Security for 5G and Beyond

Ijaz Ahmad®, Shahriar Shahabuddin, Tanesh Kumar®, Jude Okwuibe, Andrei Gurtov, and Mika Ylianttila, Senior Member, IEEE

Abstract—The development of the fifth generation (5G) wireless networks is gaining momentum to connect almost all aspects of life through the network with much higher speed, very low latency and ubiquitous connectivity. Due to its crucial role in our lives, the network must secure its users, components, and services. The security threat landscape of 5G has grown enormously due to the unprecedented increase in types of services and in the number of devices. Therefore, security solutions if not developed yet must be envisioned already to cope with diverse threats on various services, novel technologies, and increased user information accessible by the network. This paper outlines the 5G network threat landscape, the security vulnerabilities in the new technological concepts that will be adopted by 5G, and provides either solutions to those threats or future directions to cope with those security challenges. We also provide a brief outline of the post-5G cellular technologies and their security vulnerabilities which is referred to as future generations (XG) in this paper. In brief, this paper highlights the present and future security challenges in wireless networks, mainly in 5G, and future directions to secure wireless networks beyond 5G.

Index Terms—5G, security, mobile networks security, SDN security, NFV security, cloud security, privacy, security challenges, security solutions.

I. INTRODUCTION

The 5G wireless networks will provide very high data rates and higher coverage with significantly improved Quality of Service (QoS), and extremely low latency [1]. With extremely dense deployments of base stations, 5G will provide ultra-reliable and affordable broadband access everywhere not only to cellular hand-held devices, but also to a massive number of new devices related to Machine-to-Machine communication (M2M), Internet of Things (IoT), and Cyber-Physical Systems (CPSs) [2]. Such enrichment implies that 5G is not a mere incremental advancement of 4G as one might intuitively think, but an integration of new disruptive technologies to meet the ever growing demands of user traffic, emerging services, existing and future IoT devices [3]. With all these capabilities, 5G will connect nearly all aspects of the human life to communication networks, and this underscores the need for robust security mechanisms across all network segments of the 5G. Wireless networks have particularly been a major target for most security vulnerabilities from the very inception.

The First Generation (1G) mobile networks were prone to the challenges of illegal interception, cloning and masquerading [4]. Message spamming for pervasive attacks, injection of false information or broadcasting unwanted marketing information became common in the Second Generation (2G) of mobile networks. The main disruption, however, was in the Third Generation (3G) of mobile networks in which IP-based communication enabled the migration of Internet security vulnerabilities and challenges into mobile networks. The security threat landscape got further widened and complicated in the Fourth Generation (4G) mobile networks with the increased use of IP-based communication necessary for new devices and new services [5]. The amalgamation of massive number of IoT devices and the provision of new services, for example for smart homes, hospitals, transport, and electric grid systems in 5G will further exacerbate the security challenges.

The security solutions and architectures used in previous generations (i.e., 3G and 4G), apparently, will not suffice for 5G. The main reason for new security solutions and architecture is the dynamics of new services and technologies in 5G [6]. For example, virtualization and multi-tenancy in which different, and possibly conflicting, services share the same mobile network infrastructure were not common before. The latency requirements, such as authentication latency in vehicular communication or Unmanned Aerial Vehicles (UAVs) were not that much critical. Succinctly put, the security architectures of the previous generations lack the sophistication needed to secure 5G networks. The Next Generation Mobile Networks (NGMN) consortia suggests that 5G should provide more than hop-by-hop and radio bearer security, which was common in 4G and prior generations of cellular networks [7].

Furthermore, there are new technological concepts or solutions that will be used in 5G to meet the demands of increasingly diverse applications and connected devices. For example, the concepts of cloud computing [8], Software Defined Networking (SDN) [9], and Network Function Virtualization (NFV) [10] are considered to be the potential
problem solvers in terms of costs and efficiency. However, each of these technologies have its own security challenges. For instance, the core network entities such as Home Subscriber Server (HSS) and Mobility Management Entity (MME) that hold the user billing, personal, and mobility handling information, respectively, deployed in clouds will render the whole network ineffective if security breaches occur. Similarly, SDN centralizes the network control logic in SDN controllers. These controllers will be the favorite choice for attackers to render the whole network down through Denial of Service (DoS) or resource exhaustion attacks. The same is true for hypervisors in NFV. Therefore, it is highly important and timely to bring forth the possible weaknesses in these technologies and seek solutions to those weaknesses.

This article studies the state of the art of security in 5G networks. It starts off with a dive into the security challenges and corresponding solutions for the previous generations of networks ranging from 1G to 4G. It then presents a comprehensive overview of the technologies associated with 5G with regards to their corresponding security challenges and respective solutions. This article also provides insights into security in communication networks beyond 5G, named as XG. The rest of this paper is organized as follows: Section II describes the related work followed by a brief history of security in previous generations of wireless cellular networks in Section III. A general overview of security in 5G is presented in Section IV. Section V presents the security challenges and potential security solutions for 5G networks, these challenges and solutions are categorized according to the part of access network where they affect, i.e., the backhaul network and core network. Security challenges for key enabling technologies of 5G such as SDN, NFV, cloud computing and Multi Input Multi Output (MIMO) are presented in Section VI, followed by the security solutions in the same order in Section VII. In Section VIII, privacy related issues and possible solutions are presented. Section IX presents new dimensions for security of future networks in terms of using advanced disruptive technologies to solve the existing weaknesses and forthcoming security challenges in communication networks. Section X concludes the paper. For smooth readability, the outline of the article is depicted in Fig. 1 and the most used acronyms are presented in full form in Table I.

II. RELATED WORK

The ubiquitous connectivity with below 1 ms latency and Gigabit speed aimed by 5G has gained momentum towards its realization [11]. Since 5G is yet to be deployed, “what will be done by 5G and how will it work?” as presented in a survey on 5G [11], is still a question. The vision of 5G lies in providing very high data rates (Gigabits per second), extremely low
<table>
<thead>
<tr>
<th>Acronyms</th>
<th>Full Form</th>
</tr>
</thead>
<tbody>
<tr>
<td>AI</td>
<td>Artificial Intelligence</td>
</tr>
<tr>
<td>AKA</td>
<td>Authentication and Key Agreement</td>
</tr>
<tr>
<td>AMF</td>
<td>Autonomous Management Framework</td>
</tr>
<tr>
<td>API</td>
<td>Application Programming Interfaces</td>
</tr>
<tr>
<td>ARIB</td>
<td>Association of Radio Industries and Businesses</td>
</tr>
<tr>
<td>BTS</td>
<td>Base Transceiver Stations</td>
</tr>
<tr>
<td>CCPS</td>
<td>Cloud-based Cyber-Physical Systems</td>
</tr>
<tr>
<td>CPS</td>
<td>Cyber-Physical System</td>
</tr>
<tr>
<td>DTLS</td>
<td>Datagram Transport Layer Security</td>
</tr>
<tr>
<td>DoS</td>
<td>Denial of Service</td>
</tr>
<tr>
<td>DDS</td>
<td>Distributed DoS</td>
</tr>
<tr>
<td>DPI</td>
<td>Deep Packet Inspection</td>
</tr>
<tr>
<td>EAP-AKA</td>
<td>Extensible Authentication Protocol-AKA</td>
</tr>
<tr>
<td>EPC</td>
<td>Evolved Packet Core</td>
</tr>
<tr>
<td>EPS</td>
<td>Evolved Packet System</td>
</tr>
<tr>
<td>ETSI</td>
<td>European Telecommunications Standards Institute</td>
</tr>
<tr>
<td>E-UTRAN</td>
<td>Evolved Universal Terrestrial Radio Access Network</td>
</tr>
<tr>
<td>GMM</td>
<td>Global System for Mobile Communication</td>
</tr>
<tr>
<td>GSTM</td>
<td>GSTM Association</td>
</tr>
<tr>
<td>GTP</td>
<td>GPRS Tunneling Protocol</td>
</tr>
<tr>
<td>HetNet</td>
<td>Heterogeneous Networks</td>
</tr>
<tr>
<td>HIDS</td>
<td>Host Intrusion Detection System</td>
</tr>
<tr>
<td>HSS</td>
<td>Home Subscriber Server</td>
</tr>
<tr>
<td>HTTP</td>
<td>HyperText Transfer Protocol</td>
</tr>
<tr>
<td>HTTP-DoS</td>
<td>HTTP and XML Denial of Service</td>
</tr>
<tr>
<td>IaaS</td>
<td>Infrastructure as a Service</td>
</tr>
<tr>
<td>IDV</td>
<td>Intrusion Detection Systems</td>
</tr>
<tr>
<td>IDP-S</td>
<td>Intrusion Detection or Prevention Systems</td>
</tr>
<tr>
<td>IETF</td>
<td>Internet Engineering Task Force</td>
</tr>
<tr>
<td>IP</td>
<td>Internet Protocol</td>
</tr>
<tr>
<td>IMSI</td>
<td>International Mobile Subscriber Identity</td>
</tr>
<tr>
<td>IoT</td>
<td>Internet of Things</td>
</tr>
<tr>
<td>IPsec</td>
<td>Internet Protocol Security</td>
</tr>
<tr>
<td>ITU</td>
<td>International Telecommunications Union</td>
</tr>
<tr>
<td>LTE-(A)</td>
<td>Long Term Evolution-(Advanced)</td>
</tr>
<tr>
<td>MEC</td>
<td>Multi-access Edge Computing</td>
</tr>
<tr>
<td>MIMO</td>
<td>Multiple-Input and Multiple-Output</td>
</tr>
<tr>
<td>MME</td>
<td>Mobility Management Entity</td>
</tr>
<tr>
<td>MmWave</td>
<td>Millimeter Wave</td>
</tr>
<tr>
<td>M2M</td>
<td>Machine-to-Machine Communications</td>
</tr>
<tr>
<td>MTC</td>
<td>Machine Type Communication</td>
</tr>
<tr>
<td>MVNOs</td>
<td>Mobile Virtual Network Operators</td>
</tr>
<tr>
<td>NAS</td>
<td>Non-Access Stratum</td>
</tr>
<tr>
<td>NFPV</td>
<td>Network Function Virtualization</td>
</tr>
<tr>
<td>NGMN</td>
<td>Next Generation Mobile Networks</td>
</tr>
<tr>
<td>NILS</td>
<td>Network Intrusion Detection System</td>
</tr>
<tr>
<td>NIST</td>
<td>National Institute of Standards and Technology</td>
</tr>
<tr>
<td>OPDM</td>
<td>Orthogonal Frequency Division Multiplexing</td>
</tr>
<tr>
<td>ONF</td>
<td>Open Networking Foundation</td>
</tr>
<tr>
<td>P2B</td>
<td>Platform as a Service</td>
</tr>
<tr>
<td>PDN</td>
<td>Public Data Network</td>
</tr>
<tr>
<td>QoS</td>
<td>Quality of Service</td>
</tr>
<tr>
<td>RAN</td>
<td>Radio Access Network</td>
</tr>
<tr>
<td>RA-N</td>
<td>Radio Access Technology</td>
</tr>
<tr>
<td>SaaS</td>
<td>Software as a Service</td>
</tr>
<tr>
<td>SIM</td>
<td>Subscriber Identity Module</td>
</tr>
<tr>
<td>SIPvDAS</td>
<td>Slowly-increase Polymorphic DoS Attack Strategy</td>
</tr>
<tr>
<td>SDN</td>
<td>Software Defined Networking</td>
</tr>
<tr>
<td>SN</td>
<td>Serving Network</td>
</tr>
<tr>
<td>TCP</td>
<td>Transmission Control Protocol</td>
</tr>
<tr>
<td>TLS</td>
<td>Transport Layer Security</td>
</tr>
<tr>
<td>TMSI</td>
<td>Temporary Mobile Subscriber Identity</td>
</tr>
<tr>
<td>TVDCs</td>
<td>Trusted Virtual Data Center</td>
</tr>
<tr>
<td>UAVs</td>
<td>Unmanned Aerial Vehicles</td>
</tr>
<tr>
<td>UAC-BS</td>
<td>UAV with Base Station</td>
</tr>
<tr>
<td>UDP</td>
<td>User Datagram Protocol</td>
</tr>
<tr>
<td>UE</td>
<td>User Equipment</td>
</tr>
<tr>
<td>UMTS</td>
<td>Universal Mobile Telecommunications System</td>
</tr>
<tr>
<td>USIM</td>
<td>Universal Subscriber Identity Module</td>
</tr>
<tr>
<td>VNIFs</td>
<td>Virtual Network Functions</td>
</tr>
<tr>
<td>V2V</td>
<td>Vehicle-to-Vehicle</td>
</tr>
<tr>
<td>V2X</td>
<td>Vehicle-to-Everything</td>
</tr>
<tr>
<td>XG</td>
<td>Future Generations</td>
</tr>
<tr>
<td>XML</td>
<td>Extensible Markup Language</td>
</tr>
<tr>
<td>3GPP</td>
<td>3G Partnership Project</td>
</tr>
<tr>
<td>5GPPP</td>
<td>5G Public-Private Partnership</td>
</tr>
</tbody>
</table>

Latency, manifold increase in base station density and capacity, and significant improvement in quality of service, compared to 4G systems [1]. There are a number of survey articles that thoroughly discuss 5G networks such as [1], [12]–[14]. Due to its foreseen role and impact on our lives, security of 5G is far more concerning. Therefore, there is a great effort to ensure the security of networked systems in 5G, users of the networked systems and the 5G network itself.

The limitations in 4G and how to move towards 5G by overcoming those limitations is discussed in [11]. The concerns related to security, indirectly if not directly effecting it, pertaining to 4G are the lack of mechanisms to support data traffic bursts, limited processing capabilities of base stations, and latency. These limitations, if not removed, will make the network prone to security challenges. For example, bursts in data traffic can be due to legitimate reasons such as crowd movements, or otherwise due to DoS attacks. Similarly, limited capacity of base stations in ultra-dense 5G networks will cause availability challenges for legitimate users or a weak point for resource exhaustion attacks. So also is latency, which can be problematic in authentication of vehicles in Vehicle to Everything (V2X) communication. Therefore, the survey article [11] provides some interesting insights on the limitations of the current 4G networks that must be solved in 5G.

General requirements and mechanisms for strengthening security in 5G are presented in [15]. By revisiting the LTE security requirements, the authors outline the security requirements for 5G on a high level in [15]. A survey on security of 4G and 5G networks is presented in [16]. The article focuses on existing authentication and privacy-preserving schemes for 4G and 5G networks. Security challenges and the possible mitigation techniques along with standardization efforts in 4G and older generations are presented in [17]. A survey on security threats and attacks on mobile networks is presented in [18]. The article focuses on the security threats and challenges in the mobile access and core networks. The main challenges, however, are related to the 4G network architecture.

Security challenges and the possible mitigation techniques in the wireless air interfaces are discussed in [19]. The article considers various wireless access technologies such as Bluetooth, Wi-Fi, WiMAX and LTE, and discusses the inherent security limitations and future directions for strengthening the security of each technology. The main focus, however, is on the security of the wireless air interfaces. A survey on physical layer security techniques for 5G wireless networks is presented in [20]. The main focus of the article is physical layer security coding, massive MIMO, Millimeter Wave (mmWave) communications, Heterogeneous Networks (HetNets), non-orthogonal multiple access techniques, and full duplex technology.

A study of security of 5G in comparison with current or traditional cellular networks is carried out in [21]. Here, security of 5G is studied in terms of authentication, availability, confidentiality, key management and privacy. Several challenges such as access procedure in HetNets, efficiency of security systems with respect to the stringent latency requirements, and the lack of new trust models have been highlighted. The article [21] also proposed a security architecture for 5G networks based on the findings of the study and evaluated the identity management and authentication schemes.
An interesting work on security research in future mobile networks is presented in [22]. The work aims at providing a comprehensive view of security on mobile networks, as well as open up some research challenges. A methodology is developed that can categorize known attacks, their impact, defense mechanisms for those attacks, and the root causes of vulnerabilities. The main vulnerabilities that could still cause security challenges include pre-authentication traffic, jamming attacks, insecure inter-network protocols (e.g., DIAMETER [23]), insecure implementations in network components, and signaling-based DoS attacks. However, the security challenges in 5G will be more diverse due to the amalgamation of new things or networks of new things (e.g., massive IoT), and the conglomeration of new technological concepts.

SDN, NFV and the extended concepts of cloud computing such as Multi-access Edge Computing (MEC) [24]–[26] have many benefits in terms of performance and cost efficiency, however, these technologies have their own security weaknesses. Until this time, there is no survey article that studies the security of 5G in the context of these new technological concepts and their impact on the security of 5G or future wireless networks. The articles mentioned in the related work are focused on specific areas. For instance, [21] and [22] are focused on authentication, [19] and [20] are targeted towards security of physical layer and air interfaces respectively, [18] presents security in access and core networks, presents LTE security and general security requirements of 5G, and [16] covers privacy issues in future networks.

In this article, we have focused not only on the security of 5G networks, but also on the forthcoming technologies that will be used in 5G and beyond, for example, SDN, NFV, MEC, and massive MIMO, etc. This article provides a detailed description of the security challenges within these technologies and the potential solutions to those security challenges. However, the increasing diversity and number of communicating devices such as IoT and V2X would require drastically new security solutions that will need context awareness and high degree of automation. Therefore, this article also discusses the future of security in environments replete with massive IoT, such as smart cities. Furthermore, this article outlines how new disruptive technological concepts such as Artificial Intelligence (AI), blockchain, and UAVs can be used to secure a highly connected society in the near future, termed as XG in this article.

