Untangling IoT Global Connectivity: The Importance of Mobile Signaling Traffic

Untangling IoT Global Connectivity: The Importance of Mobile Signaling Traffic

Stefan Geißler*, Andra Lutu‡, Florian Wamser*, Thomas Favale§, Viktoria Vomhoff*,
Michael Krolikowski¶, Marco Mellia§, Diego Perino‡, Tobias Hoßfeld*

*University of Würzburg, Germany
‡Telefonica, Spain
§Politecnico di Torino, Italy
¶Emnify GmbH, Germany

***Abstract*—IoT plays an important role in cellular networks, and its need for global connectivity is driving the rise of Global IoT Providers. These provide service by aggregating multiple mobile providers through roaming, complicating the understanding of the overall mobile ecosystem. This calls for lightweight monitoring solutions, which are crucial to meet the quality demanded by IoT services, and of automatic means to analyze the data, with the final goal to carry out economic and management activities. This paper provides insights from the study of two commercial, widespread IoT providers. We show how monitoring signaling traffic between mobile networks offers a unique opportunity to understand both the IoT customers' characteristics and the network functioning. Leveraging clustering, we offer the first data-driven methodology to examine large IoT signaling datasets. By analyzing over 1.3 billion signaling dialogues across two providers, we identify common signaling profiles that depend on the specific IoT vertical, likely misconfigured devices, and sudden changes that indicate potential problems. This provides actionable insights for network management decisions and service improvements, and lays the groundwork for future research on IoT traffic modeling.**

***Index Terms*—IoT, Mobile Networks, Signaling Traffic, Traffic Modeling**

I. Introduction

Recently, we have witnessed a surge in demand for global, ubiquitous cellular connectivity. Most of this comes from the massive number of connected Internet of Things (IoT) devices like smart meters, alarms, elevators, and fleet tracking devices, all of which rely on low-throughput, energy-constrained connectivity. While end-user services are currently migrating to 5G solutions, most massive IoT verticals still to this day rely on 2G or 3G connectivity [1], due to cost constraints and the long lifetime of devices.

With the surge in demand, commercial opportunities are flourishing, and a new breed of service providers – commonly called *Managed IoT Connectivity Providers* [2], or simply IoT Providers – started offering seamless, worldwide coverage and connectivity for IoT services. Similar to Mobile Virtual Network Operators (MVNOs) [3], [4], IoT providers rely on the cellular infrastructure of multiple Mobile Network Operators (MNOs) to realize their service. There are two main emerging models: *IoT Platforms* and *Dedicated IoT Operators*. IoT Platforms, on the one hand, are managed services offered by a telephone company, relying on its main base MNO, its core functions, and roaming agreements. Dedicated IoT Operators, on the other hand, manage their own core network functions while aggregating multiple MNOs worldwide via cellular roaming provided by international roaming hubs [5]. Both operational models thus overload the international mobility function that the cellular ecosystem offers to end-user devices [6]. This raises the challenge of managing the relationship between the visited MNO (that owns the local radio network), the IoT vertical (that uses the radio resource, but depends on a different core network than the IoT Provider offers), and the roaming interconnection hub. The division of managed IoT connectivity service across different domains within the mobile ecosystem makes network management, device-level monitoring and anomaly detection a challenging, still unsolved problem.

In this paper, we show that monitoring signaling traffic between the radio network provider (i.e., the visited MNO) and the core network operator (i.e., the home MNO that the IoT Provider uses or operates) can provide fine-grained insights into system health and device behavior. We build *two unique datasets*, one for each of the two different operational IoT provider models: an IoT platform based in Spain, and a dedicated IoT operator based in Germany (Section IV). These two datasets include the control traffic that flows through the corresponding roaming hubs these providers use, and allow us to identify and characterize per-device communication patterns. Though the two IoT providers support different customers in different parts of the world, we show that the emerging signaling patterns from the different populations of devices are similar. The insights gained in this work form the basic understanding required to inform the design of anomaly detection approaches that are valid for different entities, regardless of their customer base.

