# Context-Awareness Enhances 5G Multi-Access Edge Computing Reliability

**BIN HAN**[1], (Member, IEEE), **STAN WONG**[2], **CHRISTIAN MANNWEILER**[3],
**MARCOS RATES CRIPPA**[1], **AND HANS D. SCHOTTEN**[1,4], (Member, IEEE)

[1]Institute of Wireless Communication, Technische Universität Kaiserslautern, 67655 Kaiserslautern, Germany
[2]UbiXpace, London E3 4PU, U.K.
[3]Nokia Bell Labs, 81541 Munich, Germany
[4]Research Group Intelligent Networks, German Research Center for Artificial Intelligence (DFKI GmbH), 67663 Kaiserslautern, Germany

Corresponding author: Bin Han (binhan@eit.uni-kl.de)

**ABSTRACT** The fifth-generation mobile telecommunication network is expected to support multi-access edge computing (MEC), which intends to distribute computation tasks and services from the central cloud to the edge clouds. Toward ultra-responsive, ultra-reliable, and ultra-low-latency MEC services, the current mobile network security architecture should enable a more decentralized approach for authentication and authorization processes. This paper proposes a novel decentralized authentication architecture that supports flexible and low-cost local authentication with the awareness of context information of network elements such as user equipment and virtual network functions. Based on a Markov model for backhaul link quality as well as a random walk mobility model with mixed mobility classes and traffic scenarios, numerical simulations have demonstrated that the proposed approach is able to achieve a flexible balance between the network operating cost and the MEC reliability.

**INDEX TERMS** 5G, network architecture, context awareness, multi-access edge computing, MEC, network reliability.

## I. INTRODUCTION

Fifth generation (5G) mobile telecommunication network has been expected to involve Multi-Access Edge Computing (MEC) [1]–[3]. MEC has evolved from mobile cloud computing, which shifts the important computation and storage tasks from the core network to the access network. Thus, mobile services can gain superiority of low latency, power efficiency, context-awareness, and enhanced privacy protection [4], [5]. Furthermore, the security concerns become across the core network and the access network – from network elements to user equipment (UE) and from tenants to their network slices [6]. For example, the mobility management entity (MME) is set to be divided into few functions which can be co-existed in the access network and core network. This function decomposition gives the overall system flexibility, deployment efficiency and service agility.

The fourth generation (4G) long-term evolution (LTE) authentication and key agreement (AKA) has an architecturally inherited dependency on the core network, and can therefore constrain the exploitation of MEC. Nevertheless,

there has been limited contributions on the evolution of the LTE centralized security architecture despite of demands of improved AKA efficiency in MEC [7]. However, delivering a flexible next generation telecommunication system is the main objective of 5G. All services at the edge of the network should be securely provided. Hence, the authentication and authorization processes can also exist at the edge of the network to increase the subscriber authentication efficiency and the reliability to the overall system.

Generally, authentication, authorization and accounting (AAA) processes require certain amount of cost from the system per registration of a subscriber. Putting these processes at the edge of the network would increase the efficiency of subscriber authentication and authorization. Also collecting valuable information helps to understand the context of the overall network situations, the connectivity availability and the traffic pattern [8]. With such knowledge we can increase the overall system reliability with minimal operations cost. In this paper, we will use the context-awareness technique to extract the network situation and reduce the traffic in between fronthaul and backhaul when subscriber authentication and authorization take place. Particularly, we use network context information to distinguish

The associate editor coordinating the review of this manuscript and approving it for publication was Walid AL-Hussaibi.

**TABLE 1.** List of acronyms.

| Abbreviation | Definition |
|---|---|
| 3GPP | Third Generation Partnership Project |
| 4G | fourth generation |
| 5G | fifth generation |
| AAA | authentication, authorization and accounting |
| AKA | authentication and key agreement |
| AMF | access and mobility management function |
| AS | access stratum |
| BSS | business supporting system |
| CCCM | Central Cloud Connection Monitoring |
| CSSO | Central Security Service Outage |
| CSSR | Central Security Service Risk |
| EC | edge cloud |
| ES | Emergency Service |
| EPS | Evolved Packet System |
| ETSI | European Telecommunications Standards Institute |
| IKE | Internet key exchange |
| IoT-GW | Internet-of-Things Gateway |
| IPsec | Internet protocol security |
| LAA | Local Access Assistant |
| LSS | Local Subscriber Servers |
| LTE | Long-Term Evolution |
| MEC | Multi-Access Edge Computing |
| MME | Mobility Management Entity |
| MNO | mobile network operator |
| NAS | non-access stratum |
| NFV MANO | network function virtualization management and orchestration |
| OSS | operations support system |
| O&M | orchestration and management |
| RAN | radio access network |
| SA | Security Auditing |
| SDM | software-defined mobile network |
| SDMO | SDM orchestrator |
| SDM-C | SDM coordinator |
| SDM-X | SDM controller |
| SDN | software-defined network |
| SEG | security gateway |
| SMS | short message service |
| TZ | Trust Zone |
| UDM | unified data management |
| UE | user equipment |
| VIM | virtual infrastructure manager |
| VNF | virtual network function |
| V-AAA | Virtualized-AAA |
| ZM | Zone Management |

the normal from emergency situation and deliver proactive control in facing any situation.

The main contribution of this paper is a novel decentralized security approach for MEC in 5G networks, which aims at delivering a flexible and context-aware AKA executing within an edge cloud (EC), and decouples the authentication process from the central cloud when backhaul connection outages are detected. The propose approach is to make aware of network context information and able to proactively synchronize subscriber profiles to support local AKA procedures. We develop an numerical simulation demonstrates the balance in between the MEC reliability and the operations cost.

The paper is organized as follows. Section II briefly reviews the 4G LTE AKA mechanism, and analyzes its inadequacy of 5G MEC services. In Section III, we propose a novel decentralized authentication architecture in 5G. A Virtualized AAA system across the central cloud and edge cloud is elaborated in Section III-A. Section III-B formulates a MEC cognitive security system that maintains the availability of the services. Section IV-A presents a context-aware mechanism that enables the local authentication to reduce the backhaul traffic and improve the efficiency of subscriber management. Section IV-B presents a Markov chain to characterize and forecast the reliability of backhaul connections. At the end, Section V demonstrates the effectiveness of our proposed method with numerical simulations, before we conclude this paper and provide a couple of future challenges in Section VI. For the readers' convenience, Tab. 1 lists all the abbreviations used in this article in the alphabetical order.

