehicle-to-vehicle information exchange for enhanced traffic safety, but also embrace commercially available pre-standard products that enable social networking and peer-to-peer communications outside of infrastructure coverage or in case of congestion. Although in these examples social awareness and the level of trust between communicating parties are markedly different, direct communications capability remains useful in terms of reducing latency, ensuring connectivity without infrastructure, improving reliability, and, ultimately, augmenting user experience.

To reach this vision, relevant contextual elements may be utilized to identify the typical behavioral patterns and stable interpersonal relationships of humans, thus aiding in matchmaking and timely formation of trusted user groups. For example, trust can be based on social media connections, since users within such networks are more likely to acquire similar content and share it with each other. However, important questions then emerge as to what would happen in larger heterogeneous coalitions consisting of both friends and strangers. In particular, to what extent the trust is transitive (trust to a friend of a friend) and does A trusts B and B trusts C imply A trusts C (similarly, does A trusts B and A trusts C imply B trusts C)?

In turn, D2D connectivity may impact the user-initiated activities as well as provoke or encourage external interactions (e.g., a service advertisement triggered from within the proximity range). Therefore, a key challenge behind the user adoption of D2D communications lies in understanding individual trust and privacy relationships, as proximity-based connectivity may generally lead to a lack of anonymity and confidence. Interestingly, user location history (e.g., in a form of joint movement patterns) may assist in determining social ties between the communicating users to establish the level of
trust between them [8] (e.g., if users meet and travel together repeatedly, they are more likely to be familiar).

In summary, by monitoring the common contacts (including friend-of-a-friend and other weak ties), the system can augment its legacy trust establishment solutions. Yet, even these advanced approaches do not seem to completely satisfy the needs of 5G-grade trust-based D2D applications. To effectively stimulate user adoption, there has to be a meaningful value proposition for end customers. However, the existing marketing campaigns behind the next-generation D2D technology are primarily targeting operators/industry and thus do not appeal as much to masses of people. Therefore, the main question emerges: How can user adoption of D2D technology be incentivized effectively? Along these lines, we identify three possible levels of user incentives that may apply to specific D2D scenarios:

- **Pragmatic incentives**: typical user behavior is to remain egoistic, which means that the ultimate interest in using D2D technology should be proportional to the corresponding improvements in throughput, energy savings, and latency;
- **Indirect incentives**: D2D service providers potentially benefiting from the enhanced network performance may adopt new business models, where economic incentives (e.g., user’s data plan discounts) are considered as rewards offered to users for lending the resources of their personal devices;
- **Social incentives**: the key motivation that can make the user drift from its egoistic behavior to altruism or reciprocity is sociality, where users lend their resources in order to assist friends, relatives, or other relevant peers. Here, the fundamental human needs of e.g., belonging, social reputation, and social usefulness could be considered to develop novel models of creating incentives.

Importantly, to mitigate the risks of user distrust and rejection, our envisioned “social D2D” paradigm has to maintain high degrees of trustworthiness in data delivery among the connected D2D-capable UEs. This is particularly crucial whenever direct communications is utilized to extend the cellular coverage in cases when network connectivity becomes temporarily unavailable to the users (due to mobility, obstacles, disruptions, etc.). In what follows, we comprehensively outline how the above objectives can be achieved by the proposed social D2D paradigm. Then, we conduct a supportive system-level performance analysis of characteristic D2D applications, mindful of their trust requirements, that strongly emphasize the concepts of human and device sociality in the respective mobile data delivery process.

### III. Bridging Across Technology and Sociality

We firmly believe that sociality has the potential to become a core incentive across a wide range of applications and services wherein D2D communications may demonstrate non-incremental benefits. However, the social domain should not be considered as a standalone enabling factor for proximate connectivity (see Fig. 1). By contrast, it needs to carefully match the respective technology constraints and features of the physical communications domain (such as the utilized spectrum, radio technology, battery/power resources, etc.). In this regard, our vision is in that not only human users and their social interactions are to be accounted for, but also the associated interactions between the user devices with their specific notion of sociality. This expectation is well supported by the recent research developments within the IoT community, which target to embrace the social networking concepts [9] to build trustworthy relationships among the devices [8]. In our present research (see Table I), we thus consider the two distinct types of sociality as described below.

- **User-driven sociality**: in this case, humans are willing to interact and are directly controlling their social activities. The degree of how much two users are interested in exchanging data is characterized by a so-called human social relationship (HSR) factor, which may be linked to a social media tie, a family tie, etc. This measure is directly related to the level of familiarity and trust, according to which friends, relatives, or colleagues are likely to connect and share their content more frequently than the unfamiliar users. Within the same class of sociality, we may also consider the relationships based on the market pricing relational (MPR) model. The founding principle behind the MPR model is proportionality, as well as knowledge of how the relevant interactions are organized with respect to a common scale of values. In other words, the relationships established among people are driven by their willingness to interact or cooperate only in the light of achieving mutual benefits. In the literature, there are several examples that focus on smart surrounding scenarios for context-aware applications. For instance, triggers from the environment may invite and motivate people to socialize and/or cooperate, and thus take advantage of services within coverage (proximity market, gaming, advertising, etc.).

- **Device-driven sociality**: in this case, devices may autonomously interact according to the specific rules preset by the device owners and manufacturers – without an explicit user intervention during such interaction. Social relationships among the device owners are not necessarily required to foster this type of cooperation. To construct this sociality level, mobility patterns and relevant context can be considered to configure the appropriate forms of socialization [9]. Among these, the so-called co-location object relationships (C-LOR) and co-work object relationships (C-WOR) are established between devices in a similar manner as among humans, when they share personal (e.g., cohabitation) or public (e.g., work) experiences. Another type of relationships may be defined for the objects owned by a single user, which is named ownership object relationship (OOR) and may be of interest, for instance, when a number of devices belong to the same personal cloud.

Bridging across the realm of social-awareness and real-world D2D-based implementations, a factor of particular importance is dual mobility of the communicating entities. D2D application developers need to extend support for trust and
TABLE I
SOCIAL RELATIONSHIP FACTORS BETWEEN DEVICES, POSSIBLE APPLICATIONS, AND THE ASSOCIATED TRUST VALUE.

<table>
<thead>
<tr>
<th>Relationship Type</th>
<th>Typology</th>
<th>Description</th>
<th>Applications</th>
<th>Trust Value</th>
</tr>
</thead>
<tbody>
<tr>
<td>Human social relationship (HSR)</td>
<td>User-driven</td>
<td>Familiarity degree with friends/relatives/colleagues</td>
<td>Leisure applications, confidential data, eHealth, mission-critical communications</td>
<td>[0-1]</td>
</tr>
<tr>
<td>Market pricing relationship (MPR)</td>
<td>User-driven</td>
<td>Cooperative interactions with services triggered by the environment</td>
<td>Proximate marketing, proximity gaming, advertising</td>
<td>0.2</td>
</tr>
<tr>
<td>Ownership object relationship (OOR)</td>
<td>Device-driven</td>
<td>Relationship between objects owned by the same person</td>
<td>Personal cloud, smart home</td>
<td>1</td>
</tr>
<tr>
<td>Co-location object relationship (C-LOR)</td>
<td>Device-driven</td>
<td>Objects sharing personal experiences (e.g., cohabitation)</td>
<td>Information/data exchange at social aggregation points (concerts, sports events)</td>
<td>0.8</td>
</tr>
<tr>
<td>Co-work object relationship (C-WOR)</td>
<td>Device-driven</td>
<td>Objects sharing public experiences (e.g., work)</td>
<td>Information/data exchange at work aggregation points (e.g., fairs, workshops)</td>
<td>0.6</td>
</tr>
</tbody>
</table>

confidence management to ultimately enable secure proximate communications that are aware of unrestricted human/device mobility. In this regard, the most challenging use cases are those, in which the out-of-coverage cellular devices are also becoming involved into the network-assisted D2D data exchange in the absence of a reliable link to the central trusted authority (residing e.g., in the operator cloud). In order to effectively address this and other aforementioned scenarios, our study investigates how human- and device-centric social relationships can achieve trusted connectivity in relevant D2D groups under realistic mobility as well as, possibly, partial cellular network coverage. In particular, we focus on three insightful study cases:

- **Trust-based human applications (Case A).** Interactions among humans with tight trust requirements are included here. In these study cases, the end-user is willing to reliably know which person the data are exchanged with. To this end, user-driven sociality is of paramount importance and sometimes even becomes the only acceptable enabler. Examples of such applications are found in work-related environments, such as construction sites as well as transport and cargo handling facilities in harbors or airports, where stringent safety regulations dictate increased levels of trust. Other applications may include confidential and mission-critical data collection, such as that for eHealth and safety applications.

- **Leisure and entertainment applications (Case B).** Connectivity between proximate devices supports applications for users at leisure, such as entertainment and gaming, non-confidential information sharing, and similar non-critical services (e.g., map sharing for intelligent transportation systems). These applications do not necessarily need an explicit social relation between the device owners, and trusted communications may rather be driven by the sociality of devices. Typical scenarios of interest in this category may consider users distributed in a certain area and sharing similar interests, such as content dissemination in a stadium, a university campus, or a pub, where matching people (in terms of interests, age, familiarity, etc.) interact by employing their devices.

- **Critical machine-to-machine (M2M) applications (Case C).** In the situations where, by definition, there is no (or, very limited) human intervention, automated device connectivity may still benefit from some form of social awareness. One may consider hazardous working environments, such as those often met in industrial automation scenarios, where large numbers of machines, sensors, actuators, or robots communicate mission-critical data. To facilitate such information exchange, trust can be delivered by operator-enforced incentives and policies, leading to optimized communications performance with higher degrees of security.

IV. SOCIAL-AWARE FRAMEWORK FOR TRUSTED D2D

Our proposed social-aware framework aims at enabling trusted D2D-centric data delivery for proximate users in mobile environments. In these situations, direct links may (temporarily) extend or substitute cellular network connections, when the operator services become unavailable to (some of) the customers. Relevant clustering of the D2D devices can be conveniently modeled as a non-transferable utility (NTU) coalitional game $(\mathcal{N}, \mathcal{V})$, where $\mathcal{N}$ is a set of $N$ players and $\mathcal{V}$ is a function, such that for every coalition $S \subseteq \mathcal{N}$, $\mathcal{V}(S)$ is a closed convex subset of $\mathbb{R}^{|S|}$. The latter contains the payoff vectors that the players in $S$ can achieve, and $|S|$ is the number of members in the coalition $S$. The objective for the players in this NTU game is to maximize the value of the coalition they belong to. In the proposed framework, the utility for a coalition is defined as the degree of proximity and the strength of social relationships for the corresponding D2D-based cluster. To this aim, we define an NTU game, where for any coalition $S \subseteq \mathcal{N}$ the value $v_i(S)$ associated with each player $i \in S$ is determined as:

$$v_i(S) = \sum_{j=1}^{|S|} s_{i,j} \cdot p_{i,j}/|S|,$$

where $s_{i,j} \rightarrow [0, 1]$ is a function measuring the level of social relationships (or friendship) between a pair of communicating entities, whereas the second term $p_{i,j}$ is a binary function taking the value of 0 if the users $i$ and $j$ are not in proximity, and taking the value of 1 otherwise (by construction, we set $p_{i,i} = 1$). The resulting product of these two functions is then averaged across the players in a given coalition $S$, thus always yielding a value within the range of $[0, 1]$.

The actual definition of the social relationships level between the devices $s_{i,j}$ needs to allow for appropriate weighting of the contributions coming from human relationships and
device sociality. Therefore, it may be defined as a weighted function 
\[ s_{i,j} = \alpha \cdot H_{i,j} + (1 - \alpha) \cdot D_{i,j}, \]
where \( H_{i,j} \in [0,1] \) is the degree of human-to-human sociality and \( D_{i,j} \in [0,1] \) is the degree of device-to-device sociality. The social relationships between humans and devices are modeled based on the values shown in Table I, where the "Typology" field identifies which class the social relationships belong to. The "user-driven" option corresponds to relationships that are being used to determine the value of \( H_{i,j} \); the HSR and MPR relationships belong to this class. In contrast, the "device-driven" option identifies relationships that are used to determine the value of \( D_{i,j} \); the OOR, C-LOR, and C-WOR relationships belong here.

Whenever two entities can be associated to more types of relationships of the same class, we select the strongest tie having the highest value [8]. The motivation for this is that stronger social relationships lead to higher probability of "trusted" connection, thus providing improved performance. Further, the weighting term \( \alpha \in [0,1] \) is introduced into our model to adjust the respective contributions coming from the \( D_{i,j} \) and \( H_{i,j} \) terms according to a specific application and/or scenario. To this end, the two extreme cases with \( \alpha \) equal either to one or to zero, are representative of only human- and only device-driven sociality scenarios, respectively, as holds for the applications discussed under study cases \( A \) (i.e., trust-based human-to-human scenario) and \( C \) (i.e., critical machine-to-machine scenario).

In summary, the study cases \( A \) and \( B \) discussed in the article represent two illustrative examples of the extreme situations with only human- and only device-driven sociality. In the third investigated scenario, study case \( B \), the focus is on applications for users at leisure, where both human- and device-driven types of sociality are considered. In this case, the importance of the human- and device-driven sociality is assumed to be equal-weighted, which motivates the choice of \( \alpha = 0.5 \). However, other values of \( \alpha \) may also be considered based on the scenario under consideration and the application in question. While a more thorough analysis of all possible scenarios remains out of the scope of this article, here we aim at proposing a powerful model that allows to explore how the human social awareness and the D2D-enabled proximate connectivity may interact to improve the resulting communications performance and service quality.

We can now define the value of \( v(S) \) for a coalition \( S \) as the average degree of proximity and strength of social relationships for the users in the cluster:
\[ v(S) = \frac{\sum_{i=1}^{|S|} v_i(S)}{|S|}. \]
Importantly, the highest possible value associated with a certain coalition \( v(S) = 1 \) is achieved if all of the devices are located in their mutual D2D coverage, as well as all of them enjoy the maximum level of friendship. In practice, the latter seldom happens in the grand coalition incorporating all the networked devices, and thus independent and disjoint coalitions are typically formed. To control the resulting stability problems, existing solutions proposed in recent literature can be adopted [10]. For instance, an iterative application of the merge and split rules enables the much needed convergence to a stable coalitional structure of the network.

Once stable D2D-clusters are formed, the D2D connectivity within them should be secured both in the cases of full and partial cellular coverage. Whenever connected reliably to the centralized network infrastructure, the D2D clusters can establish their information security rules by employing the conventional methods, hence relying on the operator infrastructure acting as a trusted authority. However, when cellular connection becomes unavailable, secure associations between D2D partners may benefit from solutions in [11] and [8], which enforce trustworthiness of human- and device-driven interactions, respectively.

V. OUR PERFORMANCE EVALUATION CAMPAIGN

To validate the envisioned D2D framework and quantify the benefits of the proposed social-aware, secure clustering solution, a supportive system-level performance assessment has been conducted by utilizing our custom-made simulation environment, named WINTERsim\(^1\). Due to the need of modeling full-scale user mobility and application-level traffic, the underlying system-level evaluation methodology had to be streamlined, by simplifying the propagation and interference conditions, and thus employing the parameters summarized in Table II. The output metrics of interest are aggregate effective throughput and corresponding device energy efficiency, as well as degree of connectivity, which indicates the proportion of users covered by cellular and/or direct links.

Our reference scenario features a tagged cellular BS running the contemporary 3GPP LTE technology deployed within a [150m × 150m] area of interest, and having the coverage range of 100m, resulting in around 70% of reliable cellular coverage available to the users. For the sake of completeness, we later consider several alternative values for the LTE coverage range—in order to understand the effects that it has on the degree of connectivity. Further, our communicating entities (humans and their connected devices) are allowed to freely move across the considered area of interest according to the characteristic "Levy flight" mobility pattern [12]. More specifically, we investigate the performance of a multimedia application with the packet size of 100 KB and the packet inter-arrival time of 10 s (e.g., video dissemination, eHealth, etc.). As for the D2D communications technology, discovery and connection setup functions are managed directly by the LTE BS with the appropriate network assistance protocols, whereas the actual direct data transmission is performed out-of-band (e.g., over WiFi-Direct links that can operate in parallel with LTE assistance, as they utilize the unlicensed spectrum).

The following alternative communications options are compared in our system-level study:
- **Cellular (LTE) solution.** A benchmark setup, where the connectivity is available only over the conventional cellular links, without any D2D-based transmission or coverage extension possibilities;
- **Simple D2D solution.** Only mobile devices under the reliable cellular network coverage may connect directly to form the D2D pairs according to the shortest distance between them. The BS is acting as the conventional

\(^1\)WINTERsim system-level simulator: http://winter-group.net/downloads/
TABLE II
CORE SIMULATION PARAMETERS

<table>
<thead>
<tr>
<th>Application parameter</th>
<th>Value</th>
</tr>
</thead>
<tbody>
<tr>
<td>Packet size</td>
<td>100 KB</td>
</tr>
<tr>
<td>Inter-arrival time</td>
<td>10 s</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>System parameter</th>
<th>Value</th>
</tr>
</thead>
<tbody>
<tr>
<td>Cell radius</td>
<td>100 m</td>
</tr>
<tr>
<td>Maximum D2D range</td>
<td>30 m</td>
</tr>
<tr>
<td>WiFi-Direct target data rate</td>
<td>40 Mbps</td>
</tr>
<tr>
<td>LTE target data rate</td>
<td>10 Mbps</td>
</tr>
<tr>
<td>LTE BS Tx power</td>
<td>46 dBm</td>
</tr>
<tr>
<td>UE Tx power</td>
<td>23 dBm</td>
</tr>
<tr>
<td>Machine Tx power</td>
<td>0 dBm</td>
</tr>
<tr>
<td>D2D link setup time</td>
<td>1 s</td>
</tr>
<tr>
<td>Mobility model</td>
<td>Levy flight (with parameter 1.5 [12])</td>
</tr>
<tr>
<td>Number of UEs</td>
<td>[10-100]</td>
</tr>
<tr>
<td>$H_{i,j}$</td>
<td>[0-1]</td>
</tr>
<tr>
<td>$D_{i,j}$</td>
<td>[0.6,0.8,1]</td>
</tr>
</tbody>
</table>

trusted authority by guaranteeing trustworthy connectivity for all in-coverage D2D partners;
- **Advanced (social-aware) D2D solution.** Users may cluster together according to the proposed social-aware D2D framework. This may also happen under partial cellular network coverage, thus leading to D2D-based coverage extension. All connectivity (including the out-of-coverage links) is made trusted by taking advantage of the distributed information-security solution without a central trusted authority [11]. To further visualize the effects of both human- and device-driven sociality, we consider the three reference study cases and the associated $\alpha$ values as defined in Section IV: 1, 0.5, and 0 for study cases A, B, and C, respectively.

To ease further exposition, for the baseline LTE solution and the simple D2D scheme we only account for the portion of data transmitted by the users within the reliable cellular network coverage (by aggregating these effective values across individual users). In case of the advanced D2D solution, we additionally consider traffic of the out-of-coverage users enabled by our trusted, social-aware framework.

First, Fig. 2 indicates the achievable aggregate effective throughput as a function of the number of networked devices. Hence, we learn that at all times the proposed social-aware D2D solution outperforms the LTE-only alternative considered in this study, as well as the simple D2D solution. In particular, the case of $\alpha = 0$ (study case C, when only device-driven sociality is considered) achieves the best performance, followed by the cases when $\alpha = 0.5$ and $\alpha = 1$ (study case A, when only human-driven sociality is considered). This result suggests that the interactions based on the second level of sociality — those accounting for the relationships between the devices — may introduce significant benefits to the system operation, whenever the trust requirements of a running application allow for this.

Further, Fig. 3 illustrates the degrees of connectivity offered within our area of interest, when a varying percentage of such area is covered by the LTE BS (eNodeB). In particular, the considered scenario corresponds to study case B (i.e., $\alpha = 0.5$), with 100 devices residing in the system. To this end, we measure the proportion of devices being served with the proposed social-aware D2D solution and compare it against the corresponding figure as achieved with the simple D2D scheme. As observed in the left subplot of Fig. 3, the proposed approach always demonstrates higher percentage of served users, with the benefits increased even further in the face of reduced LTE coverage. In particular, the proportion of served devices more than doubles with our social-aware operation, when the LTE coverage is only available over a half of the area of interest.

Further, in the right subplot of Fig. 3 we report on the proportion of users served with a simple D2D link (i.e., in case of simple D2D solution) or a D2D cluster (i.e., considering the proposed social-aware D2D solution). Clearly, this is a subset of the entire set of served users as it represents the share of users that either (i) prefer to establish a direct link instead of downloading the content over the LTE infrastructure, or (ii) can only be served over D2D connections in the locations where there is no LTE coverage. As we learn from this plot, when the available cellular coverage area is particularly small, in case of simple D2D solution the number of users that establish a D2D connection is low. This is due to the fact...
that under-coverage users reside in proximity to the BS and thus receive higher channel quality comparing to that on the D2D link. As a consequence, a higher number of users may be served through the infrastructure links with the LTE BS. On the contrary, the percentage of users served via D2D connections is three times higher for the proposed social-aware D2D solution. The explanation of this result is in that our solution is able to provide connectivity also to those users that are outside of the cellular coverage (i.e., within D2D clusters). Note that this important outcome is achieved owing to the operation of our social-based, secure cluster formation scheme.

![Graph showing impact of social relationships on the user energy efficiency.](image)

Fig. 4. Impact of social relationships on the user energy efficiency.

Finally, performance results for the aggregate energy efficiency of user data transmissions are reported in Fig. 4. This metric has been evaluated by taking into account the relevant transmission power for each network node (refer to the values reported in Table II). Again, the social-aware D2D approach outperforms both the considered baseline LTE and the simple D2D alternatives. In particular, for the case of $\alpha = 0$ our proposed solution reaches its highest gain by contrast to the benchmark LTE operation. This is due to lower transmit power of small-scale devices (i.e., connected machines) as compared to more power-hungry handheld UEs.

To conclude, our analysis indicates that social ties among both humans and their connected devices impact the ultimate performance of the proposed social-aware scheme that enables the trusted D2D clusters. In particular, with higher levels of social relationships, the resulting effective throughput grows, also yielding positive effects on the energy consumption of the devices and their degrees of connectivity. The key reason is that having better social relationships plays in favor of having larger coalitions between proximal humans/devices, even in the cases of partial cellular network coverage. Clearly, the improved throughput performance of our social-aware D2D solution is achieved at the cost of somewhat increased latency, as compared to the simple D2D scheme. Indeed, to deliver reliable connectivity to proximate humans/devices, especially outside of LTE coverage, more time-consuming security procedures are required to be executed in the UE. For instance, handheld devices need additional time to complete the security methods from [11], which leads to slightly higher latencies with the growing number of communicating entities. However, the implementation efficiency of said security mechanisms can be optimized further to reduce the computation time, which we leave for our subsequent study.

VI. STANDARDIZATION ASPECTS AND OUTLOOK

Historically, D2D communications capabilities and respective support for proximity services were first introduced in Release 12 of the 3GPP protocol suite [13]. Correspondingly, the main targeted use cases and associated system requirements are well-captured in the feasibility report (see 3GPP TR 22.803 document). It thus serves as a solid foundation for the development of enabling technology components, including device synchronization, service and device discovery, as well as actual direct communications – both under and outside of cellular network coverage. A direct consequence of the emerging D2D interface, the so-called sidelink, is the need to ensure interoperability of the devices produced by different vendors and, possibly, served by various cellular operators. Therefore, the appearance of the sidelink is a major advancement in the 3GPP architecture, affecting physical layer procedures, higher layers, and non-access stratum protocols alike.

While the initially considered set of D2D-related scenarios and requirements has been sub-divided into public safety and general commercial use cases, Release 14 LTE networks are being prepared to additionally accommodate vehicle-to-vehicle and vehicle-to-infrastructure communications services (see 3GPP TR 22.885 document). We thus expect that as application developers, service providers, and user equipment manufacturers experiment with the rich capabilities offered by the D2D connectivity, further use cases will become attractive, including D2D-powered machine-type communications. Therefore, in 5G networks, we envision that the distinction between public safety and commercial applications will become blurred, thus making technology development (in the form of its components and end solutions) increasingly meaningful for in-coverage, partial network coverage, and out-of-coverage situations. In this context, our proposed D2D paradigm – enhanced with the involvement of social relationships established not only among familiar humans but also among familiar devices – will decisively contribute to the delivery of novel types of services over proximate links.

Among the many examples appearing on this scene, [14] already implements the discussed concepts for intelligent transportation systems by exploiting the aforementioned vehicle-to-vehicle and vehicle-to-infrastructure communications services. Complementary to the research literature on the topic, there is currently a strong need for a broader standardization campaign. This may address such issues as, for example, (i) definition of categories for inter-device social relationships as well as rules for their triggering, (ii) common social-oriented interfaces and interaction models for users in pervasive D2D scenarios, (iii) distributed methodologies to enable secure data exchange among groups of humans/devices communicating over D2D

---

2For details, see the following 3GPP specifications: TS 36.213, TS 36.300, TS 23.303, and TS 24.301.
links, possibly without a reliable connection to the centralized trusted authority, etc.

In summary, we are about to embrace the D2D communications as one of the key technologies within the rapidly maturing 5G ecosystem. It will widely enable both the owners of advanced wireless devices as well as the smart and social IoT objects across diverse, pervasive platforms to effectively become a part of the cellular landscape. This, in turn, will pave the way to improved cellular service provisioning by e.g., offering D2D-based data relaying, content distribution and caching, or other forms of cooperative communications to augment the existing spectrum usage and device energy efficiency [15]. Another exciting research direction is to develop new mechanisms that take advantage of the unique position of cellular operators – with their well-developed infrastructure and pricing methods – to create incentives, win-win collaborative strategies, and ultimately raise social awareness among spectrum owners, network operators, and wireless device users. For 3GPP networks, the basic building blocks, associated protocol structures, and physical layer procedures are already being defined, while the creation of corresponding incentives and social awareness schemes that engage users as part of the service provisioning effort remains in strong need of further research.

REFERENCES


AUTHORS’ BIOGRAPHIES

Aleksandr Ometov (aleksandr.ometov@tut.fi) received his specialist degree in information security from the St. Petersburg State University of Aerospace Instrumentation, St. Petersburg, Russia, in 2013. He has been a research assistant at the Department of Electronics and Communications Engineering of Tampere University of Technology, Finland since 2013. Currently, his major research interests are wireless communications, information security, heterogeneous networking, cooperative communications, and machine-to-machine applications.

Antonino Orsino (antonino.orsino@unirc.it) received his B.Sc. degrees in Telecommunication Engineering from University Mediterranea of Reggio Calabria, Italy, in 2009 and his M.Sc. from University of Padova, Italy, in 2012. Currently, he is a Ph.D. student at the DIIES Department, University Mediterranea of Reggio Calabria and a visiting researcher at Tampere University Technology, Finland. His current research interests include Device-to-Device and Machine-to-Machine communications in 4G/5G cellular systems. He has served as a reviewer for several major IEEE conferences and journals.

Leonardo Militano (leonardo.militano@unirc.it) is currently an Assistant Professor at the Mediterranea University of Reggio Calabria, Italy. He received his M.Sc. degree in Telecommunications Engineering in 2006 and his Ph.D in Telecommunications Engineering in 2010 from the University of Reggio Calabria. He has been a visiting Ph.D student at the Mobile Device group at University of Aalborg, Denmark. His major areas of research are wireless networks optimization, user and network cooperation, device-to-device communications and game theory.

Dimitri Molchanov (dmitri.molchanov@tut.fi) is a Senior Research Scientist in the Department of Electronics and Communications Engineering, Tampere University of Technology, Finland. He received his M.Sc. and Cand.Sc. degrees from Saint-Petersburg State University of Telecommunications, Russia, in 2000 and 2002, respectively, and Ph.D. degree from Tampere University of Technology in 2006. His research interests include performance evaluation and optimization issues in wired and wireless IP networks, Internet traffic dynamics, quality of user experience of real-time applications, and traffic localization in P2P networks. Dimitri Molchanov serves as TPC member for a number of international conferences. He authored more than 50 publications.

Giuseppe Araniti (araniti@unirc.it) is an Assistant Professor of Telecommunications at the University Mediterranea of Reggio Calabria, Italy. From the same University he received the Laurea (2000) and Ph.D. degree (2004) in Electronic Engineering. His major area of research includes Personal

Ekaterina Olshannikova (ekaterina.olshannikova@tut.fi) is a project researcher at the Department of Pervasive Computing, Tampere University of Technology, Finland. She received the Art Critic specialist degree in History and Theory of Fine Art at St. Petersburg State University of Technology and Design, St. Petersburg, Russian Federation, in 2013. Her major research interests are in Big Data, Augmented and Virtual reality, Proximity-based and Location-based services, Human-Computer Interaction and User Experience design.

Gabor Fodor (gabor.fodor@ericsson.com) received the Ph.D. degree in teletraffic theory from the Budapest University of Technology and Economics, Budapest, Hungary, in 1998. Since then, he has been with Ericsson Research, Stockholm, Sweden. He is currently a Master Researcher with a specialization in modeling, performance analysis, and protocol development for wireless access networks. He has authored around 50 papers in reviewed conference proceedings and journals and holds over 40 patents (granted or pending).