Monitoring signaling traffic allows for a much less intrusive view than monitoring application traffic. In mobile networks in particular, complex and diverse protocols let devices connect to the radio network first, and then establish the data communication channel over which application traffic is carried in an encrypted fashion. Visibility is thus much more limited compared to passive measurements in the traditional Internet. In the past, researchers have provided simple aggregated analyses of such

signaling traffic for system-wide characterization [5], [7], [8]. We deepen this analysis in two ways. First, we define a rich and detailed set of features to describe the signaling patterns any IoT device produces. Second, we engineer an unsupervised machine learning approach to identify homogeneous groups of IoT devices and applications that exhibit similar signaling characteristics. These groups give significant insights into network usage and add great value to in-network management, commercialization approaches, and issues IoT connectivity providers face.

Specifically, we make the following contributions.

- We propose features to describe the fine-grained signaling behavior of single devices, including overall statistics, device activity, message types, signaling patterns, mobility statistics, and longitudinal activity statistics (Section V).
- We design a clustering-based solution that enables common patterns to emerge, and simplifies the analysis of the data from the single device characterization to the identification of macroscopic patterns (Section VI).
- We show that these patterns highlight common threats that appear independently of the specific IoT vertical and IoT provider, as well as specific aspects and anomalies that offer the network administrator useful and actionable information to improve the service they offer.

Based on data from two real-world operators, we believe our study demonstrates the potential of analyzing signaling traffic to untangle the operations of mobile broadband networks in general. To allow other researchers to reproduce our work and continue exploring the analysis of signaling patterns, we provide the processed feature data we generated for the dedicated IoT operator that allows the reproduction of our clustering results.

The remainder of this paper is structured as follows. Section II provides a primer on the current global signaling landscape and mobile roaming. Related work regarding the characterization of devices is summarized in Section III. Section IV describes the two datasets obtained and evaluated in this work, before we detail the feature engineering process in Section V. Section VI details the results obtained from classifying IoT devices by evaluating the results obtained across the two independent datasets. We discuss the relevance of the findings made in this work for MNOs in Section VII. Finally, Section VIII concludes this work and outlines the need for additional research in the area.

II. Background and Motivation

In this section, we give an overview of the ecosystem that offers managed connectivity to the massive number of IoT devices operating worldwide with different connectivity requirements. Support for “things” operating globally has become critical for IoT verticals, from connected cars to wearables [1], [9], [10]. IoT verticals require deploying their devices world-wide, while keeping operational simplicity in terms of managing the connectivity of their devices and customers. IoT managed connectivity providers (such as IoT platforms or dedicated IoT operators such as Twilio, EMnify, or Truphone) answer to these needs by leveraging the roaming functions within the cellular ecosystem [1], [5].

Recent work [1] showed that in 2019 approximately 20% of the device population an operator connects represents inbound roaming devices, out of which at least 75% are IoT devices. Moreover, as many as 60% of these roaming IoT devices were still only 2G/3G-capable in 2019 (e.g., smart meters, fleet tracking), delaying the expected sunset of this legacy technology.

In this paper, we focus exclusively on the 2G/3G architecture (which we show in Figure 1), and the associated signaling traffic – as this is by far the dominant type of signaling originating from real-world operational IoT devices at the time of writing. Despite the rapid evolution in the Radio Access Network (RAN) over the past years, with 5G and next generation services now being deployed by vendors, recent studies [1], [5] showed that different IoT verticals still largely rely on 2G/3G legacy radio technologies, even as some operators pledge to retire these technologies in some countries. This is due to the low cost of the hardware that responds efficiently to the demands of massive IoT deployments, as well as the heterogeneous radio coverage penetration world-wide of more recent 4G or 5G technologies. To help the reader, we list all technical acronyms in the Glossary.

A. IoT Global Connectivity Models

In Figure 1, we show how different networks in the cellular ecosystem interconnect to offer managed connectivity to IoT verticals. There are two main models for this.