III. SECURITY IN WIRELESS NETWORKS: FROM 1G TO 4G

Security of communication networks has been a difficult task due to complexities in the underlying network, the proprietary and perimeter-based security solutions that are hard to manage, and the weaknesses in identity management [27]. Moreover, the Internet architecture inherits the problems arising from the infrastructure, is ripe with security challenges and is stagnant to innovation [28]. Wireless network security has been evolving with gradual up-gradation since the inception of mobile networks [29]. The main change, however, came with the introduction of IP-based communication in wireless networks where more and more Internet-based security challenges migrated to wireless networks. Therefore, in this section we provide an overview of the changing security paradigm (summarized in Table II) in wireless networks from 1G to 4G or from non-IP wireless networks to IP-based wireless networks.

A. Security in Non-IP Networks

The 1G cellular systems used analog signal processing and were designed primarily for voice services [30]. The most successful 1G system called Advanced Mobile Phone Service that was first deployed commercially by AT&T and Bell labs during 1983 [31]. Due to the nature of analog communications, it was difficult to provide efficient security services for 1G. This advance phone service did not use encryption and thus there was no security of information or telephone conversations. Hence, practically the whole system and users were open to security challenges such as eavesdropping, illegal access, cloning, and user privacy [4], [32]. Digital mobile systems were proposed to increase the efficiency of the limited frequency bands [33], [34], and thus, Global System for Mobile (GSM) communication became the most successful and widely used standard in cellular communications as part of 2G cellular networks [31].

GSM MoU association identified four aspects of security services to be provided by a GSM system. They are anonymity, authentication, signaling protection and user data protection [35]. Anonymity is provided by using temporary identifiers so that it is not easy to identify the user of a system. The real identifiers are used only when the device is switched on and then a temporary identifier is issued. Authentication is used by the network operator to identify the user. It is performed by a challenge-response mechanism [36], [37]. The signalling and user data protection was carried out through encryption in which the Subscriber Identity Module (SIM) played an important role in the encryption keys. However, 2G
had several security limitations or weaknesses. The operators only authenticated the UEs in a unilateral mechanism, whereas the UEs had no option to authenticate the operator. Therefore, it was possible for a false operator to impersonate the original operator and perform a man-in-the-middle attack [38]. Moreover, the encryption algorithms were also reverse engineered [39] and the ciphering algorithms were subject to several attacks [40], [41]. GSM did not provide data integrity against channel hijacking in the absence of encryption, and were also vulnerable to DoS attacks [39]. Furthermore, 2G systems did not have the capability to upgrade their security functionality over time [42].

B. Security in 3G

The 3G cellular networks were developed primarily to provide higher data rates than 2G networks. 3G systems also enabled new services like video telephony and video streaming via cellular networks [43]. The 3G standard proposed an upgraded security architecture to mitigate the vulnerabilities of 2G systems. The three key principles of 3G security are specified by the 3GPP in [42], namely: (I) 3G security will inherit the essential features of 2G security, (II) 3G security will upgrade the limitations of 2G security, and (III) 3G will add more security features that was not available in 2G. Universal Mobile Telecommunications System (UMTS) is a 3G cellular technology that is developed and maintained by 3GPP [44]. The security architecture of UMTS consists of five sets of security features that are specified in TS33.102 [45], which is more commonly known as Release 99.

These set of UMTS security features ensures that the UE has a secure access to 3G services and provides protection against attacks on radio access link [46]. The UMTS Authentication and Key Agreement (AKA) protocol has been designed in such a way that the compatibility with GSM is maximized. However, UMTS AKA serves additional protocol goals like mutual authentication of the network and, and agreement on integrity key, etc. Contrary to the unilateral authentication of GSM, UMTS supports bilateral authentication which removes the threat of a false base station. The access security feature includes user identity confidentiality that ensures that a user cannot be eavesdropped on a radio access link. The user identity confidentiality also needs to support user location confidentiality and user untraceability. To achieve these objectives, the user is identified by a temporary identity or by a permanent encrypted identity. Similarly, the user should not be identified for a long period of time and any data that might reveal user identity must be encrypted [45].

C. Security in 4G

The release 10 from 3GPP, which is commonly known as LTE-Advanced (LTE-A), fulfills the requirements of the 4G standard that was specified by International Telecommunications Union - Radio Communication Sector (ITU-R) [47]. LTE-A network has two major parts; Evolved Packet Core (EPC) and Evolved- Universal Terrestrial Radio Access Network (E-UTRAN). The EPC is an all IP and packet switched backbone network. The LTE-A system supports non-3GPP access networks. LTE-A systems also introduced new entities and applications like Machine-Type-Communication (MTC), home eNodeB or femtocells and relay nodes. 3GPP defined similar sets of security features for LTE-A. They are: (I) Access security, (Evolved Universal Terrestrial Radio Access Network), (II) Network domain security, (III) User domain security, (IV) Application domain security, and (V) Visibility and configurability of security [48]. However, each of the feature has been enhanced significantly to secure the LTE-A systems. Besides, totally new security mechanisms were specified for MTC [49], [50], home eNB [51] and relay nodes [52].

The Evolved Packet System-AKA (EPS-AKA) had one major enhancement over UMTS-AKA which is called cryptographic network separation. This feature limits any security breach in a network and also limits the possibility of spreading attacks across the network. This is achieved by binding any EPS-related cryptographic keys to the identity of the Serving Network (SN), to which the keys are delivered. This feature also enables the UE to authenticate the SN. Note that, UE cannot authenticate the SN in UMTS. It can only ensure whether an SN is authorized by the UE’s home network [53]. There are some enhancements created for EPS for device confidentiality; the device identity is not sent to the network before security measures for traffic protection has been activated. The user and signaling data confidentiality also went through changes for the EPS. The endpoint of the encryption of the network side is in the base station while it is the radio network controller in 3G. Additional confidentiality protection mechanism was also introduced for signaling between the UE and the core network.

3GPP specified the security features of mobility within the E-UTRAN as well as between E-UTRAN and earlier generation or non-3GPP systems. There are two types of non-3GPP access networks. They are trusted non-3GPP access and untrusted non-3GPP access. For an untrusted non-3GPP access network, the UE has to pass a trusted evolved packet data gateway which is a part of the EPC [54]. In addition, a new key hierarchy and handover key management mechanism has been introduced to ensure secure mobility process in LTE. Due to the early termination point of the encryption, EPS structure introduced a new challenge. Due to the security weakness of the early termination point, the eNB became more vulnerable than the 3G security architecture. Besides, EPS architecture allows to place the eNB outside of network operators security domain, i.e., in physically insecure locations. Therefore, the eNB is vulnerable to physical attacks, DoS attacks or passive attacks at eavesdropping on long term keys. To tackle these vulnerabilities, 3GPP introduced stringent requirement on the eNB. 3GPP requirements include secure setup and configuration of the base station SW, secure key management inside the BTS, and secure environment for handling the user and control plane data etc.

IV. SECURITY IN 5G: AN OVERVIEW

5G will provide ubiquitous broadband services, enable connectivity of massive number of devices in the form IoT, and
entertain users and devices with high mobility in an ultra-reliable and affordable way [2]. The development towards IP-based communication in 4G has already helped develop new business opportunities, however, 5G is considered a new ecosystem connecting nearly all aspects of the society; vehicles, home appliances, health care, industry, businesses, etc., to the network. This development, however, will introduce a new array of threats and security vulnerabilities that will pose a major challenge to both present and future networks [55]. Connecting the power grid for instance, 5G will connect critical power infrastructures to the network, hence security breaches in such critical infrastructures can be of catastrophic magnitudes to both the infrastructures and the society which 5G serves. Therefore, security of 5G and systems connected through 5G must be considered right from the design phases. To elaborate the security implications in 5G, the design principles of 5G are briefly elaborated below.

A. Overview of 5G Design Principles

With new types of services and devices, and new user requirements in terms of low latency, higher throughput, and ubiquitous coverage, arises the need for new design principles for 5G [56]. The 5G design principles outlined by NGMN, presented in Fig. 2, highlight the need for highly elastic and robust systems. The radio part needs extreme spectrum efficiency, cost-effective dense deployment, effective coordination, interference cancellation and dynamic radio topologies. The network beyond radio has different requirements, which are more towards inclusion of radically new technologies. For example, the common composable core will use SDN and NFV to separate the user and control planes and enable dynamic network function placement [57]. This is targeted towards minimizing legacy networking and introducing new interfaces between the core and Radio Access Technologies (RATs).

The 5G network architecture must support the deployment of security mechanisms and functions (e.g., virtual security firewalls) whenever required in any network perimeter. As presented in Fig. 2, the operation and management need to be simplified. The most prominent technology for simplifying network management is SDN [58]. SDN separates the network control from the data forwarding plane. The control plane is logically centralized to oversee the whole network underneath and control network resources through programmable Application Programming Interfaces (APIs). However, centralizing the network control and introducing programmable APIs in network equipment also open loopholes for security vulnerabilities. Therefore, we need to analyze the security challenges associated with SDN. Similarly, NFV and network slicing have security challenges such as inter-federated conflicts and resource hijacking. Therefore, the security challenges associated with all technologies used by 5G need proper investigation. In the following subsection, we provide a brief overview of 5G security architecture, focusing mainly on the security domains defined by 3GPP.

B. Overview of 5G Security Architecture

According to ITU-T [59], a security architecture logically divides security features into separate architectural components. This allows a systematic approach to end-to-end security of new services that facilitates planning of new security solutions and assessing the security of existing networks. The 5G security architecture has been defined in the latest 3GPP technical specification release (release 15) [60] with different domains. The security architecture is shown in Fig. 3, except domain (VI), and has the following main domains.

- **Network access security (I):** Comprises the set of security features that enables a UE to securely authenticate and access network services. Access security includes security of 3GPP and non-3GPP access technologies, and delivery of security context from SN to the UE.
- **Network domain security (II):** Comprises a set of security features that enables network nodes to securely exchange signaling and user plane data.
- **User domain security (III):** Consists of security features that enable secure user access to UE.
using the Extensible Authentication Protocol-AKA (EAP-AKA) or improved EAP-AKA for authentication. For untrusted non-3GPP access networks, the UE uses the evolved packet data gateway IPsec tunnel establishment to connect to the EPC [54]. Both of these mechanisms have many benefits such as short message size, and requires only one handshake between the UE and SN, and between the serving and home networks [15]. Furthermore, the symmetric-key-based protocol in these schemes makes the computations required in the authentication center (part of the HSS), and in the USIM (Universal SIM) very efficient compared to public-key-based mechanisms. However, one of the advantages of public-key based authentication and key agreement schemes is that the home network does not need to be contacted for each authentication.

D. General Overview of Security in 5G

The Next Generation Mobile Networks (NGMN) has provided recommendations for 5G based on current network architectures and the shortfall in security measures that are either not developed or developed but not yet put to use [7]. The recommendation highlights the cautionary notes which includes factors such as the infancy of 5G with many uncertainties, lack of defined design concepts and the unknown end-to-end and subsystems architectures. The recommendations highlights the security challenges in access networks, and cyber-attacks against users and the network infrastructure, as depicted in Fig. 4. The details of the security limitations and recommendations can be found in [7] and the key points are summarized below.

Flash network traffic: It is projected that the number of end user devices will grow exponentially in 5G that will cause significant changes in the network traffic patterns either accidentally or with malicious intent. Therefore, the 5G systems must efficiently handle large swings in traffic and provide resilience whenever such surges occur while maintaining acceptable level of performance.

Security of radio interface keys: In previous network architectures, including 4G, the radio interface encryption keys are generated in the home network and sent to the visited network over insecure links causing a clear point of exposure of keys. Thus, it is recommended that keys should either be secured first or not sent over insecure links such as SS7/DIAMETER.

User plane integrity: The 3G and 4G systems provide protection to some signaling messages but do not provide cryptographic integrity protection for the user data plane. Hence, it is recommended to provide protection at the transport or application layer that terminates beyond mobile networks. However, application level end-to-end security may involve too much overhead for data transmission in packet headers and handshakes. Therefore, an exception to this could be network level security for resource constrained IoT devices or latency sensitive 5G services.

Mandated network security: There can be certain service-driven constraints (e.g., latency) in security architectures leading to optional use of security measures. Unfortunately, such constraints undermine system-level security assumptions and

C. UE-Network Mutual Authentication

In LTE architecture, EPS-AKA was used to perform mutual authentication between the UE and the network [63], [64]. AKA, started in GSM, evolved with the next generations [65], [66], and is still considered as the most viable mechanism for authentication and authorization in 5G networks [60]. AKA is based on symmetric keys and runs in SIM. Extensible Authentication Protocol (EAP)-AKA method for 3G networks [66] was developed by 3GPP to support identity privacy and fast re-authentication. The EPS-AKA provided further security such as using multiple keys in different contexts, renewal of keys and that without involving the home network every time. EPS-AKA has some challenges such as computation and communication overhead, and latency [67], however, has no visible vulnerabilities demonstrated so far, and thus will be used in 5G [15].

For communication through trusted non-3GPP access network, the UE is authenticated through the AAA server

- **Application domain security (IV):** Includes security features that enable applications (user and provider domains) to securely exchange messages.
- **Service Based Architecture (SBA) domain security (V):** Comprises of security features for network element registration, discovery, and authorization, as well as security for service-based interfaces.
- **Visibility and configurability of security (VI):** Includes security features that inform users whether security features are in operation or not.

The 5G security architecture itself does not define particular security threats and the solutions for those threats [6]. However, there are certain defined security solutions either coming from the previous generations with modifications for enhancements or defined newly according to the realm of 5G. The LTE security concepts are the starting points, but considered as benchmarks for security of future wireless networks [61]. In any case, the high-level vision of 5G security is based on i) Supreme built-in-security, ii) Flexible security mechanisms, and iii) Automation, as described by Nokia [62].

As the basis of security, and related to the domains of security highlighted by 3GPP, in the following subsection we briefly describe how authentication is carried out in 5G.

---

**Fig. 3.** Overview of the security architecture.
could not be completely eliminated. The challenge exacerbates in multi-operator scenarios where one operator suffers due to inadequate security measures by the other. Therefore it is highly recommended that some level of security must be mandated in 5G after proper investigation to recognize the most critical security challenges.

Consistency in subscriber level security policies: User security measures must be intact when a user moves from one operator network to another. It is highly possible that security services are not updated frequently on per-user basis as the user moves from one place to another, or from one operator network to another as in the case of roaming. Hence network operators need to share security policies and some level of subscriber service information. This recommendation highlights the possibility of using virtualization techniques to enable per-service slice configuration to keep security of the user or service intact with roaming.

DoS attacks on the Infrastructure: DoS and Distributed DoS (DDoS) attacks can circumvent the operation critical infrastructure such as energy, health, transportation, and telecommunication networks. DoS attacks are usually designed in a way that they exhaust the physical and logical resources of the targeted devices. This threat will be more severe due to the possibility of attacks from machines that are geographically dispersed and are in huge numbers (compromised IoT). Hence, the network must increase its resilience through strong security measures.

Security is a multi-dimensional subject, and diversity of devices and services in 5G makes it even more complex. In the following sub-section we describe the possible security solutions on a high level, and the point of view and recommendations of various regulatory and standardization bodies. In the subsequent sections we provide detailed analysis of the security challenges in the network and the main enabling technologies in 5G.

E. The Road Ahead for 5G Security

Since 5G is not an incremental advancement to 4G, security systems should also be re-designed according to the new design principles and architectural requirements of 5G. The vision of secure 5G systems that is outlined by NGMN [7] is based on three principles. These are: i) flexible security mechanisms, ii) supreme built-in security, and iii) security automation, as highlighted in Fig. 5. The objective of the vision is that 5G systems must provide highly robust security against cyber-attacks with enhanced privacy and security assurance. The security systems should be flexible for inclusion of novel technologies and using different security, e.g., encryption, technologies at different layers or network perimeters. Since 5G will accumulate very diverse technologies, such as massive IoT, which could induce security vulnerabilities in the network, security by design will be inevitable. The same reason of diversity necessitates automated security systems to intelligently adjust and adapt itself according to the environment, threats, and security controls. This will also require a holistic security orchestration and management [7].

Although, the conglomeration of new things, e.g., in massive IoT, and diverse service will make the security landscape much more complex, the new technological concepts such as SDN,
NFV, and cloud computing also can minimize the complexity. For instance, the challenges of diversity can be eliminated through policy-based, yet, centralized network control leveraging cloud computing and SDN. A centralized network controller deployed in the cloud, for example the software-based SDN controller, overlooking the whole network from a central vintage point can be used to mitigate security vulnerabilities in the network that relies on the controller. The current controller-network APIs already enable the controller to fetch the flow statistics from the data plane and use it for various purposes such as load-balancing or risk assessment. Similarly, virtualization leveraging NFV can enable service-specific slicing to strengthen security of services through isolation.

Using massive MIMO, one can direct the beam only to legitimate users and can on run-time switch the beam from potential malicious users [68]–[72]. And if the attacker is not in the vicinity of the legitimate user, then beamforming makes the overall system more secure. Due to the high attenuation loss, beam forming is a major requirement. Hence, massive MIMO, in conjunction with channel coding [73]–[75], naturally brings more security to the overall system. In mmWave, due to the reduction of the size of the antennas, it is comparatively easy to use massive MIMO beamforming system. As a result, the scalability of the overall system achieved with massive MIMO, the systems will have more resistance to security threats in the mmWave domain. The same is the case of increasing network resource in ultra-dense networks. Naturally, when the resources are more the systems will have more resistance to security attacks, such as resource exhaustion and DoS attacks.