Sergey Andreev (sergey.andreev@tut.fi) is a senior research scientist in the Department of Electronics and Communications Engineering at Tampere University of Technology, Finland. He received the Specialist degree (2006) and the Cand.Sc. degree (2009) both from St. Petersburg State University of Aerospace Instrumentation, St. Petersburg, Russia, as well as the Ph.D. degree (2012) from Tampere University of Technology. Sergey (co-)authored more than 100 published research works on wireless communications, energy efficiency, heterogeneous networking, cooperative communications, and machine-to-machine applications.

Thomas Olsson (thomas.olsson@tut.fi) is an adjunct professor and post-doctoral researcher at the Department of Pervasive Computing in Tampere University of Technology (TUT), Finland. He received his Dr. Tech. degree from TUT in 2012 with a thesis addressing user experience and user expectations of future mobile augmented reality systems. Currently, he leads a research team focusing on the user experience aspects of social proximity-based systems that aim to enhance social interaction between co-located people. He has (co-)authored over 60 research papers on user experience, augmented reality, ubiquitous computing systems, multi-device interaction, smart environments, haptic interfaces, and user expectations of new interactive technology.

Antonio Iera (antonio.iera@unirc.it) graduated in Computer Engineering at the University of Calabria, Italy, in 1991 and received a Master Diploma in Information Technology from CEFRIEL/Politecnico di Milano, Italy, in 1992 and a Ph.D. degree from the University of Calabria in 1996. Since 1997 he has been with the University of Reggio Calabria and currently holds the position of Full Professor of Telecommunications and Director of the Laboratory for Advanced Research into Telecommunication Systems. He is IEEE Senior Member since 2007. His research interests include next generation mobile and wireless systems, RFID systems, and Internet of Things.

Johan Torsner (johan.torsner@ericsson.com) is a Research Manager in Ericsson Research and is currently leading Ericsson’s research activities in Finland. He joined Ericsson in 1998 and has held several positions within research and R&D. He has been deeply involved in the development and standardization of 3G and 4G systems and has filed over 100 patent applications. His current research interests include 4G evolution, 5G and machine-type communication.

Yevgeni Koucheryavy (yk@cs.tut.fi) is a professor and Lab Director at the Department of Electronics and Communications Engineering of Tampere University of Technology (TUT), Finland. He received his Ph.D. degree (2004) from TUT. He is the author of numerous publications in the field of advanced wired and wireless networking and communications. His current research interests include various aspects in heterogeneous wireless communication networks and systems, the Internet of Things and its standardization, as well as nanocommunications. He is Associate Technical Editor of IEEE Communications Magazine and Editor of IEEE Communications Surveys and Tutorials.

Tommi Mikkonen (tjm@cs.tut.fi) is a professor of software systems at the Tampere University of Technology. His research interests include software architectures, distributed systems, and mobile and Web software. Mikkonen has a PhD in computer science from Tampere University of Technology. Contact him at tommi.mikkonen@tut.fi.
Publication II

© 2016 IEEE. Reprinted, with permission, from


In reference to IEEE copyrighted material which is used with permission in this thesis, the IEEE does not endorse any of Tampere University of Technology’s products or services. Internal or personal use of this material is permitted. If interested in reprinting/republishing IEEE copyrighted material for advertising or promotional purposes or for creating new collective works for resale or redistribution, please go to http://www.ieee.org/publications_standards/publications/rights/rights_link.html to learn how to obtain a License from RightsLink.
Feasibility Characterization of Cryptographic Primitives for Constrained (Wearable) IoT Devices

Aleksandr Ometov†, Pavel Masek*, Lukas Malina*, Roman Florea†, Jiri Hosek*, Sergey Andreev†, Jan Hajny*, Jussi Niutanen†, and Yevgeni Koucheryavy†

†Tampere University of Technology, Finland, Tampere, Korkeakoulunkatu 10, FIN-33720
*Brno University of Technology, Czech Republic, Brno, Technicka 3082/12
†Intel Finland, Tampere, Insinoorinkatu 7, FIN-33720

Contact e-mail: aleksandr.ometov@tut.fi

Abstract—The Internet of Things (IoT) employs smart devices as its building blocks for developing a ubiquitous communication framework. It thus supports a wide variety of application domains, including public safety, healthcare, education, and public transportation. While offering a novel communication paradigm, IoT finds its requirements closely connected to the security issues. The role of security following the fact that a new type of devices known as wearables constitute an emerging area. This paper delivers an applicability study of the state-of-the-art cryptographic primitives for wearable IoT devices, including the pairing-based cryptography. Pairing-based schemes are well-recognized as fundamental enablers for many advanced cryptographic applications, such as privacy protection and identity-based encryption. To deliver a comprehensive view on the computational power of modern wearable devices (smart phones, watches, and embedded devices), we perform an evaluation of a variety of them utilizing bilinear pairing for real-time communication. In order to deliver a complete picture, the obtained bilinear pairing results are complemented with performance figures for classical cryptography (such as block ciphers, digital signatures, and hash functions). Our findings show that wearable devices of today have the needed potential to efficiently operate with cryptographic primitives in real time. Therefore, we believe that the data provided during this research would shed light on what devices are more suitable for certain cryptographic operations.

I. INTRODUCTION

The Internet of Things (IoT) creates the means for interconnection of highly heterogeneous entities and networks bringing a variety of communication patterns, including Human-to-Human (H2H), Human-to-Machine (H2M), and Machine-to-Machine (M2M) communications. IoT in general and wearable technology in particular empower the industry to develop new technology in almost unlimited numbers. Today, the term wearables stands for connected devices that collect data, track activities and improve user experience across different application domains. From the IoT point of view, wearables could be characterized as networked “smart devices” equipped with microchips (System on the Chip, SoC), sensors, and wireless communications interfaces deployed in the immediate vicinity of their owner [1] (see also Fig. 1 indicating the devices used in everyday life of tomorrow).

New findings from the leading telecommunication players, such as Juniper [2] and Cisco [3], reveal that global retail revenue from smart wearable devices will treble by 2016, therefore reaching $53.2 billion by 2019, compared to the $4.5 billion at the end of 2015. The market over the following five years is expected to be substantially driven by the sales of smart watches and smart glasses. As it is common for new and highly innovative digital technologies, wearables will also challenge existing social and legal norms. In particular, wearable technologies raise a variety of privacy and safety concerns, which should be addressed immediately. Without strong security frameworks capable of being executed directly on wearable devices, attacks and malfunctions might overshadow any of the expected benefits.

![Future secure smart home / IoT environment](image)

Fig. 1. Future secure smart home / IoT environment

Wearable devices can be secured by means of the public key cryptography, i.e., digital signature schemes providing user authentication and protecting data during their transmission over the medium. Information security specialists are targeting to design digital signature schemes that are (i) secure, (ii) computationally efficient, and (iii) have small communication overheads. Conventional digital signature schemes use standard operations and are based on mathematical assumptions, such as the discrete logarithm problem, the RSA problem, or integer factorization [4]. These conventional methods provide standard security properties, including authenticity, integrity, and non-repudiation.

In this paper, we expand our vision not only on classical cryptography but also on pairing-based algorithms by evaluating their usability for wearables and other constrained IoT devices (smart watches, smart phones, and embedded devices, see Section II-4). In particular, pairing-based cryptography is often used in modern solutions to implement privacy-enhancing features that are difficult to achieve with conventional asymmetric cryptography. Using bilinear pairing operations, it is possible to design schemes like group signa-
tures [5], [6], anonymous attribute-based credentials [7], [8], or identity-based encryption [9]. Some of those mechanisms are particularly important for the IoT system operation, such as efficient revocation of invalid devices based on dynamic accumulators [10] and identification of attackers. These would be difficult to construct without pairing-based cryptography. Inspired by that, we analyze and evaluate the most common personal/wearable devices – starting from the conventional smartphones (Samsung Galaxy S4, Apple iPhone 6, etc.) to the embedded devices (Intel® Edison, Raspberry Pi 1 Model B, Raspberry Pi 2 Model B) to smart watches (Sony Smart Watch 3, Apple Watch) – with a particular focus on their ability to execute both standard and advanced cryptographic operations. As pairing-based cryptography primitives have not been rigorously evaluated on these devices so far, we also address this type of security functionality.

The rest of the paper is organized as follows. Section II provides the description of related work dealing with the classical cryptography and pairing-based cryptography. Further, in Section II-4 we discuss the selected devices for our test scenarios that are employed and characterized in Section IV. Finally, the lessons learned and conclusions are summarized in Section V.

II. CLASSICAL AND PAIRING-BASED CRYPTOGRAPHY

In our work, we consider the classical cryptographic primitives (calculations using big integer as multiplication, division, power functions and schemes for classical elliptic curves) for constructing the RSA signature, hash functions (SHA1, SHA-256), and block cipher (AES). These are well-known and well-analyzed construction blocks and we target to assess their execution time on low-power and constrained IoT devices.

A thorough overview is given in [11], where implementation of 12 lightweight and standard block ciphers in ATMEAL AVR ATtiny45 has been described. Further, in [12], the authors focus on benchmarking the modern hash functions, where 15 hashes were evaluated on 8-bit micro-controllers. However, there has been very limited work documenting classical cryptography implemented on modern wearable devices e.g., smart watches.

To construct novel cryptographic primitives with advanced security properties, Pairing-Based Cryptography (PBC) has been developed. PBC-based solutions have exploded since 2000, when Joux [13] presented the three-party one-round Diffie-Hellman protocol based on bilinear pairings. The bilinear pairings enable efficient design of many protocols with enhanced properties, such as one-round three-party key agreement, identity-based encryption, and group signatures [5].

1) Bilinear Pairing Operations: A bilinear pairing function maps two elements of groups $G_1$ and $G_2$ onto the third cryptographic group $G_T$, i.e., $e : G_1 \times G_2 \rightarrow G_T$. The pairing function $e$ must be computable, non-degenerate and bilinear. The pairing operations work with pairing-friendly Elliptic Curves (EC), including MNT curves (Miyaji-Nakabayashi-Takano) [14], BN curves (Barreto-Naehrig) [15]. The pairing operations can be symmetric ($G_1 = G_2$) or asymmetric ($G_1 \neq G_2$). Symmetric and/or asymmetric bilinear pairing operations can be computed by pairing algorithms, such as Weil, Tate, Ate, Eta, or O-Ate. Symmetric pairings are usually more computationally efficient than asymmetric pairings [16].

2) Pairing-Based Cryptographic Schemes: In general, pairing functions can reduce a problem that is in one group by solving it in a different group where the problem is easier to solve. This property enables to design new cryptographic schemes such as identity and attribute encryption, group signatures or three-party key establishment. In addition, the use of elliptic curves allows some pairing-based signature schemes to produce shorter signatures than the conventional digital signature schemes like RSA [17]. For example, the pairing-based short signature scheme BLS [18] employing the Weil pairing [19] produces 20 B only signatures. Due to space limitations, only a short overview is provided in this section; for more information related to pairing-based cryptography, please follow [20], [21].

3) Optimization of Pairing-Based Cryptographic Schemes: There are only few studies addressing the pairing-based cryptography on smartphones. For example, the work in [22] presents several ways for the efficient implementation of pairing-based cryptographic protocols on restricted devices. The BBS04 scheme that requires one online bilinear pairing during the signing phase is used to illustrate the below optimization approaches. The paper presents the pros and cons behind the approaches applied to the BBS04 scheme. The first approach uses the "Shamir's trick" [23], which reduces the time complexity of scalar multiplication (or modular exponentiation). The second method replaces an expensive pairing operation by less complex operations. The pre-computation of pairing operations is utilized, but the final efficiency depends on the number of components in the multi-exponentiations. The third approach delegates the computation of bilinear pairings to a more computationally stronger entity. Only the public parameters can be delegated and the communication between a constrained device and said entity must be fast.

The work in [24] presents the first Java wrapper of a pairing-based cryptography library. The authors implement jPBC that is a Java port of the PBC library written in C. As an example, the implementation of the BLS signature scheme [18] is described. The paper contains benchmarks of bilinear symmetric pairings and exponentiation operations for two devices (Samsung I9000 Galaxy S and a PC machine). However, the paper does not offer any performance results for asymmetric pairing operations that are required by many PBC signature schemes.

The challenge of PBC scheme optimization conventionally refers to decreasing the number of pairing operations. Optimization techniques, such as a pairing precipitation and pairing collapsing, can thus be applied to the PBC mechanisms. In addition, the verification phase of signature and group signature schemes can be optimized by using a batch verification trick [25]. The batch verification method can be used only in the case, when a verifier is able to verify more signatures in one batch. Some PBC schemes on constrained devices are able to delegate expensive pairing operations to more powerful nodes. But, this trick can only be done if the pairing operation...
does not map secret and private parameters as inputs. More about the optimization tricks can be found in [4].

4) **Performance Requirements of Pairing-Based Cryptographic Schemes:** The bilinear pairing operations are considered to be more computationally expensive than other modular arithmetic operations, such as scalar multiplication and modular exponentiation. The computation time of a pairing operation depends on the type of a pairing algorithm, the type and the length of EC parameters, the implementation of the PBC methods, as well as on the hardware and software specifications of a device. For example, according to MIRACL benchmarks, one 512-bit pairing operation by the Tate algorithm takes 20 ms and one modular exponentiation with 1024-bit numbers takes 8.8 ms [26]. Unfortunately, some devices, such as smartphones and smart cards, need more time for computing a single pairing operation. For example, the results in [24] indicate that one symmetric pairing operation takes about 254 ms on smartphone (Samsung I9000 Galaxy S). Moreover, the results in [4] show that asymmetric pairing operations are more time consuming and require around 3 seconds on current Android smartphones. However, no similar study is presently available for general public in the area of wearable devices.

III. SELECTED WEARABLE DEVICES

Broadly, terms “wearable technology”, “wearable devices”, and “wearables” all refer to electronic technologies or computers that are incorporated into items of clothing and accessories, which can be worn on the body [1]. Following the fact that the computational performance is constantly growing (see Table I for a comparison of hardware parameters), contemporary wearable devices are becoming able to perform similar computing tasks as handheld devices or even laptop computers.

For the purposes of our research work, we have selected today’s pioneers as well as already widely used devices from three main categories: (i) smartphones, (ii) smart watches, and (iii) embedded devices, see Fig. 2.

As representatives of the first group, we chose devices built on two main mobile platforms: Android and iOS. More specifically, we used Samsung Galaxy S4 (SGH-I337) and Jiayu S3 Advanced (JY-S3), both running Android 4.4.2, Apple iPhone 4s (MD128CS/A) running the iOS 7.1.2, and Apple iPhone 6 (MG4F2CN/A) with the latest iOS 9.1.

To provide a comprehensive evaluation at par with the selected smartphones, we also employed smart watches running Android Wear and Apple WatchOS. The utilized devices are correspondingly Sony Smart Watch 3 (SWR-50) with Android Wear 5.1.1 and Apple Watch 42mm Sport edition with WatchOS 2.0.

Following the fact that most of today’s embedded devices (often named the IoT development boards) are intended to be used also as wearables, we decided to additionally evaluate the well-known examples from this class: Intel® Edison [27], Raspberry Pi1 (Model B), and Raspberry Pi2 (Model B). Both Raspberry Pi devices run the latest version of Raspbian OS (Jessie, v 8.0) together with the latest version of Oracle JDK (1.8.0-b132), Edison features a Ubilinux 3.10.17-yocto-standard-r2 build equipped with JDK (1.8.0-b132)1. In more detail, Edison is a small-sized computing module aiming to enable the next generation of wearables and IoT devices, where size and power consumption are extremely important factors. In addition, Edison may be attached to a number of different extension boards, for example, to enable Arduino compatibility. Hence, Edison empowers a range of different use cases, whereas Raspberry Pi might be more suitable for graphics and multimedia related applications and products.

IV. PERFORMANCE EVALUATION

To adequately evaluate the performance of the devices listed in Table I, we decided to implement the above described security primitives in a unified framework. For Raspberry Pi, Android, and Apple Watch devices, it has been executed as a standalone Java application. To run the framework on Apple devices (iPhone 4s, iPhone 6, and Apple Watch), we have ported the logic and created a standalone application written in Objective-C programming language. To make our assessment conditions even more equivalent, we terminated all unnecessary background processes and enabled the flight mode whenever possible. To execute our application on the restricted Intel® Edison board, we followed the manual [28] to prepare a Linux build equipped with JRE. Further, an executable jar file was designed, deployed, and executed on the device.

We split all of the tested devices based on their performance metrics into two groups: Smart devices and IoT boards. As the main evaluation criteria to characterize this equipment, we have selected the security primitive execution time. This is due to the unification and well-acceptance of this approach in addition to the fact that some of the devices are hardware restricted and, therefore, could not provide any other valuable

---

1 Ubilinux stands for embedded Linux distribution based on Debian Wheezy and enables to run JVM/JDK. The targeted application domain for those platforms is embedded devices with limited memory and storage capacity; the image is currently available for Intel® Galileo Gen 1/Gen2 and Edison.
and unified evaluation metric. Further, we compared the classical cryptographic primitives and the bilinear pairing on both Intel® Edison development board and “off-the-shelf” devices available on today’s market. The following results have been obtained as an average of 1000 executions for each operation to achieve statistically-reliable data.

A. Classical Cryptographic Primitives Evaluation

First, Fig. 3 indicates the average time overhead for encryption and decryption operations of the conventional non-optimized RSA schemes with correspondence to different decimal digits. Public and Private keys were generated using OpenSSL with default parameters. Adopting a security value of 1024 or 2048 bits and default public exponent (3 bytes) (which is reasonable for the constrained wireless devices [29]), the RSA Encryption operation remains under 1 ms on a typical Android smartphone, around 2.5 ms for a Smart Watch, and less than 12 ms on Intel® Edison and Raspberry Pi 2.

Decryption time looks less optimistic and, therefore, for an Android phone it takes around 25 ms, but up to 100 ms for an iPhone. Similar behavior is observed for Android Wear and Apple Watch – here, the values are 35 ms and 200 ms correspondingly. On the IoT boards, the execution may take up to half a second, which may still be feasible for delay-tolerant applications. Concerning smart devices, we can state that Sony Watch is demonstrating high performance even though it is not classified as a standalone device. Interestingly, here and further on, iPhone 4s is sometimes showing better results than iPhone 6 or Apple Watch, which may be due to the lack of the power consumption optimization feature on the version of iOS that was introduced only starting 9.0.1. Hence, CPU utilization is able to approach 90 %, while for the latest models it remains well below 50 %.

Taking into account such basic operation as Hashing function, we evaluate the execution of SHA1 and SHA2 (SHA-256) on all of the devices. The corresponding results are summarized in Fig. 4. We can conclude that for all of our test devices SHA1 and SHA2 are hardware optimized and mainly depend on the utilized equipment. As an example of the data encryption, we used AES 128 cipher. The corresponding results still follow the execution time pattern of public-key cryptosystems and hashing functions for all of our devices.

B. Pairing-based Primitives Evaluation

Further, we present the results for bilinear pairing operations. The curve types A and D (175) are assessed utilizing jPBC-benchmark framework [24]. To this end, Fig. 5 shows one pairing operation with curve A. The most efficient device here is Intel Edison with JDK 1.8.0 that computes a single pairing operation in 580 ms. At the same time, Intel Edison needs over 8 seconds for the pairing initialization. Moreover, the modular exponentiation (EC point addition) operation takes...
about 4x more time than that on Android devices. EC multiplication follows a similar pattern. Our results confirm that some curve operations on smartphones and smart watches are more efficient than on the single-board computers (Raspberry Pi 1/2 B). Further, Fig. 6 depicts the results of the PBC operations with D curves. Interestingly, Android devices with a more powerful CPU take as much time for one pairing operation as the less powerful devices (Intel Edison, Raspberry Pi). Hence, JRE seems to be more efficient than the Android platform. Our results indicate that optimized PBC schemes with only few pairing operations (i.e. < 2), several exponentiation, and scalar multiplications can be deployed in the security layers of non-real time IoT applications that run on current smartphones and wearables.

<table>
<thead>
<tr>
<th>Device</th>
<th>Cryptographic operations</th>
</tr>
</thead>
<tbody>
<tr>
<td>Apple Watch</td>
<td>SHA 1/2; Curve operations</td>
</tr>
<tr>
<td>Sony SmartWatch 3</td>
<td>RSA 1024, 2048 E/D; Curve operations</td>
</tr>
<tr>
<td>Intel® Edison</td>
<td>RSA 2048 E/D; AES; SHA 2</td>
</tr>
<tr>
<td>Raspberry Pi 1 model B</td>
<td>RSA 1024 E/D</td>
</tr>
<tr>
<td>Raspberry Pi 2 model B</td>
<td>RSA 1024, 2048 E/D; AES; SHA 1</td>
</tr>
</tbody>
</table>

V. LESSONS LEARNED AND CONCLUSIONS

In this section, we discuss the important aspects that we faced during the implementation of our experimental framework, as well as outline our conclusions and future steps. In the process of developing the said framework, we have addressed a number of challenges regarding the very different requirements by the selected operating systems (i.e., Android, iOS, Android Wear, Apple WatchOS).
with the pairing-based solutions is such that pairings consume from several hundreds of milliseconds to few seconds on current handheld and IoT devices. We identified the most resource-consuming operation of pairing-based cryptography, that is, the bilinear pairing operation. The time necessary to compute this operation is several orders longer than that for the other operations on the elliptic curve. Therefore, in practical implementation, we highly recommend using cryptographic schemes [30, 31], in which an IoT device executes only basic operations on the curve and offloads the pairing operations to some central device with more computation power. On the other hand, some smartphone applications that send data in real-time and must secure data integrity and authenticity (e.g. for remote control systems) should avoid using the PBC schemes. These applications need to be secured by classical cryptographic primitives (SHA2, AES, RSA) that take only several milliseconds.

Finally, we can conclude that modern wearable electronics has already reached the computational power of a two-year-old smartphone and, thus, IoT world fulfills the security requirements of today. Constrained but powerful IoT devices, like Intel Edison, are designed so that the energy consumption is minimized. Due to that fact, the computational power is somewhat lowered, but this class of devices appears to be an attractive enabler for the required levels of information security. Importantly, the Raspberry Pi board, which is often nicknamed “a tiny and affordable computer” is demonstrating security. Importantly, the Raspberry Pi board, which is often nicknamed “a tiny and affordable computer” is demonstrating security.

We would like to thank to Intel Finland for the possibility to perform the research, infrastructure of the SIX Center was used. For the research, infrastructure of the SIX Center was used. We would like to thank to Intel Finland for the possibility to evaluate Intel Edison IoT development boards.

ACKNOWLEDGMENT

The described research was supported by the Academy of Finland, the Ministry of Interior under grant VI20162018003, the National Sustainability Program under grant LO1401, and the Foundation for Assistance to Small Innovative Enterprises (FASIE) within the program “UMNIK” under grant 8268GU2015 (02.12.2015).

REFERENCES

Publication III

© 2017 IEEE. Reprinted, with permission, from


In reference to IEEE copyrighted material which is used with permission in this thesis, the IEEE does not endorse any of Tampere University of Technology’s products or services. Internal or personal use of this material is permitted. If interested in reprinting/republishing IEEE copyrighted material for advertising or promotional purposes or for creating new collective works for resale or redistribution, please go to http://www.ieee.org/publications_standards/publications/rights/rights_link.html to learn how to obtain a License from RightsLink.
Facilitating the Delegation of Use for Private Devices in the Era of the Internet of Wearable Things
Aleksandr Ometov, Sergey V. Bezzateev, Joona Kannisto, Jarmo Harju, Sergey Andreev, and Yevgeni Koucheryavy

Abstract—The Internet undergoes a fundamental transformation as billions of connected “things” surround us and embed themselves into the fabric of our everyday lives. However, this is only the beginning of true convergence between the realm of humans and that of machines, which materializes with the advent of connected machines worn by humans, or wearables. The resulting shift from the Internet of Things to the Internet of Wearable Things (IoWT) brings along a truly personalized user experience by capitalizing on the rich contextual information, which wearables produce more than any other today’s technology. The abundance of personally identifiable information handled by wearables creates an unprecedented risk of its unauthorized exposure by the IoWT devices, which fuels novel privacy challenges. In this paper, after reviewing the relevant contemporary background, we propose efficient means for the delegation of use applicable to a wide variety of constrained wearable devices, so that to guarantee privacy and integrity of their data. Our efficient solutions facilitate contexts when one would like to offer their personal device for temporary use (delegate it) to another person in a secure and reliable manner. In connection to the proposed protocol suite for the delegation of use, we also review the possible attack surfaces related to advanced wearables.

Index Terms—Attack surfaces, delegation of use, Internet of Wearable Things (IoWT), personally identifiable information, privacy challenges, unauthorized exposure, wearables.

I. INTRODUCTION

THE INTERNET as we know it today has undergone a fundamental transformation over the last several decades (see Fig. 1). Back in the early 1990s, it was a fixed network of computers that allowed the first million of Internet users to communicate via e-mail. The Internet access points in people’s homes and public spaces were, in essence, connected places, which offered limited connectivity supply outnumbered by the population of potential users. This has changed as of 2000s—driven by the proliferation of mobile phones and tablets—with a possibility to connect several billion more wireless Internet users. Engaged into rich social media opportunities, these connected people did not suffer anymore from a lack of available connectivity. It is then when we started to also connect various machines, objects, and devices to the Internet infrastructure. This ongoing phenomenon, known as the Internet of Things (IoT) [1], [2], promises to add another several tens of billion or more connected items by 2020 and beyond.\footnote{Cisco visual networking index: global mobile data traffic forecast, 2016.}

Employing a plethora of wireless access technologies [3], billions of connected “things” (such as sensors, actuators, smart meters, and robots) surround us and embed themselves into the fabric of our everyday lives. However, this is only the beginning of true convergence between the realm of humans and that of machines. Beyond that, an exciting innovation develops that promises to revolutionize our society thus opening a new Internet era. Connected machines worn by humans, or wearables, produce countless opportunities for their users, by helping them manage their personal lives,
health, and safety. The rapid advent of wearables, with global sales already exceeding 20 million per quarter according to the International Data Corporation, brings along an avalanche of personal devices with new feature sets and functionalities that can be worn on a person. As worldwide wearables market soars, we are standing on the brink of another decisive Internet transformation—from the IoT to the Internet of Wearable Things (IoWT).

While today’s first-generation wearables are still rather limited in what they can do, the emerging IoWT devices promise to deliver a truly personalized user experience by capitalizing on the rich contextual information [4]. Complementing contemporary smart watches, fitness trackers, wristbands, on-body cameras, and eyewear, future wearable technology comprises innovative textiles, smart clothes, augmented and virtual reality gear, as well as enterprise wearable equipment. Early adopters of the next-generation wearables are envisioned to focus on self-quantification, and in fact, a recent survey by Ericsson revealed that over 70% of respondents had the same level of interest in self-quantification as in wearables.\(^2\) Obtaining the individual’s health and wellness information is now increasingly simple with a wide variety of dedicated wearables, from heart rate monitoring rings, digital health networks, and posture sensors, to commuting ecometers, clean air bracelets, water quality checkers, and city microclimate monitors. All in all, modern IoWT technology already provides a range of useful functions, features, and services to its users, from simple fitness tracking to smartphone-like experience.

However, despite their promising potential, wearable devices have inherent constraints and limitations. First, owing to their slim form-factors, power efficiency is more important to wearables than to any other product. Second, the very high numbers of interconnected body-worn devices and the resultant personal user networks give rise to system scalability issues [5]. Third, it is still not common for a wearable device to interact with any nearby devices since they are operated in various platforms especially when devices are manufactured by different vendors. Indeed, wearables have the capability to connect and communicate to the IoWT infrastructure either directly through embedded cellular connectivity or via another device, primarily a smartphone, by using short-range wireless technology (see Fig. 2). Here, the relevant contextual information may be stored and processed on the device locally or forwarded via a gateway (i.e., the user’s smartphone) to a remote IoWT server. In the latter case, the gateway may not have a continuous (reliable) connection to the network due to obstacles, difficult propagation conditions (in tunnels, lifts, etc.), and unpredictable user mobility.

The above wearable-specific constraints—and primarily their limited computation power and intermittent network connectivity—accentuate the need to rethink the conventional approaches to maintaining data security, integrity, and reliability [6]. This is further aggravated by the fact that the information that wearable devices are targeting to store and process is highly sensitive, while the devices themselves are naturally more exposed to public compared to the handheld user equipment. Indeed, with the increasing adoption of advanced wearables, we may end up “wearing” some of the most personal aspects of ourselves, including our conversations, relationships, and even health. To this end, wearables uniquely become both the most private and the most public devices, and protecting the personal user information they handle becomes a growing concern. This is particularly true for any medical data and those likely to adopt wellness services early on value the integrity of that information more than others.