IoT Platforms lease the core network function services from (one or several) existing mobile network operators. For this reason, several MNOs (and their sibling telco providers) operate their own IoT platforms [1] that exploit their already operational cellular infrastructure. IoT platforms usually rely on a single (home) MNO to provision the global IoT SIM. They then leverage the international roaming function of the home MNO to provide IoT businesses with the global cellular connectivity they require.

Dedicated IoT Operators (e.g., Twilio, EMnify) choose to run (some) network core functions in their own premises to be able to aggregate various arbitrary MNOs worldwide. They leverage a hybrid way of operating, where they rely on mobile roaming to use the radio networks of MNOs, while controlling the core network functions required to manage, track, and bill the global presence of their users. In this case, the IoT connectivity provider builds a global footprint through the use of one or several “roaming hubs” or IP Packet Exchange (IPX) providers (Figure 1).

B. The Role of Roaming Hubs

Roaming hubs seamlessly enable IoT providers to expand their international footprint. With a single agreement with one roaming hub, IoT providers (represented in Figure 1 by the ”Home Network”) can deploy connectivity services by configuring network routing to a designated point of a public mobile network, regardless of how many roaming partners they actually contract with (i.e., the ”Visited Network” in Figure 1). The roaming hub peers (via private interconnects or public peering) with other hubs to extend their footprint

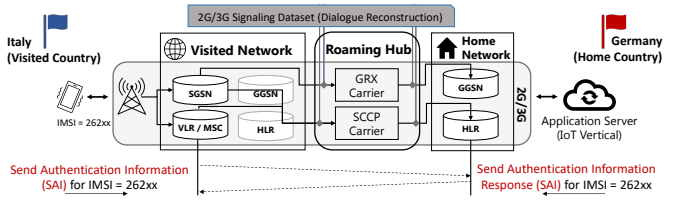


Fig. 1: Overview of the IoT cellular ecosystem and the role of roaming for global IoT connectivity. We show the interfaces we monitor to build the dataset of signaling traffic (MAP and GTP signaling dialogues). In the lower part, we show an example of MAP signaling dialogue between a VLR in Italy and an HLR in Germany to authenticate the IoT device with the IMSI 262xx.

worldwide, and forming the IPX Network – a private network, separate from the public Internet, that meshes together the infrastructures of the hubs [6]. The resulting IPX network enables the transport of global roaming data between mobile networks, with inter-operability of different implementations and standards.

Some international carriers (e.g., Syniverse, BICS) also operate as major roaming hubs in the IoT ecosystem, providing easy access to all MNOs that connect to their roaming hub. By connecting to one (or multiple, for redundancy) roaming hub, the IoT provider gains access to hundreds of MNOs worldwide at the same time.

C. Signaling in Mobile Networks

After interconnecting the IoT provider (i.e., Home Network in Figure 1) to roaming partners (i.e., Visited Network in Figure 1) via the roaming hub, the cellular ecosystem allows IoT devices to establish radio and data connectivity.

In this paper, we focus our analysis exclusively on the 2G/3G architecture (which we show in Figure 1), and we collect statistics regarding signaling dialogues between the Visitor Location Registry (VLR) and Home Location Registry (HLR), as well as between Serving General Packet Radio Service (GPRS) Support Node (SGSN) and Gateway General Packet Radio Service (GPRS) Support Node (GGSN) regarding the activity of the roaming IoT devices. The Signaling Connection Control Part (SCCP) carrier function lets IoT devices attach to the 2G/3G radio access network, while the General Packet Radio Service (GPRS) Roaming eXchange (GRX) function lets them establish the data communications.

Table I contains a summary of all Mobile Application Part (MAP) and GPRS Tunneling Protocol (GTP) signaling messages that are relevant to the analyses conducted in this work.

1 - Network Access and Device Authentication. Signaling between the VLR and HLR functions enables the authentication of the IoT devices when they connect to a network in the visited country, thus allowing them to access the radio resources of the visited network. We show an example in Figure 1. When an IoT device with a German SIM card visits Italy, it first tries to connect to a locally available radio network. This prompts the VLR to authenticate the device by querying the corresponding HLR of the home network with a Send Authentication Information (SAI) request for the International Mobile Subscriber Identity (IMSI) of the roaming SIM. The HLR then replies with a SAI Response to allow/deny the request. Other examples of dialogues between the VLR and HLR functions include Update Location (UL) or Cancel Location (CL) procedures, and are further specified in the corresponding standards [11]. Further, similar messages are exchanged between the SGSN and the HLR (cf. Table I).