To provide a thorough analysis of the security of future wireless networks, there are some basic concepts and definitions that cover the whole picture, irrespective of the underlying technologies. These are provided by ITU-T in the form of security recommendations and provide a stepping stone framework in security of communication networks.

F. Security Recommendations by ITU-T

The International Telecommunication Union (ITU) agency Telecommunication Standardization Sector (ITU-T) security recommendations provides a set of security dimensions to provide protection against all major security threats [59]. The eight security dimensions are not limited to network only, but covers applications and end user information as well. Furthermore, the security dimensions apply to enterprises offering services or service providers. These security dimensions are listed in Table III with brief descriptions. How the security dimensions are covered in 5G is discussed as follows.

ITU-T Study Group 17 (SG17) is assigned for security related work and potential recommendations. ITU-T has been working on security recommendations for various relevant areas in telecommunications and Internet technologies such as for the Next Generation Network (NGN), IoT and cloud computing among others [76]. For example in the case of 5G networks, two kinds of recommendation are given by ITU-T in the context of network authentication with services. It includes Push mode where network access control device is familiar with the application layer protocol, and thus share the key authentication features directly with the service platform. The other one is Pull mode, where access control device does not understand the application layer protocol, and thus the service platform has to take authentication results from 5G network [77].

ITU-T SG20 is dedicated for drafting the standards and recommendations for IoT technologies, smart cities and communities [78]. In the context of security, it is working with SG17 to draw the security requirements and standards [79], [80]. Recently ITU-T Y. 4806, has developed recommendation for safe execution of various IoT based infrastructures such as smart transportation and cities, industrial automation and wearables among others. The core aim is to enlist the potential threats to the safety of IoT enabled infrastructure and provide relevant recommendations to deal with such security attacks [81]. Apart from this, ITU-T X.1361 presents recommendation for IoT security architecture and ITU-T X.1362 for IoT encryption procedure using associated mask data (EAMD) [82], [83]. ITU-T is also working to draft of the security requirements and recommendations for narrow band-IoT.

Furthermore, ITU-T SG17 is also involved in security standards for the related enabling technologies. For example the major recommendation from ITU-T in SDN security
domain are given in the areas: Data Confidentiality, Data Integrity, Access Control, Authentication, Non-Repudiation, Communication Security, Availability and Privacy [27], [84]. In addition to SDN security, same group from ITU-T propose security guidelines for NFV security. ITU x.805 presents the security architecture for end to end network and that is based on security layers and planes. Most of the security requirements and recommendations revolve around these two concepts [59], [85]. ITU-T X.1600-1699 discuss the security requirements for cloud computing systems. Out of them, ITU-T X.1601 deals with the security framework for clouds and presents the potential attacks and related solutions [86]. Moreover, ITU-T also proposed the guidelines for security concerns among consumer and cloud service providers [87], [88].

G. Security Standardization

With the anticipation of 5G, various actors; even outside the telecom sector such as automotive are indulging in evaluating the security impacts of 5G. Hence, different key organizations are providing immense contributions for the rapid development of security standards, the most crucial ones being highlighted in Table III. However, the standardization is still in the drafting phase. In March 2015, 3GPP has set the deadline for defining the standards of 5G in or around 2020. Next Generation Mobile Networks (NGMN) published its white paper [7] on 5G in the same year that covered wide range of topics such as virtualization, IoT, radio architecture, privacy, and availability, etc. Regarding the 5G security standardization, the NGMN P1

WS1 5G security group is mainly gathering requirements and providing their suggestions.

In January 2016, the 3GPP security group SA3 [89] started work to standardize the security aspects of 5G and provide contributions to 5G Public-Private Partnership (5GPPP) initiated projects. The major task was to propose a 5G security architecture, analyze threats and outline requirements. Similarly, the 5G PPP phase 1 security landscape [90] produced by the 5G PPP Security WG highlights the new major security requirements and risks, presents the 5G security architecture along side privacy and access control in 5G, describes the security impacts of new technologies such as slicing and virtualization, and highlights the security standardization efforts. Overview of 5G security in 3GPP is also published in [91]. The work presented in [91] discusses 5G architecture enhancements and the security enhancements accordingly.

There are also several technology-specific standardization and specification bodies. For instance, the Open Networking Foundation (ONF) [92] is dedicated to the accelerated adoption of SDN and NFV technologies. ONF publishes technical specifications including specifications for security of these technologies [27]. The European Telecommunications Standards Institute (ETSI) Industry Specification Group for NFV [93] is working on security with a dedicated group called the Industry Specification Group NFV Security group (ISG NFV Sec). The group has published its latest specification on security management and monitoring specification in release 3 [94]. Regarding security, Release 3 specifies security requirements for automated, dynamic security policy management and security function life-cycle management, as well as, Security Monitoring for NFV systems. Furthermore, the ISG NFV Sec group has highlighted the need for a standard interface in the ETSI NFV architecture to enable adding security functions that can react to potential security threats in real-time.

In 2014, the ESTI MEC ISG [95] was initiated to look into Multi-access Edge Computing (MEC) security standards and empower NFV capabilities within the Radio Access Network (RAN) to deliver security and robustness. The National Institute of Standards and Technology (NIST) [96] is working to develop standards to improve IoT security. Similarly, the GSM Association (GSMA) IoT security guidelines and assessment [97] program has delivered guidelines and assessment schemes to provide proven and robust end-to-end security for IoT systems. NGMN 5G security group is working on identifying the security requirements for MEC and proposing the corresponding recommendations. Regarding privacy, subscription privacy is one of the core security areas of focused in the 3GPP SA3. For example privacy enhanced identity protection deals with safeguarding the International Mobile Subscriber Identity (IMSI) from adversaries on the air interface. SA3 is also taking valuable inputs from the Fraud and Security Group of GSMA to identify subscriber privacy challenges [90]. Furthermore, the standards suggested by the Internet Engineering Task Force (IETF) will be critical because 5G will use various Internet protocols. The ITU-T continuously gathers contributions from regional organizations

<table>
<thead>
<tr>
<th>Security Dimension</th>
<th>Brief Explanation</th>
</tr>
</thead>
<tbody>
<tr>
<td>Access Control</td>
<td>Protects against unauthorized use of network resources. It also ensures that only authorized persons or devices access the network elements, services, stored information and information flows.</td>
</tr>
<tr>
<td>Authentication</td>
<td>Confirms identities of communicating entities, ensures validity of their claimed identities, and provides assurance against masquerade or replay attacks.</td>
</tr>
<tr>
<td>Non-Repudiation</td>
<td>Provides means for associating actions with entities or user using the network and that an action has either been committed or not by the entity.</td>
</tr>
<tr>
<td>Data Confidentiality</td>
<td>Protects data from unauthorized disclosure. ensures that the data content cannot be understood by unauthorized entities.</td>
</tr>
<tr>
<td>Communication security</td>
<td>Ensures that information flows only between the authorized end points and is not diverted or intercepted while in transit.</td>
</tr>
<tr>
<td>Data integrity</td>
<td>Ensures the correctness or accuracy of data, and its protection from unauthorized creation, modification, deletion, and replication. It als provides indications of unauthorized activities related the data.</td>
</tr>
<tr>
<td>Availability</td>
<td>Ensures that there is no denial of authorized access to network resources, stored information or its flow, services and applications.</td>
</tr>
<tr>
<td>Privacy</td>
<td>Provides protection of information that might be derived from the observation of network activities.</td>
</tr>
</tbody>
</table>
like ETSI and ARIB and proposes recommendations for the standardization organizations.

A number of industry initiated efforts are in place to find the potential security challenges and seek solutions to those challenges. The Security, Identity, and Mobility alliance called the SIMalliance [98] is a global, non-profit SIM card industry association. The association advocates the protection of sensitive connected and mobile services in wireless networks. The alliance has published several reports and white papers focusing on how to improve the security of 5G networks [99], its use-cases such as smart homes security [100], and an analysis of the security needs of the 5G market [101]. A consistent point that remains in the SIMalliance is that: "Security in 5G is use case dependent".

Other related players in the industry also publish their concerns, statistics, challenges and solutions regarding security in 5G. For example, mobile network vendors such as Huawei, Ericsson and Nokia have published several reports and white papers. Huawei has been publishing various architectural and transformational aspects of security from earlier generations to 5G, and focus on the 5G security challenges and blueprint is presented in [102]. In its white paper [103], Huawei has provided interesting insights in the 5G security requirements and challenges, the security architecture transformation and promotion of 5G security standardization and their security ecosystem. Ericsson has provided interesting insights into 5G security scenarios and solutions in the white papers [104], [105]. Nokia also has also published several documents [106] and research articles [15] on 5G security. Similarly, Cisco has presented the potential of 5G and the security challenges involved with the new technologies that will be enabled by 5G in [107].

All these major vendors also participate in the standardization process and thus are providing interesting insights into security challenges and possible solutions in the respective forums such as ETSI, NGMN, and ITU. In the following section (Section V), we discuss the security challenges in 5G networks and the potential solutions or possible security approaches to counter the security challenges. In the subsequent section (Section VI), we thoroughly discuss the security challenges in the main enabling technologies in 5G along with the corresponding security solutions for those challenges.

V. SECURITY IN 5G NETWORKS: CHALLENGES AND SOLUTIONS

To properly investigate the security perspectives of the overall network in a systematic way, security in the network architecture is described in three-tiers, i.e., i) access networks, ii) backhaul network, and iii) the core network. For clarity, we have first highlighted the security challenges (also depicted in Table III) and then presented the potential solutions.

A. Heterogeneous Access Networks

1) Security Challenges: The main requirement of 5G networks is very higher data rates with ubiquitous availability and extremely low latency. New use cases of MTC, IoT, and V2X etc. will impose very diverse requirements on the network [108]–[110]. For instance, V2X and mission-critical MTC applications will need latencies on the order of 1 ms or less. On top of such requirements, the needed reliability and availability of services will be orders of magnitude higher than the current networks [111]. The current networks, however, are already prone to many Internet-based threats that can target the access nodes such as eNBs in LTE and low powered access nodes, as detailed in [54]. With the amalgamation of diverse IP devices in 5G, the security threats will further increase [5].

With a fast increase of a massive number of new devices and services, network capacity demand is increasing faster than ever. Besides improving link budget and coverage, Het Nets will contain nodes with different characteristics such as transmission power, radio frequency, low power micro nodes and high-power macro nodes, all under the management of the same operator [112], [113]. However, the diversity of nodes and access mechanisms will also bring about some new types of security challenges. For example, open access supplemental networks such as wireless local area network or even Femtocells are usually preferred by network operators to increase network capacity [114]. However, such open

<table>
<thead>
<tr>
<th>TABLE IV</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>SECURITY ACTIVITIES OF VARIOUS STANDARDIZATION BODIES</strong></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Standardization bodies</th>
<th>Workgroups</th>
<th>Major security areas in focus</th>
<th>Milestones</th>
</tr>
</thead>
<tbody>
<tr>
<td>3GPP</td>
<td>Service and System Aspects Security Group (SA3)</td>
<td>Security architecture, RAN security, authentication mechanism, the subscriber privacy, network slicing</td>
<td>TR 33.899 Study on the security aspects of the next generation system, TS 33.501: Security architecture and procedures for 5G System</td>
</tr>
<tr>
<td>5GPP</td>
<td>5GPP Security WG</td>
<td>Security architecture, the subscriber privacy, the authentication mechanism</td>
<td>5G PPP Security Landscape (White Paper) June 2017</td>
</tr>
<tr>
<td>NGMN</td>
<td>NGMN 5G security group (NGMN P1 WSI 5G security group)</td>
<td>Subscriber privacy, Network slicing, MEC security</td>
<td>5G security recommendations: Package 1 and 2, and 5G security: Package 3</td>
</tr>
<tr>
<td>NIST</td>
<td>Security working group</td>
<td>IoT security guidelines and assessment</td>
<td>Draft Interagency Report, NISTIR 8200</td>
</tr>
</tbody>
</table>
Small cells, such as femtocells [116], [117] and picocells [118], [119] using low power access points have gained momentum due to many benefits such as low-cost and indoor coverage, higher date rates and data offload relief to macro base stations, and improved user satisfaction, etc [120]. However, the low power access points will require low complexity and highly efficient handover authentication mechanisms. Such fast and reliable handover mechanisms are yet to be developed for 5G, whereas the previous cryptographic methods will not suffice for low-powered access points [121]. This gape can open the network to security vulnerabilities such as man-in-the-middle attacks and phishing attacks.

Handover between different access technologies, for instance 3GPP and non-3GPP has been another challenge from security perspective as outlined for 4G HetNets in [122]. For example, session replay attacks through recovering session keys and the possibility of malicious point of access being non-3GPP secured are the key challenges. The later has more relevance to 5G due to the increased number of access points and different access technologies in 5G HetNets. An instance of that in terms of vulnerability with roaming has been revealed in [123] regarding the 5G-AKA protocol specified within 3GPP TS 33.501 v0.7.0. [124]. Furthermore, the encryption keys for the radio interfaces are usually computed in the home core network and transmitted to the visited radio network over SS7 or DIAMETER signaling links. NGMN has pointed out in [77] that these keys can be leaked out and making a clear point of exposure in the network. The basic techniques for improving their security include improving the SS7 and DIAMETER security by introducing firewalls [77]. However, design of secure key management protocols is still an open challenge in 5G networks.

The current 3GPP networks require the UE to provide its IMSI over the air in an unencrypted form during the initial attach phases. This enables passive attackers to identify a user from the IMSI by observing the traffic [125]. This also enables the attacker to track the user during roaming from one network to another. Future mobile network operators will use less trusted or non-3GPP networks alongside the trusted or 3GPP networks. Roaming from one network to another, i.e., non-3GPP to 3GPP will be common. During roaming, the UE also has to provide its IMSI to the SN for authentication, which is another challenge for IMSI security or user’s privacy [125].

Since, 5G will accommodate a huge number of user devices, and smart things (e.g., IoT), one of the key challenges in 5G will be the massive number of nodes sending and receiving data simultaneously, practically jamming the radio interfaces. The situation can be worsened by malicious nodes sending excessive signaling traffic to cause scalability challenges or in other words, DoS attacks. According to the 5G security recommendations by NGMN [77], a malicious attack can trigger massive simultaneous and continuous acquisition of radio resources to cause the signaling plane overload. A critical analysis of physical layer based security approaches are highlighted in [126] suggesting that the existing solutions for jamming attacks, channel prediction, and passive and active eavesdropping will not suffice for the enhanced physical layer techniques adopted for 5G.

2) Security Solutions: In ideal situation, the access network security should protect the user, the network infrastructure and services from all the possible threats that could originate in the radio part of the network. 5G will use a diverse variety of access technologies for extended coverage, higher throughput and lower latencies. Hence, 5G will leverage virtualization, SDN, and cloud technologies to adapt execution logic to specific services with composition and instantiation of access in different network locations [127]. With such capabilities, 5G will improve the systems’ robustness against various types of security challenges arising in diverse access technologies.

Communication security can be provided through multiple varying methods, usually in the upper layers. Physical layer security, however, usually reduces the design complexity besides ensuring information security [128]. Initiated mainly by Shannon [129], and further strengthened with theoretical basis using information-theoretic approach by Wyner [130], physical layer security of wireless communications does not rely on higher-layer security systems or encryptions [131]. The fundamental principle behind physical layer security is to utilize the randomness of the noise and communication channels to restrict the amount of information that can be revealed by unauthorized receivers [131]. Hence, the use of physical layer security schemes makes it very difficult for attackers to decipher or access information under transmission [132]. Physical layer security has been a hot research topic that has various dimensions and viewpoints as described in [126]. Therefore, there are many surveys and magazine articles [20], [131], [133]–[135] that cover various aspects of physical layer security.

The modification of transmitted waveforms for secure transmission gained a lot of attention in recent years. In [136], the authors presented a method that eliminates the need of Cyclic Prefix (CP) insertion between successive Orthogonal Frequency Division Multiplexing (OFDM) symbols [137], [138]. The proposed technique adds an alignment signal on top of each transmitted OFDM symbol that cancels the interference of adjacent symbols. Due to the alignment signal function, the eavesdroppers at different locations experience more interference. In [139], a secure Orthogonal Transform Division Multiplexing (OTDM) waveform is designed to diagonalize the channel matrix of the legitimate user while degrades the channel condition of the eavesdropper. The proposed method uses orthogonal transform basis functions which are extracted from the legitimate user’s channel to shape the waveform securely.

A physical layer security technique called OFDM with Subcarrier Index Selection (OFDM-SIS) is presented in [140]. The technique uses the SIS and adaptive interleaving to provide a two-fold security. The SIS technique maximizes the SNR only towards the legitimate user while the interleaving performed on the legitimate user’s channel is different from eavesdropper’s channel. In [141], a chaotic Discrete Hartley Transform (DHT) is proposed to enhance the physical layer security of optical OFDM in passive optical networks. Typically, a data
encryption approach is adopted to enhance the physical layer security for optical OFDM which mostly disregard the transmission performance improvement. The chaotic DHT for the optical OFDM of [141] enhances the physical layer security and improves the transmission performance. A frequency pre-coder and post-coder for OFDM-based systems for physical layer security design is proposed in [142]. The diagonal matrix of the channel of the legitimate user is decomposed into two unitary orthonormal matrices with this technique. The first orthonormal matrix is used as a pre-coder before the IFFT at the transmitter while the other matrix is used as a post-coder at the receiver.