Given that wearables sense, process, and transmit data about their users, they generate more personally identifiable information than any other today’s technology. That data includes, but is not limited to wearer’s location, activity, movement, and vital signs. Therefore, the biggest security risk associated with wearables becomes the unauthorized exposure of the personally identifiable information associated with them. According to [7], the information privacy is guaranteed if “the data can only be accessed by the people who have authorization to view and use it.” Presently, a traditional enabler to achieve data privacy is secure authentication [8]. However, such approaches are complex to apply in large-scale distributed scenarios [9], especially when storing and sending the sensitive data occurs across a heterogeneous environment [10] not only via direct connections, but also remotely.

More broadly, privacy involves control over one’s data, which incorporates privacy by design and contractual privacy that, in turn, implies trust. Secure authentication naturally touches upon the notion of trust [11], that is, ensures that the communications partners are actually who they claim to be. This concept of trust, together with the respective tools to manage it, have also been evolving alongside with the transformation of the Internet shown in Fig. 1. Initially, while the Internet mostly consisted of the terminal computers, these were able to authenticate their users with local accounts. Since then, the Internet has grown to connect people with one another by taking advantage of centralized remote services. Most recently,
the IoT and thus the IoWT promise to connect our surroundings with each other and with us. The accompanying information security protocols have also traveled a long path to reach the current state of their evolution.

1) In the early Internet era, access to the desired resources was granted to a particular computer/human based on the corresponding authentication procedure. In contrast, now each device has to complete such a procedure and, additionally, prove its association with its owner so that the data privacy could be guaranteed.

2) Early Internet was based on the assumption that access to the target resource could only be granted based on the resource owner decision, that is, by providing a certificate or a password to the end user. Today, there is a need to provide such a certificate for each connected device explicitly by its owner, and the private data is often stored distribute across several devices.

In light of the above transformative changes, this paper targets to propose efficient means for the delegation of use [12] applicable to a variety of constrained wearable devices, so that to guarantee privacy and integrity of their data. Accordingly, since most wearables are inherently limited, our specific solutions are mindful of their restricted computation and communication budget. In particular, we target to facilitate temporary exchange of the IoWT devices belonging to different persons, groups, organizations, or companies in a secure and reliable manner to protect the contextual and personalized information they handle. The rest of this paper is structured as follows. Section II surveys the related work in the target area and establishes that a comprehensive solution for the delegation of use is not yet available in the existing literature. To this end, Section III proposes a novel protocol suite to facilitate such delegation in various contexts, while Section IV details the actual protocols that comprise it. In connection to that, Section V reviews the available attack surfaces related to wearables in general and the proposed solutions in particular. Section VI concludes this paper with useful numerical results and an accompanying discussion.

II. STATE-OF-THE-ART AND RELATED WORK OVERVIEW

The existing literature is still rather scarce on the topic of the delegation of use for resource-constrained devices and services. Most of the available papers focus primarily on the challenges related to authentication between the user and an unfamiliar service/data, while having a connection to the trusted authority. Other works propose information security primitives to solve the task at hand, but do not offer effective protocols employing the out-of-the-box structures on the constrained devices. In what follows, we summarize our literature review moving all the way down the protocol stack and then toward more conceptual approaches.

A broad overview on security-centric challenges in IP-based networks for the IoT could be found in [13]. Here, the authors survey the key IoT-specific architecture and network deployment issues by focusing on a superdense IoT scenario, as well as address the technical limitations of the conventional protocols. The main conclusion is that IPv6 has the potential to solve the identification and transport challenges, thus facilitating communication between the devices, but somewhat downgrading user’s privacy. The paper in question also introduces an automation control center, which acts as a trusted authority and network assistance unit by monitoring the lifecycle of the involved IoT devices. Finally, the authors speculate on the pros and cons behind distributed versus centralized architectures and the corresponding systems operation.

Another study in [14] concerns IP-based scenarios for the IoT devices with a particular emphasis on the offloading of the delegation-related traffic to the remote server in the cloud (assuming its uninterrupted availability) [15]. Here, the initial connection initialization procedure is separated from the data protection itself by utilizing DTLS protocol [16]. This is in addition to the usage of a public handshake while establishing a connection between the devices. Further, Hummen et al. [17] reduced the protocol operation overheads by offloading the handshake procedure to a more powerful device and thus arrived at the protocol that is applicable for resource-constrained devices. Complementing this, Seitz et al. [18] discussed a framework that enables simple authorization and access control procedures for resource-constrained equipment. The main focus of said paper is to evaluate the impact of using the public key infrastructure (PKI) cryptography on the RAM and ROM utilization. The authors claim that their approach is suitable for communication between the application server and the constrained IP-based end device.

With regards to the access controls schemes, the work in [19] surveys distributed privacy-preserving access arbitration mechanisms for sensor networks. The respective protocol implementation assumes that the users have to be provided with tokens by the device owner (e.g., factory) in order to access the needed data from a device. The main focus of this research is on protecting the privacy of a user toward the device and hence preventing from the reuse of specific tokens. Another access rights delegation platform in [20] represents a complete framework based on the premise that the users in the IoT use cases are allowed to manage the access control system to their services and information, therefore contributing an authorization model employing the capability-based security [21]. The result in question is suitable for anonymous services using the individual’s token to access any of the owned information. Similar approaches could be found in [22] and [23].

The work in [24] offers a set of cryptographic primitives for the PKI-based encryption taking advantage of the time-release cryptography. The authors elaborate on a solution that allows to encrypt a message in such a way that the receiver cannot decrypt the ciphertext until a certain target time in the future. Correspondingly, privacy of the user can be maintained. Then, the paper in [7] considers privacy in the medical body area networks, from the viewpoint of the distributed data storage utilization. This research outlines an important set of requirements for distributed data storage systems as well as reviews possible attacks on the body area networks. Hence, the discussed publication reiterates on the need for fine-grained data access control that follows the concept of preset rights management, but relies on a role-based model [25].
Finally, Morchon and Baldus [26] as well as Garcia-Morchon and Wehrle [27] detailed the secret key cryptography-based schemes capable of solving a distributed access control task in wireless medical sensor networks. The proposed approach utilizes a Blundo’s key predistribution method to support the role-based access control. Owing to the pregenerated and distributed polynomial key shares, the user can easily establish a pairwise key with any authorized entity, and then encrypt a copy of sensitive data utilizing the corresponding key for the target entity. Even though patients can exert individual control over the exact access rights of the communicating entities, they would need to know the actual set of authorized users when distributing the data, and thus encrypt one copy for each user in the set, which is hardly practical.

In summary, we establish that the challenge of the delegation of use remains a sound research problem for already more than a decade. A lack of corresponding procedures for wearable devices in current academic research is profound, with only a handful of primitives and a few generic solutions. At the same time, the market predictions reviewed in the previous section and the use cases discussed in what follows corroborate the prompt need for having efficient enablers to facilitate the delegation of use for wearable devices. We bridge the indicated gap in the rest of this paper by outlining our own comprehensive protocol suite to support such operation.

III. PROTOCOL SUITE FOR THE DELEGATION OF USE

In this section, we begin with discussing the attractive practical scenarios for the application of our proposed protocol suite followed by its general description as well as the relevant underlying assumptions.

A. Use Cases and Market Overview

Today, there are two highly contrasting opportunities in case one would like to offer their personal device for temporary use (that is, delegate it) to another person. First, there is a formal process requiring interaction with, e.g., a notary officer often followed by expensive, cumbersome, and time consuming paperwork. However, in this case the owner is guaranteed that the concerned device is delegated according to the word of law and will be returned after use. Second, a more widespread case is when one is willing to lend a device to a familiar person, but without any confirmed guarantee that it will be returned except for natural human trust. In this paper, we propose and advocate for a novel solution that extends the notion of “casual” delegation of use (case 2) to offer certain guarantees on the device return (similar to case 1).

In fact, a recent survey established that over a half of those who buy a wearable will stop utilizing it after only six months [28]. In connection to this, there already exist companies offering more advanced IoWT devices for a “try-before-buy” period. For example, Lumoid introduces this opportunity: for $20, anyone can try out as many as five different wearables for seven days. By the end of the trial period, customers can either buy the wearables of their choosing, or return all of them to the company. While a week is not particularly long of a trial period, it could in principle be extended for as long as there is no business need for the wearables being tested by their current users.

Another company, named ByeBuy, adopts a “pay-as-you-go,” on-demand model for the gadgets they offer, which effectively means that the user does not actually need to purchase the latest tech products. Available first to the customers in Germany and the U.K., the initial lineup of available high-end products for rental includes the Xbox One, Apple Watch, and the Parrot Bebop Drone. Interestingly, ByeBuy management maintains that there will be neither up-front payments nor minimum contract periods in their business model.

To this end, we see that the IoWT market is just at the beginning of a long journey, with a rapidly growing list of possible scenarios for (sub-)renting high-end wearable devices by the owner to the temporary user. Hence, security, privacy, and user experience aspects in this new type of context have to be carefully evaluated and Table I gives a quick overview of the candidate use cases that are both attractive and challenging for future IoWT rental business. Summarizing these, the pay-as-you-go model may soon take off rapidly in the wearables market. The bigger picture behind this thinking is that many Americans already prefer to access information and things through Netflix, Spotify, Uber, and other means rather than actually own them [29]. In this regard, we believe that many more objects and services may eventually adopt the flexible all-subscription model.

### Table I

<table>
<thead>
<tr>
<th>Scenario</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>Golf Club</td>
<td>Renting smart golf equipment, such as cart, swing, glasses, etc. that are fully customizable for their temporary owner and may adjust to the personal parameters.</td>
</tr>
<tr>
<td>Scuba Diving</td>
<td>Smart wetsuit computers, cameras, and spear fishing gun may be rented directly on the boat.</td>
</tr>
<tr>
<td>Skiing</td>
<td>Renting smart skis, boot sensors, body armor, augmented reality glasses, etc. while being on a distant resort.</td>
</tr>
<tr>
<td>In-flight</td>
<td>Providing a virtual reality headset for on-board customers, both naval and airborne, potentially with connectivity to the Internet.</td>
</tr>
<tr>
<td>Keyless Remote Access</td>
<td>Receiver or provide access to the door merely by being in its proximity even without the Internet connection for a customer, medical staff, or a police officer.</td>
</tr>
</tbody>
</table>


Despite bringing forth more flexible usage models, all of the device renting companies in our survey utilize conventional notary-like solutions when offering temporary access to wearables. Moreover, a user may have difficulty to receive timely digital support in situations when the Internet connection to the company servers is not available. From the communications perspective, remote connectivity with the rented IoWT device could be arranged via a gateway (user’s smartphone) whenever possible. In cases of a guaranteed stable connection to the owner, simpler solutions including secure time and/or hash chains may become useful to control the devices [30]. However, these may be challenged to provide all of the desired functionality for advanced devices that require dynamic feedback (e.g., for policy updates as well as to extend the lease time on-demand). This problem becomes particularly involved when the rented wearables do not have a reliable Internet connection to the owner while only maintaining access to the gateway over a direct link outside of the cellular network coverage.

Our novel protocol suite detailed below offers this much needed functionality.

B. General Description of This Paper

Whenever a person or a company is willing to provide the use of a device to a “trusted” or known person, the respective solution is rather straightforward. However, there is no confirmed guarantee that the device in question will be returned on time. Our proposed solution employs a trusted authority that is involved into the process of lending the device and thus can provide guarantees on its successful return. In particular, said authority may be made responsible for controlling the duration of the temporary device delegation as well as for the corresponding interactions between the owner and a temporary user.

More specifically, we assume that people and their personal wearable devices proceed through the initialization phase while connected to the trusted authority. Further, they may invoke the actual delegation phase at a later time, even without a reliable connection to the authority, which is especially convenient in cases of intermittent or unavailable Internet connectivity. The developed model and the corresponding set of protocols (named here the protocol suite) are specifically designed in such a way that they can accommodate most of the constrained wearable devices without imposing significant computation or transmission overheads.

Further in this paper, we concentrate on the following system structure, where the overall IoWT system may comprise the distributed data storage in the cloud, the local wireless networks for communication with remote servers, and the personal IoT networks of individual users (see Fig. 2). A personal user network within the IoWT consists of the following components.

1) The primary data aggregation gateway represented by, e.g., the user’s smartphone (or a smart gateway in home networks, etc. [31])—the most essential required features of such a gateway are superior to wearables computation power and energy resource.

2) The actual constrained IoWT wearables that have less resources than the gateway.

3) Various other devices that can store, process, and transfer data, but are neither a part of the personal IoWT network nor that of the remote cloud.

In this paper, we also assume that all the wireless communications channels are secure, and thus the aspects related to the corresponding well-known attacks on them are not discussed. Further, the proposed protocols could be instantiated with the specific cryptographic primitives (encryption, hashing functions, signatures, etc.) according to the effective system specifications and based on certain target requirements, as well as the particular IoWT devices in question and their limitations. For instance, as a hashing function we may utilize any of the existing alternatives [32]: SHA-2, SHA-3, BLAKE2 [33], etc. For the certificate authority operation (in the PKI case), we may use the conventional primitives, such as: 1) RSA (factorization) [34] and 2) ElGamal/Diffie-Hellman [35] or elliptic curve cryptography [36].

Given our assumption on secure communications medium, it is possible to utilize the classic Diffie-Hellman [37] or elliptic curve Diffie-Hellman [38] protocols. This is because using the PKI-based solutions for gateway-to-wearable connections is computationally-hungry and hence should be avoided. As a widely-used contemporary alternative, we may employ symmetric solutions, where additional (e.g., visual) channels are utilized to establish a secure link. In this case, the required entropy (128 or 256 bits) needs either a QR code or a shorter symmetric token matched with a password-authenticated key exchange utilizing asymmetric cryptography [39].

The issues related to the actual delegation rules, including environment, biometry, positioning, etc., are not considered in this paper either. Therefore, the main focus in what follows remains on the composition and operation of the user’s personal network, that is, data aggregation gateway and the associated wearable devices. The main goals of our protocol suite are to provide continuous possibility for: 1) authentication between the users and their wearable devices; 2) software and/or hardware integrity; and 3) data security.

C. Protocol Suite Assumptions and Composition

We further assume that the IoWT network features a certificate authority employing trusted relations in accordance with the “trusted tree” principles. Every user gateway (i.e., smartphone) has a pregenerated secret key SKA and a certificate signcloud(IDA, PKA) on its public key received from the IoWT certificate authority in the cloud. Here, signcloud is obtained by using any appropriate cryptographic signature primitives with the secret key SKCA of the IoWT certificate authorities. Each gateway has a certificate certcloud = PKCA obtained from the IoWT certificate authority, while each wearable device (wi) has a unique hardware-locked serial number (IDi) and a factory-preset PIN (can be changed manually by using

TABLE II
KEY CONSTRUCTS UTILIZED BY THIS PAPER

<table>
<thead>
<tr>
<th>Construct</th>
<th>Container</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>$A, B, \text{cloud}$</td>
<td>–</td>
<td>Names of the cooperating parties: Alice, Bob, and Cloud (IoWT network).</td>
</tr>
<tr>
<td>$u_i$</td>
<td>–</td>
<td>$i^{th}$ wearable device.</td>
</tr>
<tr>
<td>$SK_A, PK_A$</td>
<td>–</td>
<td>Secret and public keys of user Alice.</td>
</tr>
<tr>
<td>$\text{sign}_{\text{cloud}}(PK_A)$</td>
<td>–</td>
<td>Here, $PK_A$ was signed by the root cloud certificate.</td>
</tr>
<tr>
<td>$t_f, t_d$</td>
<td>–</td>
<td>Delegation and reset timers.</td>
</tr>
<tr>
<td>$S_A$</td>
<td>–</td>
<td>Secret key generated by user Alice to communicate with her wearable device.</td>
</tr>
<tr>
<td>$\text{hash} (SW_i)$</td>
<td>–</td>
<td>A cryptographic hash extracted from the wearable device software by the user</td>
</tr>
<tr>
<td>$\text{cert}_{\text{cloud}}$</td>
<td>$\text{sign}_{\text{cloud}}(u_i, PK_A, /D_A, \text{hash} (SW_i))$</td>
<td>Certificate generated by the cloud for data integrity reasons.</td>
</tr>
<tr>
<td>$\text{cert}_A$</td>
<td>$\text{sign}<em>A(\text{cert}</em>{\text{cloud}})$</td>
<td>User envelope to be used in the wearable certificate storage.</td>
</tr>
<tr>
<td>$m(D_A)$</td>
<td>$\text{sign}_A(u_i, t_f, D_A, \text{ID}_B, {\text{delegation rules}})$</td>
<td>An “initialize delegation” message sent to the wearable device by the owner.</td>
</tr>
<tr>
<td>$m(R_A)$</td>
<td>$\text{sign}_A(u_i, R)$</td>
<td>Envelope verified by the certificate authority.</td>
</tr>
<tr>
<td>$m(C[S_A])_A$</td>
<td>$\text{sign}_A(u_i, C[S_A])$</td>
<td>A “secret removal” message sent to the wearable device by the current user.</td>
</tr>
</tbody>
</table>

Further, every “out-of-the-box” wearable device $w_i$ already has the necessary factory software preinstalled. At a later time, the current state of this software can be reset back to the “factory default” state, that is, the trusted image provided by the manufacturer [40]. The communication between $w_i$ and the gateway is carried out over a secure channel. As mentioned above, we assume that network connectivity is already protected against any possible malicious or “person-in-the-middle” attacks. The gateway, in turn, has a pregenerated $SK_A$ and a certificate $\text{sign}_{\text{cloud}}(PK_A)$ on its public key received from the IoWT certificate authority. Finally, each user additionally has the IoWT authority certificate $\text{cert}_{\text{cloud}}$ to verify the transmitted data as well as the validity of the devices. In case of a lost or stolen device, the user may setup a reset timer $\text{sign}_A(t_f)$, which is also assumed secure.

In summary, we provide a complete list of constructs employed for the composition of our proposed protocol suite in Table II. We continue in the following section with a detailed description of the protocols comprising this suite.

IV. PROTOCOL DESCRIPTIONS WITHIN PROPOSED SUITE

In this section, we offer a detailed description of the individual protocols comprising the proposed suite, which has been introduced in the previous section. To this end, Table III outlines the state machine corresponding to our solution from the viewpoint of the wearable device to be delegated. Here, a user is represented as a personal network with a data aggregation gateway (smartphone). Names Alice and Bob refer to the two users. The main phases of operation to be discussed further on are presented in Fig. 3. Please note that numbering of the protocol iterations is according to the algorithm description and may be not in incremental order.

A. Association (State $1 \rightarrow State 2$)

Here, Alice purchases a completely new wearable device from the manufacturer and is willing to add it to her personal IoWT network. In other words, we describe the procedure of adding a wearable device $w_i$ to the personal network of the owner Alice. As the device belongs to Alice, it is associated with her by utilizing the unique ID (alice@address.com) with the assistance from the application center in the IoWT cloud. The key steps of the proposed association protocol are summarized in Fig. 4 and Algorithm 1. In practice, this construction may take advantage of already existing transport layer security primitives [41], [42].

B. Delegation (State $2 \rightarrow State 3$)

Here, Alice is willing to lend her wearable device to Bob for some time, that is, the device owner is delegating a wearable device to another temporary user. Importantly, we differentiate between two main scenarios: 1) when both Alice and Bob have a reliable wireless connection to the

![Fig. 3. Example wearable device lifecycle while delegating the use.](image)
IoWT cloud and 2) when at least one of them does not have it. Conveniently, our delegation procedure may in principle be executed even in situations when Alice and Bob are not geographically close to each other (in case of the door lock access delegation, for example). The key steps of the proposed delegation protocol are summarized in Fig. 5 and Algorithms 2 and 3.

1) Both Alice and Bob Have Reliable Network Connection: This scenario requires that the gateway has a reliable wireless connection to the certificate authority, so that it could validate all the involved operational procedures.

Algorithm 1 Wearable Device Association Protocol

1: Alice generates $S_A$ for the wearable $w_i$ and sends it to $w_i$ securely
2: $w_i$ sends the hash of the factory software to the cloud ($\text{hash}(\text{SW})$)
3: Alice also sends her PK$_A$ and ID$_A$ to the cloud
4: Cloud generates the certificate $\text{cert}_{\text{cloud}} = \text{sign}_{\text{cloud}}(\text{PK}_A, \text{ID}_A, \text{hash}(\text{SW}))$
5: Cloud sends $\text{cert}_{\text{cloud}}$ to Alice
6: Alice signs $\text{cert}_{\text{cloud}}$ and obtains $\text{cert}_A = \text{sign}_A(\text{cert}_{\text{cloud}})$
7: Alice sends $\text{cert}_A$ to $w_i$

Algorithm 2 Wearable Device Delegation Protocol: Reliable Connection to the Cloud

1: Alice sets delegation timer $t_d$ on $w_i$ using a message $m[D]_A = \text{sign}_A(w_i, t_d, \text{ID}_A, \text{ID}_B, \text{delegation rules}).$
2: Alice sends $m[D]_A$ to the cloud.
3: Cloud checks the validity of $m[D]_A$ by using PK$_A$. If it is not valid $\rightarrow$ exit.
4: Cloud signs $m[D]_{\text{cloud}} = \text{sign}_{\text{cloud}}(m[D]_A)$.
5: Cloud sends $m[D]_{\text{cloud}}$ and $\text{cert}_A$ to Bob.
6: Alice deletes $S_A$ on $w_i$ using $m[C(S_A)]_A$.
7: if Bob does not trust Alice then
8: Device is reset keeping the original certificate stored and Bob checks the hash($\text{SW}$) from the $\text{cert}_A$ and hash calculated from the $w_i$-th software directly. If both are equal – we may proceed; otherwise, the $w_i$ is considered malicious $\rightarrow$ exit. In this case, $w_i$ may not be used by Bob (factory software was modified by the owner, i.e., it is not the same as the default). Importantly, resetting to factory defaults in this case keeps the certificate storage and the trusted timer unchanged.
9: else
10: All the applications are kept unchanged and Bob may use the software of user Alice that is free or has been previously owned by Bob.
11: end if
12: Bob generates new $S_B$ for the $w_i$.
13: Bob sends $S_B$ to $w_i$ securely.
14: To ensure software integrity, Bob signs $\text{sign}_B(w_i, \text{SW})$.
15: if the delegation time is expired then
16: Device is reset to factory default state saving the original certificate. The timer can be reset while connected to the cloud over Bob’s gateway, but it requires interaction with the original owner Alice as $m[D]_A = \text{sign}_A(w_i, t_d, \text{ID}_A, \text{ID}_B, \text{delegation rules}).$ This could also be done via a direct connection.
17: end if

2) Both Alice and Bob Do Not Have Reliable Network Connection: This scenario does not require that the gateway has a reliable wireless connection to the certificate authority (in/on tunnels, boats, mountains, etc.). Alternatively, the user(s) may decide to block their wireless connection intentionally.

C. Reclaiming (State 3 $\rightarrow$ State 2)

Here, the temporary user Bob returns the previously rented wearable device to its original owner, Alice. The key steps of the proposed reclaiming protocol are summarized in Fig. 6 and Algorithms 4 and 5.

1) Both Alice and Bob Have Reliable Network Connection: See Algorithm 4 for details.

2) Both Alice and Bob Do Not Have Reliable Network Connection: See Algorithm 5 for details.

D. De-Association (State 2 or 3 $\rightarrow$ State 1)

1) Manual De-Association (Disposal or Sale): Here, the owner Alice is willing to sell or dispose of her wearable
Fig. 6. Wearable device reclaiming protocol. (a) Reliable connection to the cloud. (b) No reliable connection to the cloud.

Algorithm 3 Wearable Device Delegation Protocol: No Reliable Connection to the Cloud

1: Alice sets delegation timer \( t_d \) on \( w_i \) using a message \( m[D_A] = sign_A(w_i, t_d, ID_A, ID_B, \{delegation\ rules\}) \).
2: Alice sends \( cert_A, m[D_A] \) to Bob securely.
3: Bob checks if \( cert_A \) and \( m[D_A] \) are valid by \( cert_{cloud} \).
4: Alice deletes \( S_A \) on \( w_i \) using \( m(C(S_A)) \).
5: if Bob does not trust Alice then
6: Device is reset keeping the original certificate stored and Bob checks the hash\( (SW_i) \) from the \( cert_A \) and hash calculated from the \( w_i \)-th software directly. If both are equal – we may proceed; otherwise, the \( w_i \) is considered malicious \( \rightarrow \) exit. In this case, \( w_i \) may not be used by Bob (factory software modified by the owner, i.e., it is not the same as the default).
7: else
8: All the applications are kept unchanged and Bob may use the software of user Alice that is free or has been previously owned by Bob.
9: end if
10: Bob generates new \( S_B \) for the \( w_i \).
11: Bob sends \( S_B \) to the \( w_i \) securely.
12: To ensure software integrity, Bob signs \( sign_B(w_i, SW_i) \).
13: if the delegation time is expired then
14: Device is reset to factory default state saving the original certificate. The timer can be reset while connected to the cloud over Bob’s gateway, but it requires interaction with the original owner Alice as \( m[D_A] = sign_A(w_i, t_d, ID_A, ID_B, \{delegation\ rules\}) \). This could also be done via a direct connection.
15: end if

device, that is, she wants to remove all the personal data from the device including any keys and certificates. The main steps of the corresponding de-association protocol are summarized in Algorithm 6.

Algorithm 4 Wearable Device Reclaiming Protocol: Reliable Connection to the Cloud

1: Bob generates a message \( m[R]_B = sign_B(w_i, R) \).
2: Bob sends \( m[R]_B \) to the cloud.
3: Cloud checks the validity of \( m[R]_B \) by using \( PK_B \). If it is not valid \( \rightarrow \) exit.
4: Cloud signs \( m_{cloud} = sign_{cloud}(m[R]_B) \).
5: Cloud sends \( m_{cloud} \) to Alice.
6: Bob deletes \( S_B \) on \( w_i \) using \( m(C(S_B)) \).
7: if Alice does not trust Bob then
8: Device is reset keeping the original certificate stored and Alice checks the hash\( (SW_i) \) from the \( cert_A \) and hash calculated from the \( w_i \)-th software directly. If both are equal – we may proceed; otherwise, the \( w_i \) is considered malicious \( \rightarrow \) exit. The owner Alice should reset her device using the factory PIN.
9: else
10: All the applications are kept unchanged and Alice may use new software of the previous user Bob which is free or has been purchased by Alice while Bob was using the device.
11: end if
12: Alice sends generated during the association \( S_A \) to \( w_i \).
13: To ensure software integrity, Alice signs \( sign_A(w_i, SW_i) \).

As a result, now \( w_i \) has only \( cert_A, cert_{cloud} \).

Algorithm 5 Wearable Device Reclaiming Protocol: No Reliable Connection to the Cloud

1: Bob generates a message \( m[R]_B = sign_B(w_i, R) \).
2: Bob sends \( m[R]_B \) to Alice over a direct link.
3: Bob deletes \( S_B \) on \( w_i \) using \( m(C(S_B)) \).
4: Alice checks if \( m[R]_B \) is valid by \( cert_{cloud} \).
5: if Alice does not trust Bob then
6: Device is reset keeping the original certificate stored and Alice checks the hash\( (SW_i) \) from the \( cert_A \) and hash calculated from the \( w_i \)-th software directly. If both are equal – we may proceed; otherwise, the \( w_i \) is considered malicious \( \rightarrow \) exit. The owner Alice should reset her device using the factory PIN.
7: else
8: All the applications are kept unchanged and Alice may use new software of the previous user Bob which is free or has been purchased by Alice while Bob was using the device.
9: end if
10: Alice sends generated during the association \( S_A \) to \( w_i \).
11: To ensure software integrity, Alice signs \( sign_A(w_i, SW_i) \).

As a result, now \( w_i \) has only \( cert_A, cert_{cloud} \).

2) Automatic De-Association (Loss or Damage): Here, we consider the situation when the wearable device in question is lost, damaged, or stolen, that is, any private data should be removed to prevent a potential third party from accessing it. The main steps of the corresponding de-association protocol are summarized in Algorithm 7. Note that this construction is similar to the case of manual de-association above, but device reset in this case is triggered based on the preset timer value.
Algorithm 6 Manual Wearable Device De-Association Protocol
1: Owner Alice sends a signaling message to \( w_i \): \( m[F|A] \).
2: \( w_i \) is reset to the factory defaults thus removing all data, including the certificate storage.
3: Device can be restored by only using factory (or modified) PIN, and the connection to the cloud is required according to the association phase.

Algorithm 7 Automatic Wearable Device De-Association Protocol
1: If \( w_i \) leaves the personal network coverage of its current user, the timer \( t_f \) is initialized.
2: If reset timer \( t_f \) expires, \( w_i \) is automatically reset to the factory defaults thus removing all data, including the certificate storage.
3: Device can be restored by only using factory (or modified) PIN, and the connection to the cloud is required according to the association phase.

V. POSSIBLE ATTACKS ON WEARABLE DEVICES

As a further evolution of the IoT, the IoWT and its wearable devices are susceptible to similar threats as the machine-type equipment, which served an attractive target for "hackers" for decades [43]. Contrary to the IoT devices, as we discussed in Section I, wearables are additionally vulnerable to unauthorized exposure of the personally identifiable information associated with them. Therefore, attackers could be after the physical assets of the users (i.e., the wearable devices themselves) or they could attempt to access the user’s data directly on a wearable device. In addition, an attacker could be interested in the metadata about the user, which would mean, for example, any information about past device delegations.