## II. SECURITY ARCHITECTURE OF CELLULAR NETWORKS
### A. AUTHENTICATION AND KEY AGREEMENT IN THE 3GPP EPS

Security in the Third Generation Partnership Project (3GPP) Evolved Packet System (EPS) is divided into access stratum (AS) and non-access stratum (NAS) security. Security procedures rely on the EPS architecture; the most important elements are shown in Fig. 1 [9], [10].

When a user equipment (UE) attaches to a LTE network, the AKA is executed between UE and Mobility Management Entity (MME). Using the same so-called key of access security management entities ($K_{ASME}$), keys for securing the NAS signaling are derived. Using $K_{ASME}$, UE and MME also derive a key $K_{eNB}$, which is also passed to the evolved Node B (eNB) that the UE is connecting to. $K_{eNB}$ is used in the communication session as the basis of AS security, i.e. security on the radio link between the UE and the eNB. Three further keys are derived from $K_{eNB}$ for control plane integrity protection, control plane encryption, and user plane encryption, respectively. A fourth key for user plane integrity protection is derived from $K_{eNB}$ only when required. These traffic protection keys are not used for an infinite amount of data. The eNB periodically initiates intra-cell handover to derive a new $K_{eNB}$ and new protection and encryption keys.

LTE specifies three pairs of cryptographic algorithms, each pair comprises a confidentiality protection algorithm and an integrity protection algorithm. This redundancy provides fallback options in case a pair fails during the expected system life-cycle. Also, the backhaul link is protected by Internet key exchange (IKE) / Internet protocol security (IPsec). Generally, the IPsec tunnels are terminated by a dedicated security gateway (SEG), but they may also be terminated at the MME or serving gateway, for the control and user plane, respectively.

Since the eNB is considered as a physically exposed entity, with a notable risk of being compromised by an attacker with physical access, it has no access to the NAS signaling. However, it has access to the user plane traffic that is decrypted and re-encrypted at the eNB between radio interface and backhaul link. Also, 3GPP requires all network elements installed in a ''secure environment'', unauthorized access to eNBs is strictly prohibited. The eNB is also required to protect the keys and other important processes.
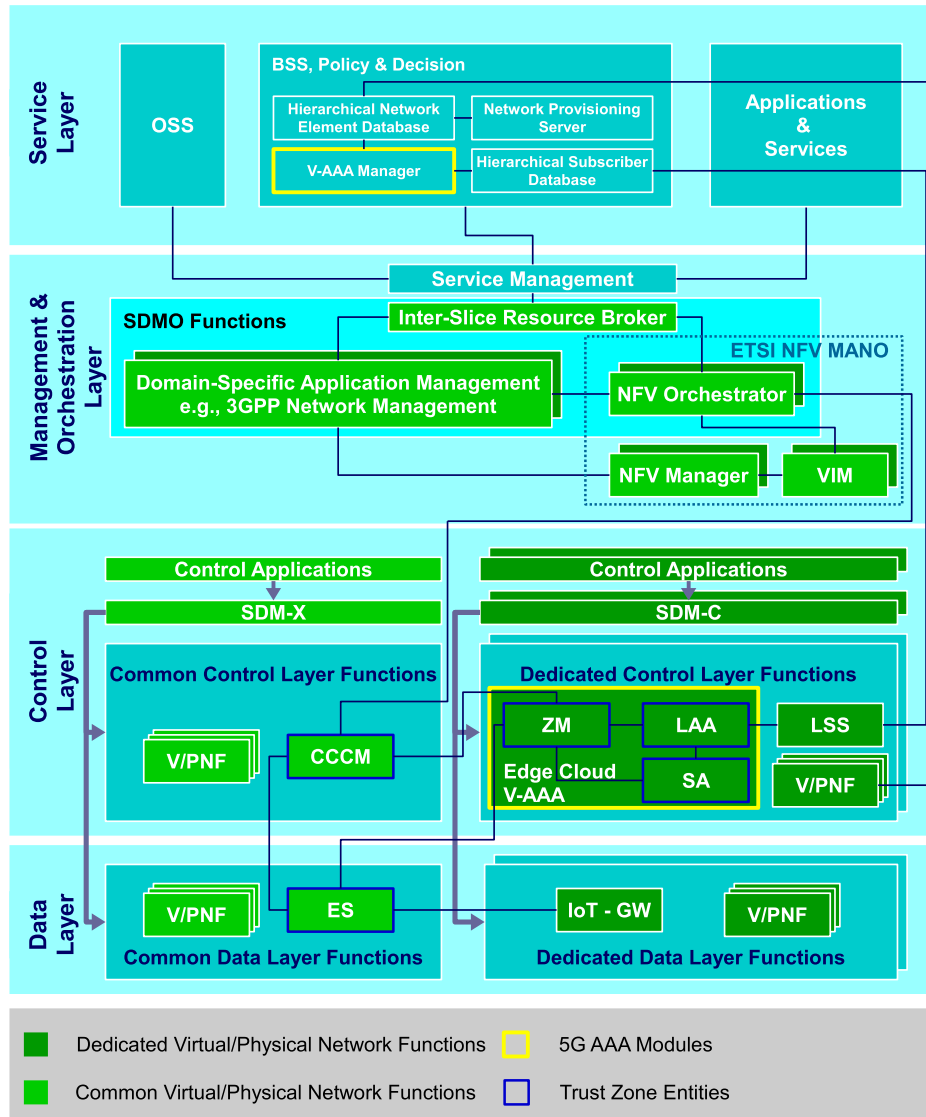
**FIGURE 1.** Security architecture of 3GPP LTE / EPS [11].

For an overall security concept for an LTE network, additional non-standardized network security measures must be implemented, among them traffic filtering and separation between network-internal security zones, secure operation and maintenance (O&M), and secure operation of services such like DNS, NTP, IP routing, etc.

### B. CHALLENGES IN 5G MULTI-ACCESS EDGE COMPUTING
The 3GPP security architecture specified for LTE/EPS cannot fulfill new requirements of emerging services in 5G. 5G networks requires to improve the flexibility of security functions with the MEC services.

The concept of MEC was proposed by the European Telecommunications Standards Institute (ETSI), suggesting to use the base stations instead of centralized cloud servers for offloading computation tasks from mobile users, in order to provide "*IT and cloud-computing capabilities within the radio access network (RAN) in close proximity to mobile subscribers*" [12]. Differing from traditional mobile cloud computing systems, MEC systems with hierarchical controlling and management are more near to the end users, less coupled with the central cloud, less dependent on the backhaul network, and therefore with a better support to latency-critical applications [4]. The exploitation of MEC, however, can be constrained by the centralized 3GPP LTE/EPS security architecture.