2 - Data Channel Setup. Once a device has gained access to the mobile network, it eventually establishes data tunnels to use for Internet data connectivity. This is the result of a signaling dialogue between the visited network SGSN and the home network GGSN before IP packets can be routed towards the open Internet using an encrypted GTP tunnel.

Dataset. Our datasets contain millions of 2G/3G signaling dialogues for real-world IoT devices operating across the world, corresponding to different verticals. We capture two large-scale datasets, one for each operational IoT provider we introduced in Section II-A. Though each IoT provider serves different IoT customers, they all rely on the capability of the supporting MNOs to use the same signaling protocols in order to enable the IoT devices to connect globally. Our work focuses on analyzing the patterns in these dialogues, and look for emerging trends that transcend the immediate characteristics of the IoT vertical application.

III. Related Work

IoT device characterization and classification has been investigated with different approaches at different network locations. Table II provides a summary of the research discussed in this section. Similar works leverage information available from the data plane in home network or testbed scenarios, such as MAC addresses, listening services, ports and protocols, application layer responses or hostnames (e.g., [12], [13]). Similarly, other approaches use detailed statistics to perform the classification since packet level information or WiFi probes are typically available for local networks (e.g., [14], [15], [8], [16]). In contrast, another line of work uses statistics available at higher network levels (e.g., ISP, Internet Exchange Point, DNS servers) such as DNS or TLS-level information and remote IP/port (e.g., [17], [18], [19]). Here, we take a different approach, and are among the first to characterize IoT traffic from the point of view of IoT providers – not by leveraging the data packets – but the signaling information, which offers a very different set of information than prior work.

Standardization bodies and working groups have been defining both network structure and services for IoT platforms [20], [21], [22], [23]. Considering mobile networks, two trends currently coexist, pushing towards re-purposing 2G/3G to serve IoT, and adopting 4G/5G [24], [25], [26]. Differently from this literature, we take a data-driven approach, focusing on the technologies we see being used for IoT in live networks.

Regarding the scope of prior research, literature on cellular network traffic has mainly focused on either human-to-human communication, or machine-to-machine communication within a single MNO [27], [28], [29], [9]. Studies including mobile roaming also often focus on specific operators without global perspective [30], [31], [32]. All these works

TABLE I: Overview of 2G/3G signaling dialogues relevant to this work.

Dialogue	Abbr.	Protocol	Src.	Dest.	Description
sendAuthenticationInfo	SAI	MAP	VLR SGSN	HLR	SAI is sent to get encryption triplets so a visited network can authenticate the SIM card that tries to connect.
updateLocation	UL	MAP	VLR	HLR	UL is sent whenever a subscriber migrates to another VLR. This may for example occur due to mobility or switching operators.
updateGprsLocation	UL_GPRS	MAP	SGSN	HLR	UL_GPRS is sent whenever a subscriber migrates to another SGSN. This may for example occur due to mobility or switching operators.
cancelLocation	CL	MAP	HLR	VLR SGSN	CL is sent to the old VLR or SGSN after and UL has occurred to notify that the subscriber has migrated to a new location.
createPdpContext	PDP_CREATE	GTPv1	SGSN	GGSN	PDP_CREATE is sent to establish a new PDP tunnel for data transmission.
updatePdpContext	PDP_UPDATE	GTPv1	SGSN	GGSN	PDP_UPDATE is sent to update parameters of an existing PDP tunnel.
deletePdpContext	PDP_DELETE	GTPv1	SGSN	GGSN	PDP_DELETE is sent to close an existing PDP tunnel.

measure the "Internet data plane" of mobile deployments. Only recently, Lutu et al. [6], [5] analyze the IP eXchange (IPX) ecosystem and an operational IPX provider by measuring the "signaling plane" of 3G/4G networks. Moreover, these past efforts have mostly focused on roaming on traditional human communication, while the authors of [1] present the only characterization of an IoT platform, and analyze the traffic of different verticals.