According to the USA Federal Trade Commission, a comprehensive classification of the attack surfaces for wearables is illustrated in Fig. 7. Hence, we learn that the conventional attack areas are somewhere between the gateway and the network cloud. These are well researched upon already, whereas wearable-specific attacks call for a more detailed discussion. In the rest of this paper, we review possible wearable-specific attacks and compare those against the existing alternatives. This information should help protect the actual instantiations of the proposed protocol suite with the practical primitives, when implemented.

Fig. 7. Classification of the attack surfaces for wearables (according to the U.S. Federal Trade Commission).

A. Privacy

Protocols that employ signatures, including the one proposed above, are particularly vulnerable in terms of privacy, since they typically also enable nonrepudiation. This important property means that the user cannot at a later time deny the fact of the delegation or assertion. More specifically, noninteractive protocols rely on this property for their security, which causes a conflict between the security and the privacy [44].

B. Phishing

Phishing attacks target to exploit the weak bindings between the digital and the physical identities [45]. For example, Eve masquerading as Bob initiates a delegation from Alice to Bob, but then presents her own identity. If Alice cannot verify that Bob is ID_A instead of ID_E, a phishing attack succeeds. Opportunities for phishing are aggravated by the intrinsic properties of wearables, including the one that they often have small or no displays. Phishing cannot usually be prevented completely (residual error and finite user effort), but it can be controlled to a desired extent (i.e., how small differences in authenticity a human user has to notice). Finally, resistance to phishing may also be in contradiction with privacy, i.e., more attributes make users more recognizable, but leak information about them.

C. Relay Attacks

It also includes the conventional person-in-the-middle attacks [46]. Here, Eve asks \( m[D|A] \) for Bob (ID_B) from Alice, and later on introduces herself as Alice to Bob, also offering the \( m[D|A] \) to Bob at that time. Alice cannot use the wearable device herself, but can observe delegations, and may convince Bob to believe that she is in fact ID_A.

D. Downgrade

As actually employed signature primitives are not discussed as part of the proposed protocol, the general problem of "downgrade" concerns mostly the key distribution stage [47].
TABLE IV
POWER CONSUMPTION OF DIFFERENT RADIO INTERFACES

<table>
<thead>
<tr>
<th></th>
<th>WiFi</th>
<th>BLE</th>
<th>ZigBee</th>
</tr>
</thead>
<tbody>
<tr>
<td>Consumption (mW)</td>
<td>220</td>
<td>147</td>
<td>71402</td>
</tr>
</tbody>
</table>

Fig. 8. Relative transmit power consumption of the proposed protocol suite.

Accordingly, if a user has multiple public keys, they all need to withstand prolonged attacks against them. Another less severe downgrade attack happens when communication with the cloud is prevented by a malicious party, or reachability of the cloud is not verified by one of the parties in advance.

E. Malicious Wearable

After observing a valid protocol message $m[D]_A$ for the wearable device $w_k$ from Alice to Bob, Eve crafts a malicious wearable device that reports the identity $w_k$ and the integrity hash($SW_k$). Then, the wearable in question can, for example, log Bob’s activity. This attack looks similar to any malicious device attack, but—due to the fact that most wearables are constrained devices—can be performed mostly on the factory side.

F. Wearable Device Compromising

The devices in a personal user network are subjected to compromising [48], as they are relatively easy to be lost, stolen, or forgotten. If the entire piece of sensitive data is directly encrypted and stored inside a wearable device together with its encryption key, the compromise of this device will lead to the disclosure of data.

G. Network Dynamics Threats

Naturally, a user operating the aggregation gateway (smartphone) along with the personal wearable devices is mobile throughout the day. Due to accidental failures or malicious activities, wearable devices may join or leave the network frequently [49]. This may also happen due to the battery constraints. To this end, attackers may attempt to place fake sensors in order to masquerade the authentic devices, and can then acquire legitimate devices deliberately. The important user-related data, if not well-kept in more than one device, could be lost accordingly as a result of high network dynamics.

VI. Conclusion

As we discover in the previous section, one of the likely attacks on the proposed wearable-specific device delegation protocol is phishing, where Eve masquerades herself as Bob. If Alice does not trust Bob’s certificate issued by the IoWT certificate authority (or Eve’s certificate in case of an attack), we may utilize the following procedure. Accordingly, Alice sends a symmetric delegation key to Bob encrypted with Bob’s public key. The delegation key for Bob can be, for instance, challenge||KDF($k_A$, challenge), which the wearable device can verify during Bob’s communication attempt. Then, the wearable device does not have to employ public key cryptography to associate the user. Here, the challenge has to have structure, which binds it to the actual delegation. Also, it is desirable to change the key $S_A : w_i$, which is the symmetric key between Alice and the device $w_i$.

Further, we assess the power consumption performance of our proposed protocol suite, as this should become a major limiting factor in its ultimate practical operation. This discussion is not presented in absolute numbers due to the fact that the transmission overheads depend on the practical networking scenario, the interference picture, and other unpredictable factors. Therefore, we analyze the case where network conditions remain similar for all the underlying wireless technologies. More specifically, the power consumption figures for the cellular interface are taken from [50]. For the power consumption of short-range wireless technologies, we refer to [51]–[53]. Based on the obtained numerical results, we estimate the transmission overheads when using our proposed protocol suite for different phases, while having equal data packet payloads.

In Fig. 8, the comparison of relative communication overheads for both in- and out-of-coverage cases is presented, whereas the calculations are based on Table IV. We learn that the association and the delegation phases of the proposed protocol suite consume the most power, as they generally involve more signaling messages to travel between a wearable device and the network. At the same time, the reclaiming phase is relatively more lightweight. In addition, we observe that running the protocols over short-range WiFi radios consumes more power than executing them over less power-hungry Bluetooth low energy (BLE) and ZigBee technologies.

In summary, this paper has comprehensively outlined a number of important aspects related to privacy of advanced wearables within the IoWT ecosystem that they construct. To this end, we started with a thorough review of contemporary trends behind the evolution of next-generation wearables, surveyed the corresponding security research background, reviewed the emerging device rental market, as well as offered a comprehensive overview of potential use cases. Further, we outlined a complete protocol suite enabling the delegation of use for wearable devices, whenever their owner is willing to lend a device for temporary use.

The proposed solutions are described at length, both when the personal user network has a reliable wireless connection to the IoWT infrastructure, as well as when such connection is not available. Finally, we have analyzed the associated attacks on wearable devices themselves, as well as our designed protocols, and discussed some of the important practical
implications, including protection from phishing and relative power consumption. We believe that the proposed protocol suite and the accompanying discussion will become a useful consideration facilitating the delegation of wearables across multiple casual and business IoT scenarios.

REFERENCES

Aleksandr Ometov received the specialist degree in information security from St. Petersburg State University of Aerospace Instrumentation, Saint Petersburg, Russia, in 2013. He has been a Research Assistant with the Department of Electronics and Communications Engineering, Tampere University of Technology, Tampere, Finland, since 2013. His current research interests include wireless communications, information security, heterogeneous networking, cooperative communications, and machine-to-machine applications.

Sergey V. Bezzateev received the Diploma degree in computer science and the Ph.D. degree in information theory from the Aerospace Instrumentation Institute of Leningrad (now the St. Petersburg State University of Aerospace Instrumentation), Saint Petersburg, Russia, in 1980 and 1987, respectively. From 1980 to 1993, he was with Aerospace Instrumentation Institute. From 1993 to 1995, he was a Researcher with Prof. Y. Iwadare’s Laboratory, Nagoya University, Nagoya, Japan. From 1995, he was an Associate Professor with the Department of Information Technologies and Information Security, State University of Aerospace Instrumentation (SUAI), Saint Petersburg, Russia. From 2004 to 2007, he was a Project Leader with Joint Laboratory Samsung-SUAI on Information Security in Wireless Networks. In 2010, he became a Professor and the Head of the Department of Technologies of Information Security, SUAI. His current research interests include coding theory and cryptography.

Joona Kannisto received the M.Sc degree from Tampere University of Technology, Tampere, Finland, in 2011. He is a Researcher and Ph.D. candidate with Tampere University of Technology and has been involved in security related research ever since. His research interests include secure protocols, usable security, and reputation and trust management.

Jarmo Harju received the Ph.D. degree in mathematics from the University of Helsinki, Helsinki, Finland, in 1984. From 1985 to 1989, he was a Senior Researcher with the Technical Research Center of Finland, Espoo, Finland. From 1989 to 1995, he was a Professor with Lappeenranta University of Technology, Lappeenranta, Finland. Since 1996, he has been a Professor of telecommunications with Tampere University of Technology, Tampere, Finland, where he is leading a research group concentrating on network architectures and network security.

Sergey Andreev received the specialist and Cand.Sc. degrees from St. Petersburg State University of Aerospace Instrumentation, Saint Petersburg, Russia, in 2006 and 2009, respectively, and the Ph.D. degree from Tampere University of Technology, Tampere, Finland, in 2012. He is a Senior Research Scientist with the Department of Electronics and Communications Engineering, Tampere University of Technology. He has co-authored over 100 publications on wireless communications, energy efficiency, heterogeneous networking, cooperative communications, and machine-to-machine applications.

Yevgeni Koucheryavy received the Ph.D. degree from Tampere University of Technology (TUT), Tampere, Finland, in 2004. He is a Full Professor and the Laboratory Director with the Department of Electronics and Communications Engineering, TUT. He has authored numerous publications in the field of advanced wired and wireless networking and communications. His current research interests include various aspects in heterogeneous wireless communication networks and systems, the Internet of Things and its standardization, and nanocommunications.

Dr. Koucheryavy is an Associate Technical Editor of IEEE Communications Magazine and an Editor of the IEEE COMMUNICATIONS SURVEYS AND TUTORIALS.
Publication IV

© 2017 MDPI. Reprinted, with permission, from


Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (http://creativecommons.org/licenses/by/4.0/).
Article

Secure and Connected Wearable Intelligence for Content Delivery at a Mass Event: A Case Study

Aleksandr Ometov 1,*, Dmitrii Solomitckii 1, Thomas Olsson 1, Sergey Bezzateev 2, Anna Shchesniak 2, Sergey Andreev 1, Jarmo Harju 1 and Yevgeni Koucheryavy 1

1 Departments of Electronics and Communications Engineering, and Pervasive Computing, Tampere University of Technology, FI-33720 Tampere, Finland; dmitrii.solomitckii@tut.fi (D.S.); thomas.olsson@tut.fi (T.O.); sergey.andreev@tut.fi (S.A.); jarmo.harju@tut.fi (J.H.); evgeni.koucheryavy@tut.fi (Y.K.)
2 Departments of Cyber Physical Systems Security, and Wireless Telecommunications, ITMO University, 197101 Kronverksky pr., 49, St. Petersburg, Russia; bsv@aanet.ru (S.B.); anna.schesnyak@scaegroup.com (A.S.)
* Correspondence: aleksandr.ometov@tut.fi

Academic Editor: Hakima Chaouchi
Received: 31 January 2017; Accepted: 11 May 2017; Published: 22 May 2017

Abstract: Presently, smart and connected wearable systems, such as on-body sensors and head-mounted displays, as well as other small form factor but powerful personal computers are rapidly pervading all areas of our life. Motivated by the opportunities that next-generation wearable intelligence is expected to provide, the goal of this work is to build a comprehensive understanding around some of the user-centric security and trust aspects of the emerging wearable and close-to-body wireless systems operating in mass events and under heterogeneous conditions. The paper thus intends to bring the attention of the research community to this emerging paradigm and discuss the pressing security and connectivity challenges within a popular consumer context. Our selected target scenario is that of a sports match, where wearable-equipped users may receive their preferred data over various radio access protocols. We also propose an authentication framework that allows for delivery of the desired content securely within the considered ecosystem.

Keywords: wearables; security; authentication; WiGig; mass event; wireless; challenges

1. Introduction and Scope

Today, smart and connected wearable systems, such as on-body sensors, head-mounted displays and other small form factor capable personal computers are rapidly pervading all areas of our life as a more personal part of the Internet of Things (IoT) paradigm [1,2]. Such emerging wearables open new avenues for fundamentally different forms of both user-centric contextual services and interactive multi-user applications based on sharing data and resources locally [3]. This trend provides immense opportunities but, on the other hand, constitutes a vast unexplored area, riddled with numerous research challenges for both academia and industry [4]. One of the key challenges relates to the ‘big three’: security, privacy, and trust [5], i.e., how to ensure that wearable-specific information is produced and consumed appropriately by a multitude of devices and users [6].

Wearables are developing rapidly to become the next major information and communications technology paradigm, while manifesting ubiquitous computing and intelligent information technology on the most personal level [7]. Wearable devices are the pinnacle of miniaturized computation and communication technology for tracking, storing, processing, and reporting important human activity, such as physiological parameters, social interactions, and events [8]. They enable the next...
generation of wearable intelligence, where the communications chain may interconnect not only the user smartphones (acting as gateways) but also a wide range of new services (see Figure 1).

Particularly in urban environments, the proliferation of mobile wearable technology [9], as well as the rise of smart and automated cities are opening up new opportunities for improved public safety and security [10]. Even though there has been significant deployment experience of diverse IoT systems, the understanding behind these systems and the corresponding implications in the context of safety and security have only just scratched the surface [11].

![Figure 1. User-centric applications and services enabled by next-generation wearable intelligence.](image)

1.1. Applications of Wearables

Wearables have become a decisive innovation offered by a plethora of accessories and clothes (smart glasses, smart watches, fitness bands, augmented reality glasses, cameras, gadget gloves, etc.), which dramatically augment human capabilities [12]. Underpinning the topicality of this area, large international business around wearables is now in turbulence, which is evident from, for example, recent acquisitions of corresponding start-ups by larger companies like Intel (Recon Instruments, Replay Technologies, etc.) as well as fascinating new releases, such as Samsung brainBAND [13], to measure specific health aspects in mass sports events. As a result, wearable devices are already taking over the market with an increase in shipments of over 40% in 2016 as compared to the previous year, while the total shipments are expected to exceed 200 million units by 2019 [14]. While wearable technology clearly demonstrates much promise for novel exciting innovations, a lot remains to be studied and understood before the full potential of large-scale wearable ecosystems can be harnessed beyond individual user applications [15]. The key challenges to address relate to information management within the device ecosystem, as well as the user-centric security, privacy, and trust aspects therein [16]. The rapid adoption rate of the paradigmatically new wearable technology poses multiple novel challenges along the lines of its security and privacy [17]. The sheer diversity of the devices leads to increased dynamics and complexity in terms of user management of the access rights and accentuates the need for wireless 5G-grade connectivity within the user’s “personal cloud” [18].
The breadth of the kinds of wearable devices available to an individual person is steadily increasing, which calls for new solutions for managing the “personal cloud” security [19]. The wearable devices of one user being in physical proximity to each other creates opportunities for establishing trusted networks between smart devices and building meta-level intelligent services based on multiple devices operated by different users [20]. Additionally, the constraints of wearables in terms of, for example, their battery life, connectivity range, and computation power increase the complexity of management as the user’s personal ecosystem remains constantly in a state of change [21]. To date, these aspects of future wearables have remained largely unexplored from the end user perspective. As most security threats are difficult for a consumer to understand or even identify, and the users should be allowed to focus on their desired activities in the real world, it is unclear how the management of security and, for example, access to information in the personal cloud should occur [22]. In what follows, we provide a concrete example of what the ecosystem of intelligent wearables could look like and what kind of security and privacy challenges the aspects of proximity, dynamics, and constraints of wearables bring about in this vision.

Considering the use cases of wearables, much of the prior research has focused on facilitating automation, healthcare, and other applications with pragmatic business prospects [23] and a clear return-on-investment [24]. From the research viewpoint, such contexts can often be readily modeled and controlled, since the tasks as well as the contextual factors are often well-understood, whereas handling mass contexts remains challenging to predict due to their behavior. This can lead to solutions that function satisfactorily in a specific context but cannot be generalized or transferred to other contexts, like Augmented and Virtual Reality (AR/VR) cases [25].

In stark contrast, the focus of this paper is on highly dynamic and complex contexts related to mass consumer applications in the leisurely use of wearables [26,27]. This emerging area sets unprecedentedly high requirements to: (1) understand the privacy- and security-related threats; and (2) develop scalable connectivity solutions that are acceptable for mass consumer markets. Indeed, consumers are particularly interested in knowing that their communicated data are protected and privacy is maintained. As of today, the information security ecosystem for such a wide range of intelligent wearables has not yet been established.

1.2. Structure of This Manuscript

The structure of this work is as follows. The future of wearable devices in the context of a mass event is discussed in the next section. A survey on possible market-available and next-generation wireless technologies with respect to AR/VR limitations is offered in Section 3. Further, our developed simulation framework allowing for improvement of the connectivity planning and security assessment, including the focus on the use of higher communication frequencies (such as 60 GHz bands), is outlined in Section 4. The produced results based on ray-based pass loss modeling within the characteristic scenario of a hockey match are given in Section 5. Further, in Section 6 we propose an authentication framework that secures the mass content delivery in the target scenario, and conduct a brief security analysis of the framework. The last section concludes the manuscript.

2. Background and Motivation

2.1. Related Historical Overview

For years now, sports players have been wearing tracking devices in training, so that the coaches could see who is in adequate condition and who has been “burning the candle at both ends” [28]. So far, such devices have not been allowed during competitive play [29]. In football, the only wearable that is permitted as of today is the referee’s watch that buzzes to let them know whenever a goal has been scored [30].

At the same time, hockey is one of the most dynamic global sports [31]. However, it still has not caught on to the technology boom that is changing the ways that teams track the progress of their
players (like in football). The hockey teams are thus not yet producing the data to be processed and, therefore, the mass consumer cannot obtain them through the conventional channels (TVs, radio, etc.) nor with the next-generation AR/VR equipment. However, this may soon change as new wearable devices have been in the development process for a long time [32–34] and are currently being released into the market [35].

In the “2015 All Star game”, the National Hockey League (NHL) placed tracking devices on the players’ jerseys and inside the puck [36]. Putting technology on ice allowed coaches and players alike to focus on the game-play performance. Some teams are also considering various hockey stick add-ons that measure the power and the speed of slap shots, or the amplitude and execution speed of each swing [37]. With these advancements, it is not expected to be long until we see comprehensive tracking technology enter the NHL, providing the mass consumers with a completely new level of experience.

Technology has profoundly altered the way we do sports by capturing the attention of spectators for hours. With today’s technology, we are now able to make grounded conclusions on the team’s performance based on statistics made available to the audience through broadcasting [38]. Today, the data obtained from players, coaches, etc., are only available to a very limited circle of people. However, in the world of tomorrow, a hockey match spectator may enjoy a range of different services based e.g., on the type (level) of the purchased ticket, fan club membership, and/or the place on the tribune [39].

2.2. Market-Available Professional Products

Since wearable technology is becoming increasingly integrated into professional sports, various metrics can now be taken into account and utilized throughout training, thus allowing for real-time decisions to be made subsequently. Many known tech and clothing companies are attempting to bridge the gap between the state-of-the-art technology and the pace of the evolution. We hence list some of the professional wearable equipment already employed in sports training:

Adidas miCoach [40] is an ecosystem of connected wireless sensors with gear or apparel that is capable of quantifying athletic performance, including acceleration, speed, distance, power, heart rate, etc. After collection, all of the essential data are sent to the coach instantly, so that the performance of an individual could be monitored to make conclusions on potential concussions and injuries.

Viper Pod is a device widely utilized in the sports world by more than 10 globally-known teams, such as the football teams of Barcelona, Arsenal, and Manchester United, as well as the rugby team England National [41]. With a weight of under 50 g, this chest-mounted device is equipped with a Global Positioning System (GPS) module, accelerometer, gyroscope, digital compass, and heart rate monitor. The corresponding metrics are then transferred to other devices, thus enabling the coach to conduct real-time analysis depending on the team performance. Similarly to football, the National Hockey League (NHL) has embarked onto the Viper Pod with teams such as the Chicago Bulls, the Cincinnati Bengals, and the Carolina Panthers, all making use of this innovative technology.

Catapult OptimEye G5 is a piece of equipment suitable for goalkeepers [42]. The device in question allows the coach to track goalkeeper’s movements together with a host of other statistics [43]. It is equipped with a set of sensitive accelerometers, a heart rate monitor, and a wireless module, thus providing close to real-time bio-mechanical and tactical analysis. The lifetime of the device is 5 h and the post-game analysis is also available as one of the features. The company also offers a variety of devices for the NHL, National Basketball Association (NBA), and National Collegiate Athletic Association (NCAA).

The E39 performance shirt by Armour is a high-tech T-shirt equipped with a removable computer that features a triaxial accelerometer, processor, and 2 GB of storage supplied with additional monitors to measure the wearer’s heart rate and breathing [44].

ShotTracker is a basketball wearable consisting of the wrist and net sensors, which intends to improve the statistics of the players during the game [45]. This device was the first to be adopted by the basketball league with increasing intensity.
These are but a few of the professional sports wearables used by the leagues and the international teams across the globe. Some of the monitoring devices were also developed not for the public market but for the professional-targeted training, including the ones presented in [46–48]. The use of wearable technology is undeniably a major game-changer, while increased adoption of professional sports wearables during the games becomes another testament of this effective technology.

2.3. Proposed Model

Our envisioned service could be included with the initial game ticket and/or enabled by utilizing micro-transactions during the match itself. We further provide a non-exhaustive list of possible applications within this context:

- Obtaining video content made available by the proprietary sources (team players, opponents, referees, hockey gates, main cameras, etc.);
- Accessing general information related to the club (history, events, players, etc.);
- Monitoring critical information (warnings, evacuation plans, etc.);
- Advertisements and promotions (closest fast-food venue, order of a drink, taxi, etc.).

Complementing the above avenues of available monetization opportunities, the game organizers (owners of the stadium) may also acquire anonymized data related to the actual number of seats taken, distribution of spectators, amount of specific requests, and necessary feedback. The collected statistics could be utilized e.g., to improve the general levels of physical security as well as increase the effectiveness of future event planning. To this end, Figure 2 details our characteristic scenario depicting a wearable-enhanced ice hockey match, where we can differentiate between the following categories of participants (named here ‘roles’):

- **Mass spectators** (purchased their personal ticket; have access to a personalized set of AR-based services; are main producers and consumers of data; have the possibility for on-demand content acquisition; engage into direct interactions);
- **Support personnel** (broadly includes technical, medical, maintenance, advertising, and other specialists with access to their specific and AR-based data; access detailed information on players/spectators);
- **Competing teams** (including players and coaches with data possibly affecting tactics and strategy of the team; the requirement of long-term protection against misuse of such more dynamic and context-oriented information).

Summarizing, the main target here is to benefit the following stakeholders: regular audience (spectators, fans, etc.), event venue (e.g., sports stadium), competing teams (coalitions) or their owners/sponsors, first responders and maintenance personnel, and other services (advertising companies, wearable equipment vendors, etc.).

The pragmatic example output that we assume when considering the ice hockey business setting is a secure, wearable-aware data streaming system. As players carry around multiple wearable devices and sensors, such as heart rate monitors, lung capacity, metabolism, and location monitors, collision sensors and cameras [49], this equipment streams relevant data, where more sensitive information is only delivered to the authorized nodes, while game-oriented data is made available to the mass public. Moreover, we may consider an “intelligent puck” system that tracks the puck location in the ring and communicates it to e.g., the ice hockey arena’s public node.

The league organizing the games may have its own dedicated node and, depending on the agreement between the team and the league, the data might be aggregated and abstracted away before delivery by means of masking [50,51]. This can therefore fuel third-party services running on the spectator’s smart display [52] (AR glasses or mobile on-demand TV screen). In this context, specific security demands arise for: (1) signatures to prove the origin and the authenticity of all data; (2) encryption for improved confidentiality while accessing the parts of data in
transit; (3) randomized inspection protocols for data validity assessment; (4) restricted Application Programming Interfaces (APIs) to develop applications that perform computations over private data, where the application owner does not have the actual data but only the result; and (5) mechanisms to prevent from covert channels or to limit their bandwidth. The following section sheds light on the scenario-related security concerns.

3. Security Context in Public Events with Wearable Intelligence

In this section, we bring the reader’s attention to the key dimensions that we argue as underpinning the most crucial challenges pertaining to the security aspects of wearable intelligence:

- The notion of close proximity between several wearable and carryable devices, which can be mobile;
- Higher dynamics of personal user environment, where the components in the user “personal cloud” depend on the situation and where particular devices may often (de-)associate in real time;
- Tighter constraints on the available processing capabilities and energy supply of contemporary wearables as a result of their reduced form-factor and functionality.

These dimensions comprise a solid foundation that allows for a rapid advancement toward next-generation wearable intelligence, which has to be made secure, privacy-friendly, and trustworthy.

3.1. Proximity

The rapidly diversifying ecosystem of wearable devices that reside geographically close to each other—either belonging to one or to several users—opens new opportunities for developing secure connectivity solutions based on physical proximity [53,54]. For example, in the aforementioned scenario, proximity-based networking can be used to form data dissemination channels that are only
available at the mass event. Additionally, wearable technology could be utilized by other users as a neighboring resource: various sensors on a user’s smart jacket could provide him/her with customized and personalized service, while they could also accommodate other people nearby (proximate users), possibly with limited quality of service.

However, there are critical challenges that are required to be resolved before these applications can truly take off. The utilization of proximity as a security feature is vulnerable to some extent against attackers using specialized equipment [55]. Attackers can also spoof their presence using sybil attacks, particularly if radio resources or other physical resources are not attested [56]. However, whenever proximity can be guaranteed through the use of distance-bounding protocols [57], it could also be utilized to set-up security sensitive boundaries. Furthermore, the possible social and collective use of “personal clouds” of the neighboring users broadens these challenges from an individual user’s perspective to that involving several users.

3.2. Dynamics

This dimension relates to unpredictable mobility of human users in real time as well as the highly-dynamic composition of their respective personal networks even though the actual movement could generally be tracked [58]. The research field is suffering from a significant lack of activities in the area of dynamic privacy and trust management. Earlier research focused on industrial solutions for hospitals [59] or controlled body-area networks (BANs) [60], where devices are assumed to always have stable connectivity to the control unit. Many new challenges arise in light of mobility, where privacy of the device location is one of the important requirements [61].

Besides the issues in defining initial trust anchors, trust relationships evolve continuously [62]. Hence, trust management could become a burden for the end-user equipment unless these tasks can be largely automated through the use of authorization protocols. Changes in device management transform physical device ownership more toward a responsibility and a liability, rather than absolute control over the device. This means that an understandable security model is needed for distributed management functions, in order to, for example, maintain dynamic user privacy. Furthermore, trust decisions of humans as well as relations between physical devices and their hosts have to be supported in verifiable ways. This requires identity management frameworks that support identities beyond these of only users and wearable devices (e.g., virtualization and sandboxing).

3.3. Constraints

Limitations of small form-factor and battery-powered wearable devices constitute one more, “internal” dimension, which imposes constrains on the respective complexity of cryptographic protocols suitable for next-generation wearables [63]. First, constraints in the device’s user interfaces create challenges for the provisioning of the devices and communicating the security status together with the current trust relationships. In addition to verifying the configuration of wearable devices, detection of potentially malicious applications and actors in the “personal cloud” environment needs to be supported. Second, the coordination takes place on the lower levels of hardware. Hence, the real challenge is how to coordinate the actual hardware resources of each platform.

Developers require new programming frameworks and open-source platforms that help “get the most out of” hardware. Moreover, the corresponding security procedures are not standardized for the connectivity between the device and the gateway, and should be studied in more detail [64]. Further, the challenges of authenticating and authorizing wearable devices become pronounced, as their average densities around a user grow. Mutual authentication of wearables becomes therefore a glaring problem, including the risk of mismatch (accidental or premeditated). Finally, most of the wearables of today are optimized based on their energy consumption [65]. To this end, utilizing the conventional RSA-like information security solutions may be unacceptable for battery-constrained devices and thus new lightweight primitives should be proposed and developed.
4. Implementation of the Target Scenario

This section details the mass sports event within our selected target scenario, together with the delivery method (AR/VR) application requirements, potential wireless solutions, and the corresponding setup implementation. We have chosen the 20,000-seat hockey stadium as our reference design case. As with any high-density venue, specific deployments may have slightly different requirements. However, the principles outlined here are generally applicable for the venue of any size. The application type is mainly downlink data (video) streaming to the mass consumers. The goodput of the studied wireless technologies is not considered in this manuscript due to the static broadcast-like behavior. We base our research on the assumptions adopted from the industrial works by Ericsson [66,67]. Our custom simulator utilized in this work was previously calibrated with the real-life measurements in [68].