In [143], a resource allocation problem for the orthogonal frequency division multiple access (OFDMA) systems is formulated which takes multiple-antenna eavesdropper, dynamic power consumption, artificial noise injection for secure communication and secrecy data rate requirements were taken into consideration. The closed-form solution of the problem is derived by the dual decomposition which maximizes the number of securely delivered bits per joule. The simulation results provide the achievable maximum energy efficiency in presence of a multiple-antenna eavesdropper. The simulation results also reveal that the system energy efficiency decreases as the capability of the eavesdropper increases. A physical layer technique using time domain scrambling technique is proposed in [144] which improves the confidentiality and security of OFDM systems. The technique proposes scrambling the sample sequence within each time domain OFDM symbol to eliminate the inherent signal statistic features. The proposed scheme increases time complexity for cracking and thus demonstrates promising security capability. In [145], a secure OFDM scheme is proposed which is based on channel shortening. The channel of the legitimate user is made shorter than the CP length, while the channel of the eavesdropper remains longer than the CP. This causes loss of overall performance of the eavesdropper. Simulation results show a significant performance difference between the legitimate user and eavesdropper. A list of the techniques to shape the

<table>
<thead>
<tr>
<th>Security threats</th>
<th>Potential targets</th>
<th>Affected network segments</th>
</tr>
</thead>
<tbody>
<tr>
<td>DoS attack on signaling plane</td>
<td>Centralized control elements</td>
<td></td>
</tr>
<tr>
<td>Hijacking attacks</td>
<td>SDN controller, hypervisor</td>
<td>✓</td>
</tr>
<tr>
<td>Signaling storms</td>
<td>5G core network elements</td>
<td>✓</td>
</tr>
<tr>
<td>Un-authorized access</td>
<td>Low-power access points</td>
<td>✓</td>
</tr>
<tr>
<td>Configuration attacks</td>
<td>Low-power access points</td>
<td>✓</td>
</tr>
<tr>
<td>Saturation attacks</td>
<td>Ping-pong behavior in access points, and MME</td>
<td>✓</td>
</tr>
<tr>
<td>Penetration attacks</td>
<td>Subscriber information</td>
<td>✓</td>
</tr>
<tr>
<td>User identity theft</td>
<td>User information data bases</td>
<td>✓</td>
</tr>
<tr>
<td>Man-in-the middle attack</td>
<td>Un-encrypted channels, e.g. in IoT</td>
<td>✓</td>
</tr>
<tr>
<td>TCP level attacks</td>
<td>Gateways, router and switches</td>
<td>✓</td>
</tr>
<tr>
<td>Key exposure</td>
<td>Radio interfaces</td>
<td></td>
</tr>
<tr>
<td>Session replay attacks</td>
<td>Session keys in non-3GPP access</td>
<td>✓</td>
</tr>
<tr>
<td>Reset and IP spoofing</td>
<td>Control channels</td>
<td>✓</td>
</tr>
<tr>
<td>Scanning attacks</td>
<td>Radio interfaces interfaces</td>
<td>✓</td>
</tr>
<tr>
<td>IMSI catching attacks</td>
<td>Roaming and UE</td>
<td>✓</td>
</tr>
<tr>
<td>Jamming attacks</td>
<td>Wireless channels</td>
<td>✓</td>
</tr>
<tr>
<td>Channel prediction attacks</td>
<td>Radio interfaces</td>
<td>✓</td>
</tr>
<tr>
<td>Active eavesdropping</td>
<td>Control channels</td>
<td>✓</td>
</tr>
<tr>
<td>Passive eavesdropping</td>
<td>Control channels</td>
<td>✓</td>
</tr>
<tr>
<td>NAS signaling storms</td>
<td>Bearer activation in core network elements</td>
<td>✓</td>
</tr>
<tr>
<td>Traffic bursts by IoT</td>
<td>Saturation of GTP end-points</td>
<td>✓</td>
</tr>
</tbody>
</table>

Fig. 6. High-level architecture presentation of 5G networks.

TABLE V
SECURITY CHALLENGES IN 5G NETWORK SEGMENTS
waveform to secure the communication is presented in [146].
A comprehensive survey on various OFDM based physical layer techniques is presented in [147]. We invite interested readers to go through [146] and [147] to learn more about physical layer security by modifying the potential waveforms that are expected to be used in 5G.

In multi-tier HetNets, the security performance of multiple diverse RATs can be increased by deploying more low power base stations in high path loss environments [148]. To mitigate the security risks during handover between low and high-power base stations, [149] proposes SDN-based authentication handover that enables HetNet management through global visibility of user behavior and resource management. The SDN-based approach shares user-dependent security context information among related access points, as well as ensures efficiency in delay-constrained 5G environments. Duan and Wang [121] proposes SDN enabled weighted secure context information transfer in order to minimize the authentication delay during handover among multiple RATs in HetNets.

Simple public key infrastructure based distributed access control is proposed in [150] to enable access control and authentication in 5G HetNets. The proposed authentication and registration scheme uses elliptic curve zero knowledge proof to achieve secure certificate generation and enable secure end-to-end communication. The certificate of Simple public key infrastructure based handover schemes ensures seamless and secure mobility among different cells within the latency requirements. SDN-based centralized control platform monitors the whole network activity in [150]. A deterministic low-jitter scheduling and lightweight layer 2 encryption coupled with deterministic network slices is presented in [151]. The mechanism in [151] uses SDN for centralized monitoring and dynamic reconfiguration of slices to achieve strong security and privacy for M2M and IoT communication.

Solutions for roaming between trusted and non-trusted networks include the use of an identifier (temporary identity) instead of sending IMSI to non-trusted networks. A possibility is that the identifier is generated by the trusted network and shared between UE and non-trusted networks [125]. 3GPP has, so far, used the EAP-AKA for authentication between UE and non-3GPP access networks. In 5G, multiple methods based on EAP can be used, based on the type of service or preferences by the service provider.

B. Backhaul Networks

1) Security Challenges: The backhaul network comprise network elements and communication channels between the base station and the core network [152]. Therefore, the backhaul network is not as exposed security threat as the access network. The communication can be either wireless [153], usually microwave and occasionally satellite links, wire-line through dedicated copper or fiber optics [154], or a converged one using both [155]. The diversity in RAN technologies has also evolved the backhaul networks, more so until LTE [156]. The security of backhaul is different in a sense that it involves both radio and core part of the network. For instances, the security parameters are, usually, adjusted in the elements of the access network such as eNB, and MME in the core network.

For traffic towards the Internet or external network, the eNB sends the traffic to the serving gateway through GPRS Tunneling Protocol (GTP). The serving gateway sends the traffic to Public Data Network (PDN) gateway which communicates with external networks or the Internet. GTP is again used over the S5/S8 interface between the serving gateway and PDN-gateway. The LTE the backhaul enhanced network security through introducing Internet Protocol Security (IPSec) based GTP tunnels for the X2 interface between eNBs, and S1 interface between eNBs and MMEs [156]. One of the basic challenges with such tunnels is that a tunnel must be setup when a UE enters active mode, or even starts a session with a new service. This will be highly challenging when a huge number of IoT devices transmits small and sporadic traffic [157], opening doors for DoS attacks on the end-points.

With the introduction of SDN, the backhaul network will be simplified due to the control-data planes functional split, as described in [158]. The control plane of the devices such as serving or PDN gateways will be transformed into the core network, whereas simple forwarding devices will be used to route traffic to and from the Internet. However, the simplified devices such as OpenFlow switches and the communication channels have their own security challenges such as weakness of Transport Layer Security (TLS), and resource exhaustion attacks as detailed in [27] and described in Section VII-B. In the case of wireless backhaul, massive MIMO will be used due to the increasing traffic demands. Massive MIMO has its own security challenges, which are described in Section VII-A.

2) Security Solutions: Due to its close nature and as a medium between two secure nodes, there are few security challenges in the backhaul network. Although GTP has several weakness, it also has some benefits regarding security. Since GTP works on top of User Datagram Protocol (UDP), it operates well with firewalls [159]. Even though GTP used in the operator’s internal network, the support for firewalls might be needed to curb resource exhaustion attacks on GTP end-points. Furthermore, the signaling traffic due to short traffic burst due to changing active-sleep modes, or increased number of IoT devices, can be reduced by using the same tunnel for multiple IoT devices with similar service requirements [157]. Moreover, the backhaul network is transforming into virtualized and SDN-based data plane, whereas most of the control plane functionalities will be shifted to the core network. Working as simple forwarding devices, the security threats in backhaul will drastically minimize. Since, SDN based backhaul will have its security challenges, the solutions for those challenges are discussed in Section VII-B.

C. Core Network

1) Security Challenges: The core network of LTE or 4G, called EPC, comprised different entities such as MME, serving gateway, PDN gateway, and HSS [160]. In 5G, the core network elements are represented by network functions [161]. The detailed architecture of 5G core network with description of network functions is available in the latest 3GPP
The core network is IP based and ensures end-to-end service delivery, security and QoS, and maintains subscriber information. The 5G core network is more dynamic compared to the previous generations leveraging NFV, SDN and cloud technologies as described in Section VI. However, it is the main target of security threats and prone to security vulnerabilities as well.

The massive penetration of IP protocols in the control and user planes for different network functions make the 5G core network highly vulnerable. The network must be highly resilient and ensure availability with the increasing signaling traffic. The increasing types of communication services and devices can lead to high traffic volumes for signaling purposes. The signaling procedures for attach/detach, bearer activation, location update, and authentication occurring at the Non-Access Stratum (NAS) layer of 3GPP protocols can cause NAS signaling storms [1]. This will be particularly more challenging in 5G due to possible integration of billions of IoT devices. As a precaution, Nokia published [163] that signaling traffic is increasing 50% faster than the data traffic, yet the data traffic is increasing by 56 percent each year [164]. Using excessive signaling to create a DoS attack on LTE has been demonstrated in [165].

Small cells with huge number of mobile devices moving around will increase mobility handovers, thus adding to the signaling traffic. Hence, the signaling load will not only increase on MME, but on other control elements (network functions in 5G) such as HSS, PDN gateways and serving gateways as well to maintain QoS. A DoS or resource exhaustion attack on any of these network elements will be easy to materialize. Furthermore, the 3GPP NAS layer protocols used for UE attach or detach functions, authentication, bearer activation, and location update can also result in signaling storms [166]. The 3GPP also recommends the use of IPSec encryption for LTE interfaces such as S1-MME, X2, S5 and S6, etc. Such massive encryption-based tunnel establishment also increases the signaling costs in the core network elements. The main objective here is to outline the vulnerability of the core network to DoS attacks.

In 5G, huge number of infected IoT devices can overload the signaling plane as an attempt to gain access or perform a DoS attack [77]. Resource constrained IoT devices, probably in billions [167], will require the resources in the clouds to perform processing, storing or sharing of information. Their limited capabilities also make these devices an easy target to masquerade or operate in a compromised environment for attacks on the network in the form of DoS attacks. Thus, IoT will bring many challenges for the signaling plane or the core network of 5G networks. In LTE, the HSS has been described in [168] as the main point of attacks in terms of requests for authentication and authorization.

3GPP details the subscription security policy update in its study on the “Feasibility Study on New Services and Markets Technology” [125]. The point of concern is the suggestion that IoT devices should periodically update the subscription security credentials, even devices with less frequent communication with the network should update the credentials from the core network. The reasons provided are that, IoT devices will be huge in number, have low capacities, and thus can be compromised which necessitate the update to avoid security lapses in the network side. However, the massive burden on the core network entity that will be involved in the security policy update is not discussed. Such updates involving massive number of IoTs can potentially make the relevant core network element or function a bottleneck.

Moreover, the 5G core network is expected to utilize SDN and NFV to provide higher flexibility and scalability with cost-effectiveness [111]. These technologies have their own security challenges that must also be properly investigated due to their vital role not only in the core network, but also in the security of the overall network and services. The main security challenges in these technologies are described below.

2) Security Solutions: The 5G core network introduces new features and utilizes new technological concepts compared to EPC [91]. According to the 3GPP release 15 [169], the most obvious changes are: 1) Control-User plane separation, 2) Network slicing, 3) Service based architecture, and 4) Flexible Non-3GPP access internetworking. All these changes have been made possible through technological concepts such as SDN, NFV and virtualization, and cloud or mobile edge computing and their instantiation in the wireless network domains to achieve cost efficiency and simplicity in network management. Therefore, their security challenges and solutions are properly described in Section VII and Section VIII respectively. In this section, we describe the security solutions for those challenges that are highlighted in Section V, considering the core network as holistic IP based logically centralized core network.

For the logically centralized core network the main challenge would be the signaling overload due to the huge number of diverse devices. However, one of the main change in 5G is that the core network elements, e.g., MME are represented by network functions such as Access and Mobility Management Function (AMF) [169]. There are also clearly stated protocols and reference points for interaction among them [161]. Hence, to effectively handle the signaling overload challenge, two approaches are discussed in 5GPPP [90]. First, using lightweight authentication and key agreement protocols for communication of massive IoT. Second, using protocols that allow to group devices together through various types of group-based AKA protocol [90].

A group based authentication scheme for narrow-band IoT devices has been proposed in [170]. IoT devices with similar attributes are grouped (temporarily or permanently) together, a group leader is chosen, sensitive information of IoT in the group is aggregated by the group leader, and sent to the core network where the credentials of each IoT device can be verified separately by the corresponding node in the core network. The groups are formed using anonymous attribute-based group establishment techniques. The propose mechanism also improves identity privacy preservation besides minimizing the authentication overhead and signaling in the core network. A similar unified group-based authentication scheme using SDN has been proposed for V2X in [171] that minimizes the handover signaling overhead in future networks.
Since 5G is supposed to provide higher flexibility and agility, the two concepts that are most prominent in this regard are Virtual Network Functions (VNFs) and software-based network control. These features will be enabled by NFV and SDN. The basic philosophy of NFV is to implement network function in software called VNFs and deploy them on high-end servers or cloud platforms instead of specialized function-specific hardware. SDN, on the other hand, separates the network control plane from the data forwarding plane. The network control plane can be logically centralized and the forwarding devices can be rendered simple to increase innovation in network control platforms and simplify the network management. The two concepts can also be used to improve network security.

The SDN based core network is described in [172]. A hierarchical SDN-based control is proposed for the whole network that would allow different grade performance of the core network and provide service differentiation in 5G. The proposed core network is a unified approach that takes care of mobility, handoff and routing management with hierarchical control planes. The hierarchical control plane that handles different functions of the core network such as mobility or QoS management, increases the scalability of the network to meet the requirements of diverse services. Even though the proposal is not for strengthening the network security, the approach increases the capability of the network to ensure availability for diverse services through modularity.

NFV also increases the network capabilities in terms of security, as described in [173] for sharing network resources dynamically. For instance, in the case of mobility the virtual AMF can be placed in different network sections (edge nodes) at run time where mobility is higher or instantiated on multiple hardware near the highly mobile devices. The aim is to eventually divide the traffic among multiple VNFs that could potentially exhaust the resources of an AMF on a single node.

A detailed article on the benefits and challenges in using NFV in mobile networks [174] discusses how the foreseen signaling costs can be reduced in mobile networks. The authors propose a grouping criterion to bundle together various functions of the core network in order to minimize the signaling costs and improve the performance of the overall network. Similarly, by using the GTP and avoiding some protocols, such as the DIAMETER protocol, the signaling traffic is drastically reduced. The DIAMETER protocol relies on Transmission Control Protocol (TCP) and stream control transmission protocol. Both are known to downgrade the network performance during small bursts of traffic [175].

### VI. Security in Key 5G Technologies

The security challenges in 5G can be effectively outlined by considering the main enabling technologies of 5G. As described in Section IV, the main enabling and disruptive technologies compared to the previous generations are massive MIMO antennas, SDN, NFV and the concepts of cloud computing such as Multi-access Edge Computing (MEC). SDN, NFV and cloud computing have been used in non-wireless networks and thus have a rich literature in terms of security. In this section, we describe their security challenges and solutions mainly from the scope of their use in 5G network. First the security challenges are discussed (also highlighted in Table VI) and then the security solutions or proposals are described.

#### A. Security in Massive MIMO

**1) Security Challenges in Massive MIMO:** Massive MIMO is considered as one the most promising and disruptive technologies for 5G [176]. The key idea of massive MIMO is to equip the base station with a large number of antenna elements that can serve a large number of user terminals with the same frequency band [3]. The large number of antenna
elements can be used in various modes to increase the data rates or to enhance the reliability, coverage or energy efficiency. In addition, random matrix theory demonstrates that the effect of small-scale fading and uncorrelated noise start to fade as the number of antennas goes towards infinity [177]. Despite its promises, the base-station needs to estimate the Channel State Information (CSI) either through feedback or channel reciprocity schemes to reap the benefits of massive MIMO. The use of non-orthogonal pilot schemes for a multi-cell Time Division Duplex (TDD) networks introduce the concept of pilot contamination due to the limitations of coherence time [178]. The effect of pilot contamination is much more profound on massive MIMO. Pilot contamination is considered as one of the major performance limiting factor of a massive MIMO system [179].

The security vulnerabilities in massive MIMO: passive eavesdropping and active eavesdropping, are described in [20]. In the passive eavesdropping, the attacker tries to intercept the transmitted signals. The passive eavesdropper does not transmit any signal itself. In the active eavesdropping, the attacker also transmits signals to disrupt the legitimate user’s transmission. If the only goal of the active attack is to disrupt the legitimate transmission, it can be called a jamming attack [180]. Another intelligent form of the active attack is based on pilot contamination. It is called pilot spoofing where the attacker pretends to be a legitimate user.