In this work, for the first time to the best of our knowledge, we provide a detailed analysis of signaling traffic corresponding to two separate production, large-scale IoT deployments that serve devices all over the globe. Though the IoT verticals connecting to these IoT providers differ (both in terms of geographical placement and application), we show that similar signaling behavior emerges. We are thus able to generalize signaling traffic patterns for IoT devices by comparing results across these two independent providers.

IV. IoT Signaling Traffic

We provide a first characterization of the datasets we collect from two different providers serving IoT verticals with different approaches: an IoT platform and a dedicated IoT operator. Our work is unique in contributing measurements and operational insights on IoT signaling traffic from commercial live deployments of cellular core networks from two independent real-world systems. We argue that our analysis helps shed light on the complexities of the systems, and motivates the need for automatic processing pipelines to provide actionable findings to network administrators and network design architects.

A. Datasets

Though both providers operate systems based on the structure in Figure 1, and address the same global customer base (i.e., IoT verticals), we note that the underlying infrastructure and the general approach for providing services differ (cf. Section II).

In this work, we capture signaling traffic corresponding to the IoT devices each provider manages at their respective mobile core ingress. We identify each IoT device via its encrypted unique identifier and capture the traffic corresponding to two signaling services, namely SCCP signaling and GTP signaling.

Table III summarizes the two datasets we build in this work. The dataset for the IoT platform (DAT1) was captured during February 2021 and contains approximately 310 000 unique devices. The dataset for the dedicated IoT operator (DAT2) was collected in January 2020 and includes approximately 270 000 unique devices. In total, the dataset contains more than one billion signaling interactions that we analyze using Apache Spark. Note that each such interaction, called a dialogue, consists of one or multiple request-response pairs. Note further that due to the selected measurement points, as depicted in Figure 1, the datasets contain only information regarding the signaling plane, and no information about the lower layers (i.e. the physical layer) has been collected.

Figure 1 shows schematically where and on which interfaces we collect the data. For both providers, we capture the raw signaling traffic and process it to re-build the signaling dialogues between core network functions in the home network and those in the visited network whose RAN is used by devices. There are no ethical concerns regarding the datasets we analyze, as no identifying information is processed.

Considering SCCP signaling, we monitor the MAP protocol that supports end-user mobility and allows network elements (e.g., the HLR, VLR, Mobile-services Switching Centre (MSC)) to communicate. We collect traffic corresponding to the procedures of each device: i) authentication and security, and ii) location management (update/cancel location).

For the GTP-c, we monitor dialogues required to manage data tunnels between roaming partners. Here we capture the create, update, and delete Packet Data Protocol (PDP) context procedures that devices trigger before and after a data communication. Specifically, we generate one record for each create session exchange and retain basic information such as the tunnel ID, which enables mapping of create and delete procedures.

Due to the high amount of traffic and processing involved in collecting and creating these statistics, we only collect information for IoT platform customers connecting to Point of Presences (PoPs) in a few locations relevant for operations

TABLE II: Taxonomy of related work.

Ref	Scope	Setup	Information Used	Goal
[12]	83M home devices from 16M households	Home and lab deployment	MAC addresses, ports, protocols, application layer responses, hostnames	IoT device classification, security issues
[13]	285K unique devices	Campus deployment	MAC addresses from unencrypted 802.11 traffic	IoT device classification via MAC addresses
[14]	33 unique devices	Lab deployment	Packet-level traces	IoT device classification
[15]	81 unique devices deployed in the US and UK	Lab deployment	Packet-level traces	Identification of information exposure
[8]	28 unique devices	Lab deployment	Statistical attributes of packet level traces	IoT device classification
[16]	3 measurements with between 32 and 108 devices	Environment with static device number	WiFi probes	Crowd density estimation
[17]	254k unique devices	Campus deployment, IPX transit network	DNS or TLS-level information	IoT device identification
[18]	US-based ISP hosting more than 40M devices	ISP deployment	DNS information	IoT device identification
[19]	14 days of captures, appr. 1M devices	ISP and IXP deployment	Sparsely sampled flow headers	IoT device identification
[6], [5], [1]	21 days of captures, approx. 140M devices	MNO and IPX deployment	Signaling information	Analysis of IP eXchange (IPX) ecosystem, IoT platform characterization, global perspective
This Work	Two datasets with 580k unique devices across two global mobile deployments	Ingress and egress of two commercial deployments	143 features based on signaling information	Detailed analysis of signaling traffic for IoT devices, generalization of signaling traffic patterns across different IoT providers