4.1. General Application Requirements

First, consider the data delivery though the smart glasses or portable televisions, where AR adds (computer-generated) supplemental elements to the user’s viewpoint, whereas solid VR recreates the entire scene that the user sees based on multiple sensory sources (e.g., cameras and other sensors). Both technologies operate in real time, and the main requirements of these applications are throughput (up to 1.5 Gbps total or 6 Mbps per stream), latency that ranges from the sub-millisecond level to tens of milliseconds, and jitter of below 1 ms [69]. Importantly, in our scenario the challenges of mobility are not discussed due to nearly static behavior of spectators. The application range is limited by the dimensions of the arena. Power consumption requirements do not appear to be a major issue, since the deployment scenarios allow the equipment to be either powered or recharged in a timely manner due to the limited duration of the event.

One of the major issues in our scenario is, however, scalability [21]. Both capturing and viewing equipment are deployed at high density, and the actual numbers will depend on the complexity of the scene e.g., the size of population as well as the numbers of cameras/sensors in order to acquire/transmit all aspects of the match. At the same time, recent academic activity demonstrates that the actual user density within the stadium scenario is 200,000 users/km² on average [70]. In this work, we solve the scalability challenge by applying a solution from the field of information security, which allows for dynamic and secure content delivery. The proposed technique is discussed in the latter part of this manuscript.

The following subsection overviews the potential solutions that satisfy the requirements of the mass content delivery scenario.

4.2. Candidate Connectivity Solutions

Today, communications technology researchers and vendors are competing to fulfill the requirements of efficiency, flexibility, and simplicity of coexistence [71]. This subsection briefly surveys a number of widely adopted wireless protocols that are suitable for the utilization in cases of a highly dense environment.

We first focus on the market-available solutions that satisfy the above listed requirements. The most adopted and widespread technology is represented by Institute of Electrical and Electronics Engineers (IEEE) 802.11 protocols, or WiFi. As a prominent example, IEEE 802.11ac-based devices are offering high data rates acceptable for the AR connectivity of today. A conventional WiFi medium access protocol (MAC) was designed to efficiently support up to around 25 nodes communicating their bursty content simultaneously, due to its characteristic operation based on the random channel access i.e., the binary exponential backoff (BEB) protocol. Ultimately, it can offer up to 90% of spectral efficiency for as many as 5–10 devices [72]. In the scope of this research, conventional operation of WiFi does not fulfill the requirement of a high number of nodes from the efficiency or from the interference points of view [73].
Since we focus primarily on the AR/VR as a demanding wearable-based application, one of the key performance requirements is to provide at least tens of Mbps per client. Hence, Bluetooth (up to 20 Mbps, 15 m, and 8 clients per host) and WiFi technologies (up to 400 Mbps for IEEE 802.11ac, 50 m, and 15 clients per host), which are currently deployed on most consumer devices, need to scale by several orders of magnitude. As one of the possible market-ready options, a solution by Wireless Gigabit Alliance (WiGig) may be utilized to employ the specifics of millimeter-wave (mmWave) communications [74].

Practically, conventional radio technologies that serve up to 25 demanding devices have a considerable probability of receiving low quality of service (QoS) levels, while doubling this number may degrade the performance altogether. The reasoning behind this is that WiFi has been designed for private indoor use, where the number of served users per access point varies below 10. Any increase in this number would dramatically impact the collision probability. However, the standards do not specify the exact BEB parameters and such a setup is left entirely at the discretion of vendors, which may cause faulty operation of the devices. Similar situations could be observed for the upcoming generation of short-range wireless technologies [75].

The second group of connectivity solutions considered here can be referred to as “the next generation”. Today, a large portion of wireless research and development is targeting the use of extremely high frequency bands in the range of 60 GHz. A brief overview of such technologies is given below.

The first considered solution was presented in 2008 and named Wireless HD [76]. Its target utilization is in home theaters and media centers. The main feature brought along by this standard is the use of both random and scheduled channel access, where one controller has a complete picture of the served nodes in its coverage. The main issue here is the lack of knowledge between the neighboring networks, which brings challenges of uncontrollable interference and hidden node problems [77] that primarily affect scheduled transmissions.

The second step made by the wireless industry is IEEE 802.11ad standard, widely known as WiGig [78]. The main requirement here is to enable the throughputs of at least 1 Gbps on top of the MAC layer. It offers a number of flexible protocol solutions, especially for low-cost devices. The scheduling is implemented similarly to that in Wireless HD, thus bringing along the same interference and management issues—the number of possible networks that may coexist in space and time is four. Another solution is ECMA 387 [79], which provides mobility support by dynamically resolving the cross-cloud collisions with a novel approach: soft channel switch and coordination. Basically, whenever a collision is detected, the communication channel is switched, hence reducing the subsequent collision probability. If there are no free channels left to utilize, the beacons time is reassigned by keeping the networks operational, so that the beacons of neighboring clusters would be transmitted one after another. Unfortunately, there are no vendors supporting this standard as of today.

Summarizing, we can conclude that the best practical option to be utilized today is WiGig technology due to its desirable properties and market support. The general challenge here is radio propagation due to the inability of 60 GHz wireless signals to penetrate almost any material at such high frequencies. On the other hand, this construction allows to assume that the receivers are mounted as part of the AR/VR heads-on devices [80] and thus the line-of-sight communication is delivered. An inherent feature of the 60-GHz solution is to adopt beam-forming that increases e.g., the levels of security while delivering the user-specific content at high rates. Further in Section 5, we compare the results of its utilization with the conventional IEEE 802.11n and .11ac operation.

4.3. Scenario Details and Simulation Description

To study wireless propagation in our characteristic scenario, we consider a 3D stadium grid presented in Figure 3a. A large number of receiving nodes (RXs) are located on the tribunes according to the mass user positions. This considered layout is common for any large sports event [81]. In this work, we first assess the conventional WiFi-like solutions provided, for example, by Cisco and then extend
the evaluation to embrace the next-generation mmWave technology. The geometrical parameters of
the scenario in question are summarized in Table 1.

<table>
<thead>
<tr>
<th>Parameter</th>
<th>Value</th>
</tr>
</thead>
<tbody>
<tr>
<td>Overall scenario size</td>
<td>200 m × 160 m</td>
</tr>
<tr>
<td>Scenario height</td>
<td>40 m</td>
</tr>
<tr>
<td>Ice ring size</td>
<td>61 m × 37 m</td>
</tr>
<tr>
<td>Number of receivers</td>
<td>515</td>
</tr>
<tr>
<td>Number of transmitters</td>
<td>1–3</td>
</tr>
</tbody>
</table>

The site-specific deterministic Ray-Launcher (RL) tool was utilized in this work, which models the
multi-path propagation of a wavefront within the wireless medium [82]. This principle is implemented
in the geometrical engine of the RL tool, which is based on the ray-casting methods, where the
continuous wavefront is replaced with the discrete one. Multiple 3D rays (or beams) outgoing from
the transmitting (TX) node propagate to the RX through the line-of-sight links and on the reflected
paths. At the same time, the physical engine of the RL tool is based on the geometrical optics (GO) and
the uniform theory of diffraction (UTD) techniques [83].

In this work, we concentrate on a first-order evaluation and thus disregard the relatively small
objects due to several reasons. First, simulation of a highly detailed scenario requires powerful
computing resources and significant computation time. Second, diffuse scattering produced by the
objects that are electrically small with respect to the wavelength does not offer a considerable impact in
terms of power. Based on that, only bulky and electrically large objects feature in our 3D reconstruction
of the stadium, which are presented in Figure 3. Here, Figure 3a is an original model with the high
level of details, while Figure 3b is a preprocessed and simplified model acceptable for the purposes of
our intended evaluation. The 3D simplification utilized in our work is to reduce the number of vertices
on a mesh by lowering the maximum angle to 15°.

Further, the RXs are carefully positioned to cover the entire area of interest. As the RXs are located
around the hockey arena, an empirical consideration to place the TX below the tableau was applied.
We then assumed that the TX and the wearable devices (i.e., RXs) of the spectators are vertically
polarized, that is, the polarization mismatch is insufficient. To avoid additional complexity at the
MAC-layer, an isotropic radiator was selected as the reference antenna design for both the TX and
RX, which has a uniform gain in the spherical coordinate system. Thereby, each RX observes the
signal combined by summing up different rays that propagate on the various paths and with different
power levels.

In this paper, we study three representative broadcast scenarios: (1) one TX placed in the middle
of the stadium below the tableau; (2) two TXs on the opposite sides of the stadium; and (3) a combined
scenario with all three TXs, whereas the locations are marked in Figure 3a.

![Figure 3. Model utilized for calculation.](image-url)
4.4. Key Performance Metrics

In this study, we evaluate to what extent the utilization of different wireless technologies could support the broadcast data delivery during a mass event (e.g., a hockey match). More specifically, we focus on the coverage optimization target [84]; hence, our main metric of interest is the path loss (PL). It can be obtained as follows:

\[ PL = P_{TX} - P_{RX}, \]  

where \( P_{TX} \) is the radiated power from the TX and \( P_{RX} \) is the total received power at the RX.

In this work, each of the RXs collects its own portion of power after multi-path radio propagation. Taking into account the carrier frequency, which leads to the corresponding attenuation per distance, each \( P_{RX} \) must be different at 2.4 GHz, 5 GHz, and 60 GHz.

5. Selected Numerical Results

Evaluating the most widely utilized radio technologies that operate in unlicensed spectrum (IEEE 802.11n at 2.4 GHz and IEEE 802.11ac at 5 GHz bands), we also address the benefits brought along by the potential use of mmWave communications technology (IEEE 802.11ad or WiGig at 60 GHz). We emphasize that both the RXs and TXs have zero antenna gains. The calculation error at the RX side is within the 3dB range, while the calculation time for our model with 8000 faces takes approximately 4 h with 1 GB RAM. The resulting PL maps are collected in Figure 4.

In order to validate our simulation results, we also compared these with the free-space PL model at each frequency:

\[ FSPL = 20 \log_{10}(d) + 20 \log_{10}(f) + 20 \log_{10} \left( \frac{4\pi}{c} \right) - G_t - G_r, \]

where \( d \) is the average distance from the TX to RX, \( f \) is the frequency, \( c \) is the speed of light, and \( G_t, G_r \) are the RX and TX gains that are set to 0, respectively. All of the simulation data fell within the acceptable bounds of \( +/- 5\% \) compared to the analytical results across the three scenarios of interest. Reporting on the results of our evaluation in Figure 4, we provide the average theoretical Signal-to-Noise Ratio (SNR) values for all the scenarios: (a) for 2.4 GHz this is 72.0854 dB; (b) for 5 GHz this is 78.4606 dB; and (c) for 60 GHz this is 100.0442 dB. It could be observed in the color map of the plots that the theoretical values fit well within the bounds of the simulated results. The main thinking behind the utilization of the RL techniques for our study is the fact that the pseudo-random paths of rays would provide similar picture independently of the frequency. However, the received power figure should vary based on the TX, as can be seen in the plots, for example, by considering a set of Figure 4a. This allows us to utilize our custom framework for the analysis of different scenarios by consuming less time and fewer computation resources.

Importantly, from the information security perspective, these obtained results offer useful functionality by aiming to avoid a number of threats. We further list some of the applications to be considered during the mass event at the network planning phase:

- The PL map allows for predicting the levels of transmit power required for the path-based denial-of-service attacks, such as jamming [85]. By doing so, the detection of potential malicious activity becomes more straightforward.
- The users with the lowest RX power are also vulnerable to the distributed denial-of-service attack. The PL value allows for deriving a lower bound on the throughput, which can compromise such (edge) nodes [86].
- Even though general mobility levels in this mass system are considered to remain low, the PL map allows for addressing the moments of a possible handover between the access points, thus outlining the risk zone of rogue access points [87].
Figure 4. Receiving node (RX) path loss for the selected technologies: (1) Transmitting node (TX) placed in the middle; (2) Two TXs on the opposite sides; (3) Three TXs: one in the middle, and two on the opposite sides.

In summary, the utilized RL technique may be employed at the network planning phase, thus allowing for not only solving the conventional scenario-related connectivity challenges but also understanding the security issues pertaining to it.

6. Authentication Methods for Massive Content Delivery

In the previous section, we modeled the broadcast system operation suitable for content delivery at a mass event of a hockey match e.g., when the spectators are willing to access information encoded as the broadcast data stream. Typically, users attending such an event segregate not only based on the teams that they follow, but also subject to the ticket price, which directly transforms into the event observation quality, that is, better experience at higher price. Based on this fact, we further propose an authentication mechanism that may be utilized to offer the next-generation services that incorporate the AR/VR content delivery based on the subscription, which could be resolved with a dynamic authentication mechanism.

Generally, authentication protocols may be classified into three main groups [88]: (1) based on symmetric cryptosystems; (2) based on asymmetric cryptosystems; and (3) hybrid. The majority of those utilize hash functions as their basis, thus allowing to combine the secret information “shares” into one specific secret key. As the most trivial and well-known example, we may recall the exclusive disjunction (XOR) function that makes it possible to group a set of identifiers $ID_i$: 

$$ X \oplus Y = Z $$
The possible utilization of the public key infrastructure (PKI) is an example of applying the asymmetric cryptosystems \[89\]. It allows for implementation of flexible authentication frameworks suitable for serving higher numbers of active users with a variety of access control mechanisms and features.

For our target scenario, the first requirement of the appropriate authentication protocol is to consider a set of simultaneously available secret shares i.e., unique identifiers that depend on the seat number, ticket number, etc. The second important requirement is the association of the secret/public keys with each specific content-consuming user that could thus be verified. The simultaneous provision of the set of required shares within a specified time interval is a solution to prevent reuse of the same identifiers by different users. One of the well-known solutions that enables such a functionality is the hybrid Yoking–Proof protocol \[90,91\]. Our proposed example is given in what follows.


We propose an adequate technological solution based on the stadium equipment availability: (1) both seats and tickets are supplied with the radio-frequency identification (RFID) tags \[92\]; (2) the spectator has a smartphone equipped with the near field communication (NFC) technology. The required identification data are obtained based on the simultaneous verification of both the ticket and the seat by utilizing the broadband access deployed at the stadium.

In order to provide with the needed functionality, the following requirements are to be satisfied: (1) each seat is numbered (equipped with a one-time sticker containing its unique identifier), all of the tickets have the corresponding unique identifier; and (2) each spectator has a smartphone. In this simple setup, the authentication code may be delivered through a trusted cellular network via the SMS code.

By doing so, the company responsible for the event in question has an opportunity to obtain the following information from each attendee: (1) the seat identifier \((ID(s)_k)\); (2) the main ticket identifier \((ID(t)_i)\); and (3) an additional subscription identifier \((ID(a)_j)\). The level of service \((S_i)\) provided to the \(i\)th user could thus be estimated based on the received information. The system would need to provide a unique result of the hashing function:

\[
h = H(ID(s)_k||ID(t)_i||ID(a)_j||...||ID(a)_m||Ac),
\]

where \(k, i, j\) are the counters for different components, \(m\) is the maximum value for the counter \(j\), and \(Ac\) is the authentication code. The result is further signed with the unique user’s \((ID_i)\) secret key \((SK_i)\) as:

\[
s_v = sign(h_z, SK_i),
\]

where \(h\) is obtained by the Equation (4) corresponding to \(z\)th result of the function, and \(SK_i\) is a unique user’s secret key.

Therefore, each \(v\)th subscription level acquires the corresponding unique pair of the secret \(SK_i\) and the public \(PK_i\) keys. Hence, each user has an anonymized (from the third-party perspective) ticket, \(t_i = (i||h_z||s_v)\),

\[
where \(h_z\) is produced by the Equation (4) and \(s_v\) is given by the Equation (5).

Based on the above, each ticket is composed of a unique sequence including the specific seat, the level of service provided, and the signature that allows to validate the previous fields based on the public key stored in the service provider cloud. The event-organizing company is assumed to act as a trusted certificate authority. In our scenario, there may be two cases of interest:
The organizer provides its services to the customers in an anonymized way. To achieve anonymity in relation to the service provider (stadium administration), the following addition to the authentication protocol could be utilized:

- Conventional PKI authentication and integrity protocols need to be replaced with ID-based formulations [93,94];
- Certificate authority in the modified scheme is represented by the private key generator (PKG);
- The secret key $SK_i$ is not directly “connected” to a unique user $ID_i$, but rather links with the ticket number and/or the seat number, and the event parameters (name, date, time, etc.). The $SK_i$ is to be obtained by the PKG with the use of any ID-based key generation protocol.

- At the signature verification stage $sv$, the event-organizing company requires only the ticket number and the event parameters. Therefore, it is not necessary for a user to provide any personal information (for realizing the verification procedure) directly to the event organizers. However, the authority may still obtain these data if necessary.
- In cases of, for example, Public Protection and Disaster Relief (PPDR) [95] or mass riots during the event, the administration has an opportunity to acquire the data on each user and forward it to the dedicated security units.

### 6.2. Framework Security Analysis

The main challenge behind the proposed solution lies in the very structure of the Yoking–Proof protocol and is related to the timeouts [96]. Generally, during simultaneous verification, the main device utilizes a preset timer to enable the said check. This is mainly due to the inability of scanning two or more sources at the exact same time. Therefore, a threshold value is defined and an attack could be executed if it provides, for example, a too-long validation interval not related to the actual source reading times.

Further, we briefly elaborate on the security analysis of the proposed solution. In our framework, security is based on the RSA assumptions, similarly to [97]. It utilizes primitive arithmetic operations at the user equipment (UE) side, such as the Add function, XOR operation, random number generator, and hash function. Therefore, the container data can be XOR-ed with the random values to prevent private data leakage.

Tag anonymity in the proposed solution is not considered by this work, since the tags are distributed across a publicly available area and thus could be temporarily accessed by an eavesdropping user. The IDs are accessible in plaintext and each of them is associated with the corresponding secret key on the owner’s side to perform meaningful computation. Note that the plain IDs can be eavesdropped but the security robustness of the meaningful data in the transmitted messages will not be compromised, and thus confidentiality can be guaranteed [98].

Another important issue to solve for the Yoking–Proof protocol is a replay attack [99]. At the stage of querying the smartphone tags, each of them responds with its corresponding message. An attacker can eavesdrop on the transmitted information over an insecure channel and store the messages locally. Next, the attacker may utilize the intercepted messages to complete the reader’s authentication. One of the solutions to overcome this threat is to utilize timestamps [100] and/or pseudo-random numbers [101] along with every communicated message, in exchange for additional space and connectivity requirements. Generally, this makes the replay attack more cumbersome for the attackers.

Another attack to be considered is the so-called counterfeit-proof attack [102]. Similarly to the method used against the replay attack, a timeout mechanism may be utilized to ensure that all of the proof-involved tags coexist for a specific and limited time period.

Finally, the notorious person-in-the-middle attack may also take place [103]. Here, an attacker can eavesdrop on the messages transmitted between the tag and the smartphone, and then modify the information to counterfeit a legitimate role. This challenge is solved by utilizing secure cellular assistance mechanisms by means of an extra SMS verification code, which solves most of the pressing
authentication issues. To this end, the operation of our proposed protocol primarily relies on the assumption of a secure cellular channel. Since security-centric analysis is not the main goal of this paper, the proposed protocol may require a deeper evaluation and testing in field scenarios.

7. Conclusions

Today, the rapidly expanding deployments of wearable technology as well as the rise of smart cities are underpinning new opportunities for wearable paradigm adoption. While there have been multiple attempts to deploy different IoT systems, our understanding of those and the corresponding implications in the context of safety and security have only scratched the surface, especially in wearable scenarios.

Particularly, we surveyed wireless technologies suitable for wearable-equipped consumers with the emphasis on the AR/VR applications as well as the corresponding security challenges in case of a mass sports event (e.g., a hockey match). Then, we utilized our developed ray-based simulator employing the ray-launching principles to study some of those technologies at various frequencies to conclude that WiGig (a 60-GHz solution) is the most appropriate choice for the broadcast content delivery in terms of its achieved path loss. We also elaborated on how the utilization of our study may improve security within the target scenario at the network planning phase. The proposed tool could be further utilized for both indoor and outdoor radio network planning.

Further, we also proposed an authentication technique based on the Yoking-Proof protocol that allows for secure content dissemination in the presence of simultaneous access to multiple unique identifiers, such as the ticket, seat, SMS, etc., therefore enabling a secure ecosystem within our representative scenario of interest.

The ultimate goal of this work is not to answer all of the pressing questions, but rather to bring the community’s attention to the challenges of mass wearable scenarios. The state-of-the-art in wearables is just at the beginning of a long journey, but there is already a lot to consider before making the next step.

Acknowledgments: The described research was partially supported by the Foundation for Assistance to Small Innovative Enterprises (FASIE) within the program “UMNIK” under grant 8268GU2015 (02.12.2015).

Author Contributions: A.O., D.S., and S.B. conceived and designed the experiments; A.O. and D.S. performed the experiments; A.S. and D.S. analyzed the data; T.O., S.B., S.A., J.H., and Y.K. contributed analysis tools; A.O., D.S., T.O., S.B., and A.S. wrote the paper.

Conflicts of Interest: The authors declare no conflict of interest.

References


Publication V

© 2018 IEEE. Reprinted, with permission, from


In reference to IEEE copyrighted material which is used with permission in this thesis, the IEEE does not endorse any of Tampere University of Technology’s products or services. Internal or personal use of this material is permitted. If interested in reprinting/republishing IEEE copyrighted material for advertising or promotional purposes or for creating new collective works for resale or redistribution, please go to http://www.ieee.org/publications_standards/publications/rights/rights_link.html to learn how to obtain a License from RightsLink.
Safe, Secure Executions at the Network Edge: 
Coordinating Cloud, Edge, and Fog Computing

Niko Mäkitalo, Aleksandr Ometov, Joona Kannisto, Sergey Andreev, Yevgeni Koucheryavy, and Tommi Mikkonen

Abstract

System design where cyber-physical applications are coordinated from the cloud may simplify the development process. However, all private data is then pushed to these remote ‘swamps’, and human users are losing the actual control as compared to when the applications are executed directly on their devices. At the same time, computing at the network edges is still lacking support for such straightforward multi-device development, which is essential for a wide range of dynamic cyber-physical services. In this work, we propose a novel programming model as well as contribute an associated connectivity framework for leveraging coordinated device proximity as an additional degree of freedom between the remote cloud and the network edge.

This article proposes a programming model and an associated secure-connectivity framework for leveraging safe coordinated device proximity as an additional degree of freedom between the remote cloud and the safety-critical network edge, especially under uncertain environment constraints.

1 Motivation and Background

Today, the cloud has evolved into a ubiquitous solution for enterprises in their quest for a unified digital platform. To this end, it represents a centralized infrastructure that has become the control point to manage computing power, storage, processing, integration, and decision-making assets for modern corporations. Boosting these processing and decision-making capabilities even further, there is a need to offer novel high-confidence technologies that will have the capability to support billions of networked devices, as we are stepping into the era of interconnected Cyber-Physical Systems (CPSs) [1].

The particularly challenging operation conditions of future CPSs are represented by the areas of poor or unavailable Internet connectivity, which need to be handled to maintain sustainability and enable faster decision-making based on localized data intelligence across heterogeneous system components. These considerations are underpinning the recent trend to transition from the centralized cloud platforms to the network edge, which is essentially dispersing the cloud back to the origin – the end devices [2]. With advanced microservices, containers, and APIs, it is increasingly feasible to execute these smaller, self-contained, and purpose-driven services that specifically target certain dedicated functions required on the edge.

The distinct computing paradigms, such as “cloud” (computation on a remote server), “edge” (computation on end devices), and “fog” (computation at the local-area-network level) computing, act similarly under conventional use cases but become fundamentally different when considering safety-critical CPSs that operate in dynamic and uncertain conditions [3]. Examples include overtaking control of a moving vehicle and hacking an industrial robot, where various vulnerabilities may be exploited to hijack remote access to the capable factory equipment.
To mitigate these vulnerabilities, new systems and software engineering methods are required where networked machines can interact more freely, by forming connections according to the actual demand and not because this is requested by the central managing entity. This calls for revisiting the ways of interaction between the devices and their operating environment, which becomes the target of this article. Its main contributions thus are the Action-Oriented Programming (AcOP) model and an associated framework that can dynamically adapt to the edge and the cloud according to particular environment and connectivity conditions. Further, AcOP is compared to mobile-app-based and cloud-based CPS deployments. Finally, a framework to enable secure coalitions and dynamic management of collective executions is also outlined.

2 Sensing and Actuation Executions at the Network Edge

Billions of smart CPS devices at the network edge require proximity-based communication together with the cloud connectivity, but these two aspects have traditionally been addressed in isolation [4]. At the same time, the lion’s share of the CPS interactions is still about their users (e.g., the concept of quantified self [5]), and the human element is tightly involved into the decision-making process. The capability to switch from the cloud to the edge (or the app) dynamically, based on the operation conditions and user requests, allows to control device behavior more efficiently. In what follows, we focus on advancing software development efforts so that the said activities are executed on the network edge to boost the dynamic adaptation of a CPS to complex environmental constraints, while still being mindful of human perception.

2.1 Action-Oriented Programming Model

The AcOP model’s roots are in the so-called Social Devices concept [6], and the original idea is tailored here to the context of fog computing as driven by CPS evolution. The new model is realized on top of the JavaScript programming language and includes the following constructs.

**Sensation – An input coming from the physical, cyber, or social world.**

Instrumental to CPSs is observation of various events coming from the outer world, and then acting upon these events. In AcOP, these observed events are named *sensations*. The abstraction level of the sensations may vary, and in addition to observing physical world’s phenomena, the processes in cyber and social worlds can be monitored as well. A concrete example of a sensation is the changed sensor value, while a more abstract sensation is, e.g., when a friend is nearby, which combines data from different worlds (Facebook friendship and Bluetooth signal strength values).

**Capability – Physical objects as programmable JavaScript objects.**

In AcOP, physical objects and digital (micro)services constitute programmable JavaScript objects. They are described with AcOP *capabilities* that define the ways in which a certain machine can interact with other machines and humans. An example here is the *talking* capability where the device is able to translate text into speech. The capabilities produce abstract *capability sensations*, such as “temperature has changed” or “coffee is ready”, which are derived from raw sensations. These help the developers define scheduling policies for collective executions.

**Action – Joint behavior of machines and humans.**

In the heart of our model are *actions* that define joint operations across multiple devices and people. Typically, an action is a modular unit that determines how a predefined set of devices interact with each other over a certain period of time. The actions are defined with JavaScript and comprise two parts, an *enabling condition* and a *body*, used for programming
the interactions by utilizing the AcOP capabilities as well as the basic programming logic. The modularity of the actions helps in making them more generic so that they may be exploited in many different executions; similarly, the device capabilities may be employed by many actions.

**Collective Execution – Coalition of trusted entities that sense and act towards a common goal.**

In AcOP, a set of machines and humans form coalitions by engaging into trust negotiations. These coalitions then collectively execute software where machines and humans interact and cooperate. The key idea of the collective execution is detecting and maintaining information about the sensations coming from the various worlds as well as from the coalition participants. Then, collective execution attempts to schedule an action for a set of devices, which are selected for their roles in this action based on their capabilities and properties. In practice, collective execution works so that one device in a coalition at a time assumes the role of a coordinator and then executes the code that is responsible for scheduling the actions. At the same time in the background, all of the other devices in a coalition contribute by exchanging information that is essential for that specific execution via secure connections.

![AcOP model example operation in an emergency scenario.](image-url)
2.2 Traffic Emergency Context: AcOP Operation at the Network Edge

As a characteristic use case, we study the scenario where contemporary vehicles can call 911 in emergency situations. Consider Fig. 1 depicting a traffic accident, in which the car involved would immediately report to the traffic monitoring system. The vehicle or the traffic system also communicates the supporting information – or a sensation – to a set of executions. For instance, Fire Department’s execution may detect the accident and schedule an action to leverage the car’s talking capability to poll passengers whether they are unharmed as well as provide instructions. The execution can also schedule further actions, e.g., command a drone to fly over and analyze if a fire has started, call an ambulance if an injury was suspected [7], etc.

Clearly, similar functionality might be achieved with conventional mobile applications and – to some extent – with cloud-based services. However, these alternative solutions fall short of providing adequate security and functional safety guarantees. Furthermore, they also require that the coalition forming capabilities are included into each enabled application. Further pain points of the traditional approaches are summarized in Table 1.

3 Establishing Secure Communicating Coalitions

To liberate the programmers from considering coalition forming as part of the application logic, an appropriate framework is required to enable operation of communicating peripherals, e.g., in case of a traffic accident described in the previous section. Certain known approaches exist already, thus bringing attention to the challenge of sandboxed executions in the emerging CPSs (e.g., FlowFence [10]). However, sandboxed collective executions of the same piece of software on the network edge devices have not gained sufficient attention thus far.

Today, mobile devices may establish and utilize a direct link only if they have a reliable connection to the server that is responsible for the key and connection management, or if they trigger the connection themselves. In the latter case, no security and safety guarantees can be provided by the operator. To mitigate this limitation, the Public Key Infrastructure (PKI) is commonly used to enable secure and authenticated communication when a connection to the centralized authority is not always available [11]. Without it, many applications might become disabled if a single user leaves the network coverage. This particularly occurs in cases of disaster and/or when a cellular connection is unreliable (the network is overloaded) or unavailable (on a train, airplane, elevator, etc.).