Typically, the CSI is used at the base station to precode the transmission so that a composite beam, which is created from the signals transmitted through different antennas, can be focused towards a specific user. The CSI is obtained through the channel estimation process which is generally based on the pilot signals sent by the legitimate users. A pilot spoofing scheme that exploits the pilot contamination is presented in [181]. The eavesdropper sends the same pilot signals to confuse the base station. Therefore, the base station designs the precoder incorrectly which benefits the reception of the eavesdropper. The pilot training sequence is fixed and repeated over time and therefore, it can be obtainable by an attacker. Due to this attack, the estimated channel between the base station and legitimate user becomes inaccurate while it helps the attacker to detect the transmitted signals from the base station. The important assumption made in [181] is that the transmission of the attacker and legitimate user is synchronized. The pilot spoofing attack detection and countermeasures are also presented in [180], [182]. We discuss the solutions in the next subsection.

According to [183], the jamming attacks are more difficult to deal with than the spoofing attacks for a massive MIMO receiver. The attacker tries to create maximum jamming possible unlike the spoofing attack. The jamming attacks are typically dealt with designing receiver that consider the jamming signals as additive noise. However, the jamming is not noise-like for massive MIMO as the legitimate channel is correlated with the jamming channel [183]. The concept of beamforming, where several antennas serve a specific user, make the massive MIMO systems inherently robust against passive eavesdropping attacks. However, the eavesdropper can take countermeasures by exploiting the high channel correlation in the vicinity of the user or the weakness of channel estimation. An overview of passive eavesdropping and active attacks on massive MIMO and possible detection techniques for such threats is presented in [184]. The channel estimation procedure in MIMO has been one of the soft targets for security attacks, and it has been properly demonstrated in [185]. Incorrect channel state information can also be used for jamming attacks [186], which is also demonstrated in [185].

2) Security Solutions for Massive MIMO: To attain full benefits of MIMO, the system must be secured against the major security challenges. Two different schemes to detect an active eavesdropper are presented in [184]. One method is to exploit controlled randomness by transmitting random pilots to detect active eavesdroppers. The legitimate user transmits a sequence of random symbols of random phase-shift keying that enables the base station to detect the eavesdropper. The drawback of this method is that it incurs the overhead of transmitting additional random sequences. In the second method, the beamformer is constructed in such a way that the received sample by the legitimate user equals to an agreed value. In the case of an active eavesdropper, the legitimate user will observe a much smaller value. The detection of active eavesdropping can be tackled by a co-operative base station. In such scenarios, different base stations can exchange information and thus presents an opportunity to jointly estimate the level of legitimate user induced pilot contamination. Machine learning methods can also be employed for detecting active eavesdropping attacks.

As the massive MIMO base station can serve a large number of users at the same time, it is necessary to secure a message from all the users other than the intended one. The precoder algorithm used in the base station has to be designed in a way to achieve this objective. It is also worth considering the possibility of an eavesdropper employing powerful massive antenna arrays to intercept the information. In [187], a physical layer security approach called original symbol phase rotated secure transmission scheme is proposed to defend against such a scenario. The basic idea of this phase is to rotate the phase of the original signal randomly to confuse the eavesdropper. On the other hand, the legitimate users are able to correctly infer the phase rotation and take proper inverse operations necessary to recover the original transmission.

A jamming resistant receiver is proposed in [183], [188] to counter the jamming attacks on massive MIMO up-link. The authors exploited the unused pilots sequences to estimate the jamming channels. The receive filters are designed based on both legitimate user channels and the jamming channel to combat the jamming signals. The filters are based on conventional Minimum Mean-Square Error (MMSE) and Zero Forcing (ZF) filters. In [189], the authors use a generalized likelihood ratio test to detect a jamming attack. The performance of the detector increase with the number of BS antennas. An anti-jamming strategy based on pilot re-transmission is proposed in [190]. The authors proposed counter attack strategies for random and deterministic jamming attacks.

A two-way training based scheme against the pilot spoofing attack is proposed in [182]. The authors proposed a scheme
where the uplink and downlink both channel estimates are available at the base station receiver which can be used to find the difference between two estimation results. The detection outcome will be fed back to the transmitter together with the down-link channel estimates. The proposed scheme achieves a high detection probability and a positive secrecy rate [182]. In [180], an enhanced scheme is proposed to combat spoofing attacks which is based on the recent approach of superimposing a random sequence on the training sequence at the legitimate receiver. The authors consider two scenarios where the spoofer transmits the pilot signals only and the spoofer transmits both pilots and random sequences. The proposed scheme is based on random matrix theory based source enumeration [180].

In [191], a data-aided secure massive MIMO transmission with an active eavesdropper is presented. The authors show analytically that decreasing the signal power of the legitimate user is an effective approach to combat against a strong active eavesdropper. The authors proposed a data-aided secured down-link transmission scheme with an achievable secrecy sum-rate precoding. In [192], a linear precoder for massive MIMO eavesdropper's wiretap channel with finite alphabet input is investigated. An upper bound on the secrecy rate for the Gaussian Singular Value Decomposition (GSVD) design is derived which reveals that GSVD leads to a significant performance loss. The authors proposed and analyzed Per-Group-GSVD (PG-GSVD) which eliminates the performance loss of GSVD. Another secure massive MIMO transmission scheme is investigated in [193]. The authors derived an achievable secrecy rate analytically when the number of transmit antennas approaches infinity. The derivation also assumes matched filter precoding and artificial noise generation in the transmitter side. The authors proved that the effect of the active eavesdropper can be completely eliminated when the transmit correlation matrices of the eavesdropper and the users are orthogonal. In addition, the authors derived a closed-form expression for optimal power allocation for secure communication for a single antenna active eavesdropper.

B. Security in SDN

SDN separates the network control plane from the forwarding plane and centralizes the network control into software-based network control platforms. The softwarized network control functions are logically centralized that interact with forwarding devices through programmable APIs. This achieves simplicity in network control, management and operation, and accelerates novelty in network feature development and deployment. Therefore, using the concepts of SDN in wireless networks has been a major focus of researchers and industry. Thus, there are many proposals for SDN-based wireless networks [194]–[197]. The SDN architecture has three functional layers with interfaces between the layers, as shown in Figure 7. OpenFlow is the first viable implementation of SDN that follows the three tier architecture of SDN with OpenFlow applications, OpenFlow controller and OpenFlow switches.

1) Security Challenges in SDN: In SDN, network functions can be implemented as applications deployed in the SDN application plane. The controller provides an abstract view of the network to applications. Thus, in SDN applications can manipulate the network according to the application requirements. However, SDN has many security challenges, more so already demonstrated about the OpenFlow implementation of SDN. For instance, centralizing the network control make the control platform a favorable choice for DoS attacks. Similarly, exposing critical APIs to unintended software will expose the network to security threats. The number of security challenges have grown since the inception of OpenFlow. Below we highlight the main security challenges related to SDN.

Security Challenges in SDN applications: SDN has two principle properties that make the foundation of innovation in communication networks on one hand, and the basis of security vulnerabilities on the other. These are, first, the control of a network with software, and second, centralized network intelligence [198]. In SDN, most network functions will be implemented as applications, thus, a malicious application if granted access, can spread a havoc across the network. The security challenges introduced by applications can be due to open APIs in network equipment, lack of trust mechanisms between applications and controllers, and lack of proper authentication and authorization techniques for applications, mainly third party applications [198]. Since 5G will implement most network functions as applications, solutions to such challenges need to be sought out before enabling SDN applications to manipulate the network.

Security Challenges in SDN controllers: The centralized control plane of SDN (e.g., OpenFlow controller) makes it a highly targeted point for compromising the network or carrying out malicious activities in the network due to its pivotal role in decision making. The main types of threats will be DoS and DDoS attacks. However, the centralized control plane as implemented in OpenFlow have even more security consequences. For example, controller visibility, controller-applications interaction and controller scalability can be adversarially used to compromise the security of the whole network. Regarding controller visibility, a DoS attack is demonstrated in [199] that uses the control-data plane separation logic to recognize the controller and send specifically crafted flows to
exhaust the control plane resources. Regarding the controller-applications interaction, there are no compelling mechanisms to secure the controller from malicious applications opening the challenges of application authorization, and resource usage auditing and tracking [200]. By exploiting the controller scalability, an attack is demonstrated in [201] in which IP packets with random header fields are continuously sent to the controller making it unavailable to legitimate flow setup requests.

Security Challenges in the data plane: The SDN data plane comprises simple forwarding devices having flow tables that are used by SDN controllers to install flow forwarding rules. The flow tables are limited in capacity, and thus can be exhausted by maintaining a large number of unsolicited flows. Thus malicious flows with different field headers will easily exhaust the flow tables. This principle can be used for saturation attacks. During such attacks, legitimate flows will be discarded due to the limited capability of the switch to buffer TCP/UDP flows. Since SDN switches are dumb and are not capable to differentiate between genuine and malicious flows, they can be used for attacks against other switches and even controllers. The dependency on the controllers also make the data plane security dependent on the controllers, and thus, if a controller is compromised a network of SDN switches will be unwittingly compromised.

Security Challenges in interfaces: SDN has two main interfaces, i.e., the north-bound interface between controllers and applications, and the south-bound interface between controllers and SDN switches. The north-bound interface is mainly a challenge for remote applications due to the availability of standardized interfaces. The south-bound interface uses Transport Layer Security (TLS) and Datagram Transport Layer Security (DTLS). However, due to the configuration complexity in the use of TLS and DTLS, their use is left optional. This makes the controller-switch communication open to a multitude of attacks, including eavesdropping or even attacking the control plane.

2) Security solutions for SDN: The security of SDN is multi-dimensional. In this article we focus on two main disciplines. First, the security of SDN due to its inherent nature that makes SDNs vulnerable to security threats. The second is how SDN or the principles of SDN, such as centralized network control and greater control and visibility of traffic flows, can be used to increase network security. In this subsection we focus on the first principle, i.e., security mechanisms proposed to increase the security within SDNs or its individual planes.

Security solutions for the SDN application plane: Malicious applications must not be granted access to the network or the network control plane in SDN. Therefore, there are various proposals that perform rigorous verification of SDN applications before they are granted access for network configurations through the control plane. For example, PermOF [202] is a fine-grained permission system that puts boundaries on applications to work within its defined privileges. The design of PermOF provides read, write, notification, and system permissions to different application to enforce permission control. Thus, it secures the control platforms from malicious applications. Another permission system for SDN applications is described in [203] that ensures that the control plane operations are available only to trusted applications. Similarly, FortNOX [204] is a security enforcement kernel proposed as a solution for malicious applications that implements role-based authorization for each OpenFlow application. The ROSEMARY controller [205] also proposes a permission system for applications to secure the controller operation from buggy or malicious applications.

Security solutions for the SDN control plane: Due to the pivotal role of the control plane, there are many proposals and approaches to strengthen its security. The Security-Enhanced (SE) Floodlight controller [206] extends the security of the original floodlight controller [207]. To secure the SDN control layer, the SE-Floodlight controller provides mechanisms for privilege separation by adding a secure programmable north-bound API to the SDN controller. It operates as a mediator between the application and data planes by verifying flow rules generated by applications. The ROSEMARY controller [205] is a robust network operating systems for the SDN controller to secure it from malicious applications.

To mitigate the controller scalability issues and improve resilience against DoS attacks, AVANT-GUARD [208] limits the flow requests (failed TCP sessions) to the control plane using a connection migrating tool. There are also various approaches to improve the control plane security against DoS attacks. For instance, self-organizing maps [209] are used in [210] against lightweight DoS and Distributed DoS (DDoS)attacks. Further approaches to increase controller resiliency against security lapses include distributed control plane [211], [212], [213], controller placement [214], [215], control plane redundancy [216], and reactive vs proactive controller flow rule setup and updates [217].

Security Solutions for the data plane: The data plane transports the actual packets, and thus needs proper security mechanisms. Since, the configurations in the data plane can be changed by applications, it must be secured from unauthorized applications. Therefore, security mechanisms such as authentication and authorization are used for applications that change the flow rules in the forwarding elements in the data plane. FortNox [204] provides the mechanism in the controller to check contradictions in flow rules that are generated by applications. The FlowChecker [218] can check and identify inconsistencies in flow rules in OpenFlow switches. FlowChecker can also detect intra-switch misconfigurations through binary decision diagrams.

Security Solutions for the SDN interfaces: The north-bound interface in SDN is still an open security challenge though there are many proposals to secure the controller interface from malicious applications. The main challenge that needs further research is the security of the interface when remote applications would like to access the control plane or configure the data plane. Security mechanisms such as those for the south-bound interface that use TLS are not yet available. In the south-bound interface, the OpenFlow protocol supports TLS and DTLS for TCP and UDP traffic respectively. TLS provides privacy and data integrity for the user communication, whereas DTLS secures UDP traffic between applications.
TLS uses symmetric cryptography for data encryption. Their detailed use is available in the OpenFlow specification [219].

C. Security in NFV

The main idea behind virtualization is to decouple a system’s service model from its physical realization to use logical instances of the physical hardware for different purposes. Using the same idea, NFV separates network functions from the underlying proprietary hardware [174]. By transferring network functions from hardware to software applications, NFV makes the foundation for enabling run-time network function placement at different network locations [174]. Thus, NFV provides a futuristic demand-based network functions placement in different network perimeters and eliminates the need for function or service-specific hardware [220], [221]. However, there are a number of security challenges that have sparked serious concerns about the security of user information, the services and the network itself. These challenges are briefly described below.

1) Security Challenges in NFV: With the deployment of NFV, a number of security challenges will surface mainly due to the possibility of functions or service migration from one point to another or from one resource to another [222]. Since the number of services or virtual functions will grow, a growing concern is related to the manual configurations of the virtual systems or VNFs [223] that can lead to potential security breaches due to the increased complexity with the growth of the systems. Similarly, the increased number of VNFs is also a major concern for unauthorized data access, traffic eavesdropping, and theft of services [18]. Moreover, there are security challenges that are inherent to virtualized or NFV systems as described below.

Challenges in virtual systems: Virtualized systems cannot always be secured like the physical systems [224]. Many virtualized systems can run on the same network component but each might not need the same security, thus the same security procedures cannot be applied on the whole machine. A scenario is described in [224] in which a server hosting virtual machine is divided into multiple zones, each having different security levels. In this case, a zone that need to conform with specific security level cannot be moved to another physical server since the other computer may not offer the same security. On top of that, service chaining of various NFVs will make the analysis of root causes of security threats even further complicated [225].

Challenges in hypervisors: In network virtualization, hypervisors are used to map various network functions to logical instances or physical hardware of network components [226]. A hypervisor can create and execute multiple guest operating systems and controls the necessary CPU scheduling and memory partitioning for those systems. Thus the hypervisor, also called the virtual machine monitor, is the main entity in the whole of hypervisor-based virtualized eco-system. Hence, if a hypervisor is hijacked the whole system can be compromised [224]. A number of attacks on the hypervisor are discussed in [227]. In brief, the hypervisor can be targeted for a number of attacks such as exploiting host operating system to damage the isolation of a network slice, DoS attack on VMs, and VM hopping attacks.

Challenges due to dynamicity: An interesting benefit of VMs is that they can be created, deleted and moved around a network easily. The same reason makes VMs more a security concerns since tracking a malicious virtual machine would be much more complex. The same challenge will also be carried into the 5G networks domain [5]. The main challenge will be the dynamic nature of VNFs that will be more prone to configuration errors [228]. A study on security implications of virtualization [229] reveals several challenges that must not be present in 5G. For instance, “the loss of uniqueness of devices and data” in virtualized environments will be specifically more challenging in 5G due to the integration of diverse things (IoT) or networks of things. Furthermore, tracking malicious devices and developing trust system between hypervisors, VMs or management modules will be further complicated [230].

Challenges for MVNOs: Mobile Virtual Network Operators (MVNOs) [231]–[233] leveraging NFV to lease and operate the network face the security challenges mainly due to limitations in the current NFV systems as described in [233]. The network infrastructure owners will expose APIs on their network hardware platforms to third parties to deploy their services on the same hardware [220]. Since 5G will have very diverse actors, let alone in services their can be several service providers. For example, multiple MVNOs and Communication Service Providers (CSPs) or other service providers will share the same network infrastructure. All of these service providers will have different security and privacy policies. The synchronization or otherwise mismatching security and privacy policies will be another pressing challenge in such shared environments [234]. Furthermore, the dynamic nature of VMs will make it difficult for MVNOs to track the usage of network resource and binding it to specific users and devices for the purpose of billing.

2) Security Solutions for NFV: Virtualization can highly increase the user, service and network security. A basic mechanism is to use slicing to separate traffic of different services (Fig. 8) or network segments based on security priorities [235]. Second, distributed VNFs can be deployed that increases the scalability and availability of the system and resolve DoS and DDoS attacks [236]. By adding more intelligence regarding self-protection, these VNFs can substantially improve the security of 5G networks [236]. In [237], the authors leverage...
NFV besides SDN to improve network security. Slicing the network for different services is considered as a strength of 5G compared to the previous generation networks. Leveraging NFV, 5G thus provides a unique opportunity to provide security as a service in cellular networks [237]. General best practices for improving the security of NFV infrastructures is presented in [238] and below we present the existing solutions to some of the common security threats in NFV.

Solutions for virtual systems: Even though virtual systems can be difficult to monitor compared to physical systems, virtual systems have their own benefits in terms of security. For instance, a virtual system can be easily migrated or replicated to minimize the effect of security attacks. To cope with the challenge of inconsistency, various solutions are proposed that ensure consistent security policies in virtualized environments, as discussed in [228]. A policy manager that enforces security policy in dynamic VNF environments is proposed in [239]. The proposed software component for the NFV framework provides an easy and effective way for users to specify their requirements and enforce it within the network without dealing with complex security configurations.