TABLE III: Dataset overview: managed connectivity for IoT devices.

Dataset	Infrastructure	Procedures captured	Scope
DAT1: IoT Platform	Telco-owned platform operating via international roaming, using the core of a Spanish MNO (the home MNO).	MAP signaling traffic; location management, authentication and security. GTP-c control data; Create/Delete PDP Context/Session, Timeout Procedure.	310k devices February 2021
DAT2: Dedicated IoT Operator	IoT-focused independent provider, deploying own cloud-based virtual mobile core, and aggregating RANs worldwide.		270k devices January 2020

(including Spain, US, Brazil, Argentina, Colombia, Peru, Costa Rica, Uruguay, Ecuador). Conversely, for the dedicated IoT operator we capture the entire IoT device customer base, located in 195 countries, with a focus on the US, Germany, and Mexico.

Naturally, due to the origins of the two independent datasets, there are certain limitations we need to keep in mind when working with the data.

Missing Information. Due to the source of both datasets being live, productive environments, information regarding detailed use cases of specific devices is usually not available as operators do not keep track of customers' applications within the aggregated traffic mix.

Independent Deployments. Although the evaluated datasets are obtained from IoT-focused platforms, the two systems operate independently and serve different customers. Each have their own bias due to the providers' operating models. While target customers operate in similar areas, differences regarding physically deployed devices need to be considered.

Geographical Bias. As a result of investigating two distinct platforms, their customer bases exhibit geographical tendencies that may impact the observed behavior. The dedicated IoT operator (DAT2) services devices deployed in 180 countries, with a strong bias towards the US and Germany. The traffic investigated for the IoT platform (DAT1) contains mostly devices deployed in Central and South America.

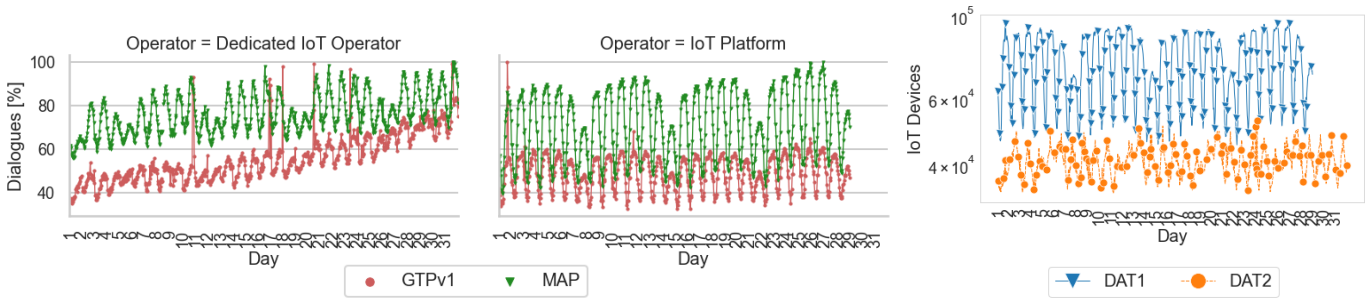
B. Signaling Timeseries

Figure 2a shows the timeseries of total number of dialogues per hour for both datasets.[1]

For both DAT1 and DAT2, we note that MAP dialogues account for a larger volume than GTP dialogues (Figure 2a). This is expected because the MAP protocol must allow the RAN operator to authenticate the end-user with the home network before setting up data communications.