To augment edge- and fog- computing technologies, we propose to employ a secure communication framework that we developed in a series of trials within a live cellular network [12]. Our system is built upon the advanced security protocols contributed by 3GPP specifications. This novel framework applies the knowledge of distributed solutions to enable secure communication as, e.g., in the discussed traffic emergency example. Accordingly, execution in the devices of people in the emergency scene enables them to seamlessly join and leave a coalition without disrupting collective execution.

The main operation phases of the considered approach are illustrated in Fig. 2. The only procedure that requires stable connectivity to the cloud is the coalition initialization. First, the involved mobile devices receive their certificates with the corresponding secret and public keys. These are utilized to establish secure direct connectivity with each relevant device. When a device is willing to create a secure coalition with its “neighbors”, a request containing the public identifiers of the future coalition members is issued to the corresponding server in the network.

A polling procedure is then triggered by the network to ensure that the subject devices are actually willing to join this coalition. After the confirmations have been received, both coalition certificate and coalition secret (based on the Lagrange polynomials technique) are
Table 1: Comparing mobile app (M) and cloud service based (C) approaches with our AcOP model (A) for CPS development.

<table>
<thead>
<tr>
<th>Detecting and handling contingencies</th>
<th>M</th>
<th>C</th>
<th>A</th>
</tr>
</thead>
<tbody>
<tr>
<td>Detecting and handling contingencies</td>
<td>Easy to catch errors with try-catch notation. How to handle errors that occur in coordination or on other devices?</td>
<td>Easy to catch errors that occur on the server side. Some contingencies on the device side can be sent to the server side. Internet connectivity, however, is not available in emergency zones, air planes, or crowded locations. How to manage errors in between cloud and devices?</td>
<td>Contingency handlers for remote and coordination related issues. Within the handler methods, developers can define ways for recovering from unwanted behavior and, e.g., replace a disconnected device with another one, or change the connectivity type. It is also possible to reschedule an action, or take a completely different action.</td>
</tr>
<tr>
<td>Detecting sensations</td>
<td>Easy to implement detection, e.g., when device location and/or orientation change (for instance, with the use of delegate methods). How to detect state changes on other devices?</td>
<td>Data coming from multiple devices and sensors can be streamed to the cloud services and processed there. How does this approach scale when there are thousands of devices streaming data continuously?</td>
<td>Collective execution is designed to be used for detecting when a state changes on a device that is participating in the same execution. Then, for the task at hand, the sensations can be combined and processed from as many sources as required.</td>
</tr>
<tr>
<td>Reacting to sensations</td>
<td>When a sensation has been detected, the device can be instructed to act upon this event. How to select and command another device or a group of devices?</td>
<td>Cloud-based approaches are typically used for coordinating the CPS devices. The coordination, however, relies on the Internet connectivity, and communication in latency-sensitive systems can easily become a bottleneck. How to ensure coordination without adequate Internet connectivity?</td>
<td>After collective execution on the network edge, a device detects certain sensation(s) and attempts to schedule an action. Actions are designed to serve as the output to the world, for commanding joint operations between one or many devices when performing a certain task.</td>
</tr>
<tr>
<td>Distributing computation and cooperation</td>
<td>Particular tasks can be posted to be executed by the background cloud service. For instance, recent serverless approaches (e.g., Google’s Cloud Functions, Azure Functions, and AWS Lambda) gained certain popularity. While the developer is liberated from server management, the computation distribution still takes place on the servers [8].</td>
<td>The executions in cloud-based approaches can be either centralized or distributed when executed by several services and machines [9]. How to harness other (nearby) devices to perform computation?</td>
<td>Collective executions and actions can run on any of the edge devices capable of handling JavaScript. It is natural to motivate such distribution: privacy and security of content and data to be utilized on a specific device.</td>
</tr>
<tr>
<td>Deployability of cooperation</td>
<td>Mobile apps are deployed onto the devices via application stores. How to deploy cooperation?</td>
<td>Tasks can be deployed onto cloud services, where the devices are instructed to cooperate. How to deploy a certain well-defined task to be executed by particular devices?</td>
<td>AcOP components for executing certain tasks can be downloaded/installed from a repository in advance, or dynamically acquired at runtime whenever needed.</td>
</tr>
</tbody>
</table>

Figure 2: The structure of the coalitions operating behind the collective executions of AcOP.
delivered. After these steps, secure direct interaction may continue over any conventional network protocols. The members of an existing coalition have the possibility to invite new devices as well as remove the existing ones based on the flexible voting system, i.e., when \( k \) out of \( M \) devices in the coalition agree on a particular decision (runs automatically for machines and can be manual for humans). This allows the coalition to be updated dynamically to manage collective executions in various scenarios. For example, predefined “hidden” coalitions may be utilized in cases of disaster, thus enabling operational stability.

4 Discussion and Conclusion

Edge computing is increasingly demanded due to the CPS requirements for increased scalability and functional safety – if the entities are coordinated from the cloud, a risk remains that without reliable Internet connectivity the functional safety cannot be guaranteed. In cooperation at the network edges, devices need to be able to trust each other, thus calling for dynamic coalitions with secure and trusted topology. This, in its turn, improves functional safety since trusted entities can cooperate and act as back-up options for one another in various CPS applications (see Fig. 3): if one device fails, others are there to stand in.
In order to achieve this, we proposed the AcOP model and the associated framework that can dynamically adapt to the cloud, the fog, and the edge.

5 Acknowledgments

The work of N. Mäkitalo and T. Mikkonen was supported by the Academy of Finland (project 295913). The work of S. Andreev was supported in part by a Postdoctoral Researcher grant from the Academy of Finland and in part by a Jorma Ollila grant from the Nokia Foundation.

References


NIKO MÄKITALO is a postdoctoral researcher in the University of Helsinki’s Department of Computer Science. His main interests are web technologies in the context of fog computing and Internet of Things programming. Mäkitalo received his doctorate in computer science from Tampere University of Technology. He’s a member of IEEE and the IEEE Computer Society. Contact him at niko.makitalo@helsinki.fi; nkm.io.

ALEKSANDR OMETOV is a PhD candidate at Tampere University of Technology’s Laboratory of Electronics and Communications Engineering. His research interests include wireless communications, information security, heterogeneous networking, cooperative communications, and machine-to-machine applications. Ometov received his MSc with distinction in telecommunications from Tampere University of Technology. Contact him at aleksandr.ometov@tut.fi.

VOGENI KOUCHERYAVY is a full professor at Tampere University of Technology’s Laboratory of Electronics and Communications Engineering. His research interests include various aspects of heterogeneous wireless communication networks and systems, and nanocommunications. Koucheryavy received his PhD in technology from Tampere University of Technology. He is an associate technical editor of IEEE Communications Magazine and an editor of IEEE Communications Surveys and Tutorials. Contact him at evgeni.kucheryavy@tut.fi.

JOONA KANNISTO is a PhD student at Tampere University of Technology’s Laboratory of Pervasive Computing. His research interests include cryptographic protocols, identities, and authentication, with an emphasis on undesired side effects such as threats on privacy and denial of service. Contact him at joona.kannisto@tut.fi.

TOMMI MIKKONEN is a professor of software engineering at the University of Helsinki. His research focuses on software architectures, agile methodologies, web technologies, and connected devices. Mikkonen received his doctorate in information technology from Tampere University of Technology. Contact him at tommi.mikkonen@helsinki.fi.

SERGEY ANDREEV is a senior research scientist at Tampere University of Technology’s Laboratory of Electronics and Communications Engineering. Andreev received a PhD in technology from Tampere University of Technology. Contact him at sergey.andreev@tut.fi.
Multi-Factor Authentication: A Survey †

Aleksandr Ometov 1,*, Sergey Bezzateev 2, Niko Mäkitalo 3, Sergey Andreev 1, Tommi Mikkonen 3 and Yevgeni Koucheryavy 1

1 Laboratory of Electronics and Communications Engineering, Tampere University of Technology, FI-33720 Tampere, Finland; sergey.andreev@tut.fi (S.A); evgeni.kucheryavy@tut.fi (Y.K.)
2 Department of Security of Cyberphysical Systems, ITMO University, St. Petersburg RU-197101, Russia; bsv@aanet.ru
3 Department of Computer Science, University of Helsinki, FI-00014 Helsinki, Finland; niko.makitalo@helsinki.fi (N.M.); tommi.mikkonen@helsinki.fi (T.M.)

* Correspondence: aleksandr.ometov@tut.fi
† This manuscript is an extended version of work by A. Ometov and S. Bezzateev titled “Multi-factor Authentication: A Survey and Challenges in V2X Applications” presented at the 9th International Congress on Ultra Modern Telecommunications and Control Systems (ICUMT) on 6 November 2017.

Received: 30 November 2017; Accepted: 18 December 2017; Published: 5 January 2018

Abstract: Today, digitalization decisively penetrates all the sides of the modern society. One of the key enablers to maintain this process secure is authentication. It covers many different areas of a hyper-connected world, including online payments, communications, access right management, etc. This work sheds light on the evolution of authentication systems towards Multi-Factor Authentication (MFA) starting from Single-Factor Authentication (SFA) and through Two-Factor Authentication (2FA). Particularly, MFA is expected to be utilized for human-to-everything interactions by enabling fast, user-friendly, and reliable authentication when accessing a service. This paper surveys the already available and emerging sensors (factor providers) that allow for authenticating a user with the system directly or by involving the cloud. The corresponding challenges from the user as well as the service provider perspective are also reviewed. The MFA system based on reversed Lagrange polynomial within Shamir’s Secret Sharing (SSS) scheme is further proposed to enable more flexible authentication. This solution covers the cases of authenticating the user even if some of the factors are mismatched or absent. Our framework allows for qualifying the missing factors by authenticating the user without disclosing sensitive biometric data to the verification entity. Finally, a vision of the future trends in MFA is discussed.

Keywords: survey; authentication; SFA; 2FA; MFA; evolution; vision

1. Introduction

The continuous growth in the numbers of smart devices and related connectivity loads has impacted mobile services seamlessly offered anywhere around the globe [1]. In such connected world, the enabler keeping the transmitted data secure is, in the first place, authentication [2–4].

According to the fundamental work in [5], authentication is a process where a “user identifies himself by sending x to the system; the system authenticates his identity by computing F(x) and checking that it equals the stored value y”. This definition has not changed significantly over time despite the fact that a simple password is no longer the only factor for validating the user from the information technology perspective [6].

Authentication remains a fundamental safeguard against illegitimate access to the device or any other sensitive application, whether offline or online [7–9] (see Figure 1). Back in time, the transactions were authenticated primarily by physical presence, i.e., for example, by applying the wax seal [10].
Closer to present days and with the advancement of our civilization, it was realized that the validation based on the sender identification only is not always adequate on the global scale [11].

Initially, only one factor was utilized to authenticate the subject. By that time, Single-Factor Authentication (SFA) was mostly adopted by the community due to its simplicity and user friendliness [12,13]. As an example, the use of a password (or a PIN) to confirm the ownership of the user ID could be considered. Apparently, this is the weakest level of authentication [14,15]. By sharing the password, one can compromise the account immediately. Moreover, an unauthorized user can also attempt to gain access by utilizing the dictionary attack [16], rainbow table [17], or social engineering techniques [18]. Commonly, the minimum password complexity requirement is to be considered while utilizing this type of authentication [19].

Further, it was realized that authentication with just a single factor is not reliable to provide adequate protection due to a number of security threats [20]. As an intuitive step forward, Two-Factor Authentication (2FA) [21–23] was proposed that couples the representative data (username/password combination) with the factor of personal ownership, such as a smartcard or a phone [24,25].

Today, three types of factor groups are available to connect an individual with the established credentials [26]:

1. **Knowledge factor**—something the user knows, such as a password or, simply, a “secret”;
2. **Ownership factor**—something the user has, such as cards, smartphones, or other tokens;
3. **Biometric factor**—something the user is, i.e., biometric data or behavior pattern.

Subsequently, Multi-Factor Authentication (MFA) was proposed to provide a higher level of safety and facilitate continuous protection of computing devices as well as other critical services from unauthorized access by using more than two categories of credentials [27–29]. For the most part, MFA is based on biometrics, which is automated recognition of individuals based on their behavioral [30,31] and biological characteristics [32]. This step offered an improved level of security as the users were required to present the evidence of their identity, which relies on two or more different factors [33]. The discussed evolution of authentication methods is shown in Figure 2.

![Figure 1. Conceptual authentication examples.](image-url)
Today, MFA is expected to be utilized in scenarios where safety requirements are higher than usual [34,35]. According to SC Media UK, 68 percent of Europeans are willing to use biometric authentication for payments [36]. Consider the daily routine of ATM cash withdrawal [37,38]. Here, the user has to provide a physical token (a card) representing the ownership factor and support it with a PIN code representing the knowledge factor to be able to access a personal account and withdraw money.

This system could be easily made more complex by adding the second channel like, for example, a one-time password to be entered after both the card and the user password were presented [39,40]. In a more interesting scenario, it could be done with the facial recognition methods [41,42]. Moreover, a recent survey discovered that 30 percent of enterprises planned to implement the MFA solution in 2017, with 51 percent claiming that they already utilize MFA, and 38 percent saying that they utilize it in “some areas” of operation [43]. This evidence supports the MFA as an extremely promising direction of the authentication evolution.

As one of the interesting future trends, authentication between a vehicle and its owner or a temporary user may be considered. Based on the statistics [44], a vehicle is stolen every 45 s in the U.S. The current authentication method that allows for starting and using the vehicle is still an immobilizer key [45,46]. The MFA may significantly improve access to most of the electronic devices from both security and user experience perspectives [47,48].

Generally, MFA applications could be divided into three market-related groups: (i) commercial applications [49,50], i.e., account login, e-commerce, ATM, physical access control, etc.; (ii) governmental applications [51,52], i.e., identity documents, government ID, passport, driver’s license, social security, border control, etc.; and (iii) forensic applications [53,54], i.e., criminal investigation, missing children, corpse identification, etc. Generally, the number of scenarios related to authentication is indeed large. Today, MFA becomes an extremely critical factor for:

- Validating the identity of the user and the electronic device (or its system) [55,56];
- Validating the infrastructure connection [57];
- Validating the interconnected IoT devices, such as a smartphone, tablet, wearable device, or any other digital token (key dongle) [58].

Presently, one of the main MFA challenges is the absence of correlation between the user identity and the identities of smart sensors within the electronic device/system [59]. Regarding security, this relationship must be established so that only the legitimate operator, e.g., the one whose identity is authenticated in advance, can gain the access rights [60,61]. At the same time, the MFA process should be as user-friendly as possible, for example:

1. Customers first register and authenticate with the service provider to activate and manage services they are willing to access;
2. Once accessing the service, the user is required to pass a simple SFA with the fingerprint/token signed in advance by the service provider;
3. Once initially accepted by the system, the customer authenticates by logging in with the same username and password as setup previously in the customer portal (or social login).
For additional security, the managing platform can enable secondary authentication factors. Once the user has successfully passed all the tests, the framework automatically authenticates to the service platform;

4. The secondary authentication occurs automatically based on the biometric MFA, so the user would be requested to enter an additional code or provide a token password only in case the MFA fails.

Biometrics indeed significantly contribute to the MFA scheme and can dramatically improve identity proofing by pairing the knowledge factor with the multimodal biometric factors [62,63], thus making it much more difficult for a criminal to eavesdrop on a system while pretending to be another person. However, the utilization of biological factors has its challenges mainly related to the ease of use [64], which largely impacts the MFA system usability.

From the user experience perspective, fingerprint scanner already provides the most widely integrated biometric interface. This is mainly due to its adoption by smartphone vendors on the market [65]. On the other hand, it is not recommended to be utilized as a standalone authentication method [66]. However, the use of any biometrics often requires a set of separate sensing devices. The utilization of already integrated ones allows for reducing the authentication system costs and facilitate the adoption by end users. A fundamental trade-off between usability and security is one of the critical drivers when considering the authentication systems of today [67].

Another challenge is that the use of biometrics relies on a binary decision mechanism [68]. This was well studied over past decades in classical statistical decision theory from the authentication perspective [69,70]. There are various possible solutions to control a slight mismatch of the actual “measured” biometrics and the data stored in previously captured samples. The two widely utilized techniques are: false accept rate (FAR) [71] and false reject rate (FRR) [72]. Manipulations with the decision criteria allow adjusting the authentication framework based on the predefined cost, risks, and benefits. The MFA operation is highly dependent on FAR and FRR, since obtaining zero values for both of the metrics is almost infeasible. The evaluation of more than one biometric feature to establish the identity of an individual can improve the operation of the MFA system dramatically [73].

Since the currently available literature faces a lack of detailed MFA analysis suitable for non-specialists in the field, the main contributions of this work are as follows:

1. This work provides a detailed analysis of factors that are presently utilized for MFA with their corresponding operational requirements. Potential sensors to be utilized are surveyed based on the academic and industrial sources (Section 2);
2. The survey is followed by the challenges related to MFA adoption from both the user experience and the technological perspectives (Section 3);
3. Further, the framework based on the reversed Lagrange polynomial is proposed to allow for utilizing MFA in cases where some of the factors are missing (Section 4). A discussion on the potential evaluation methodology is also provided;
4. Finally, the vision of the future of MFA is discussed (Section 5).

2. State-of-the-Art and Potential MFA Sources

Presently, the authentication systems already utilize an enormous number of sensors that enable identification of a user. In this section, we elaborate on the MFA-suitable factors, corresponding market-available sensors, and related challenges. Furthermore, we provide additional details on the ones that are to be potentially deployed in the near future.

2.1. Widely Deployed MFA Sensors/Sources

Today, identification and authentication for accessing sensitive data are one of the primary use cases for MFA. We further list the factors already available for the MFA utilization without acquiring additional specialized equipment.
2.1.1. Password Protection

The conventional way to authenticate a user is to request a PIN code, password, etc. [74]. The secret pass-phrase traditionally represents a knowledge factor. It requires only a simple input device (at least one button) to authenticate the user.

2.1.2. Token Presence

The password could then be supplemented with a physical token—for example, a card, which is recommended as a second factor group—the ownership [75,76]. From the hardware perspective, a user may present a smartcard, phone, wearable device, etc., which are more complicated to delegate [77]. In this case, the system should be equipped with a radio interface allowing for two-way communication with the token [78,79]. On the other hand, the most widely known software token is one-time software generated password [80]. The main drawback of the above is the problem of uncontrollable duplication.

2.1.3. Voice Biometrics

Most of the contemporary smart electronic devices are equipped with a microphone that allows utilizing voice recognition as a factor for MFA [81,82]. At the same time, the technology advancement of tomorrow may allow special agencies not only to recognize the speakers but also to mimic their voices including the intonation, timbre, etc., which is a serious drawback of utilizing voice as a primary authentication method [83,84].

2.1.4. Facial Recognition

As the next step, facial recognition could be considered. At the beginning of its development, the technology was based on the landmark picture analysis, which was relatively simple to replicate by supplying the system with a photo [85]. The next phase was by enabling three-dimensional face recognition, i.e., by asking the user to move head during the authentication process in a specific manner [86,87]. Finally, the advancement of this system reached the point of recognizing the actual expressions of the user [88]. To enable facial recognition, it is required to equip the system with at least one output device and a camera [89].

2.1.5. Ocular-Based Methodology

The iris recognition techniques are on the market for more than 20 years [90]. This approach does not require the user to be close to the capture device while analyzing the color pattern of the human eye [91]. Retina analysis is another attractive technique [92]. Here, a thin tissue composed of neural cells that are located in the posterior portion of the eye is captured and analyzed. Because of the complex structure of the capillaries that supply the retina with blood, each person’s retina is unique. The most prominent challenges in those methods are the need for high quality capture device and robust mathematical technique to analyze the image [93].

2.1.6. Hand Geometry

Some systems employ the analysis of the physical shape of a hand to authenticate the user. Initially, pegs were utilized to validate the subject, but the usability of such methods was low [94]. Further on, the flatbed scanner was used to obtain the image without the need to fix the user’s hand in one specific position [95]. Today, some systems utilize conventional cameras not requiring close contact with the capture surface. This approach is, however, not very robust to the environment [96]. Some vendors apply so-called photoplethysmography (PPG) to determine whether a wearable device (e.g., a smartwatch) is currently on its user’s wrist or not [97,98]. The process is similar to the one followed when measuring heart rate [99].
2.1.7. Vein Recognition

The advances in fingerprint scanners offer an opportunity to collect the vein picture of the finger as well [100]. More complicated devices utilize palm print recognition to acquire and store the shape/movement of the entire hand [101,102]. At the current stage of development, vein biometrics are still vulnerable to spoofing attacks [103,104].

2.1.8. Fingerprint Scanner

Utilizing fingerprint scanner as the primary authentication mechanism is currently being pushed by the majority of smartphone/personal computer vendors [105]. This solution is intuitive to use but remains extremely simple to fabricate—mainly due to the fact that our fingerprints could be obtained from almost anything we touch [106,107]. The integration potential of this method is indeed high [108], even though it is also not recommended to be used as a standalone authentication approach. Most of the smartphone vendors install an additional camera to obtain the fingerprint instead of more safe vein recognition.

2.1.9. Thermal Image Recognition

Similarly to vein recognition, thermal sensor is utilized to reconstruct the unique thermal image of one’s body blood flow in proximity [109,110]. Many challenges with this authentication method may arise due to the user conditions: sickness or emotion may significantly influence the perceived figures [111].

2.1.10. Geographical Location

Utilizing the device’s and user’s geographical location to validate whether access to the device/service could be granted is a special case of location-based authentication [112,113]. Importantly, GPS signal could be easily jammed or considered faulty due to the propagation properties; thus, it is recommended to utilize at least two location sources, for example, GPS and wireless network cell ID [114]. A smartphone could be used to support MFA from the location acquisition perspective.

2.2. Future of MFA Integration

Accelerated adoption across many industries as well as increased availability of biometric services in a wide range of readily-available consumer products is pushing the concept of tight MFA integration. Currently, researchers and early technology adopters attempt to integrate new sensors to be utilized in MFA systems.

2.2.1. Behavior Detection

Back in time, behavior recognition was utilized to analyze military telegraph operator’s typing rhythm to track the movement of the troops [115]. Today, gestures for authentication purposes may range from conventional to “hard-to-mimic” ones, since motor-programmed skill results in the movement being organized before the actual execution [116].

A modern example of such identification is the process of tapping the smartphone screen [117,118]. This approach could be easily combined with any text-input authentication methods as a typing pattern is unique for each person [119–121]. In case the MFA system is specifically developed for predefined gesture analysis [122], the user is required to replicate a previously learned movement while holding or wearing the sensing device [123–125].

A natural step of authentication for widely used handheld and wearable devices is the utilization of accelerometer fingerprinting [126,127]. For instance, each smartphone holder could be verified based on the gait pattern by continuously monitoring the accelerometer data that is almost impossible to fake by another individual [128].
For in-vehicle authentication, the integral system is expected to monitor the driver-specific features [129,130], which could be analyzed from two perspectives: (i) vehicle-specific behavior: steering angle sensor, speed sensor, brake pressure sensor, etc. [131,132]; and (ii) human factors: music played, calls made, presence of people in the car, etc. [133]. Another important blocker-factor is alcohol sensor. The engine start function could be blocked in case when the level of alcohol in the cabin is above an acceptable legal limit [134].

2.2.2. Beam-Forming Techniques

From the telecommunication perspective, Radio-frequency Identification (RFID) and Near-Field Communication (NFC) techniques have already observed widespread adoption and acceptance within the community [135]. Recent trends in physical-layer security claim that utilizing wireless Multiple-Input and Multiple-Output (MIMO) solutions to locate the source of the signal may become a significant breakthrough in validating the token on the user body [136–138].

2.2.3. Occupant Classification Systems (OCS)

Some vehicular systems already have the OCS solutions integrated in consumer cars [139]. A system of sensors can detect who is currently in the passenger/driver seat by utilizing, for example, weight or posture and automatically adjusting the vehicle to personal needs [140–142].

2.2.4. Electrocardiographic (ECG) Recognition

ECG data could be collected from the user’s smart watch or activity tracker and compared with an individually stored pattern [143,144]. The main benefit of using this factor for authentication is that ECG signals emerge as a potential biometric modality with the advantage of being difficult (or close to impossible) to mimic. The only way is by utilizing the existing personal recording [145].

2.2.5. Electroencephalographic (EEG) Recognition

This solution is based on the brain waves analysis and could be considered from the fundamental philosophical proposition “Cogito ergo sum” by R. Descartes, or “I think, therefore I am” [146]. It allows for obtaining a unique sample of the person’s brain activity pattern [147]. Formerly, EEG data capture could have been performed only in clinical settings by using invasive probes under the skull or wet-gel electrodes arrayed over the scalp. Today, the simple EEG collection is possible by utilizing market-available devices having the size of a headset [148].

2.2.6. DNA Recognition

Human cell lines are an essential resource for research, which is most frequently used in reverse genetic approaches or as in vitro models of human diseases [149]. It is also a source of unique DNA fingerprinting information [106]. Even though the process is time-consuming and expensive, it may be potentially utilized to pre-authorize the user to the highly secure facility along with other factors. Subsequently, a comparison of the main indicators for the already deployed and emerging factors [150] is given in Table 1. The factors/sensors are evaluated based on the following parameters:

- *Universality* stands for the presence of factor in each person;
- *Uniqueness* indicates how well the factor differentiates one person from another;
- *Collectability* measures how easy it is to acquire data for processing;
- *Performance* indicates the achievable accuracy, speed, and robustness;
- *Acceptability* stands for the degree of acceptance of the technology by people in their daily life;
- *Spoofing* indicates the level of difficulty to capture and spoof the sample.
Table 1. Comparison of suitable factors for MFA: H—high; M—medium; L—low; n/a—unavailable.

<table>
<thead>
<tr>
<th>Factor</th>
<th>Universality</th>
<th>Uniqueness</th>
<th>Collectability</th>
<th>Performance</th>
<th>Acceptability</th>
<th>Spoofing</th>
</tr>
</thead>
<tbody>
<tr>
<td>Password</td>
<td>n/a</td>
<td>L</td>
<td>H</td>
<td>H</td>
<td>H</td>
<td>H</td>
</tr>
<tr>
<td>Token</td>
<td>n/a</td>
<td>M</td>
<td>H</td>
<td>H</td>
<td>H</td>
<td>H</td>
</tr>
<tr>
<td>Voice</td>
<td>M</td>
<td>L</td>
<td>M</td>
<td>L</td>
<td>H</td>
<td>H</td>
</tr>
<tr>
<td>Facial</td>
<td>H</td>
<td>L</td>
<td>M</td>
<td>L</td>
<td>H</td>
<td>M</td>
</tr>
<tr>
<td>Ocular-based</td>
<td>H</td>
<td>M</td>
<td>M</td>
<td>L</td>
<td>H</td>
<td>M</td>
</tr>
<tr>
<td>Fingerprint</td>
<td>M</td>
<td>H</td>
<td>M</td>
<td>H</td>
<td>M</td>
<td>H</td>
</tr>
<tr>
<td>Hand geometry</td>
<td>M</td>
<td>M</td>
<td>M</td>
<td>M</td>
<td>M</td>
<td>M</td>
</tr>
<tr>
<td>Location</td>
<td>n/a</td>
<td>L</td>
<td>M</td>
<td>H</td>
<td>M</td>
<td>H</td>
</tr>
<tr>
<td>Vein</td>
<td>M</td>
<td>M</td>
<td>M</td>
<td>M</td>
<td>M</td>
<td>M</td>
</tr>
<tr>
<td>Thermal image</td>
<td>H</td>
<td>H</td>
<td>L</td>
<td>M</td>
<td>H</td>
<td>H</td>
</tr>
<tr>
<td>Behavior</td>
<td>H</td>
<td>H</td>
<td>L</td>
<td>L</td>
<td>L</td>
<td>L</td>
</tr>
<tr>
<td>Beam-forming</td>
<td>n/a</td>
<td>M</td>
<td>L</td>
<td>L</td>
<td>L</td>
<td>H</td>
</tr>
<tr>
<td>OCS</td>
<td>n/a</td>
<td>L</td>
<td>L</td>
<td>L</td>
<td>L</td>
<td>M</td>
</tr>
<tr>
<td>ECG</td>
<td>L</td>
<td>H</td>
<td>L</td>
<td>M</td>
<td>M</td>
<td>L</td>
</tr>
<tr>
<td>EEG</td>
<td>L</td>
<td>H</td>
<td>L</td>
<td>M</td>
<td>L</td>
<td>L</td>
</tr>
<tr>
<td>DNA</td>
<td>H</td>
<td>H</td>
<td>L</td>
<td>H</td>
<td>L</td>
<td>L</td>
</tr>
</tbody>
</table>

However, many other issues are to be addressed while integrating the MFA for the end users. In the following section, we elaborate on those challenges and formalize the recommendations for improved ease of integration.

3. MFA Operation Challenges

An integration of novel solutions has always been a major challenge for both developers and managers. The key challenges are presented in Figure 3. In the first place, user acceptance is a critical aspect for the adoption of strong identity and multi-factor authentication. While adopting and deploying MFA solutions, it is required to follow a careful and thorough approach—where most challenges arise from opportunities and potential benefits [151].