The current AKA procedure deeply relies on the MME in central cloud, the backhaul latency in network attachment and mobility management can become a bottleneck for the quality of MEC services. When data congestions occur in the core network, this delay will significantly increase, and may violate the service latency requirement, especially for some ultra-reliable and ultra-low-latency emergency communication applications [13].

Additionally, the emerging concept of "5G Islands" [14], [15] might be expected to autonomously provide local devices with central-cloud-independent MEC services.

On one hand, 5G Islands can be created in a scheduled manner to serve public events with dense requirement of local MEC traffic, or to maintain MEC services with planned disconnection from the core network. On the other hand, they can be also created upon unexpected demand, so as to provide temporary local MEC services for an improved network resilience against cyber-attacks and extreme disasters (such like fires, explosions, earthquakes etc.) that can cause infrastructural malfunctions. This concept is novel, but the development of its elements has already been launched. For example, Simsek *et al.* [16] discuss how 5G networks will be able to support highly autonomous and smart decision making capabilities at various layers of the network. Network architectures [17], [18] were proposed where MMEs autonomously perform UE management in a distributed manner. In this field, distributed and decentralized edge security methods shall be developed to cope with non-static and unreliable backhaul connections. More specifically, the AKA procedures are expected to be decentralized at the network edge, and to provide stable quality of service of MEC in the 5G.

Aiming at a decentralized security approach for reliable MEC, there are two key challenges to overcome:

- In the architectural perspective, a novel network architecture is needed to enable a flexible and secure migration of AAA functions between the central cloud and edge clouds in virtualized 5G networks.
- In the operations perspective, it is critical to keep the extra operations cost generated by the decentralization and migration of AAA functions on a reasonable level.

### III. DECENTRALIZED SECURITY APPROACH OVER EDGE CLOUDS
### A. 5G AUTHENTICATION AUTHORIZATION ACCOUNTING
Towards a flexible decentralized AKA mechanism for reliable 5G MEC services, we proposed a novel virtualized AAA

**FIGURE 2.** The 5G AAA framework enhanced with Trust Zone in the *5G PPP* architecture. The terms OSS, VIM, SDM-X, SDMO and SDM-C stand for operations support system, virtual infrastructure manager, software-defined mobile network (SDM) controller, SDM orchestrator and SDM coordinator, respectively. V/PNF denote virtual/physical network functions.

approach: the 5G AAA [19]. It combines two independent international standard systems that are the 3GPP and the ETSI, as a single platform to manage and secure the subscribers, the tenants and the network slices under the 5G flexible network environment.

Generally, 5G AAA system is constructed by a Virtualized-AAA (V-AAA) manager in core network and several V-AAA servers distributed in various ECs, and supported by hierarchical databases approaches for tenant data and tenant's subscribers synchronization. This hierarchical database approach has been divided into two levels of synchronization. Within the tenant network or dedicated resources of tenant network slice would have bi-directional replication across different regions. On the other hand, the tenant's subscriber database would be applied many-to-one synchronization

with the MNO core network subscriber database. This a directional synchronization. This logical illustration cab be found in Fig. 2. From the 3GPP view on 5G network architecture [20], the 5G AAA system is an advanced implementation of the access and mobility management function (AMF), which provides a complete generic AMF set in the central cloud with the V-AAA Manager, while offering specifically customized subsets of AMF for various 5G network slices in different local ECs with the polymorphic V-AAA servers. Similarly, the hierarchical databases and the Local Subscriber Servers (LSS) together construct the 3GPP unified data management (UDM). Taking the 4-layer-view on 5G network architecture proposed by the 5G Infrastructure Public Private Partnership (5G PPP) [21], the V-AAA Manager can be seen as part of the business supporting system (BSS) on the

service layer, and each EC V-AAA server is a virtual network function (VNF) dedicated to a particular network slice.

With such an architecture that distributes security functions and databases to all ECs, the 5G AAA approach enables a flexible and decentralized decision and application of security policies in every EC. It is able to:

- Keep the centralized governance of tenants and mobile subscribers in the central cloud;
- Allow a degree of freedom of tenant governance their subscribers at the edge cloud;
- Use the AKA generated data to track and authenticate the mobile subscribers from the edge cloud and central cloud;
- Maintain the mobile subscribers and tenants with the point of attachment in the central cloud as well as in the edge cloud;
- Retain subscriber services even when the backhaul connection is limited or unavailable;
- Provide a trust platform collecting the trust value from the mobile subscribers to the NFV-based tenants, and from the physical network entities to the virtual network entities.

### B. TRUST ZONE

The proposed 5G AAA [19] approach enables the AKA procedure in the local EC to improve the MEC services. Nevertheless, its deployment still calls for novel solutions of adaptive AMF invocation. Due to the incomplete AMF set and probably limited budget on security measures, the local V-AAA servers are usually less secure than the central V-AAA Managers. Therefore, it is generally preferred to invoke the AMF from the central cloud in common cases, and only to rely on the local version when the backhaul connection cannot satisfy the latency requirement of MEC services due to – intentional or unintentional – restrictions, congestions and collapses. To achieve this, the 5G AAA system must be able to adaptively switch between the centralized and local security modes with respect to the current backhaul connection status. The switching process must be carefully designed to mitigate data leaks.

Motivated by this challenge, we designed an EC security architecture, the Trust Zone (TZ), in order to enhance and extend the 5G AAA approach to a cognitive solution [22]. The basic principle of TZ is to let the local EC cognitively measure the quality of its backhaul connection. When the backhaul is healthy, the AMF is invoked at the V-AAA Manager in central cloud. When the backhaul connection is seriously limited or unavailable, the TZ relies on the V-AAA server that executes local AMF to maintain MEC services.

In a structural view, as illustrated in Fig. 2, a TZ system consists of five entities:

1) **Central Cloud Connection Monitoring (CCCM)** – This entity periodically visits the NFV orchestrator in central cloud to evaluate the backhaul status. Additionally, when the backhaul connection is limited or unavailable, CCCM tries to diagnose the malfunction.