Solutions for hypervisors: Due to the central role of the hypervisor, the basic choice for strengthening its security is least or restricted exposure to VMs or other systems. Acting on these grounds, the modified version of Xen called Xoar [240] increases its security. The platform breaks the control VM into multiple single-purpose VMs to make risk exposure explicit, maintain least access privileges, and thus increase the security of the overall system. The OpenVirtexX [241] presents a new interesting network virtualization platform that provides customers with virtual SDNs. Working similar to the OpenFlow [28] SDN implementation, the hypervisor acts like the controller in OpenFlow in which the slices of the users have their own control and data planes. The interesting fact in this solution is the capability of the SDN controller that can oversee the activities of the hypervisor and can easily track security vulnerabilities.

Solutions for dynamicity: The dynamic nature of VNFs and virtualized resources can be used as a strength in terms of security. For instance, the flexibility of NFV allows isolating compromised network elements, or even whole network segments through defining security zones and using traffic steering [225]. In the [225], a security oriented management and orchestration framework has been proposed that dynamically manages the entire life cycle of security functions, as well as secures NFV-based infrastructures and platforms. However, the elasticity of NFV has rarely been used to increase network security.

Solutions for MVNOs: The security of MVNOs is highly dependent on the security of the systems MVNOs use. However, that will not be enough alone. For instance, interoperation of various components in a mix-and-match system will require a security and trust monitoring framework. A security and trust framework for VNFs in SDN-based mobile networks is proposed in [234]. The proposed framework applies adaptive trust evaluation and management technologies and sustainable trusted computing technologies to ensure computing platform trust and achieve software-defined network security [234]. Similarly, a conceptual NFV-based security management and orchestration framework is presented in [225]. The proposed framework protects the resources or other VNFs from security threats originating from the Internet or other VNFs by validating their security characteristics. However, the limited deployment of such operators in real-time makes it difficult to fully understand the full range of possible security threats. Therefore, there is little research that concentrates on the security solutions for MVNOs, apart from increasing the security of VNFs or NFV.

D. Security in Clouds

The role of cloud computing in 5G mobile network has become a cardinal focus for both the industry and research organizations working on 5G and related technologies. Cloud computing provides computing resources and services to both small and large enterprises on an on-demand fashion, hence optimizing on available resources and allowing for higher abstraction of the underlying mechanisms on the side of customers. Currently, a good number of social media sites like Facebook, Netflix, YouTube, and Twitter have already embraced the cloud concept and are expanding on its possibilities.

One of the major use case of cloud computing in 5G is the massive machine-type communications [242]. This use case is characterized by massive number of connected devices, forming a network of billions of sensors and actuators, supporting a huge number of low-cost and energy-constrained devices. This use case underscores the need for cloud computing in two folds. First is on the need for real-time data sharing among devices, and second is on the need for low-storage devices to be provided with virtual storage capacity on the cloud. MEC is another key use case of cloud computing that holds great significance for 5G mobile networks and beyond. MEC extends IT and cloud computing capabilities within the RAN at the edge of mobile networks. This provides developers and content providers with direct access to real-time radio access information, hence promoting ultra-low latency, higher bandwidth and enhancing overall user experience [243], [244]. However, these technologies and use cases also make the 5G cloud computing confront major security challenges, as described below.

1) Security Challenges in Cloud Computing: The security challenges in cloud computing are highly connected with the technologies involved in cloud computing such as networking, virtualization, and the services deployed in the clouds [245], [246]. Since diverse types of services reside in the cloud and the cloud resources are highly distributed, heterogeneous and virtualized, traditional security mechanisms are no longer enough for the security of clouds [247]. There are diverse technologies involved in cloud computing, and thus, diverse security challenges exist that require specific security solutions tailored for those challenges as highlighted in [248]. Below we describe the most concerning security challenges related to cloud computing that will affect the security of 5G.

The role of cloud computing in 5G mobile network has become a cardinal focus for both the industry and research organizations working on 5G and related technologies. Cloud computing provides computing resources and services to both small and large enterprises on an on-demand fashion, hence optimizing on available resources and allowing for higher abstraction of the underlying mechanisms on the side of customers. Currently, a good number of social media sites like Facebook, Netflix, YouTube, and Twitter have already embraced the cloud concept and are expanding on its possibilities.

One of the major use case of cloud computing in 5G is the massive machine-type communications [242]. This use case is characterized by massive number of connected devices, forming a network of billions of sensors and actuators, supporting a huge number of low-cost and energy-constrained devices. This use case underscores the need for cloud computing in two folds. First is on the need for real-time data sharing among devices, and second is on the need for low-storage devices to be provided with virtual storage capacity on the cloud. MEC is another key use case of cloud computing that holds great significance for 5G mobile networks and beyond. MEC extends IT and cloud computing capabilities within the RAN at the edge of mobile networks. This provides developers and content providers with direct access to real-time radio access information, hence promoting ultra-low latency, higher bandwidth and enhancing overall user experience [243], [244]. However, these technologies and use cases also make the 5G cloud computing confront major security challenges, as described below.

1) Security Challenges in Cloud Computing: The security challenges in cloud computing are highly connected with the technologies involved in cloud computing such as networking, virtualization, and the services deployed in the clouds [245], [246]. Since diverse types of services reside in the cloud and the cloud resources are highly distributed, heterogeneous and virtualized, traditional security mechanisms are no longer enough for the security of clouds [247]. There are diverse technologies involved in cloud computing, and thus, diverse security challenges exist that require specific security solutions tailored for those challenges as highlighted in [248]. Below we describe the most concerning security challenges related to cloud computing that will affect the security of 5G.
Virtualization threats: Since virtualization would play a vital role in the design and implementation of cloud-based networks for 5G realization, the threat landscape on virtualized platforms is equally an issue of concern for service providers as well as users and application developers. Security threats in this category may range from DoS attack to VM manipulation as shown in Table VI. Though there can be logical isolation of resources in the cloud platforms, malicious entities can exploit data leaks and cross-VM manipulation [249]. Moreover, the threats in this landscape can be targeted towards the virtualized infrastructures and platforms, edge computing systems as well as the hypervisors.

Cloud-Based Cyber-Physical System security threats: Cloud-based Cyber-Physical Systems (CCPS) achieve virtualization of network components using cyber-physical clouds, such components include various sensors and actuators. Such virtualized components are provisioned as conventional cloud resources used to provide cloud services [250]. Typical security attack in this landscape include Hyper Text Transfer Protocol (HTTP) and Extensible Markup Language (XML) DoS (HX-DoS) attack. HTTP and XML Denial of Service (HX-DoS) combines HTTP and XML messages and flood them at rates preconceived to inundate the capacity of the cloud CPS infrastructures. Such attack could be launched on cloud infrastructures, Software or platforms, i.e., Infrastructure as a Service (IaaS), Service as a Service (SaaS), and Platform as a Service PaaS [251], [252]. Other threats in this landscape include is Slowly-increasing Polymorphic DDoS Attack Strategy (SIPDAS) identified in [252]. SIPDAS is a kind of DoS attack that evades pattern detection algorithms by modifying its behavior dynamically.

Cloud intrusion: Cloud intrusion is a threat on the integrity of cloud resources provided on the network. This threat also affects the availability and confidentiality of cloud resources and services. The more sophisticated the intruder the more the severity of such intrusion. In addition, threat severity is also influenced by the nature of the loopholes and weak-links on the cloud environments. Such intrusion can affect different cloud service models as shown in Table VI.

Insider attacks: This is a socio-technical form of attack that poses a major threat to the overall cloud platform [253]. Here, the term insiders implies the service provider’s staff with access to the physical servers on which user data is stored. When such insiders set to misuse or mismanage user data and information, this could constitute a major threat to the whole idea of the cloud. A classic example in recent time is the Facebook Cambridge Analytica data scandal where personal information of up to 87 million Facebook users were allegedly used for political motives [254].

2) Security Solutions for Clouds: The security solutions to security threats in clouds are also multi-faceted. For instance virtualization security and the security of VNFs residing in clouds has direct implications on the security of the clouds [248]. Therefore, below the security solutions to each of the mentioned challenges are described.

Virtualization security: Security threats related to virtualization are mostly on the part of the virtual machines. Virtual machine security has been a major project interest for companies like IBM and XEN. In 2011, IBM designed The Trusted Virtual Data Center (TVDc) [255], a technology that is designed to address the need for isolation and integrity in a bid to address the security vulnerabilities on virtualized systems. Here is important to mention that in securing virtualized platforms, it is imperative to extend all security measures implemented on the operating system of the physical machines to the operating system of the VMs, only then can the effectiveness of the security strategy be guaranteed [256]. Prior to the development of the TVDc, XEN developed the sHype system which is a secure hypervisor architecture for isolating VM systems with flexible security of mandatory access control [257]. Such access control systems would control the flow of information and communication between multiple VMs across different machines. Another possible mitigation technique for networked DoS attack in virtualized systems is the implementation of firewall proxies, this will require an ACK to be received on the client side before an attacker’s request can be forwarded.

Cloud-Based Cyber-Physical system: Cloud-Based CPS (CCPS) attacks in the form of HX-DoS, DDoS or SIPDAS can generally be mitigated by frequently checking the consumption of computational resources as well as the intensity of incoming requests. A detection of some anomalies will then trigger a more advanced control measure. For mitigating CCPS attacks in the form of HX-DoS, authors in [251] proposed the use of the so called ENDER (pre-decision, advaNce Decision, LEarNing) to identify and distinguish a legitimate CPS from an illegitimate one. To accomplish this, the ENDER system uses two decision theory methods to detect attack traffic and then using a similar technique as in traditional Intrusion Detection Systems (IDSs), is then able to identify and mark an attack message. Upon detecting such attack message, a small 1bit mark is then added to the message, and using a Reconstruct and Drop (RAD) algorithms, the system is able to remove such message before any further harm is caused to the system.

Cloud intrusion: Mitigating against cloud intrusion is mainly achieved by building IDSs to work in conjunction with other control mechanisms in the cloud computing environments. Such IDS systems are designed to constantly monitor the activities in the cloud environment and will identify any form of malicious activities and policy violations [253]. Various IDSs have been proposed to meet the varying needs of the different cloud service models, namely IaaS, SaaS, PaaS. There are also IDS systems designed for network hosts and hypervisors. The author in [258] provides a comprehensive discussion on the nature IDS for different cloud service models, and further analyzes the evolving trends towards advanced IDSs to detect intrusion in present and future cloud environments. Among such IDS are the hypervisor-based intrusion detection system, traditional Host-IDS (HIDS) and network-IDS (NIDS). Where HIDS are preferably deployed on the front and back ends of the cloud platform, while NIDS and hypervisor-based intrusion detection systems are better placed on the back-end where the CSPs operate.

A more passive approach to addressing cloud intrusion is by building intrusion-tolerant cloud applications. Such
applications are designed to identify and ignore intrusive requests from adversaries. However, this approach is more or less of temporal nature depending on the vulnerability of the cloud facility. Over time adversaries tend to circumvent the mechanisms of such controls, hence the service providers will need to keep pace with evolving intrusion patterns to keep this approach effective.

**Insider attack:** Mitigating insider attack vulnerabilities in 5G cloud-based networks is as much of a social challenge as it is a technical challenge. While the CPS works on providing users with secure interfaces and APIs, the possibility of abuse and nefarious use of the data in the cloud by authorized service provider staff equally calls for concern. This challenge is further confounded by other natural happenstances like leakages and data losses. On this aspect, the need for more robust backup facilities and multiplicity of backups across different locations and platforms comes as workable mitigating strategy. For other intentional insider attacks, implementing proper auditing and routine background checks coupled with digital time stamping and signatures on cloud data, could help to mitigate the possibility of abuse and nefarious use of cloud data by authorized insiders [253].

VII. PRIVACY CHALLENGES AND SOLUTIONS

Privacy remains a core issue in mobile and communication systems since the invention of these technologies. In recent decades, the evolution of smart devices such as mobile phones and other handheld devices have enabled more diverse services to consumers. With the advancement in mobile technologies, the services are also becoming smarter, ubiquitous and more pervasive. The current evolution towards 5G networks is quite promising not only from consumers perspective but also from the involved stakeholders point of view. This will help various stakeholders to enhance their business models for bigger and better revenues. However, the complete and successful deployment of future 5G systems need addressing the challenges of privacy. Hence in this section, the changing paradigm of privacy from 1G to 4G, and in-depth study of privacy in 5G is discussed.

A. Privacy: 1G - 3G

Traditionally, phone users have been mainly concerned about the privacy of their calls, however, 1G did not have suitable encryption mechanisms to ensure privacy. Hence, private communications can be listened to and intercepted by the adversaries even from a far away distance. 2G was then proposed as an advancement to address some of the challenges identified in 1G. In order to preserve the user’s privacy in 2G systems, there have been two major mechanisms proposed. First is radio path encryption which is mainly designed to protect the voice and data circuits from invalid interceptions. Second is the Temporary Mobile Subscriber Identity (TMSI) which is proposed to avoid attacks related to user’s identity and ensure anonymity [274]. Notwithstanding, both techniques have their shortcomings also. For instance, regarding the encryption in GSM, A5/1 and A5/2 were considered to be insecure and exposed to various attacks such as eavesdropping. The TMSI could not completely assure the anonymity of the users because IMSI catchers can still reveal the real identity of subscribers. 2G systems did not have mutual authentication approaches between mobile phone subscribers and corresponding networks. This is because GSM could only be able to authenticate the subscriber with the respective network and not vice versa. This means that it can only provide authentication and confidentiality characteristics and couldn’t guarantee the complete authorization mechanism. This also implies that it does not provide complete non-repudiation [40], [274], [275].

With further advancements in mobile communication systems, 3GPP proposed various specifications for 3G networks that could address these identified privacy short-comings. For example, mutual authentication mechanism is proposed to provide more robust privacy features. The 3GPP also defined various subscriber privacy requirements for 3G UMTs, which include; user identity confidentiality, user location confidentiality and user traceability. The user identity confidentiality refers to the globally unique IMSI and ensure that user communication over the radio access link should not be eavesdropped by any means [276]. However, it is observed that they are exposed to various attacks which were mainly targeting the identity and confidentiality of the subscribers such as IMSI paging attacks and AKA error message attacks [275], [277].

B. Privacy: 4G

4G networks are currently the most widely used mobile networks and have significant amount of enhancement in terms of data rates compared to the previous generations. As such, it can facilitate applications that require voice and video technologies. With the increased speed or user bandwidth, it is exposed to more privacy risks compared to previous generations. Keeping the context of privacy in mind, there are quite a number of vulnerabilities in 4G networks. Two of the most critical vulnerabilities being man-in-middle attacks, and eavesdropping attacks. This happens when adversaries set up fake base stations and act as real network base stations [16]. Several research works have proffered various possible solutions to mitigate man-in-the-middle attacks, popular among these is the use of cryptographic authentication protocols.

Another prominent challenge in such networks is the ability to preserve the User Equipment (UE) privacy. The exposure of the IMSI generates issues to user privacy, and this information can cause several active and passive attacks. Also the TMSI can be attacked by malicious adversaries, i.e., rainbow and brute force attacks. Thus, one of the solutions to IMSI attacks is encrypting the IMSI [278], since revealing it can lead to several passive and active attacks targeted at specific IMSIs and their respective users. Furthermore, TMSI generated by the authentication center has been found to be prone to rainbow and brute force attacks, hence an attacker who gets hold of the TMSI can be able to perform social engineering in tracing the TMSI to the corresponding IMSI of a UE [15].
### C. Privacy: 5G

The advent of new architectures, technologies and services in 5G networks will eventually generate higher privacy risks for users and other stakeholders as compared to previous generations. In addition, the integration of technologies such as SDN, NFV, cloud/edge computing with the 5G eco-system will expose the networks to even more serious privacy challenges. Table VI highlights the potential privacy concerns for various mobile generations. Three parameters are used to show the importance of various privacy features in different generations, i.e., Low (L), Medium (M), and High (H). ‘NA’ is used in the table to present if any privacy challenge is “Not Applicable”.

5G network will be a shared eco-system comprising various actors, and will provide different sets of digital services for these different actors. For example, various heterogeneous operators and service providers involved in the process may access personal data of the consumer with or without their consent. This is a critical issue because consumers/users will be mostly unaware of the entities that are gathering, processing and storing their data. The users will, in most cases, not know how their data is used. Hence, there are concerns that the personal information can be exploited by various involved stakeholders. Therefore, end-to-end data confidentiality mechanisms are required to ensure data protection [279].

5G is considered as a driving force for IoT based applications and expected to play a vital role in the successful deployment of low latency services [280]–[282]. 5G wireless networks will enable IoT communication which is crucial for future smart and digital services, such as health-care and industrial applications [283]. Hence, there will be increased privacy challenges from both perspectives. It is important that data must not be accessed by unauthorized entities and only legitimate stakeholders should have access to it. Data attacks can be made in two ways: one, attacks on data while in transit, and second, illegal access to data in storage systems. During transit, the potential attacks can be message modification, eavesdropping and man-in-the-middle attacks [284]. Strong cryptographic techniques are required to protect data in such cases. For storage systems, privacy-by-design approach, i.e., protecting privacy during the design and manufacturing phases [285], and strong authentication and authorization mechanisms are needed to protect the data from unauthorized access [286]. There are also privacy concerns regarding user location and identity [287], that require various techniques such as cryptography and obfuscation [288].