3.1. Usability

The main usability challenges emerging in the authentication process could be characterized from three perspectives [152]:

- **Task efficiency**—time to register and time to authenticate with the system;
- **Task effectiveness**—the number login attempts to authenticate with the system;
- **User preference**—whether the user prefers a particular authentication scheme over another.

In addition to the approaches discussed previously, researchers have already started an investigation of more specific effects in the authentication procedures based on a variety of human factors. The authors of [153] provided a study on how the user age affects the task efficiency in cases of PIN and graphic access mechanisms. It is concluded that younger generation can spend up to 50 percent less time to pass the authentication procedure in both cases. Interestingly, the authors of [154] have shown that gender, in the same case, does not affect the results.

Another direction in the authentication mechanisms usability is related to cognitive properties of the selected human [155]. The work in [156] offered an overview on how to make the passwords memorable while keeping them relatively usable and secure at the same time. Paper by Belk et al. [157] delivered a research on the task completion efficiency and effectiveness among the conventional passwords and the realistic ones. The results revealed that, for most of the participants, the utilization of graphic passwords requires more time than for the textual ones. However, cognitive differences between users, i.e., being Verbal or Imager [152], affect the task completion significantly. Here, Verbals complete the text-based tasks faster than Imagers and vice versa. The work by Ma et al. [158] studied
the impact of disability (Down syndrome) in the same two scenarios. It was once again confirmed that textual passwords are utilized better compared to the graphical ones.

**Figure 3. Main operational challenges of MFA.**

In addition, the properties of the authentication device play a major role in this process. The authors of [159] investigated the usability of textual passwords on mobile devices. It was proven that using a smartphone or other keyboardless equipment for creating a password suffers from poor usability as compared to conventional personal computers. Another work [160] confirmed the same theory from a task efficiency perspective.

Today, most of the online authentication services are knowledge-based [161], i.e., depend on the username and password combination. More complex systems require the user to interact with additional tokens (one-time passwords, code generators, phones, etc.). Complementing traditional authentication strategies, MFA is not feasible without biometrics. From this perspective, the work in [62] provided an analysis on how gamification and joy can positively impact the adoption of new technology. The gesture-related user experience research conducted in [162] showed that security and user experience do not necessarily need to contradict one other. This work also promoted pleasure as the best way for fast technology adoption. The reference [163] addressed the usability of the ECG solution for authentication, and it was concluded that the application of ECG is not yet suitable for dynamic real-life scenarios.

Many researchers promoted the utilization of personal handheld devices to be utilized during the MFA procedure. Michelin et al. [164] proposed using the smartphone’s camera for facial and iris recognition while keeping the decision-making in the cloud. Another work on biometric authentication for an Android device [165] demonstrated an increased level of satisfaction related to higher task efficiency achieved with the MFA solution. Reference [166] studied the usability and practicality of biometric authentication in the workplace. It was concluded that the ease of technology utilization and its environmental context play a vital role—the integration and the adoption will always incur additional and unexpected resource costs.

An extremely important problem of MFA usability roots in the fact that “not all users can use any given biometric system” [167]. People who have lost their limb due to an accident may not be able to authenticate using a fingerprint. Visually impaired people may have difficulties using the iris-based authentication techniques.

Biometric authentication requires an integration of new services and devices that results in the need for additional education during adoption, which becomes more complicated for seniors and due to related understandability concerns. One fact is clear—user experience plays a prominent role in
successful MFA adoption; some say, “user comes first” [168]. Today, research in usable security for knowledge-based user authentication is in the process of finding a viable compromise between the usability and security—many challenges remain to be addressed and will arise soon.

3.2. Integration

Even if all the usability challenges are resolved during the development phase, integration brings further problems from both technological and human perspectives.

Most of the consumer MFA solutions remain hardware-based [52]. Generally, “integrating physical and IT security can reap considerable benefits for an organization, including enhanced efficiency and compliance plus improved security” [169]. However, convergence is not so simple. Related challenges include bringing the physical and the IT security teams together, combining heterogeneous system components, and upgrading the physical access systems.

While developing the MFA system, biometrics independence should be considered carefully, i.e., assurance of interoperability criteria should be met [170]. The framework needs to have functionality to handle the biometric data from sensors other than the initially deployed ones [171]. The utilization of multi-biometrics, that is, simultaneous usage of more than one factor should also be taken into account [172].

Another major interoperability concern is vendor dependency [173]. Enterprise solutions are commonly developed as stand-alone isolated systems that offer an extremely low level of flexibility. Integration of newly introduced to the market sensors would require complicated and costly updates, which most probably will not be considered soon.

Further, it should be noted that most of the currently available MFA solutions are not fully/partially open source. This introduces the questions of trustworthiness and reliability to the third party service providers. The available level of transparency delivered by both hardware and software vendors should be taken into consideration while selecting the MFA framework in the first place.

3.3. Security and Privacy

Any MFA framework is a digital system composed of critical components, such as sensors, data storage, processing devices, and communication channels [174]. All of those are typically vulnerable to a variety of attacks at entirely different levels, ranging from replay attempts to adversary attacks [175]. Security is thus a necessary tool to enable and maintain privacy. Therefore, we begin with the attacks executed on the input device itself [176]. Letting only the legitimate controller access and process sensitive personal data exposes the community to the main risks related to MFA security that are listed further.

The first of the key risks is related to data spoofing that would be successfully accepted by the MFA system [177]. Notably, due to biometrics being used by a variety of MFA frameworks, a glaring opportunity for the attacker to analyze both the technology underlying the sensor and the sensor itself results in revealing the most suitable spoofing materials. The main goal of the system and hardware architects is to provide either a secure environment or, in case it is not possible, to consider the related spoofing possibilities in advance. A risk of capturing either physical or electronic patterns and reproducing them within the MFA system should be addressed carefully.

Conventionally, the safeguard to protect against electronic replay attacks requires utilization of a timestamp [178]. Unfortunately, a biometric spoofing attack is fairly simple to execute [179]. Even though biometrics can improve the performance of the MFA system, they can also increase the number of vulnerabilities that can be exploited by an intruder. Further risk is sensitive data theft during the transmission between the sensor and the processing/storage unit. Such theft may primarily occur due to insecure transmission from the input device through extraction and matching blocks to the database, and there is potential for an attack [180]. The required levels of data safety should be guaranteed to resist against this risk type [181,182].
Another opportunity to attack the MFA system is by capturing the secret data sample [183]. For knowledge factors, the system would be immediately compromised in case zero-knowledge solutions are not utilized [184]. Specific interest is dedicated to capturing a biometric sample that could not be updated or changed over time [185]. Hence, protection of the biometric data requires a higher level of security during capture, transmission, storage, and processing phases [186].

The following risk is related to the theft from the data storage. Conventionally, databases are stored in a centralized manner, which offers a single point of failure [187,188]. At the same time, some of the remote systems contacting the database are not always legitimately authorized to access the personal data stored. High level of isolation is required to protect the data from theft in addition to utilizing irreversible encryption [189]. Subsequent risk is related to location-related attacks. The GPS signal could be vulnerable to position lock (jamming) or to feeding the receiver with false information, so that it computes an erroneous time or location (spoofing) [190,191]. Similar techniques may be applied to cellular- and WLAN-based location services [192,193].

Finally, being an information technology system, MFA framework should deliver relatively high levels of “throughput” [194], which reflects the capability of a system to meet the needs of its users in terms of the number of input attempts per time period [195]. Even if the biometrics are considered suitable in every other aspect, but the system can only perform, e.g., one biometrics-based match per hour, whereas it is required to operate at 100 samples per hour, such a solution should not be considered as feasible. The recommendation here is to select appropriate processing hardware for the server/capture side.

The MFA security framework should also support a penetration testing panel to assess its potential weaknesses. Today, the developers are often conducting external audit to evaluate the risks and act based on such evaluation for more careful planning. The MFA system should thus be assessed to deliver a more secure environment.

3.4. Robustness to Operating Environment

Even if the security and privacy aspects are fully resolved, the biometric systems, mainly fingerprinting, were falling short of fulfilling the “robustness” requirement since the very beginning of their journey [196]. This was mainly due to the operational trials being conducted in the laboratory environment instead of the field tests. One distinct example is voice recognition, which was highly reliable in a silent room but failed to verify the user in urban landscapes.

A similar problem applies to early facial recognition techniques, which failed to operate without adequate light support, quality camera, etc. [197]. The flip side of the coin was the need for continuous supervision of the examined subject. Even today, there are either bits of advice on where to look/place fingers, or there is visual aid available during the security check. The lack of experience in machine-to-human interaction is commonly analyzed with Failure to Enroll (FTE) as well as Failure to Acquire (FTA) rates [198]. They both depend on the users themselves as well as the additive environmental noise.

Since a significant part of MFA is highly dependent on biometry, it could be classified as inherently probabilistic due to such nature [199]. The base of the biometric authentication lies in the field of pattern matching, which in turn relies on approximation. Approximate matching is a critical consideration in any MFA system, since difference between users could be crucial due to a variety of factors and uncertainty. The image captured during a fingerprint scan would be different every time it is observed because of the presentation angle, pressure, dirt, moisture, or differentiation of sensors even if taken of the same person.

Two important error rates used to quantify the performance of a biometric authentication system are FAR and FRR. FAR is the percentage of impostors inaccurately allowed as genuine users. It is defined as the ratio of the number of false matches to the total number of impostor match attempts. FRR is the number of genuine users rejected from using the system, which is defined as the ratio of the number of false rejections to the total number of genuine match attempts.
Literature further recommends the utilization of the Crossover Error Rate (CER) in addition to the previously discussed metrics [200]. This parameter is defined as the probability of the system being in a state where FAR equals to FRR. The lower this value is, the better the system performs. According to [201], “Higher FAR is preferred in systems where security is not of prime importance, whereas higher FRR is preferred in high-security applications”. The point of equality between FAR and FRR is referred to as Equal Error Rate (EER) [202]. Based on the above, it could be once again concluded that a system utilizing solely biometrics may not be considered as a preferred MFA framework.

By analyzing the above listed challenges, it is possible to evaluate and assess the entire MFA system. In what follows, we propose an approach to enable MFA for vehicular integration based on the availability of a large number of sensors in modern vehicles.

4. Enabling Flexible MFA Operation

In this work, we offer a new authentication scheme that focuses on the vehicle-to-everything (V2X) scenarios, since cars of today are already equipped with multiple sensors that could potentially be utilized for MFA. Conventionally, the user has a username/password/PIN/token [203] and will additionally be asked to utilize a biometric factor, such as facial features or fingerprints. The general overview supported by a follow-up discussion is given in Figure 4. If the authentication procedure fails to establish trust by using this combination of factors, then the user will be prompted to authenticate by utilizing another previously registered factor or a set of those. This MFA system may not only verify the accuracy of the user input but also determine how the user interacts with the devices, i.e., analyze the behavior. The more the user interacts with the biometric system, the more accurate its operation becomes.

![Figure 4. Current and emerging MFA sensors for vehicles.](image)

Another feature of the discussed scenario is the actual sensor usability in case of interaction with a car [204]. If a sensor (e.g., a fingerprint reader) is being utilized and that device is not available from where the user is attempting to log in or gain access—the user experience becomes inadequate. Having a dual-purpose device—smartphone or smartwatch (suitable for executing the information security primitives [205]), which the user already has in his or her possession—as an additional MFA factor (not only as a token) makes both the system costs and usability much more reasonable [206].

The presence of large amounts of sensor data brings us to the logical next step of its application in MFA. We further envision potential utilization of the corresponding factors to authenticate the user
without implementing a dedicated “verifier” with the actual biometric data except for the one collected in real time.

4.1. Conventional Approach

One of the approaches considered within the scope of this work is based on utilizing Lagrange polynomials for secret sharing [207]. The system secret $S$ is usually “split” and distributed among a set of key holders. It could be recovered later on, as described in [208–210] and numerous other works, as

$$f(x) = S + a_1x + a_2x^2 + \cdots + a_{l-1}x^{l-1},$$

$$f(0) = S,$$  

(1)

where $a_i$ are the generated polynomial indexes and $x$ is a unique identification factor $F_i$. In such systems, every key holder with a factor ID obtains its own unique key share $S_{ID} = f(ID)$. In conventional systems, it is required to collect any $l$ shares $\{S_{ID_1}, S_{ID_2}, \ldots, S_{ID_l}\}$ of the initial secret to unlock the system, while the curve may offer $n > l$ points, as it is shown in Figure 5. The basic principle behind this approach is to specify the secret $S$ and use the generated curve based on the random coefficients $a_i$ to produce the secret shares $S_i$. This methodology is successfully utilized in many secret sharing systems that employ the Lagrange interpolation formula [211,212].

![Figure 5. Lagrange secret sharing scheme.](image)

Unfortunately, this approach may not be applied for the MFA scenario directly [213], since the biometric parameters are already in place, i.e., we can neither assign a new $S_i$ to a user nor modify them. On the one hand, the user may set some of the personal factors, such as password, PIN-code, etc. On the other hand, some of them may be unchangeable (biometric parameters and behavior attributes). In this case, an inverse task where the shares of the secret $S_{ID}$ are known as factor values $S_i$ is to be solved. Basically, $S_i$ are fixed and become unique $\{S_1, S_2, \ldots, S_l\}$ when set for a user. In this case, $S$ is the secret for accessing the system and should be acquired with the user factor values. A possible solution based on the reversed Lagrange interpolation formula is proposed in the following subsection.

4.2. Proposed Reversed Methodology

In this work, we consider the MFA system with explicit $l$ factors $F$. Each factor $F_i$ has a unique secret $S_i$ obtained with the corresponding procedure (PIN, fingerprint, etc.) from the user. In the worst case, it is related to the biometric data—the probability that it changes over time is low. The corresponding factors and secrets could then be represented as
\[ F_1 : S_1, \]
\[ F_2 : S_2, \]
\[ \ldots \]
\[ F_l : S_L, \]
\[ F_{l+1} : T, \]

where \( S_i \) is the secret value obtained from the sensor (factor), \( l \) is the number of factors required to reconstruct the secret, and \( F_{l+1} \) is a timestamp collected at time instant \( T \).

It is important to note that providing the actual secrets to the verifier is not an option, especially in case of sensitive biometric data, because a fingerprint is typically an unchangeable factor. Hence, letting even a trusted instance obtain the corresponding data is a questionable step to make. Conversely, compared to the method considered in Section 4.1, the modified algorithm implies that \( S_i \) are obtained from the factors (only one polynomial describes the corresponding curve), as it is shown in Figure 5. In other words, the proposed methodology produces the system secret \( S \) based on the collected factor values \( S_i \) instead of assigning them in the first place.

A system of equations connected to the Lagrange interpolation formula with the factors, their values, and the secret for the system access is

\[
\begin{align*}
S_1 &= S + a_1 F_1 + a_2 F_2 + \cdots + a_{l-1} F_1^{l-1} + a_l F_l, \\
S_2 &= S + a_1 F_2 + a_2 F_2^2 + \cdots + a_{l-1} F_2^{l-1} + a_l F_2, \\
\vdots \\
S_l &= S + a_1 F_l + a_2 F_l^2 + \cdots + a_{l-1} F_l^{l-1} + a_l F_l, \\
T &= S + a_1 T + a_2 T^2 + \cdots + a_{l-1} T^{l-1} + a_l T,
\end{align*}
\]

(3)

where \( a_i \) are the corresponding generated coefficients, \( f(x) = S + a_1 x + a_2 x^2 + \cdots + a_{l-1} x^{l-1}, \) and \( f(0) = S \). The system in Equation (3) has only one solution for \( S \) and it is well known from the Lagrange interpolation formula.

Lemma 1. One and only one polynomial curve \( f(x) \) of degree \( l - 1 \) could be described by \( l \) points on the plane \( (x_1, y_1), (x_2, y_2), \ldots, (x_l, y_l) \)

\[
f(x) = a_0 + a_1 x + \cdots + a_{l-1} x^{l-1}, \{f(x_i) = y_i\}_{i=1}^l.
\]

Hence, the system secret \( S \) may be recovered based on \( l \) collected shares as given by the conventional Lagrange interpolation formula without the need to transfer the original factor secrets \( S_i \) to the verifier. Hence, the sensitive person-related data is kept private, as

\[
S = (-1)^l \sum_{i=1}^{l+1} S_i \prod_{j=1, j \neq i}^{l+1} \frac{F_j}{F_i - F_j},
\]

(4)

where \( F_{l+1} = T \). The proposed modifications are required to assure the uniqueness of the acquired data, see Figure 6.

Due to the properties of the Lagrange formulation, there can only be one curve described by the corresponding polynomial (Lemma 1); therefore, each set of \( [F_i : S_i] \) will produce its unique \( S \). However, if the biometric data collected by MFA has not been changed over time, the secret will always remain the same, which is an obvious vulnerability of the considered system. On the other hand, a simple addition of the timestamp should always produce a unique curve, as it is shown in Figure 6 for \( T, T_1, \) and \( T_2 \).
The proposed solution provides robustness against the case where all $S_i$ remain unchanged over time. This is achieved by adding a unique factor of time $T$, which enables the presence of $F_l$ with the corresponding secret. It is necessary to mention that the considered threshold scheme based on the Lagrange interpolation formula utilizes Rivest–Shamir–Adleman (RSA) mechanism or ElGamal encryption/decryption algorithm for authentication during the final step. In this case, it is proven that we obtain a secure threshold scheme related to secrets $S_i$ in [214].

4.3. Proposed MFA Solution for V2X Applications

Indeed, our proposed solution may operate out-of-the-box in case where all $l$ factors are present. The system may thus provide a possibility to identify and report any outdated factor information—for example, weight fluctuation [215]. Access to a service could be automated when some of the factors are not present [216]. We further elaborate on this feature in the current subsection.

4.3.1. Factor Mismatch

Assuming that the number of factors in our system is $l = 4$, the system secret $S$ can be represented in a simplified way as a group of

$$S \leftarrow \begin{bmatrix} F_1 & F_2 & F_3 & F_4 \end{bmatrix}.$$

Here, if any of $S_i$ are modified—the secret recovery mechanism would fail. An improvement to this algorithm is delivered by providing separate system solutions $S_i$ for a lower number of factors collected. Basically, for $I = 3$, the number of possible combinations of factors with one missing is equal to four, as follows

$$S_1 \leftarrow \begin{bmatrix} F_1 & F_2 & F_3 \end{bmatrix} ,$$
$$S_2 \leftarrow \begin{bmatrix} F_1 & F_3 & F_4 \end{bmatrix} ,$$
$$S_3 \leftarrow \begin{bmatrix} F_1 & F_2 & F_4 \end{bmatrix} ,$$
$$S_3 \leftarrow \begin{bmatrix} F_2 & F_3 & F_4 \end{bmatrix}.$$ (5)
The device may thus grant access based on a predefined risk function policy. As the second benefit, it can inform the user (or the authority) that a particular factor $F_i$ has to be updated based on the failed $S_i$ combination. Indeed, this modification brings only marginal transmission overheads, but, on the other hand, enables higher flexibility in authentication and missing factor validation.

4.3.2. Cloud Assistance

Another important scenario for MFA is potential assistance of the trusted authority in $F_i : S_i$ mismatch or loss. In case when the user fails to present a sufficient number of factors, the trusted authority can be requested to provide the temporary factor keys, as it is demonstrated in Figure 7.

For example, assume that the user forgot or lost two factors $F_2$ and $F_3$ with the corresponding keys $S_1 = f(F_1)$ and $S_2 = f(F_2)$. The trusted authority is willing to assist in authentication—two temporary keys $S_{F_1} = f(F_1)$ and $S_{F_2} = f(F_2)$ are thus generated and sent to the user via a secure channel. Obtaining these keys and applying the Lagrange interpolation formula with RSA or ElGamal encryption/decryption-based threshold authentication procedure involves the following factors and keys

\[
F_1 : S_1,
F_2 : S_2,
\ldots
F_l : S_L,
F_{l+1} : T,
\Phi_1 : S_{\Phi_1},
\Phi_2 : S_{\Phi_2},
\]

as described in [214]. This allows for gaining access to the device.

The proposed solution is designed explicitly to complete the MFA step of the authentication, that is, its usage for SFA and 2FA is not recommended. This is mainly due to the features of the Lagrange interpolation formula. Basically, in the SFA case and without the $F_{l+1} : T$ factor, the equation at hand can be simply represented as $S_1 = S + b_1 F_1$, i.e., it will become ‘a point’. Even adding a random timestamp factor will not provide any valuable level of biometric data protection, since an eavesdropper could be able to immediately recover the factor secret.
The above is not suitable for the 2FA either, since providing two factors allows the curve to have linear behavior, i.e., the eavesdropper is required two attempts to recover the secrets. However, adding a timestamp factor here allows for providing the necessary level of safety with three actual factors, as discussed below.

### 4.4. Potential Evaluation Techniques

Conventionally, authentication systems utilizing only the knowledge of ownership factors operate in pass/fail mode, i.e., the input data is either correct or incorrect. When it comes to using biometrics, the system faces potential errors during the biometric sample capturing, which was discussed previously in Section 3.4. We further elaborate on our proposed methodology from the crucial FAR/FRR perspective.

Typically, the FAR/FRR parameters of a sensor are provided by vendors based on the statistically collected data [217]. For the MFA framework, we assume two possible decisions made during the user authentication phase, as it is displayed in Figure 8: (i) $H_0$—the user is not legitimate; or (ii) $H_1$—the user is legitimate. These form the entire sample space of $P(H_0) + P(H_1) = 1$. The risk policy is assumed to be handled by the authentication system owner who also sets up the distributions of $P(H_0)$ and $P(H_1)$.

![MFA system mode](image)

Figure 8. MFA system mode. $P_{TH}$ is the selected threshold.

Generalizing, there might be $n$ biometric sensors collecting the user input data. Each individual sensor measurement from the set $Z = \{z_1, \ldots, z_n\}$ is distributed within $[0, 1]$, and this set is further analyzed under the conditions of two previously considered hypotheses. The measurements delivered from the sensors could be processed in two different ways as introduced in the sequel.

#### 4.4.1. Strict Decision Methodology

Each sensor decides whether the user is legitimate or not by returning either accept or reject. The MFA system then combines the collected results and provides a group decision based on the resulting vector. Hence, it is possible to utilize the threshold decision functions or weighted threshold functions depending on the reliability of the sensor.

For the first case, the sensor will return the value $z_i$, $z_i = [0; 1]$, which could be interpreted as either YES or NO. Then, the conditional probabilities $P(z_i | H_0)$ and $P(z_i | H_1)$ are defined by $\text{FAR}_i$ and $\text{FRR}_i$ values, respectively, for $i$-th sensor. Here, $\text{FAR}_i$ and $\text{FRR}_i$ are taken at the CER/EER point, e.g., $z_i$ is selected at the point where $\text{FAR}_i = \text{FRR}_i$. Generally, this methodology reflects the scenarios of ownership or knowledge factors from the biometric perspective.
4.4.2. Probabilistic Decision Methodology

The sensor responds with a result of its measurements as well as a probabilistic characteristics. Further, the data is merged before the final decision is made. Therefore, the entire set of the measured data could be utilized when making a group decision and, accordingly, a common result might be established based on the set collected from all sensors.

In the second case, the sensor returns a result of the measurements as well as the template comparison in the form of a match score $z_i (0 \leq z_i \leq 1)$. For each of the values $z_i$, the conditional probability $P(z_i | H_0)$ is calculated based on the FAR values at $z_i$. In addition, the conditional probability $P(z_i | H_1)$ is determined by FRR values at $z_i$.

This approach offers an opportunity to consider the strict decision methodology as a simplified model of the probabilistic one for the case where FAR and FRR are given only in one point. Here, the measurement result can only take two values, i.e., higher or lower than the selected threshold.

4.4.3. Evaluation

In this work, we consider a more general case of the probabilistic decision-making methodology, while a combination of the measurement results for the individual sensors is made similarly to the previous works by using the Bayes estimator [218]. Since the outcomes of measurements have a probabilistic nature, the decision function is suitable for the maximum a posteriori probability solution.

In more detail, the decision function may be described as follows. At the input, it requires a conditional probability of the measured value from each sensor $P(z_i | H_j), j \in \{0; 1\}$ together with a priori probabilities of the hypotheses $P(H_0)$ and $P(H_1)$. The latter values could be a part of the company’s risk policy as they determine the degree of confidence for specific users. Then, the decision function evaluates the a posteriori probability of the hypothesis $P(H_1 | Z)$ and validates that the corresponding probability is higher than a given threshold $P_{TH}$.

The measurement-related conditional probabilities can be considered as independent random variables; hence, the general conditional probability is as follows:

$$P(Z | H_j) = \prod_{z_i \in Z} P(z_i | H_j), j \in \{0; 1\}. \quad (7)$$

Further, the total probability $P(Z)$ is calculated as

$$P(Z) = \prod_{z_i \in Z} P(z_i | H_0)P(H_0) + \prod_{z_i \in Z} P(z_i | H_1)P(H_1), \quad (8)$$

where $P(z_i | H_j), j \in \{0; 1\}$ are known from the sensor characteristics, while $P(H_0)$ and $P(H_1)$ are a priori probabilities of the hypotheses (a part of the company’s risk policy).

Based on the obtained results, the posterior probability for each hypothesis $H_j, j \in \{0; 1\}$ can be produced as

$$P(H_1 | Z) = \frac{\prod_{z_i \in Z} P(z_i | H_1)P(H_1)}{P(Z)}. \quad (9)$$

For a comprehensive decision over the entire set of sensors, the following rule applies

$$P(H_1 | Z) > P_{TH} \Rightarrow \{\text{Accept}\}, \text{ else } \{\text{Reject}\}. \quad (10)$$

As a result, the decision may be correct or may lead to an error. The FAR and FRR values could then be utilized for selecting the appropriate threshold $P_{TH}$ based on all of the involved sensors.

5. Discussion and Future Prospects

Today, authentication matters more than ever before. In the digital era, most users will rely on biometrics in matters concerning systems security and authorization to complement the conventional
passwords. Even though privacy, security, usability, and accuracy concerns are still in place, MFA becomes a system that promises the security and ease of use needed for modern users while acquiring access to sensitive data.

Without a doubt, biometrics are one of the key layers to enable the future of MFA. This functionality is often regarded not standalone but as a supplement to traditional authentication approaches like passwords, smart cards, and PINs. Combining two or more authentication mechanisms is expected to provide a higher level of security when verifying the user. The expected evolution towards MFA is rooted in the synergistic biometric systems that allow for significantly improved user experience and MFA system throughput, which would be beneficial for various applications (see Figure 9). Such systems will intelligently couple all three factor types, namely, knowledge, biometrics, and ownership.

![Figure 9. Biometric MFA for the airport scenario.](image)

Since conventional single-factor systems of today are based on only one parameter (unimodality property), if its acquisition is affected in any way (be it noise or disruption), the overall accuracy will degrade. As a reminder, collecting a single type of non-knowledge related data, e.g., biometrics, could exclude part of the user population when particular disabilities are present. Moreover, spoofing this only factor is a relatively simple task.

One of the most promising directions in MFA is behavior-based biometrics providing entirely new ways of authenticating the users. The solutions that are based on muscular memory, e.g., writing or gestures, coupled with machine learning become more prominent examples. Already today, software can extrapolate user handwriting and reach the confidence levels of above 99.97 percent [219]. More forward-looking MFA sources to be utilized in the nearest future are heart and brain [220]. The attractive area of ECG and EEG analysis is also expected to provide unique identification samples for each subject.

Another military-inspired research activity already shows the capability to identify the users based on the way they interact with computer [221]. This approach takes into consideration the typing speed, typical spelling mistakes, writing rhythm, and other factors [222]. The appropriate terminology is not settled yet, and some call this methodology Passive Biometrics [223], while others name it Continuous Authentication [224]. It results in having a unique fingerprint of the user–computer interaction pattern, which is extremely difficult to replicate.

All of the discussed MFA scenarios require significant memory resources to statistically analyze the input data and store the biometric samples even if utilizing different optimization techniques [225,226]. A very promising direction of the MFA development is therefore in the area of neural networks and Big Data [227]. Here, many successful applications have been known to the community for more than a decade. Examples could be found in [228–230] where conventional factors, such as iris, retina,
fingerprints, etc., are considered. Utilizing neural networks for the next-generation biometrics is the most likely way to proceed due to presently high levels of the analysis complexity [231,232].

In summary, biometric technology is a prominent direction driven by the mobile device market. The number of smartphones to be sold only in the US is expected to reach 175 million units by 2018 with the corresponding market to exceed $50.6B in revenues by 2022 [233,234]. It is believed that a strong push towards the utilization of biometrics in many areas of life is imminent, since most of the flagman devices are already equipped with the fingerprint scanner and facial recognition technology in addition to convention PIN codes.

This work provided a systematic overview of the state-of-the-art in both technical and usability issues, as well as the major challenges in currently available MFA systems. In this study, we discussed the evolution of authentication from single- through two- and towards multi-factor systems. Primarily, we focused on the MFA factors constituting the state-of-the-art, future possible directions, respective challenges, and promising solutions. We also proposed an MFA solution based on the reversed Lagrange polynomial as an extension of Shamir’s Secret Sharing scheme, which covers the cases of authenticating the user even if some of the factors are mismatched or absent. It also helps qualify the missing factors without disclosing the sensitive data to the verifier.