This diagnosis, including information such the likely location and reason of malfunction, can be used by the software-defined network (SDN) management to assist backhaul recovery with backup resources, e.g. network redundancy and alternative routes such as satellite links. A dynamic network resource allocation in the SDN management can also be supported by the backhaul status information. When the backhaul connection becomes poor, specific network functions such as mobility management and fault management can obtain network resources with higher priorities.

2) **Zone Management (ZM)** – This is the core entity of TZ, which implements most of AMFs and is connected with all other TZ entities. It triggers and coordinates the state transitions in TZ, when it receives reports about updated backhaul status. Integrated with interfaces to the central cloud and to the UEs, it also collaborates with the Local Access Assistant (LAA) to accomplish the AS security procedures in the local EC. When the backhaul connection is healthy, ZM cooperates with the V-AAA Manager in central cloud to execute centralized AKA. When the backhaul is limited or unavailable, local authentication is needed, ZM will then collaborate with the LAA to execute local AKA instead.

3) **Local Access Assistant (LAA)** – This entity makes use of the subscriber information that is essential for authentication and stored in the LSS, in order to locally execute security measures such as AKA procedures. It can be considered as a lite AMF module migrated from the central cloud to the local EC. LAA is deactivated by the ZM when the backhaul is reliably healthy, and activated otherwise. The LAA is isolated from the ZM mainly due to security concerns: the subscriber database should be decoupled from the ZM that is the central controlling unit of TZ and the priori target of potential cyber-attacks. To mitigate the risk that the LAA is attacked and manipulated to send fake information to the central cloud, an asymmetric trust relation between the central V-AAA manager and the local V-AAA server is needed, as will be introduced in Sec. III-C.

4) **Security Auditing (SA)** – Mobile network operators (MNOs) usually keep logs of security critical operations for each subscriber. These logs can be audited to investigate all potential risks of illegal access and cyber-attacks. Both the log database and the auditing center are usually located in the central cloud, so when the backhaul connection is unavailable, they cannot keep tracking the AAA operations executed by local V-AAA servers. To close this gap and provide a continuous audit, the SA module is designed. It is activated under the local security mode to record all local AAA operations. After the backhaul recovery, these records can be either actively pushed to the central auditing center, or passively pulled upon request.

5) **Emergency Services (ES)** – we have named some extreme disasters that can lead to backhaul outages in Sec. II-B, such as fires, explosions and earthquakes. These events can be harmful not only to the network infrastructure, but also to human safety. A set of emergency services can be defined, which help users avoid personal injury and property damage under such disasters, even when their devices cannot be authenticated by the network. These services are implemented in the ES entity, including but not limited to public disaster alarm, evacuation guidance, positioning service, emergency call and short message service (SMS). Additionally, the ES is connected to the Internet-of-things gateway (IoT-GW) to collect context information about public disasters from external sensor networks. Such information can be visited by the network function virtualization management and orchestration (NFV MANO) through the CCCM to assist the network diagnose and backhaul recovery processes.

Among those entities listed above, the first three entities are dedicated to a specific tenant or network slice, and within the EC V-AAA server. However, CCCM and ES are defined as common functions and non-V-AAA server functions exist on both control plane and user plane[1]

### C. ASYMMETRIC TRUST TO SAFELY TRANSFER ACCESS MANAGEMENT

It should be noted that local V-AAA servers are less secure than the central V-AAA Manager. Cyber-attackers may initiate attacks to disconnect the central cloud and the EC, and hack the local TZ that is easier to target than the central cloud. Then they may eventually attempt to obtain access to the central cloud during the backhaul recovery, when the TZ hands the access management back over to the central cloud. To mitigate this risk, we designed an asymmetric approach of transferring the access management between the central cloud and the EC, which can be briefly summarized as follows: When a degradation of backhaul connection occurs and the TZ switches to the local security mode, all devices already authenticated by the central V-AAA Manager before the mode switching will be considered as trusted by the local V-AAA server, and retain a seamless access to the local MEC services. In contrast, devices that newly arrive after the switching must go through a local authentication and authorization process to access the MEC services. Only the users that are successfully authenticated and authorized are considered as temporarily trusted to access the local MEC services, while for the rest devices only the emergency services will be available. When the backhaul connection recovers, all temporarily trusted devices in the local EC have to reconnect to the network in a preplanned order, so that they are authenticated and authorized by the central V-AAA Manager again.

---

[1]Further details about the interfaces between TZ entities and a behavior model of TZ are available in [22].

## IV. CONTEXT-AWARENESS IN TRUST ZONE
### A. CONTEXT-AWARE SYNCHRONIZATION OF SUBSCRIBER PROFILES

The proposed local V-AAA server – or more specifically, the LAA – relies on the subscriber authentication information stored in the LSS to accomplish the local AKA procedure in the EC. However, due to security concerns, parts of such information, such like the NAS key, cannot be locally generated in ECs, but only available in the central cloud. Thus, the subscriber authentication information must be synchronized from the central cloud to the LSS before the EC switches to the local security mode, and periodically updated. Furthermore, taking the device mobility into account, users may enter an EC from outside while it is suffering from a high backhaul latency – or in the worst case, disconnected from the central cloud. Therefore, a LSS should not only synchronize the authentication information of devices in the local EC, but also that of the devices in neighbor ECs nearby, which are able to arrive in the local EC during the next synchronization interval.

Self-evidently, with respect to the mobility, various devices can move over different distances within the same synchronization interval. Thanks to the modern technologies of transportations and communications, people are nowadays able to travel over three hundred kilometers per hour, while continuously exploiting the cellular network services. Considering the limited area of a typical EC, e.g. between $5 \times 5$ km$^2$ and $15 \times 15$ km$^2$ [23], the authentication information of every such high-mobility subscriber must be synchronized to 400 to 4000 different LSSs per hour. Accounting the large amount of user devices in the entire mobile network, this synchronization process will generate a considerably significant data traffic over the backhaul network, and hence leads to extra operating cost and raises the risk of data congestion. In other words, there is a conflict between the interests of subscribers (MEC service reliability) and of MNOs (operating cost). In this concern, we designed an advanced synchronizing mechanism with awareness of the context information of both UEs and ECs, which can be explained as follows.