As mentioned, 5G will be a shared environment of multiple stakeholders such as users/subscriber, network operators, network infrastructure providers, service provider, MVNOs and Communication Service Providers (CSPs). One of the key challenges in 5G is to build and maintain trust among...
them. These multiple stakeholders in 5G systems will enable new business models [287], [289]. This is because each involved entity might have different priorities for preserving subscriber’s privacy considering their own business interests. Thus, it may lead to the situations where all involved entities do not fully cooperate to guarantee the privacy of the consumers [234]. Thus, in order to maintain subscribers trust on various stakeholders and also within all involved parties, it is highly important to have such trust models which can fulfill the privacy needs of the subscriber and protect the interest of each entity in the whole 5G system.

Managing identities in 5G networks would be a key challenge because 5G will enable rapid communication among huge number of users and devices having their own identities. Leakage in identity privacy may lead to various consequences, for example, it can reveal users information about daily life routine. One of the common attacks on identity is IMSI catching, where IMSI of mobile user is fetched [15]. Due to unavailability of TMSI, UE is unable to use IMSI as its identifier. A quicker method of fetching IMSI is by setting a fake base station which might be considered as genuine/real because of its signal strength. This base station can ask identity information from mobile users, and in response the UEs give their IMSI because they considered it as preferred base station. IMSI attacks can be both active and passive. A potential solution to overcome the IMSI catching attack is by allocating random TMSI to various mobile users at regular interval in a given network. IMSI is only used when any fault occurs or TMSI is not available [290].

Location based services are getting vast attention for various purposes such as finding nearest desired places, friends and relatives, and wearable devices for tracking and monitoring purposes [287]. However, location based services will raise several privacy concerns. For example, in certain mobile applications, location based service providers ask for personal information of users which might not be needed for that particular mobile application. By using location information, one can easily track the life schedule of any particular individual and can cause harm. In order to tackle such location based privacy issue, several approaches are already utilized such as; encryption based approaches, anonymity based techniques, obfuscation based approaches, location cloaking algorithms, privacy policies, and regulations among others [297], [299], [300], [316], [317].

Recent developments in wireless and communication technologies allow global access and connectivity of user data. Hence, if a user is using any online application in one country, his/her data can be stored or processed in any other country [302], [303]. This leads to huge user privacy concerns because each country has its own privacy regulations and polices for data protection. For example, regulation agencies may give priority to some data which are crucial in that country but may not be important or acceptable in the other country. Also various legislation might have different values for personal data privacy. With the addition of public clouds and online service providers, the storage of user information cannot be restricted to only a particular country. Therefore, it is vital to define regulations that can protect consumer data through cross border territories. European Union has recently announced the updated General Data Protection Regulation that highlights European Union laws on data protection and to preserve privacy of all users within the European Union and the European economic area [318], [319]. It also defines rules and regulations on international transfers of personal information.

**VIII. NEW DIMENSIONS IN SECURITY OF FUTURE NETWORKS: THE XG**

The dramatic increase in types and number of IoT devices, the concepts of networked smart societies or a radically new gadget-free world as presented in [320] necessitate a paradigm shift in security as well. The development in computing paradigms, such as quantum computing, will compel us to design new robust security architectures leveraging powerful computing to strengthen network security. Breaking cryptographic algorithms with super-fast or organized shared computing might become a reality. The basic idea is that the diversity and growth of computing and communication technologies must be matched with novel security systems.

In this section, we present a broader view of a future generation of wireless network, termed as the XG (Fig. 9) and discuss strengthening its security through novel technological concepts that are currently hot research topics. For example,
super-fast security mechanisms at various network nodes and locations such as network edge or fog nodes, security automation and self healing that would require the development in AI, and using the concepts of blockchain for securely providing security-critical services.

A. The Concept of XG

For presenting future directions in network security, we define XG as a secure and autonomous network of numerous smart objects in smart environments, all connected to make a smart city, with no observable latency, no restrictions on bandwidth, and available everywhere. Smart environments [321] have been the focus of research for many decades. The revolution of IoT [280] provides the basis for the computation and communication paradigms required for smart environments. By integrating smart environments, smart cities are at the forefront of future Internet, whereas IoT is the basic building block of a smart city [322]. IoT provides the foundational infrastructure for the smart world, but the foundation of IoT itself lies in smart networks [323]. Smart networks described in [323], are network environments that use software that can configure heterogeneous networks automatically to meet the user and service needs through network infrastructure abstractions [323]. Hence, the real benefits of IoT, aiming the smart cities, can be realized when the communication systems are also smart enough to intelligently and autonomously deliver the necessary information generated and needed by IoT [324]. This needs i) intelligent communication systems that are responsive to the needs of IoT in real-time, and ii) provide coverage, in ideal conditions, everywhere. XG, in our view, will be the future network having these capabilities.

B. A Glimpse of Security Challenges in XG

5G networks will integrate low power and low data rate IoT devices [325]. The number of the IOT devices are projected to reach 80 billion within the next decade. The massive number of IoT devices in massive numbers of smart spaces will introduce new security challenges, as discussed in [326]. Currently, the important communication protocols for IoT include the IEEE 802.15.4 standard [327], [328], IPv6 over Low-power Wireless Personal Area Networks (6LoWPAN) standard [329], [330], and Constrained Application Protocol (CoAP) [331], [332]. These protocols rely on cryptographic algorithms such as the Elliptic Curve Cryptosystems (ECCs) as the basis to secure the communication. With the advent of quantum computing and increasing capacities of the next generation computing devices, these protocols will no longer be secure, as thoroughly discussed in [167]. Similarly, proper security keying models for IoT are still lacking [333] and routing protocols have vulnerabilities [334]–[336].

Furthermore, in a huge number of connected devices, recognizing malicious ones will be more challenging. Devices with limited computation, battery and storage capacities, could potentially act as malicious bots in a network [326]. Mobile botnets, compromised limited capability devices, are already posing serious security challenges due to connectivity to the Internet through cellular networks [18]. On one hand, devices not needing connectivity can be interconnected or be made available online through embedding a simple software. On the other hand, a large-scale analysis of firmware images [337] reveal that embedded systems are ripe with security vulnerabilities. Such devices will also endanger public privacy [338].

However, there are potential security approaches that can be used to ensure security in such environments. For instance, service or application-based security, using strong isolation leveraging network slicing, software-based security functions that are portable, and leveraging AI for proactive security measures in different network perimeters. In the following subsections, we describe how new technological concepts such as security softwarization and virtualization, AI and blockchain technologies can be used to secure future (XG) networks.
C. Security Softwarization and Virtualization

Taking the network functions from software rather than hardware is one of the key trends in future wireless networks. Virtualization creates another level of abstraction by providing the framework to deploy network functions as software components, as promoted with the concepts of VNFs. The two concepts interwoven together, will not only provide the benefits of cost and performance efficiency, but will increase end-to-end reliability of future networks [339]–[341]. Security functions implemented in software being able to be deployed in any network perimeter based on necessity, introduces many new opportunities in strengthening network security.

For example, traffic monitoring can be used to detect and prevent intrusions. Intrusion detection or prevention systems (ID/P-S) observe the traffic, find vulnerabilities or threats, and use countermeasures to secure the network. Due to independent control planes in traditional networks, ID/P-S technologies in each network domain are usually independently configured to deal with security challenges. Hence, it is hard to update the current systems with newly identified types of attacks and threat vectors or deploy consistent network-wide security policies. Software-based ID/P-S systems using virtualization technologies will enable run-time updates, as well as run-time security policy or system deployment in any network segment, solely based on necessity.

Similarly, conventional access control mechanisms use firewalls on network boundaries to examine incoming or outgoing packets to prevent attacks and unauthorized access [342], [343]. First, the insiders are considered as trusted partners who might be previously compromised and can launch attacks or circumvent the security mechanisms. Second, changes in network policies and traffic conditions require complex configuration of firewalls. Technologies like SDN that enable programmability, centralize the network control, and equip the network management plane with global visibility of the network state can mitigate the risks involved in configurations and easily monitor the overall network traffic (ingress or egress ports). SDN is also termed as security-defined networking [344], since it enables security software (SDN applications) to be easily deployed, updated or removed from a network.

A number of firewall applications such as FLOWGUARD [345] and OpenFlow firewall software [346] can be considered as the basic step towards softwarized security for softwarized and virtualized networks. A detailed description of software-based virtualized security functions such as dynamic, reliable and scalable firewaling is described in [343]. A virtualized Deep Packet Inspection (DPI) system for intrusion detection and prevention in a softwarized network infrastructure is described in [347] and softwarized network monitoring in [348]. Even though virtualization and softwarization have their own limitations in terms of speed or latency, the flexibility and agility of such systems make these inevitable in future networks.

D. AI-Based Security

Analysis of huge amounts of data or monitoring the network traffic-in-transit for security require a paradigm of proactive, self-aware and self-adaptive intelligent systems. Such systems will employ novel algorithms and technologies of AI, and as a result, cyber-security may become one of the best application areas for AI. Conventionally, a security attack or a security lapse happens and then patching starts once the attack or the lapse is recognized. This needs to change since critical infrastructures such as electricity smart grids, transportation and health-care systems are formally accepting connectivity through communication infrastructures and reactive security measures will not suffice. The change must be towards proactive security measures due to the criticality of security of these infrastructures. Proactive security measures would require continuous intelligence gathering, and using that intelligence to mitigate the possibility of security risks or lapses. AI, with its promising algorithms and full solutions having gained appreciation in other fields, must also be used in the realm of network security.

Firewalls using DPI can be considered as an instantiation of AI in security, yet clear and specific development of AI-based security approaches and solutions need to be introduced [349]. A more enticing example is the Completely Automated Public Turing test to tell computers and humans apart that brings the two domains (security and AI) at an intersection [350]. This is widely used in commercial contexts where a human is differentiated from a bot through recognizing distorted characters or a sequence of characters [351]. As the pattern recognition software move towards more sophisticated pattern recognition, the drive to maintain security will also drive more sophisticated use of AI. Consequently, the advances in decision procedures, model checking, and recently, Boolean satisfiability have grasped the attention of cyber-security researchers [350].

A cloud monitoring model based on cooperative intelligent agents is proposed in [352]. The agents, located in different units of the cloud, learn about the environment using specific monitoring methods, communicate with each other, and make decisions based on AI. These agents detect abnormalities, misfunctions and security threats within the systems. Using the tools of AI such as neural networks and decision trees, the authors in [353] demonstrate revealing hidden communication by mobile malware or malicious software. Similarly, the authors in [354] use fuzzy logic, neural networks and trend analysis in IDSs. Moreover, AI algorithms and solutions such as Bayesian AI methods [355] dealing with uncertainty in terms of finding malicious activity, as in DPI, opens new horizons for strengthening security of networks of massive number of connected devices (e.g., IoT), applications, and diversified services. For example, in [356] the authors demonstrate using AI to secure vehicular communication and effectively counter DoS attacks.

Due to the diversity of services and devices in next generation networks, autonomous decision making in terms of security policy verification, policy conversion to configurations and subsequently deployment will require leveraging AI for the purpose. As described in [357], AI has the potential to help operators in situations where there are no prior data or experience, or the data is too complicated to understand with traditional approaches. With the conglomeration of diverse IoT devices, UAVs, V2X, wearables and smart home
apart from other benefits of AI in the automation of various processes in the education sector, it will help in the collaboration of teachers and students, leading to an overall improvement in the learning process.


e. Security Automation

Machine execution of complex functions or automation can be used in i) information acquisition, ii) information analysis, iii) decision and action selection and iv) action implementation for accuracy and reliability [354]. Due to the complexity of next generation networks in terms of heterogeneity in networks, devices, applications and services, network functions must be automated [358], [359]. Automation leveraging AI is already happening in many areas related to communication networks, for instance, in autonomous vehicles [356], [361]. As the networks are getting complex making the network management much more complicated, security parameters adjustment, intrusion detection and prevention, and security policy enforcement such as access control and authorization need to be automated. Human-machine interaction is already a major reason for the network downtime [362], and security lapses due to human errors are no exception [363].

Manual configurations of network security technologies such as firewalls and IPSec technologies on extended sets of devices are prone to configuration errors, intra- and inter-policy conflicts resulting in serious security vulnerabilities and threats [364]. A study on manual configurations of firewalls in [365] reveals that major security breaches occur due to the complexity of manual configurations. Therefore, security automation is inevitable and with the emergence of complex IoT systems and networks, human intervention for each security policy setup, system updates or even lapses will be highly challenging.

The evolving network paradigm paves the way for security automation, albeit the limited efforts for security automation. For instance, SDN automates many processes and procedures including physical and virtual network management, reconfiguration, and introduces the possibility of deploying new automated services leveraging network abstractions [366]. Similarly, using various dimensions of AI to perceive or understand and eliminate a security threat well before it causes any damage in an automated fashion is already realizable. Examples of such advancements include self-healing solutions for future mobile networks. A self-healing solution resolves outages or carry out network tasks autonomously without human intervention [367]. However, security automation will minimize the involvement of user perspectives which opens the debate of social acceptance as discussed in [368].

F. Blockchain: A New Security Perspective

Blockchain is considered as the next big revolution for future Internet and communication technologies. Blockchain, as a decentralized and distributed technology, has immense potential in various useful and critical applications such as banking, health-care, supply chain management, and IoT among others [369], [370]. The initial focus area of blockchain technology was related to bitcoin, crypto-currencies and financial/banking purposes, however, its potential applicability in various areas of daily life made it necessary to draw the focus of blockchain research to other identified applications as well. The utilization of the blockchain for current network and communication technologies are opening doors for more secure and safer means of communications. Blockchain allows various stakeholders/entities to securely share and access the critical data. A distributed ledger containing the required data is shared with all the involved stakeholders within the process. Thus, blockchain ensure more security features in the overall communication system. The emergence of AI in future communication systems in integration with the blockchain technology will be key enabler of many critical applications. A few use cases of blockchain with keeping security as the major highlight is discussed in this subsection.

Health-care is among one of the prominent application domains where blockchain will have a huge impact with its decentralized nature of computation. For example, secure sharing of health-care information is one of the crucial requirements in order to provide intelligent and better quality of health-care services [371], [372]. Health-care data is very sensitive and must be controlled by the respective patients [373]. Blockchain can provide a secure data accessibility mechanism where patient’s health data can be accessed by various relevant stakeholders in the system [374]. Blockchain can also be a vital tool for providing various means to secure storage for the patient’s records. In addition, it can also securely maintain patient’s medical history required for proper treatment or medication. Future blockchain based health-care systems are expected to improve particularly in terms of intelligence and context awareness, secure data access and sharing, better quality and management, and security and privacy. The low latency, massive data communication and intelligent services provided by the future XG technologies will be one of the key requirements in future blockchain based health-care systems.

Another useful and popular application of blockchain is in the area of IoT [375]–[377]. There will be numerous sensors/nodes/devices connected with IoT networks that require proper management with less complexity and resource utilization. Along with that, security and privacy remain the top most concerns in IoT because of its massive scale in diverse applications [378]. Traditionally, the security frameworks for IoT networks consume higher resources in terms of energy and processing power. Also, most of the related work suggest that many of the security frameworks are based on centralized approaches. Centralized security systems, however, have challenges of scalability and single point of failures. Furthermore, user privacy has many loop holes in IoT based applications with existing privacy preserving methods. Thus,
blockchain based approaches in such cases can offer decentralized means of security and privacy mechanisms for IoT applications [379], [380]. Due to decentralized and distributed nature of the blockchain technology, it can be well applied to IoT, for example to manage the configuration of IoT devices, store and share the critical data, and enhance the security and privacy [376]. The future XG technologies will provide solid platform to support blockchain for building security and privacy mechanism for low power and resource constrained IoT devices.

G. UAV Base Station for Security

The agility and resilience requirements of future services motivate the use of UAVs equipped with Base Stations (UAV-BS) for future wireless networks. The UAV-BS provides quick deployment opportunity and temporary solutions to any potentially unexpected situation or disaster management. However, strong security measures are required for UAV-BS assisted communications. There are already some existing security threats that can be used against UAV-BS too. For instance, an attacker may manipulate the transferable data by delaying the control commands. The manipulation can be carried out through corrupting, replaying and blocking the data during transmission [381]. An adversary may also passively eavesdrop on critical data that can be used for further attacks [382].

Due to stringent power and weight requirements, the BS carried by the UAVs may not be able to support complex cryptographic algorithms like a terrestrial base stations. This can cause the UAV-BS more prone to security vulnerabilities. Moreover, the UAVs will have higher possibility to fall in the hands of an attacker due to its operating location. Therefore, UAVs must support mechanisms to safeguard from physical tampering in the events it falls to the hands of an attacker. In such cases, UAVs should use basic insider-attack protection mechanisms in which physical access does not guarantee access to the internal functions of a system. Due to lack of specific lightweight cryptographic protocols for UAV-BS, the backhaul communication can also be targeted by attackers. However, the research on UAV-BS security is gaining momentum and various solutions are proposed to tackle the mentioned challenges, as discussed in [383]–[386].

Albeit the inherent security challenges, UAV-BS can also assist in secure information flow, specifically in disaster hit areas, or network segments under cyber-threats. Effective communications is the key to public safety in disaster or emergency situations [387]. A use-case presented in [385] is using drone BS to augment the operation of public safety networks in critical situations. A fleet or swarm of drone used together to provide services in various scenarios, such as described in [384], can be used to securely operate and provide services. A security attack on a whole fleet will be much more difficult due to shared intelligence and task execution. For instance, resource-exhaustion attacks will not easily affect the working of the drones when multiple drones operate. Hence, using UAV-BS in situations where the available base station is either under a security attack (e.g., DoS) or congested, that limits its availability to legitimate users, seems an interesting approach for future dense networks or smart cities with massive number of IoT devices, smart vehicles or normal mobile users.