Acknowledgments: The work of the second author is supported by the Academy of Finland.

Author Contributions: A.O. prepared the state-of-the-art; N.M. and T.M. conducted the analysis of challenges; S.B. designed the flexible MFA solution; A.O., S.B., N.M., S.A., T.M, and Y.K. wrote the paper.

Conflicts of Interest: The authors declare no conflict of interest.

Abbreviations

MFA Multi-Factor Authentication
SFA Single-Factor Authentication
2FA Two-Factor Authentication
SSS Shamir’s Secret Sharing
PIN Personal Identification Number
ID Identification Number
ATM Automated Teller Machine
FAR False Accept Rate
FRR False Reject Rate
PPG Photoplethysmography
RFID Radio-Frequency Identification
NFC Near-Field Communication
OCS Occupant Classification Systems
ECG Electrocardiography
EEG Electroencephalography
GPS Global Positioning System
FTE Failure to Enroll
FTA Failure to Acquire
CER Crossover Error Rate
EER Equal Error Rate
V2X Vehicle-to-Everything
IAM Identity and Access Management

References


Publication VII

© 2018 IEEE. Reprinted, with permission, from


In reference to IEEE copyrighted material which is used with permission in this thesis, the IEEE does not endorse any of Tampere University of Technology’s products or services. Internal or personal use of this material is permitted. If interested in reprinting/republishing IEEE copyrighted material for advertising or promotional purposes or for creating new collective works for resale or redistribution, please go to http://www.ieee.org/publications_standards/publications/rights/rights_link.html to learn how to obtain a License from RightsLink.
Challenges of Multi-Factor Authentication for Securing Advanced IoT (A-IoT) Applications

Aleksandr Ometov, Vitaly Petrov, Sergey Bezzateev, Sergey Andreev, Yevgeni Koucheryavy, and Mario Gerla

Abstract—The unprecedented proliferation of smart devices together with novel communication, computing, and control technologies have paves the way for the Advanced Internet of Things (A-IoT). This development involves new categories of capable devices, such as high-end wearables, smart vehicles, and consumer drones aiming to enable efficient and collaborative utilization within the Smart City paradigm. While massive deployments of these objects may enrich people’s lives, unauthorized access to the said equipment is potentially dangerous. Hence, highly-secure human authentication mechanisms have to be designed. At the same time, human beings desire to comfortably interact with their owned devices on a daily basis, thus demanding the authentication procedures to be seamless and user-friendly, mindful of contemporary urban dynamics. In response to these unique challenges, this work advocates for the adoption of multi-factor authentication for A-IoT, such that multiple heterogeneous methods – both well-established and emerging – are combined intelligently to reliably grant or deny access. We thus discuss the pros and cons of various solutions as well as introduce tools to combine the authentication factors, with an emphasis on challenging Smart City environments. We finally outline the open questions to shape future research efforts in this emerging field.

I. INTRODUCTION AND RATIONALE

The vision of the Internet of Things (IoT) opens a new era of technology penetration into the human lives, which touches upon a wide range of use cases: from Smart Home to Smart City and from Smart Grid to Factory Automation [1]. The numbers of IoT devices that are able to collect, store, combine, and analyze the massive amounts of data around them by producing valuable knowledge and making relevant actions is growing uncontrollably in an attempt to offer decisive societal benefits while handling both routine and critical tasks across multiple verticals [2].

As it simplifies the lives of people, the IoT also brings unprecedented security and privacy risks, since close to any object around us becomes interconnected with others to collect and process sensitive information [3]. The conventional massive IoT involves numerous low-cost devices (e.g., sensors, actuators, and smart meters), with limited computational capabilities and stringent power constraints; hence, the traditional security and privacy solutions had to be reconsidered and adjusted to the specifics of massive IoT. Over the recent decades, security and privacy in IoT remained a major research topic subject to heated discussions e.g., in the area of lightweight cryptography [4], secure connection and trust establishment [5], and privacy-preserving data processing [6]. While there are multiple open problems yet to be solved, the current progress in this field promises to provide the demanded levels of security to these massive IoT deployments.

Meanwhile, in contrast to the massive and low-cost IoT solutions, an emerging trend in today’s IoT is a rapid proliferation of high-end IoT equipment that features more capable connected devices. These include sophisticated wearables (including augmented, virtual, and mixed reality systems), smart vehicles, and consumer drones (see Fig. 1) – that may collectively be named Advanced IoT (A-IoT). These relatively high-cost devices have more abundant performance, memory, and battery resources to execute full-scale security and privacy protocols; thus, the establishment of secure machine-to-machine connections may not be a challenging problem for the A-IoT.

![Fig. 1. Human-centric Advanced IoT (A-IoT) applications in a Smart City.](image-url)

At the same time, a number of specific security and privacy concerns emerge in connection with such systems, since unauthorized access to these powerful devices may lead to severe risks that range from theft of this high-cost equipment (drones or cars) and up to putting human lives in danger by e.g., manipulating with the information projected to the augmented reality grasses or maneuvering smart vehicles uncontrollably [7]. Therefore, reliable assessment of the fact of ownership for the A-IoT devices that belong to both personal and collective
use becomes one of the key challenges that is being faced today [8], which is very different from massive IoT.

In this work, we first systematically review the unprecedented research challenges related to determining the human ownership of the A-IoT systems. We then classify the specific features of the A-IoT that can be employed to securely verify the fact of ownership of the A-IoT devices and map them onto the challenges by illustrating how the features can complement each other while covering the potential issues. We also discuss the concept of multi-factor human authentication to the A-IoT system, where multiple heterogeneous factors are intelligently combined to achieve higher levels of security while not compromising the usability levels of the A-IoT services. We finally enumerate the important practical matters to be resolved on the way towards successful implementation of the introduced concept.

II. CHALLENGES OF DETERMINING OWNERSHIP IN A-IoT

As unauthorized access to A-IoT systems brings severe security threats, the challenge of reliable access control becomes one of the most crucial research problems for securing the A-IoT solutions. An access control procedure can generally be decomposed into user authentication and authorization. The second stage is relatively simpler and can be implemented by conventional discretionary, mandatory, or role-based access control methods. However, the first stage introduces a number of A-IoT-specific research questions that we carefully review in this section.

A. Multi-Modality of Human-Computer Interaction

Today, most of the conventional ICT systems are equipped with advanced input devices, such as keyboards and touch-screens, as well as output devices, most commonly, LCD screens used for human-computer interaction (HCI). Since textual input remains the dominating form of HCI, these systems have historically been adopted for authentication purposes: memorable textual or numerical passwords, possibility to display a hint or visual advanced instructions, etc.

In contrast, the very nature of authentication does not imply text-based commands or responses. Very few of the emerging IoT devices are controlled by a keyboard; hence, the authentication methods based on textual passwords will need to evolve accordingly for them to continue being usable on the mass IoT market.

B. Robustness to Environment and User Behavior

The authentication process of today is typically applied in dedicated, comfortable, and stationary environments. Many such actions occur indoors, where neither weather conditions nor other unpredictable factors can impact the authentication decisions. Even when this process happens outdoors, the input devices to enter the security credentials acquire additional protection to resist the environmental changes up to a certain extent.

However, the A-IoT systems in Smart Cities are mobile by design. Their interactions with a user are spontaneous and occur in uncontrolled and unpredictable environments. Moreover, even under regular weather/environment conditions, the initial state as the user begins interacting with the A-IoT system may be notably different. For example, the user opening a vehicle may be wearing gloves during winter time, such that the fingerprint scanner installed on the door handle may be not be available. Therefore, authentication of A-IoT devices must be made robust to both dynamic environmental conditions and flexible user behavior.

C. High Levels of Reliance and Trust

The wide penetration of the ICT systems on the consumer market and their role in the daily human life have always been associated with a level of trust that people grant to these systems. High trust is impossible to achieve without appropriate authentication and authorization procedures [9].

At the same time, the A-IoT systems are more elaborate than the ICT platforms of today. They are often granted direct access to sensitive personal information; hence, the data they collect and handle should not be made available to potential third parties. On the other hand, large vehicles, drones, and industrial robots represent more capable platforms, sometimes termed as sources of increased danger. This recognizes that they may become hazardous as long as health and even lives of humans are concerned. Therefore, A-IoT systems have to be featured with more secure and reliable authentication procedures, so that they are capable of distinguishing their valid user from an unauthorized adversary.

D. Constrained Response Times and Usability

Regardless of their stringent security demands, the response levels of A-IoT authentication are also crucial for its successful adoption. Previously, authentication process was a dedicated phase of the HCI thus making users prepare for it both physically and mentally: recall the secret phrase, bring the token key, etc. With further development and penetration of the A-IoT systems, they become more ubiquitous and omnipresent. In future Smart Cities, users will be interacting with various A-IoT devices multiple times a day; hence, they cannot afford to spend several second by authenticating to each of those and tolerate second-long delays in acquiring access.

In response to these demands, A-IoT authentication must evolve to become capable of operating within stringent time intervals, preferably in an inconspicuous form, i.e., transparent to the user. For multi-functional A-IoT systems, this may even bring the need to temporarily provide access to certain basic functionality sooner, while more rigorous authentication is performed in the background. This is because the users are unlikely to require sensitive actions from the very first moments of their interaction with the target A-IoT platform.

From the above, it follows that designing adequate A-IoT authentication mechanisms is challenging. However, the more advanced capabilities and functions of A-IoT devices can be beneficial when coining novel authentication schemes and we review these in the next section.
III. ENABLERS FOR IMPROVED A-IoT AUTHENTICATION

Reliable human user authentication by the A-IoT system is a complex task due to a number of challenges as discussed previously. Fortunately, modern A-IoT platforms feature a number of dedicated input devices as well as rich sensing, communication, and computation capabilities, which altogether can be employed during the authentication stage. Utilizing this diverse functionality, various user authentication methods become suitable for new A-IoT systems. In this section, we discuss these authentication methods and their applicability in the A-IoT systems. For convenience, we order them by following their mass adoption: from well-known to emerging, see Table I.

![Table I: Authentication factors suitable for A-IoT](image)

<table>
<thead>
<tr>
<th>Factor</th>
<th>Type</th>
<th>Action</th>
<th>Duration</th>
</tr>
</thead>
<tbody>
<tr>
<td>PIN code</td>
<td>K</td>
<td>A</td>
<td>S</td>
</tr>
<tr>
<td>Password</td>
<td>K</td>
<td>A</td>
<td>M</td>
</tr>
<tr>
<td>Token</td>
<td>O</td>
<td>P</td>
<td>S</td>
</tr>
<tr>
<td>Voice</td>
<td>BI/BE</td>
<td>A/P</td>
<td>S/M</td>
</tr>
<tr>
<td>Facial</td>
<td>BI</td>
<td>A</td>
<td>S/M</td>
</tr>
<tr>
<td>Ocular-based</td>
<td>BI</td>
<td>A</td>
<td>S/M</td>
</tr>
<tr>
<td>Fingerprint</td>
<td>BI</td>
<td>A/P</td>
<td>S</td>
</tr>
<tr>
<td>Hand geometry</td>
<td>BI</td>
<td>A/P</td>
<td>S</td>
</tr>
<tr>
<td>Geographical location</td>
<td>BE</td>
<td>P</td>
<td>L</td>
</tr>
<tr>
<td>Vein recognition</td>
<td>BI</td>
<td>A/P</td>
<td>S</td>
</tr>
<tr>
<td>Thermal image</td>
<td>BI/BE</td>
<td>P</td>
<td>S/M</td>
</tr>
<tr>
<td>Behavior patterns</td>
<td>BE</td>
<td>P</td>
<td>L</td>
</tr>
<tr>
<td>Weight</td>
<td>BI</td>
<td>P</td>
<td>S</td>
</tr>
<tr>
<td>Electrocardiographic recognition (ECG)</td>
<td>BI/BE</td>
<td>P</td>
<td>S-L</td>
</tr>
</tbody>
</table>

A. Review of Possible Enablers

1) Hardware tokens: The automotive cluster has its own legacy security mechanisms, primarily centered around the use of hardware tokens that represent the ownership factor. Recently, such tokens have been complemented by increasingly popular software-based replacements installed on smartphones1. By leveraging this concept, the A-IoT systems can make a step forward and utilize the tokens placed not only in the smartphones but also on wearable devices.

2) Memorable passwords/PINs: Utilization of conventional PINs is currently acceptable worldwide owing to widespread adoption of ATMs and early-mobile phone era. A combination of button presses to unlock a feature (e.g., engine start) or to access a restricted area in addition to the key are typical solutions. Finally, knowledge-based approaches are used widely to access a web-service. The A-IoT systems may intelligently utilize similar solutions as well, where password inputs can effectively become replaced by the use of touchscreen (where applicable) or e.g., audio forms of input.

3) Fingerprint/palm/eye scanner: While core technology principles for fingerprint and palm recognition have been known for already a while, the recent achievements in the respective miniaturization made them accessible by a wide range of consumer products, namely, smartphones. Installation of biometric scanning devices within a conventional input interface (e.g., Home button in Apple iPhones) or behind a touchscreen is not a science fiction anymore2. Hence, authentication process can become transparent for the user, thus improving the overall system usability.

4) Facial recognition: The methods of facial recognition by built-in video cameras originally started with landmark picture analysis, which appeared to be vulnerable to trivial attacks of e.g., presenting a photo instead of the real face. Over the last two decades, these tools have significantly developed towards three-dimensional face and expression recognition that is much more resilient to such attacks3. The security levels can be enhanced further by prompting the user to move the head in a specific manner, so that a particular pattern to follow is not known in advance [10]. Solving this task from another angle, a drone can fly around the user to construct a 3D map of face/body without making the user move.

5) Voice recognition: All of the considered A-IoT devices are typically equipped with a microphone that enables voice recognition. The recently announced implementations are capable of distinguishing millions of different voices after capturing only a short phrase. These solutions are however more vulnerable to presentation attacks than the facial recognition. Therefore, pseudo-random generation of phrases to be pronounced is a pressing demand. While it is technically possible for an adversary to construct a phrase based on the recorded pronunciation of syllables and sounds, the A-IoT systems are likely to have sufficient computational power for timely recognition of the corresponding attacks.

6) Data from wearables: The A-IoT devices may also employ their advanced communication capabilities. Particularly, if the authenticating user holds wearable devices, they could act as providers of the authentication factors. Being connected to the A-IoT system via a short-range radio, wearables can present the security credentials of their user, such as heart rate or electrocardiogram. The utilization of this method requires support from appropriate security protocols, so that the platform may trust the data collected by the user-controlled equipment on the one hand, and the users can be certain that their sensitive personal information is not disclosed, on the other.

7) Behavioral patterns: The A-IoT system can utilize one or several input interfaces to record and analyze the individual

---

features of user behavior: response time to typical requests, typing rhythm, micro- or macro-scale mobility, etc. Here, the choice of particular factors to monitor highly depends on the form-factor of the A-IoT device: for wearable electronics these could be accelerometer fingerprinting, for drones they are the control operations, while for smart vehicles there are plenty of options that range from brake pressure and position of hands on the wheel to musical and radio preferences.

B. Mapping Enablers onto Challenges

While each of the A-IoT-specific authentication methods can bring its additional benefits, none of them alone is capable of efficiently solving all of the discussed A-IoT challenges. To this end, Table II offers a mapping of the authentication methods onto the challenges introduced in Section II.

Notably, knowledge-based methods have their most severe limitations with usability and security requirements [11], since the user is expected to create, remember, and timely update the secret passwords for all A-IoT devices. In this case, it is very likely that the same password will be selected for multiple systems, which degrades the levels of security. In contrast, hardware tokens are more scalable to be used for multiple A-IoT systems. However, the security levels may still be insufficient as the token(s) can easily be stolen.

Biometrics allow to be authenticated without an additional device or knowledge but the fingerprint, ocular scanning, or voice recognition may require further effort from the user (e.g., remove gloves or glasses, say a particular phrase, etc.) as well as remain not fully robust to the environmental conditions. Finally, the risk of losing a biometric template has to be considered. Then, authentication with wearable data has a significant advantage over the conventional voice/face recognition, since the user is not required to perform any explicit action. Meanwhile, this method has similar drawbacks as do the tokens, where the user has to continuously carry the necessary devices, always turned on and charged.

The methods of behavior recognition allow for mitigating most of the constraints by observing the user behavior over a certain period of time. However, the amounts of time necessary for such monitoring are at least an order of magnitude higher than those for other methods, which may become a severe usability concern in delay-sensitive A-IoT applications. In addition, behavior recognition is a complex task from the algorithm design perspective, as there should be a constructive differentiation between a valid deviation in the monitored factor by the actual user and invalid patterns by adversaries.

As can be concluded from our analysis and Table II, neither of the presented methods alone is sufficient to effectively authenticate the user over a broad range of possible scenarios related to the A-IoT systems. In the following section, we propose a novel approach to construct reliable authentication solutions for A-IoT devices by intelligently combining multiple potentially unreliable methods, which follows the multifactor authentication (MFA) paradigm.

IV. USE OF MULTI-FACTOR AUTHENTICATION FOR A-IoT

Since no single authentication method is likely to be suitable to resolve all of A-IoT challenges, the use of MFA is a natural approach to construct compound solutions (see Fig. 2). At the same time, designing adequate MFA mechanisms is a complex matter, which calls for careful selection, harmonization, and combination of various individual methods, such that the resulting solution could outperform its component elements in terms of both security and usability. Below, we summarize the four key design principles to be considered when building A-IoT-ready MFA solutions.

A. Means to Compare

Before combining several heterogeneous authentication methods together, one needs to harmonize across them, such that knowledge-based methods could be integrated with e.g., biometric and ownership schemes within a single-stop A-IoT authentication mechanism. Importantly, the output of the overwhelming majority of individual authentication solutions is binary: either accept or reject, i.e., \{0; 1\}. In rare cases, a continuous variable that characterizes the “likelihood” \([0; 1]\) could be retrieved from certain biometric systems. However, most vendors do not provide with access to those values but rather convert the likelihood factor into a binary decision internally.

In addition to the output data format, alternative methods can be characterized by their accuracy, which is typically estimated with two probabilities: (i) false accept rate (FAR), the probability that an unauthorized user is accepted; and (ii) false reject rate (FRR), the probability that a valid user is rejected. These reflect two major qualities of an authentication system: security (FAR) and usability (FRR). We here advocate their generalization to knowledge and ownership methods.

For instance, in password-based protection, FAR may correspond to the probability of guessing the secret, while FRR may characterize the possibility of making an accidental mistake during input. In turn, FAR and FRR may also reflect the chances for a token to be stolen or lost for ownership factors. Therefore, we conclude that all of the discussed authentication methods can be well-represented in a unified output format and supplemented with their suitable FAR/FRR values.

B. Means to Combine

The use of several individual authentication methods does not offer immediate advantages, since it still remains unclear how to combine them efficiently. At the first glance, one may come up with either of the two extreme strategies: “A user
TABLE II
COMPARING A-IoT AUTHENTICATION METHODS

<table>
<thead>
<tr>
<th>Authentication method</th>
<th>Non-text input</th>
<th>Short contact time</th>
<th>Stringent usability</th>
<th>Environmental robustness</th>
<th>High security level</th>
</tr>
</thead>
<tbody>
<tr>
<td>Hardware tokens</td>
<td>+</td>
<td>+</td>
<td>+</td>
<td>+</td>
<td>-</td>
</tr>
<tr>
<td>Password/PIN</td>
<td>-</td>
<td>+</td>
<td>-</td>
<td>+</td>
<td>-</td>
</tr>
<tr>
<td>Fingerprint/Palm scanner</td>
<td>+</td>
<td>+</td>
<td>+/-</td>
<td>-</td>
<td>+</td>
</tr>
<tr>
<td>Facial recognition</td>
<td>+</td>
<td>-</td>
<td>+</td>
<td>-</td>
<td>+</td>
</tr>
<tr>
<td>Voice recognition</td>
<td>+</td>
<td>-</td>
<td>+/-</td>
<td>+</td>
<td>+/-</td>
</tr>
<tr>
<td>Data from wearables</td>
<td>+</td>
<td>+</td>
<td>-</td>
<td>-</td>
<td>+</td>
</tr>
<tr>
<td>Behavior patterns</td>
<td>+</td>
<td>-</td>
<td>+</td>
<td>-</td>
<td>+</td>
</tr>
</tbody>
</table>

should successfully pass all the checks to get access” (All) and “A user should successfully pass any of the checks to get access” (Any).

Below, we present a typical example that numerically illustrates the inherent weaknesses of these extreme strategies as well as emphasizes importance of a certain level of intelligence when deriving the resulting decision from a number of individual components by the component methods. We assume a number of factors, each characterized with its own FAR and FRR values. For simplicity, we require that all the FARs are equal to 0.03% whereas all the FRRs are equal to 2%. The resultant values for FAR/FRR are then derived by the Law of Total Probability.

Observing Fig. 3, the All approach has the lowest FAR, thus yielding the best security level. However, its FRR is higher than with other approaches, by reaching over 12% with 7 independent factors combined. Hence, the usability of All approach remains low, which makes it non-applicable in the A-IoT context. Further, we notice the opposite trend for the Any approach that increases the FAR value at the expense of much better FRR. Therefore, Any solution is not applicable either. Consequently, none of the trivial MFA combinations are directly usable in the challenging A-IoT scenarios.

In contrast, a more intelligent Balanced approach – “A user should successfully pass most of the checks to get access” – constitutes a viable compromise between security and usability, by decreasing both FAR and FRR indicators. The quantitative gains highly depend on the input parameters and reach $10^4$ vs. $10^8$ when 7 factors are combined. This example also highlights the importance of threshold value selection, since incorrect combining may often result in rapid system performance degradation [12]. The same holds true for any other values of FAR/FRR, even though they may actually vary for different factors.

C. Means to Evaluate

Given that A-IoT scenarios are highly heterogeneous, the results delivered by the individual devices should not lead to blind accept/reject decisions. Instead, additional data must be considered when comparing the output of the authentication function against a threshold value.

1) Binary decision: The first and foremost sub-factor to be considered is a binary decision delivered by the individual device.

2) Vendor-specific metrics: The second sub-factor is the level of accuracy, which is directly related to FAR/FRR parameters. For example, the data collected with cameras by various vendors may deliver different probabilities during a facial recognition of the same user.

3) Level of trust: Many factors may impact user and device trust. Here, trust in the “owned” devices (e.g., built-in cameras) should be valued higher than that in external equipment. Further, historically familiar devices may have higher trust levels than stranger nodes, see the paradigm of Social IoT [13]. A significant benefit may be made available by utilizing social networks, since the devices owned by a friend or a colleague may also be considered as more trustworthy.

The set of selected sub-factors can significantly affect the operation of the authentication solution. However, the above three factors are relatively stable – the overall changes in the A-IoT system from these perspectives are not as abrupt and thus could be determined in advance. Conversely, the authentication system designer should be provided with a higher level of flexibility for a given application. This could be achieved by adding another dimension – specific factor weight per application (or even per user).

Accordingly, the general authentication function is to be considered as $\sum_i \delta_i \mu_i \tau_i \varphi_i > T$, where $i$ is the factor number, $\delta_i$ is a binary decision, $\mu_i$ is the accuracy level provided by the vendor, $\tau_i$ is the trust level to the selected source, $\varphi_i$ is the factor weight, and $T$ is the system threshold set by
<table>
<thead>
<tr>
<th>Factors</th>
<th>Preliminary</th>
<th>Active</th>
<th>Continuous</th>
</tr>
</thead>
<tbody>
<tr>
<td>General video</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Hardware tokens</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>PIN, password</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Fingerprint, palm</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Facial recognition</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Voice recognition</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Wearable data</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Behavior patterns</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

**Authenticity phases**

- **Preliminary**
- **Active**
- **Continuous**

**Factors**

- Knowledge
- Ownership
- Biometrics
- Behavior

**Time**

- General video
- Hardware tokens
- PIN, password
- Fingerprint, palm
- Facial recognition
- Voice recognition
- Wearable data
- Behavior patterns

---

**Fig. 4.** Considered phases of time-separated MFA for A-IoT.

---

the designer. Hence, the system may be adjusted per device, while the ultimate decision can be made flexibly based on statistical analysis and machine learning techniques. Finally, the use of various factors consumes different amounts of time, see Table I.

D. Means to Evolve

The conventional ICT systems typically exploit a single-stage authentication method, such that the user is either granted or denied access as the result of authentication. In contrast, the more stringent time constraints of A-IoT authentication dictate the need to complement the main authentication phase with additional checks that happen before and/or after it. Here, the considered MFA solution may benefit from a range of sensing devices widely deployed in Smart Cities as well as exploit the very nature of the human interaction with the A-IoT system. Therefore, the overall authentication process can be divided into several phases and, consequently, the level of trust to the user begins to evolve in time.

1) **Pre-authentication phase:** This phase is the most dynamic and unpredictable as a person ‘approaches’ the target vehicle. Here, the surrounding environment plays a crucial role by providing with additional information. The only option during this phase is to utilize passive authentication strategies i.e., ‘observe’ the user biometrics/behavior that could be delivered by user-worn wearables, user-carried deceives, and other vehicles/infrastructure.

2) **Active authentication phase:** The most conventional phase relies upon active interaction. Hence, the user provides relevant input to the system directly. The most suitable authentication methods are knowledge- and biometrics-based.

3) **Continuous (post) authentication phase:** Another key part of the envisioned A-IoT authentication process is continuous monitoring of the fact that the user remains legitimate to operate the system even after the previous phases are completed successfully [14]. Monitoring and analyzing the subject by the smart vehicle, infrastructure, and other cars become a preferred option. Consider a case where the driver has provided all of the tokens, passed all of the biometric tests but faced a seizure during a highway trip. In this case, the vehicle may automatically overtake the control, connect with neighboring cars, and safely stop by the wayside. As an example, recent works confirm that it is necessary to monitor the driver for just under 2.5 minutes in order to validate the behavior with 95% accuracy [15].

V. ECOSYSTEM OF MFA-POWERED A-IoT

The previous section summarized the underlying design principles of the MFA solutions in A-IoT, at large. However, even if these principles are followed, further development and mass adoption of MFA-powered A-IoT systems should be considered in perspective. This section brings the community’s attention to the most significant questions to be answered in this context.

1) **How to weight factors?** While the MFA concept offers sufficient flexibility to adapt the authentication system to a wide range of possible scenarios, the choice of particular numerical weights and threshold values requires an extensive study, which needs to carefully balance the FAR and FRR values of the resulting system depending on the target use case. The system should also be made reconfigurable, such that its internal parameters are updated appropriately whenever an A-IoT device is e.g., sold to another person with different attributes.
2) How to adapt decisions?: Another key challenge is dynamic system adaptation in relation to a number of factors involved in the authentication process. For instance, recognition based on a video camera may be unavailable at nighttime or in bad weather. Hence, the decision function should dynamically adjust the weights of the factors that are available during the authentication process based on contextual data. This task is much more challenging as compared to conventional single- and two-factor authentication with only a few static factors involved.

3) How to earn user trust?: The next question is related to making a legitimate user trust the system in its operations. For example, the user had a video surveillance camera at the parking near home, which contributed 20% to the overall authentication process, while the threshold was configured to grant access. Then, the user moved the car to another address and cannot open it anymore without an additional weight from the infrastructure, since there is no external camera nearby to participate in the authentication process. Hence, it is crucial that decision-making process is at least partially transparent to the user.

4) How to receive assistance?: The A-IoT framework involves not only in-built authentication factors but also data from proximate sources. Therefore, the question remains of how secure and trusted such assistance from the neighboring devices could be. Our illustrative example considered above receives additional data from the wearable devices owned by the human user; the camera mounted on a lamp post; a surveillance drone patrolling the street, etc. Hence, designing secure and reliable methods to deliver the sensitive authentication data from these dissimilar Smart City devices to the target A-IoT system – while not compromising the user privacy for third-party entities – is an open problem.

5) How to delegate A-IoT devices?: Users tend to share their devices both privately (family) and publicly (car rent). However, secure collective delegation of use is not straightforward for the A-IoT systems. Conventional landing of a physical token is not a sufficient option anymore, since it does not necessarily confirms the right to operate the A-IoT device. From the A-IoT platform perspective, most of the factors related to its temporary user will be received for the first time. Therefore, effective delegation of A-IoT systems is another important open challenge on the way to their mass adoption.

VI. Conclusion

Reliable and secure human authentication by various smart devices is one of the key drivers in the Advanced IoT era. We reviewed the existing research challenges and possible enablers for user authentication within the A-IoT ecosystem. We introduced a concept of multi-factor authentication for A-IoT as an attractive alternative to existing single-factor solutions with limited potential. The key design principles of MFA for A-IoT were highlighted by providing useful insights into facilitation of the future MFA applications for the A-IoT. Finally, key open questions related to the development, practical implementation, and adoption of MFA for diverse A-IoT systems were discussed together with potential use cases, thus laying the foundation for further research in this emerging area.

REFERENCES