The local security mode is activated only when the central security services fail to respond in time, the gain of subscriber authentication information synchronization is therefore not significant in the ECs with negligible probability of Central Security Service Outage (CSSO). Note that we refer to the term CSSO as the cases where the security network functions in central cloud fail to respond within the desired latency, instead of a complete unavailability of service. Generally, to estimate this outage probability for a certain device in different ECs, the following factors should be taken into account:

1) **User arrival**: depending on the device-to-edge-cloud spatial distance and the user mobility, every EC has its individual expectation about the user arrival probability in the next synchronization interval. The ECs located far from the device have less chance to serve the device than those located nearby. For still and low

mobility users, the user arrival estimation can be simply accomplished though the handover prediction procedure of legacy networks. However, high-mobility users such like vehicles on highway and railway passengers can travel over multiple ECs during one backhaul outage, so that an additional position prediction in macro scale can be necessary.

2) **User stay time**: upon arrival, depending on the user mobility, the coverage area of the EC and the traffic model in that area, every EC has its individual expectation about the duration of user's stay in it. Some "important" ECs covering larger or more traffic-congested areas (e.g. the city center of Paris) are expected to serve a user for longer time, when compared to those "unimportant" ones (e.g. a small village aside a high-speed railway). The risk of CSSO during the user's stay in these ECs, therefore, is also higher.

3) **Backhaul reliability**: depending on the current status of backhaul connection, every EC has its individual expectation about the chance that a CSSO occurs during a certain period in the future. Generally, the ECs with worse current backhaul connection quality are more likely to suffer from CSSO.

Thus, from real-time information of devices (mobility and position) and ECs (coverage area, traffic model and backhaul throughput), we are able to extract high-level context information such as user motion pattern and backhaul outage probability, as illustrated in Fig. 3. These context information can be used to estimate the CSSO risk in different ECs for every certain device.
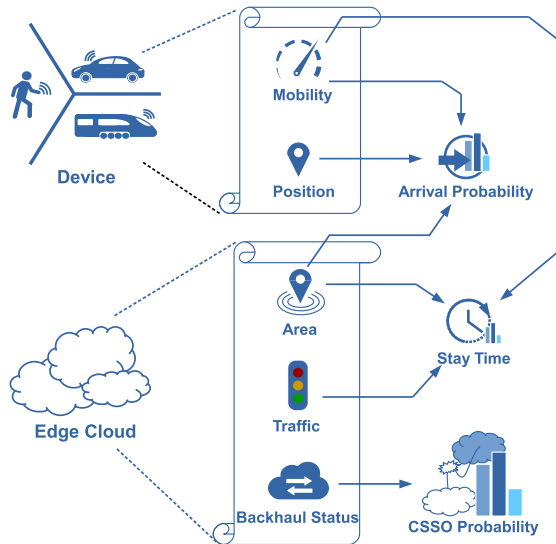


**FIGURE 3.** Trust Zone extracts the context information of UEs & EC to estimate the central security service outage probability.

More specifically, given a certain EC $C$, let $T$ denote the synchronization interval, the duration of $u$ suffering from CSSO in $C$ in the next period of $T$ is

$$E\{t_{o,u,C}\} = p_{o,C} \int_0^T f_{arr,u}(t) \int_0^{T-t} f_{stay,u}(\tau)\tau \,d\tau \,dt \quad (1)$$

where $p_{o,C}$ is the predicted probability of CSSO in $C$ over the next period $T$, $f_{arr,u,C}(t)$ denotes the probability density that $u$ arrives in the coverage of $C$ after $t$, and $f_{stay,u,C}(\tau)$ is the probability density function of $u$'s stay time $\tau$ in the coverage of $C$ upon arrival. The decision of synchronizing the subscriber authentication information of $u$ to $C$ can be then made following:

$$\begin{cases} \text{To synchronize} & E\{t_{o,u,C}\}/T \geq \mu; \\ \text{Not to synchronize} & E\{t_{o,u,C}\}/T < \mu, \end{cases} \quad (2)$$

where $\mu$ is a pre-defined threshold, so that less redundant backhaul traffic is generated. By setting $\mu$ to different levels, the MNO is able to balance its interests in operating cost and MEC service reliability.

Moreover, inspired by Han *et al.* [14], if the MNO is able to evaluate the loss of $u$ in suffering from CSSO per unit time, call it $l_u$, and the operations cost $c_{s,C}$ to synchronize the subscriber authentication information of a UE to $C$, there is a cost-optimal strategy that minimizes the sum of potential CSSO loss and operations cost:

$$\begin{cases} \text{To synchronize} & E\{t_{o,u,C}\} \cdot l \geq c_{s,C}; \\ \text{Not to synchronize} & E\{t_{o,u,C}\} \cdot l < c_{s,C}. \end{cases} \quad (3)$$

Therefore, we can individually solve the optimal threshold for every $C$:

$$\mu_{opt,C} = \frac{c_{s,C}}{lT}. \quad (4)$$

### B. MODELING THE BACKHAUL CONNECTION STATUS

Although mature models and techniques are available to estimate user mobility in mobile networks [24], which can be applied in the CSSO estimation procedure discussed above, a simple and efficient model for the backhaul status is still called for.

Generally, a backhaul connection between the central cloud and the edge cloud can be evaluated with its average delay and packet error rate, which are easy to measure. With these metrics, we are able to estimate the probability that the central security server responses in time to support a certain edge computing service, when given its specific latency requirement. We define this probability as the Central Security Service Reliability (CSSR), which exhibits stochastic behavior in long term and can be modeled as a Markov chain according to [25], as shown in Fig. 4.

In our model, the backhaul network at any time instant is considered to be in one among several predefined states. Every state describes a specific combination of CSSR range and network state. According to the CSSR level, we defined three basic network states: *healthy*, *unhealthy* and *disconnected*. Additionally, the network operator may recognize network disasters and take necessary measures to repair the backhaul connection, for which we set an extra state *under recovery*.
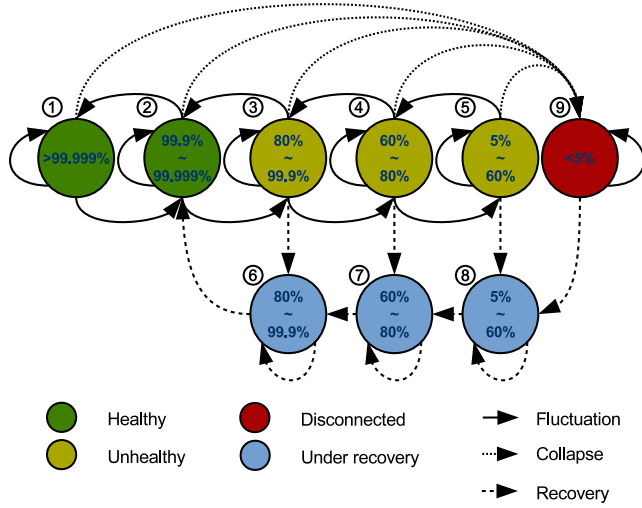
**FIGURE 4.** Central security service reliability as a Markov chain, circled numbers are state indexes.