H. Privacy

As highlighted above, the XG technology will be key enabler of massive and critical applications in various domains such as smart health-care, industries automation, transportation/ Vehicle-to-Vehicle (V2V) and massive IoT. Along with security, the privacy landscape will also vary according to different requirements of applications. Massive amount of data is expected to be generated by these future applications, because xG will provide the underlying network for ubiquitous and pervasive digital services. Therefore, there will be a wider scope for adversaries to attack the consumer’s privacy, as described below with examples.

Future digital services will merely be not only acquired through gadgets but gadget free users can also attain the desired services anytime and anywhere. The concept of gadget-free smart environment (also called as the Naked World) is introduced in [320], [388], [389], where gadget-free users can get digital services from the nearby intelligent environment and without using hand-held gadgets. Digital services and computation capabilities will be embedded in the smart environment. This kind of open and shared environment will be highly exposed to users’ privacy risks, i.e., risks in leakage of data, location and identity [390]. Also all other involved stakeholders such as infrastructure provider, network operator and service providers need to ensure that users’ personal information should not leak during various phases such as user interaction with environment, identification process and data storage among others. In addition to technological solutions to protect the privacy, there will also be need of strong regulations and polices for such smart environments [391].

Smart spaces will also play a key role in implementing the vision of future smart cities and XG is going to play vital role by providing faster means of communication. Enabling technologies related to smart cities along with the corresponding big data challenges will make privacy one of the topmost concern for users in the coming future [392], [393]. Since smart cities will require high level of interconnectivity where multiple stakeholders can provide various services, the pervasive nature of services and their delivery will add more woes for privacy. For examples authors in [394] highlighted five types of privacy requirements needed for smart cities’ applications and technologies: i) privacy of location, ii) privacy of state of body and mind, iii) privacy of social life, iv) privacy of behavior and action, and v) privacy of media. Hence, there are more dimensions for adversaries to compromise the privacy of the whole eco-system. However, these intelligent or smart environments will also be context-aware and capable to take AI-based decisions. Therefore, AI-based privacy mechanisms should be adapted to ensure intelligent and appropriate privacy solutions by sensing the specific context and needs of applications.
IX. Discussion and Future Research Directions

The network security has evolved as the networking technologies matured. There are also new disruptions due to novel technological concepts such as SDN that furthered new paradigms in security, such as Software Defined Security. However, there are still areas in which further research is highly important to ensure robust security in 5G and beyond. Few such important areas, based on the lessons learned from the existing literature, are discussed below.

A. 5G Networks: Access, Backhaul, Core Networks

5G will bring connectivity to almost all aspects of life, e.g., connecting hospitals, schools, vehicles, home appliances, and more, through increasing network capacity and efficiency of the technologies. New technological concepts are added from the access to the core networks. In the access network, ultra-densification and offloading, moving towards mmWave and unlicensed spectrum, and leveraging the advances in MIMO are the key technologies to fulfill the data rate and service requirements [3]. However, these technologies must also be investigated from the point of view of security, since the security challenges in these technologies, e.g., MIMO are much different than conventional security challenges. For instance, ensuring secrecy in resource allocation in MIMO as elaborated in [305] needs further investigations. Using the concepts of AI or big data in physical layer as described in [306] for improving the data or network security needs further research. The concepts of AI or big data are currently used mainly in the upper layers. Similarly, beamforming for improving the traffic or network security is also an interesting research area. Currently various methods are used to protect the information from eavesdropping, however using the strength of MIMO to securely perform other tasks, for example wireless power transfer [307], constitute interesting future research.

The backhaul network connecting the access and core networks has a major change in terms of splitting the control and data planes. The split is mainly advocated with the emergence of SDN, taking the control plane out of the data plane and deploying it in the clouds. Similarly, the major change in the core network is also the split of the control plane functionalities from the forwarding plane. For example, the PCRF from PDN gateway will reside in the cloud-based core network. Even though the split has many benefits such as global network resource visibility and programmability, there are many challenges that still need effective solutions. In the backhaul network, the GTP tunnel processing for small sessions generated by IoT devices will make the interaction of the control-data planes extremely high. In such scenarios, malicious or even legitimate but compromised nodes will be capable of causing saturation attacks. Furthermore, the processes of recognizing malicious or compromised nodes will further increase the interaction between the control and data planes. Hence, the security challenges due to the split in backhaul networks need thorough analysis from the point of view of the new services and devices, particularly, the resource constrained devices which can be easily compromised to act as bots. From the core network perspective, the traffic generated by massive IoT deployment will make the authentication, mobility, and policy control entities in the core network easily exhaustible by malicious actors. One way forward, although needs further investigation, is the functional split that some functions, e.g., firewalls and authentication mechanisms must reside near the edge as virtualized security functions to recognize malicious activities at the entrance points. Such capabilities require the technologies of SDN, NFV and MEC in the network. However, these technologies-albeit all the promising capabilities, need further research in terms of security as described below.

B. Key 5G Technologies: SDN, NFV, MEC

The use of SDN, NFV and the advanced concepts of cloud computing such as MEC or even fog computing [308] will be inevitable in future networks. The main concern, though, is that the security challenges in each of these technologies are handled or solved in a solitary fashion usually. Similarly, the security concerns are solved only looking at the security of the technology, paying little attention to the use case or applicability of the technology. For example, the security of the centralized control for resource exhaustion is considered mainly in the form of either increasing its resource capacity [217] or distributing the control planes [212], [213]. However, looking at the dynamism of future wireless networks, both of these solutions will be sub-optimal. On one hand, a high resource controller in each corner of the network will not be cost effective. On the other hand, the dynamism or more so, the user behavior in different parts of the network can change sporadically. Hence, resource extension in certain parts of the network is not the only way. Similarly, the dynamism of future networks might require speedy distribution of certain or partial control procedures in different parts of the network. Research on service dynamics-based or traffic volatility-based control procedures distribution is still needed. Furthermore, the security of SDN must be investigated mainly on use-case basis specifically for future wireless networks, where the changes in service functioning, user node or traffic behavior can be frequent.

NFV brings interesting opportunities to the wireless networks domain. When the basic security challenges in NFV are mitigated, NFV can strengthen the network security in a variety of ways. For instance, NFV-based softwarized security functions that can be mitigated from the centralized cloud to different network perimeters will enable highly dynamic security, as par with the promises of dynamic networks. Further research is needed to highlight the basic challenges in moving security functions throughout the network, and to investigate the impact on the network, security functionality, and the challenges related to such movement of security functions. Network slicing will enable resizing, shrinking and extending of network resources for various services and network functions. Security policies related to such adjustments of resources among various stakeholders need to facilitate security of the overall system. Security policies in this case and service level agreements for curtailing illegitimate use of network resources among lessor and the lessee parties must come forward.
The security challenges faced in different cloud computing paradigms are mostly related to virtualization and VNFs, as well as the security policies related to these technologies [242]. The security challenges in the smaller forms of clouds, e.g., fog computing, are much different and are highly dependent on the resource capacities and networking [399]. It is important to note that the services that use these technologies will have direct impact on the security of these platforms. For instance, compromised IoT devices sending sporadic messages to the MEC or fog platforms will cause resource saturation easily compared to thickly resourced cloud platforms. However, the advanced concepts of cloud computing such as MEC and fog computing will facilitate robust security besides bringing services near to the users. For example, bringing the authentication mechanisms near the edge or in other words, the user, will facilitate faster authentication in latency critical applications. Since the security of the cloud computing is highly related to the technologies using these concepts, such as NFV and SDN, therefore, it is highly important to study the mutual impact of these technologies on each other.

C. Mutual Impact of the Key Technologies on Overall Security

The combination of the key enabling technologies such as SDN, NFV and cloud computing has two sided effects on security. This combination can either decrease security or increase the security of the overall system. Furthermore, the security of the overall system can be compromised due to lapses in security in one technology. For example, in future networks the control plane can be placed in a cloud when the control-data planes separation occur. Consider a situation of security breaches in the cloud platform such that un-authorized access to the control platform, e.g., SDN controller, is granted to a malicious user or application. The malicious application in this case can change the forwarding behavior of the entire network connected to that controller. In such cases the security of the cloud will have a direct impact on SDN or the underlying network. Similarly, the compromised cloud will have threatening effects on NFV or the whole NFV management systems that usually reside in clouds [222]. Moreover, a compromised SDN controller can leave the clouds disconnected. In addition, compromised virtual machines or slices can also compromise the security of the clouds as described in [400], [401].

Strong security of each technology has positive implications for the other technologies and the overall system. For example, a secure cloud platform means only the inherent security of the technology can have adverse implications on that technology. If there is a proper isolation enforced in virtualized resources, the overall cloud platform will be more secure. Similarly, if the SDN data plane or the control planes are properly secured, the security of virtual networks over the data plane and cloud applications interacting with the SDN controller will be more secure [27]. Furthermore, there is also a need to develop integrated security solutions where coordination among various security controls, possibly in different layers, is required to deliver integrated security to the overall system [245]. However, the implications of security of these three technologies on each other are not properly studied, and thus, need further investigation.

D. Future Privacy Measures

5G and beyond networks will create heaps of privacy concerns than ever before. These challenges are not merely from consumer’s perspective but also from the view point of service providers, network operators, Vendors and CSPs in order to ensure the successful deployment of such networks. The enabling technologies such as blockchain and edge computing/clouds are integrated to future 5G and beyond networks and are expected to fulfill the privacy requirements to some extent. However, it will not ensure the complete privacy protection as future networks consists of multiple stockholders with diverse business interests. This makes privacy a challenging task for each stakeholders in the network. Therefore, it is a key for the regulatory bodies to draft the privacy regulations in such a way that it should protect consumer’s privacy and at the same time maintain the interests for each of the involved entities. Apart from this, existing cryptographic techniques can be utilized with new security algorithms to ensure the authentication and authorization of valid entities. Approaches such as privacy-by-design and customized privacy are going to play crucial role for future privacy solutions. Furthermore, AI based privacy solutions will be necessary for future network as the privacy should be given by sensing the particular context.

E. New Trust Models

In previous generations of networks, the devices connected to mobile networks were assumed to be trustworthy, and thus the trust models were straightforward [104]. Now that new devices (massive IoT) ranging from home appliances to surveillance to industrial equipment will span across the network, the existing trust models are not sufficient. New trust models must be erected for the new actors entering the domain. That will cover the extended requirements in authentication between various actors, accountability and non-repudiation. Furthermore, the new technologies that enable sharing the infrastructure resources through slicing require mechanisms to dynamically expand or contract their leased resources securely and timely. Similarly, constrained devices such as the tiny sensors used for climate monitoring might open security loopholes into the system. Therefore, new trust models are necessary which might be difficult at this time, but can be defined gradually as the actors adopt 5G.

F. New Opportunities

5G networks will amalgamate diverse services and technologies unlike previous generations. Thus, the security challenges and solutions will be as different and diverse as the services and technologies used in 5G. Albeit the complexities of services, the technologies used in 5G will also enable new security technologies that were practically absent in the previous generations. For example, using the disciplines of AI such as neural networks for security in 4G were not easily realizable due to the resource constraints, such as limited
computation near the sources of traffic. 5G delivers high computation near the sources of traffic through the concepts of cloud computing near the edge, e.g., MEC. MEC resources near the edge can be used to learn the traffic behaviour, detect possible security threats, and stop security breaches at the edge before they occur, this will mostly be facilitated using the disciplines of AI. As it is now known that the traffic over the network will further exacerbate, the data can also be used for network analysis [402], and thus its role in decision making will also be considered for improving network security. Similarly, the concepts of Blockchain can improve the security of sensitive services over the network [403]. Therefore, 5G will bring new opportunities to utilize the recent development in computing technologies to strengthen the security of the overall network, as well as the services that use the network.

X. Conclusion

Wireless communication networks have been evolving from connecting simple mobile phones in 1G towards connecting almost all aspects of life in 5G. During this evolution, security landscape has equally evolved from simple phone tapping to diverse attacks on mobile devices, network equipment and services. For integrating new things (IoT) and services into the network, 5G will use new technologies such as advanced cloud computing concepts (e.g., MEC), SDN, NFV, and massive MIMO etc. These technologies have their own inherent security challenges which can further complicate the network security landscape. Therefore, in this paper we have discussed the security challenges that exist in different parts of the network such as access network, core network, and within the technologies that will be used in 5G networks. The conglomeration of diverse devices, services, and new networking technologies does increase the security threat landscape, and thus new security solutions must be sought for efficient and secure connectivity. Hence, we have thoroughly discussed the security challenges in different parts and technologies of 5G networks, and have outlined the possible security principles, techniques and proposals for the mentioned security challenges. Since privacy of users and user information is going more towards the hands of the infrastructure owners and systems’ operators, for instance in cloud storage systems, privacy has grasped a lot of research attention. Therefore, we also outline the weaknesses in the privacy of wireless networks and discuss the potential solutions for ensuring user and data privacy.

Due to the increasing diversity of devices and emergence of new services, we have provided a vision of future connected world, i.e., the XG. Since the connected systems in near future will be complex, the security threat landscape will also be complex, and thus new modes of security operations will be inevitable. For instance, blockchain securely shares information among the intended users, thus how to use the technology to strengthen user security has been highlighted in this article. To sum it up, it is highly likely that new types of security threats and challenges will arise along with the deployment of novel communication technologies and services. However, considering these challenges right from the initial design phases to the deployment phases will minimize the likelihood of potential security and privacy lapses.

References


Technical specification group services and system aspects: study on security aspects of machine-type communications (MTC) and other mobile data applications communications enhancements, release 12,” 3GPP, Sophia Antipolis, France, Rep. 33.868, Jun. 2014.


An Analysis of the Security Needs of the 5G Market


Andrei Gurtov received the M.Sc. and Ph.D. degrees in computer science from the University of Helsinki, Finland, in 2000 and 2004, respectively. He is a Full Professor with Linköping University, Sweden. He is also an Adjunct Professor with Aalto University, the University of Helsinki, and the University of Oulu. He visited ICSI in Berkeley multiple times. He has coauthored over 200 publications, including 4 books, 5 IETF RFCs, 6 patents, over 50 journal, and 100 conference articles. He has supervised 15 Ph.D. theses. He serves as an Editor of IEEE INTERNET OF THINGS JOURNAL. He is an ACM Distinguished Scientist, an IEEE ComSoc Distinguished Lecturer, and the Vice-Chair of IEEE Sweden Section.

Jude Okwuibe received the B.Sc. degree in telecommunications and wireless technologies from the American University of Nigeria, Yola, in 2011 and the master’s degree in wireless communications engineering from the University of Oulu, Finland, in 2015, where he is currently pursuing the Doctoral degree in communications engineering with the Graduate School. His research interests are 5G and future networks, IoT, container orchestration, SDN, network security, and biometric verifications.

Mika Ylianttila (M’99–SM’08) received the Doctoral degree in communications engineering with the University of Oulu, Finland, in 2005. He is a full-time Professor with the Centre for Wireless Communications, Faculty of Information Technology and Electrical Engineering, University of Oulu, where he was the Director of the Center for Internet Excellence from 2012 to 2015, and an Associate Director of the MediaTeam Research Group from 2009 to 2011, and a Professor (pro tem) in information networks from 2005 to 2010 and has also been an Adjunct Professor in computer science and engineering since 2007. He has coauthored over 100 international peer-reviewed articles on broadband communications networks and systems, including aspects on network security, mobility management, distributed systems, and novel applications. His research interests include also 5G applications and services, software-defined networking and edge computing. He is an Editor of Wireless Networks journal.

Tanesh Kumar is a Doctoral Researcher with the Centre for Wireless Communications, University of Oulu, Finland. He has coauthored over 35 scientific publications. His research interests include security and privacy in the Internet of Things, network security, MEC, and blockchain. He was a recipient of several prestigious awards, including the Nokia Foundation and the HPY Grant Awards.

Ijaz Ahmad received the B.Sc. (Engineering) degree from UET Peshawar, Pakistan, and M.Sc. (Technology) and Ph.D. degrees in wireless communications from the University of Oulu, Finland. He is currently a Research Scientist with VTT Technical Research Centre of Finland. He has been a Visiting Scientist with Aalto University, Finland, since 2018 and visited TU Vienna, Austria, in 2019 to work with Prof. T. Sauter on AI for wireless networks. He has coauthored over 30 publications, including a patent application and one edited book with Wiley. His research interests include 5G and 6G, 5G security, SDN security, software-defined mobile networks, and machine learning for wireless networks. He was a recipient of several awards, including the Nokia Foundation, Tauno Tönnönen and Jorma Ollila Grant Awards, and two IEEE Best Paper Awards.

Shahriar Shahabuddin received the M.Sc. (Distinction) and Ph.D. degrees from the Centre for Wireless Communications, University of Oulu, Finland, in 2012 and 2019, respectively, under the supervision of Prof. M. Juntti. In 2015, he was with the Computer Systems Laboratory, Cornell University, USA, in Prof. Christoph Studer’s Group. He is currently with Nokia, Finland, as an SoC Specialist. He has received several scholarships and grants, such as Nokia Foundation Scholarship, University of Oulu Scholarship Foundation Grant, Taunu Tönnening Foundation Grant during the Ph.D. degree. His research interests include VLSI signal processing, MIMO detection and precoding, 5G security, and machine learning applications for wireless communications. He was a recipient of the Best Master’s Thesis Award from the Department of Communications Engineering, University of Oulu in 2012.