Usually, the CSSR does not vary fiercely but smoothly, so only fluctuations between neighbor CSSR levels are considered, as we illustrate with solid arrows in Fig. 4. Nevertheless, as mentioned in Sec. II-B, some disasters and cyber-attacks, although rarely happen, can cause sudden disconnection of backhaul. In this case, the backhaul connection will fail to autonomously recover by itself, unless manual repair is taken. Therefore, in the model we enabled unidirectional transitions from the *healthy* and *unhealthy* states to the *disconnected* state, which are illustrated with dotted arrows in Fig. 4. Furthermore, the network operator will take autonomous or manual measures to recover the backhaul network only when it is not healthy, and the recovery process will continuously improve the CSSR until it becomes healthy again, as denoted with the dashed arrows in Fig. 4. By training the model with historical log data of the network operator, we are able to estimate the transition probabilities and predict the future CSSR. A demonstrative example output of our Markov chain is depicted in Fig. 5, which is generated from an artificial state transition probability matrix $\mathbf{T}$ of

$$
\begin{bmatrix}
0.8 & 0.1999 & 0 & 0 & 0 & 0 & 0 & 0 & 0.0001 \\
0.5 & 0.49 & 0.0099 & 0 & 0 & 0 & 0 & 0 & 0.0001 \\
0 & 0.25 & 0.5 & 0.15 & 0 & 0.0999 & 0 & 0 & 0.0001 \\
0 & 0 & 0.25 & 0.5 & 0.15 & 0 & 0.0999 & 0 & 0.0001 \\
0 & 0 & 0 & 0.2 & 0.6 & 0 & 0 & 0.1 & 0.1 \\
0 & 0.5 & 0 & 0 & 0 & 0.5 & 0 & 0 & 0 \\
0 & 0 & 0 & 0 & 0 & 0.5 & 0.5 & 0 & 0 \\
0 & 0 & 0 & 0 & 0 & 0 & 0.5 & 0.5 & 0 \\
0 & 0 & 0 & 0 & 0 & 0 & 0 & 0.3 & 0.7
\end{bmatrix}.
$$

Here, $T_{i,j}$ denotes the probability of transition from the state $i$ to $j$, for all $\{i, j\} \in \{1, 2, \ldots, 9\}^2$.

## V. NUMERICAL SIMULATION

We conduct numerical simulation to validate the effectiveness of the TZ approach, as to describe below.
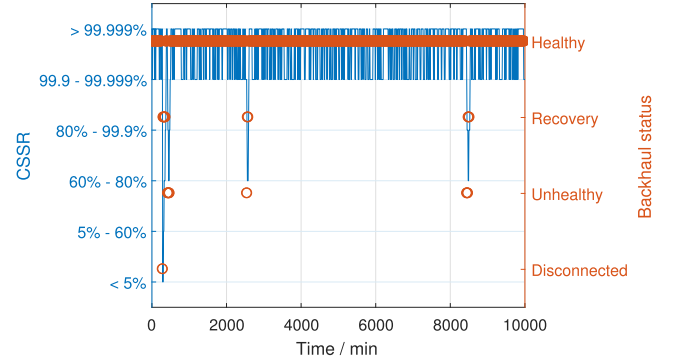


**FIGURE 5.** An example simulation of CSSR generated by our Markov chain (the blue curve), and the corresponding network state (the red circles).
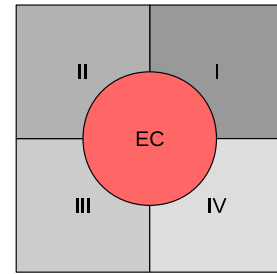


**FIGURE 6.** Geographical map of the simulated region, EC is the coverage of the local EC, while I – IV are neighbor suburban and rural areas.

**TABLE 2.** Basic speed of different UE mobility classes (in m/s).

| Mobility Class | STILL | LOW | MEDIUM | HIGH |
|---|---|---|---|---|
| Basic Speed | 0 | $\sim \mathcal{N}(1.5, 0.5)$ | $\sim \mathcal{N}(10, 2)$ | $\sim \mathcal{N}(40, 5)$ |

### A. ENVIRONMENT SETUP

An $8 \times 8 \text{ km}^2$ region with UE density of 6250/km² is considered. An EC covers a circular urban area with radius of 2 km that locates in the center of the region, which is surrounded by four suburban and rural areas, as illustrated in Fig. 6.

The simulation initializes with number of UEs that are uniformly distributed across the region. Also, every UE is assigned to a mobility class that is randomly selected from *HIGH*, *MEDIUM*, *LOW* and *STILL* with equal probabilities. To generate the mobility of individual users, a variety of models have been proposed that can be classified into four categories, namely the random walk model, the random waypoint model, the fluid flow model and Gauss-Markov model, respectively [24]. Considering the dependency of user mobility on the terrain and traffic environment, here we have extended the random walk model with 600 s steps, where a Gaussian distributed "basic speed" is generated for every moving UE according to its mobility class. This basic speed is then linearly scaled according to both the mobility class and the area where the UE is located, and combined with an uniformly distributed random direction, in order to obtain the UE's movement. Details are listed in Tabs. 2 and 3. Once a UE leaves the region, a new arriving UE will be created to keep a constant overall UE density. We let the UEs move for 40 h so

**TABLE 3.** Speed scaling factors for different mobility classes and areas, still UEs omitted as they do not move.

| Mobility Class \ Area | EC | I | II | III | IV |
|---|---|---|---|---|---|
| *LOW* | 1 | 1 | 1 | 1 | 1 |
| *MEDIUM* | 0.7 | 1 | 0.9 | 0.8 | 0.9 |
| *HIGH* | 0.2 | 1 | 0.9 | 0.85 | 0.8 |
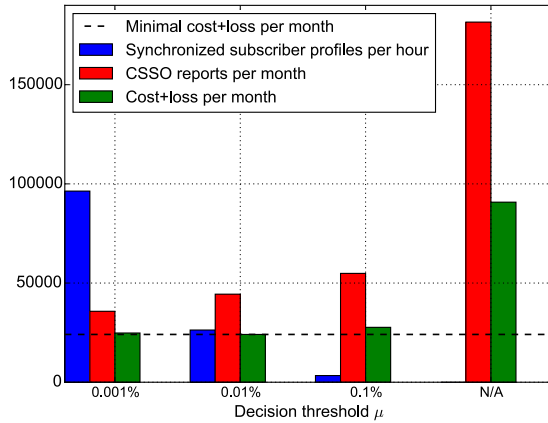


**FIGURE 7.** The performance of TZ under different specifications of decision threshold. Note that N/A stands for no TZ deployment, which is the baseline of LTE systems.

that the UE densities in all five areas converge to consistent levels.

Meanwhile, we take the Markov chain described in Sec. IV-B to generate the CSSR of unstable central security services in the EC as a random process, which is updated by every simulation step. In every simulation step, depending on the current CSSR, there is a specific chance that a CSSO occurs in the EC.

### B. TESTING THE TRUST ZONE

We consider a TZ implemented in the EC serving an urban area. It is assumed to have complete knowledge about the Markov chain parameters of the EC's CSSR. After the convergence of UE densities in different areas, the TZ tracks the positions of UEs for 24 h to learn about the statistical motion patterns, in order to predict the arrival probability and stay time of every UE in the EC. After the training phase, the TZ is activated to estimate the CSSO risk criterion $\mathrm{E}\{t_{\mathrm{o},u,C}\}/T$ in the next $T = 30$ min for every UE $u$ in the region. If and only if the risk exceeds a pre-defined threshold, the UE's subscriber profile will be synchronized from the central cloud to enable TZ operation upon CSSO, and a backhaul traffic will be thereby generated. This operation repeats every 30 min. By every simulation step, if a CSSO occurs in the EC, every UE that has not had its subscriber profile synchronized will suffer from the CSSO once, and therewith generate a CSSO report. To demonstrate the potential of cost optimization indicated by (4), we assume a normalized loss rate $l = 1/\mathrm{hour}$ and $c_{\mathrm{s},C} = 5 \times 10^{-5}$ so that $\mu_{\mathrm{opt},C} = 1 \times 10^{-4}$. This testing phase continues one month, and the performance of TZ under different specifications of decision threshold are illustrated in Fig. 7.

It can be observed that the deployment of TZ can effectively reduce the loss of user experience caused by CSSOs. Especially, by configuring the decision threshold to different levels, we can achieve a flexible balance between the service reliability and the operations cost. Especially, the sum of CSSO-caused loss and operations cost is minimized at the analytical optimum of decision threshold.

### VI. CONCLUSION AND FURTHER CHALLENGES

In this paper, we have indicated the future 5G telecommunication networks with dense application of MEC services, and the current LTE/EPS security architecture in 3GPP Release 14 drawbacks with its centralized AKA mechanism. We proposed a solution integrated with our 5G AAA approach which enables the execution of AKA functionality at the ECs, in order to support better deployment of 5G MEC services. An innovative EC security architecture, the Trust Zone, has been introduced to enhance the 5G AAA with a cognitive access management with respect to the backhaul connection quality. Furthermore, we designed a context-aware mechanism of synchronizing subscriber authentication information to local subscriber databases, which helps in reducing the backhaul network traffic generated by the Trust Zone. Numerical simulation showed that our proposed method can effectively improve the reliability of EC services in the context of unstable security network functions in central cloud, while being capable to keep a flexible balance between the MEC reliability and the operations cost.

We will continue to investigate better methods to improve the authentication process in 5G. It shall be noted that some assumptions and simplifications, such as the uniform distribution of UE and the extended random walk model, can be inaccurate in practical use cases. Advanced models such as the work reported in [26] will be applied to simulate more realistic user mobility. In terms of increasing the service reliability, we will apply machine learning and deep learning techniques to build precise models from field measurements of CSSR, in order to improve the accuracy and efficiency of centralized security. Furthermore, the proposed decentralized security mechanisms need to be evaluated with respect to their applicability of local authentication and authorization in hybrid private-public mobile networks as expected for 5G, e.g. 5G enterprise networks inter-working with 5G public land mobile networks.

### REFERENCES

[1] S. Barbarossa, S. Sardellitti, and P. D. Lorenzo, "Communicating while computing: Distributed mobile cloud computing over 5G heterogeneous networks," *IEEE Signal Process. Mag.*, vol. 31, no. 6, pp. 45–55, Nov. 2014.

[2] D. Sabella, A. Vaillant, P. Kuure, U. Rauschenbach, and F. Giust, "Mobile-edge computing architecture: The role of MEC in the Internet of Things," *IEEE Consum. Electron. Mag.*, vol. 5, no. 4, pp. 84–91, Oct. 2016.

[3] P. Corcoran and S. K. Datta, "Mobile-edge computing and the Internet of Things for consumers: Extending cloud computing and services to the edge of the network," *IEEE Consum. Electron. Mag.*, vol. 5, no. 4, pp. 73–74, Oct. 2016.

[4] Y. Mao, C. You, J. Zhang, K. Huang, and K. B. Lataief, "A survey on mobile edge computing: The communication perspective," *IEEE Commun. Surveys Tuts.*, vol. 19, no. 4, pp. 2322–2358, 4th Quart., 2017.

[5] N. Abbas, Y. Zhang, A. Taherkordi, and T. Skeie, "Mobile edge computing: A survey," *IEEE Internet Things J.*, vol. 5, no. 1, pp. 450–465, Feb. 2018.

[6] S. Lal, T. Taleb, and A. Dutta, "NFV: Security threats and best practices," *IEEE Commun. Mag.*, vol. 55, no. 8, pp. 211–217, Aug. 2017.

[7] H. Li, R. Lu, J. Misic, and M. Mahmoud, "Security and privacy of connected vehicular cloud computing," *IEEE Netw.*, vol. 32, no. 3, pp. 4–6, May/Jun. 2018.

[8] A. Klein, C. Mannweiler, J. Schneider, and H. D. Schotten, "Access schemes for mobile cloud computing," in *Proc. 11th Int. Conf. Mobile Data Manage.*, Kansas City, MO, USA, May 2010, pp. 387–392.

[9] *3GPP System Architecture Evolution (SAE); Security Architecture (Release 15)*, document 3GPP TS 33.401 v15.0.0, 3GPP Technical Specification Group Services and System Aspects, Jun. 2017.

[10] *3GPP System Architecture Evolution (SAE); Security Aspects of Non-3GPP Accesses (Release 14)*, 3GPP Technical Specification Group Services and System Aspects, document 3GPP TS 33.402 v14.2.0, Jun. 2017.

[11] I. L. Pavon *et al.*, "5G NORMA network architecture—Intermediate report," EU H2020 5G NORMA, Tech. Rep. Deliverable D3.2, Jan. 2017.

[12] M. Patel *et al.*, "Mobile-edge computing—Introductory technical white paper," ETSI, Sophia Antipolis, France, Tech. Rep. Issue 1, Sep. 2014.

[13] *Deliverable 1.5: Updated Scenarios, Requirements and KPIs for 5G Mobile and Wireless System With Recommendations for Future Investigations*, document ICT-317669-METIS, Mobile and Wireless Communications Enablers for the Twenty-Twenty Information Society (METIS), 2015.

[14] B. Han, M. R. Crippa, and H. D. Schotten, "5G island for network resilience and autonomous failsafe operations," in *Proc. Eur. Conf. Netw. Commun.*, Ljubljana, Slovenia, Jun. 2018.

[15] J. Kochems and H. D. Schotten, "AMMCOA—Nomadic 5G private networks," in *Proc. 23rd VDE/ITG Fachtagung Mobilkommunikation)*, Osnabrück, Germany, May 2018, pp. 105–108.

[16] M. Simsek, D. Zhang, D. Öhmann, M. Matthé, and G. Fettweis, "On the flexibility and autonomy of 5G wireless networks," *IEEE Access*, vol. 5, pp. 22823–22835, 2017.

[17] H. Yang, N. Wakamiya, M. Murata, T. Iwai, and S. Yamano, "An autonomous and distributed mobility management scheme in mobile core networks," in *Proc. 9th EAI Int. Conf. Bio-Inspired Inf. Commun. Technol.*, 2016, pp. 35–42.

[18] S. Jeon, S. Figueiredo, R. L. Aguiar, and H. Choo, "Distributed mobility management for the future mobile networks: A comprehensive analysis of key design options," *IEEE Access* vol. 5, pp. 11423–11436, 2017.

[19] S. Wong, N. Sastry, O. Holland, V. Friderikos, M. Dohler, and H. Aghvami, "Virtualized authentication, authorization and accounting (V-AAA) in 5G networks," in *Proc. IEEE Conf. Standards Commun. Netw. (CSCN)*, Sep. 2017, pp. 175–180.

[20] *System Architecture for the 5G System; Stage 2 (Release 15)*, document 3GPP TS 23.501 V1.5.0, 3GPP Technical Specification Group Services and System Aspects, Nov. 2017.

[21] "5G PPP architecture working group, view on 5G architecture," White Paper, Version 2.0, Jul. 2016.

[22] B. Han, S. Wong, C. Mannweiler, M. Dohler, and H. D. Schotten, "Security trust zone in 5G networks," in *Proc. 24th Int. Conf. Telecommun. (ICT)*, May 2017, pp. 1–5.

[23] A. Checko *et al.*, "Cloud RAN for mobile networks—A technology overview," *IEEE Commun. Surveys Tuts.*, vol. 17, no. 1, pp. 405–426, 4th Quart., 2015.

[24] X. Ge, J. Ye, Y. Yang, and Q. Li, "User mobility evaluation for 5G small cell networks based on individual mobility model," *IEEE J. Sel. Areas Commun.*, vol. 34, no. 3, pp. 528–541, Mar. 2016.

[25] P. Pukite and J. Pukite, *Markov Modeling for Reliability Analysis.* Hoboken, NJ, USA: Wiley, 1998.

[26] J. Ye, X. Ge, G. Mao, and Y. Zhong, "5G ultradense networks with nonuniform distributed users," *IEEE Trans. Veh. Technol.*, vol. 67, no. 3, pp. 2660–2670, Mar. 2018.

**BIN HAN** received the B.E. degree from Shanghai Jiao Tong University, China, in 2009, the M.Sc. degree from Technische Universität Darmstadt, Germany, in 2012, and the Ph.D. (Dr.-Ing.) degree from the Kalsruhe Institute of Technology, Germany, in 2016. Since 2016, he has been with the Institute of Wireless Communication, Technische Universität Kaiserslautern, where he is currently a Senior Lecturer. His research interests include the broad area of communication systems and signal processing, with a recent special focus on multi-access edge computing and the fifth-generation network slicing.

**STAN WONG** received the B.Eng. and M.Sc. degrees from the King's College London (KCL) and the Ph.D. degree from the University of London, in 2009. In 2014, he joined an IT consultancy firm, Hong Kong, as a System Consultant, where he is responsible for metro Ethernet 2.0 network infrastructure design. In 2015, he rejoined KCL, where he focused on the EU-funded project covering the fifth-generation security. He has deep knowledge in fifth-generation security, machine learning data analytic, and different standard Internet of Things air interfaces and application. He received the Collaborative Spectrum Sharing Prize from the European Commission, in 2016.

**CHRISTIAN MANNWEILER** received the M.Sc. (Dipl.-Wirtsch.-Ing.) and Ph.D. (Dr.-Ing.) degrees from Technische Universität Kaiserslautern, Germany, in 2008 and 2014, respectively. Since 2015, he has been a member of the Cognitive Network Management Research Group, Nokia Bell Labs, where he focuses on network management automation and SON for the fifth-generation systems. He has authored or co-authored numerous articles and papers on wireless communication technologies and architectures for future mobile networks. He is involved in several nationally and EU-funded projects covering the development of cellular and industrial communication systems.

**MARCOS RATES CRIPPA** received the B.Sc. degree in computer science from the Federal University of Rio Grande do Sul, Brazil, in 2010, and the M.Sc. degree from Technische Universität Kaiserslautern, Germany, in 2013. Since 2013, he has been a Research Associate with the Institute of Wireless Communication, Technische Universität Kaiserslautern. His main interests include network function virtualization and cellular networks architecture. He is involved in German and EU-funded project in the areas of spectrum sharing, software-defined networks, network function virtualization, and mobile network architectures.

**HANS D. SCHOTTEN** received the Diploma and Ph.D. degrees in electrical engineering from the Aachen University of Technology RWTH, Germany, in 1990 and 1997, respectively. Since 2007, he has been a Full Professor and the Head of the Institute of Wireless Communication, Technische Universität Kaiserslautern. Since 2012, he has been the Scientific Director of the German Research Center for Artificial Intelligence, where he is also the Head of the Intelligent Networks Department.

● ● ●