Starting from MAP, we note that both providers show an increase in the volume of signaling messages at the end of the period of analysis. To check if this growth is due to an increase of active devices, Figure 2b reports the number of active devices per hour. We observe little to no increase here. Furthermore, we note that – even if the timeseries of the number of active devices follows weekly patterns for both providers – the corresponding trends differ. Namely, in the weekly cycle, the number of active IoT devices varies much more irregularly throughout the week for DAT2 than DAT1. Curiously, for DAT1, we note the clear drop during the weekend (especially on Sundays), with an otherwise stable number of active devices during the week. We conjecture that these differences stem from the disparity in IoT verticals and geographical regions that the two providers serve.

[1]We show normalized values to the corresponding maximum value per data feed, for confidentiality reasons.

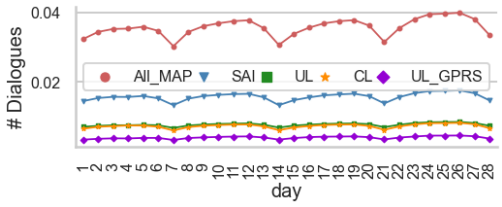


(a) Signaling dialogues per hour.

(b) Number of active devices per hour.

Fig. 2: Timeseries of aggregated MAP and GTP signaling traffic volume (dialogues) per hour (a), and number of active IoT devices per hour (b), for each of the two datasets (Table III).

Regarding the GTP signaling trend in Figure 2a, the two operators show once again different patterns. The IoT platform exhibits a consistent periodic trend with daily and weekly seasonality, while the dedicated IoT provider shows an increasing trend. This comes as an effect of the different manner in which these two providers operate. Specifically, the dedicated IoT provider enforces strict monthly data quota limits. These come into play toward the end of the month, where the number of failed create PDP context requests increases.



(a) MAP signaling traffic.

(b) GTP signaling traffic.

Fig. 3: MAP and GTP signaling breakdown per type of procedure. We show DAT1, results are consistent across datasets.

C. Signaling Procedures

Figure 3 shows the breakdown of dialogue types in a) MAP and b) GTP. We only show the timeseries for DAT1 in Figure 3, as the patterns are similar for DAT2. We show on the y-axis the fraction of a type of dialogues from the entire volume of dialogues across all dialogue types. Hence, a value of 0.01 means that 1% of all observed dialogs have occurred in the form of this respective dialogue type.

Specifically, we show the timeseries of signaling traffic for the SAI, UL, Update General Packet Radio Service (GPRS) Location (UL-GPRS), and CL dialogue types for MAP traffic. Overall, the MAP signaling accounts for 3-4% of the total volume of traffic signaling traffic in DAT1. The trend throughout the month follows the daily and weekly patterns we observe in the number of active devices in DAT1 (Figure 2b).

The IoT devices generate twice as many SAI dialogues compared to UL or CL dialogues. As expected, the majority of MAP dialogues (45%) are authentication requests (SAI procedures). Interestingly, the volumes of UL and CL dialogues are comparable, while the UL-GPRS corresponding to data communications are less than 1% of all signaling traffic.

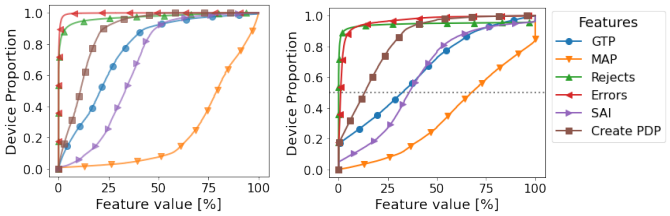
For GTP signaling traffic, we show the distribution of create PDP context, delete PDP context, and timeout messages. We note that most dialogues are requests for bringing up the PDP context for data communications (i.e., PDP_CREATE). These usually appear in pairs with the requests to tear down these tunnels (i.e., PDP_DELETE) once the data communication finalizes. Yet, due to possible failures, a terminal may issue more create than delete messages (e.g., to re-establish a tunnel once it has failed). We also observe a constant fraction of timeout messages, likely due to device or connectivity issues.

Takeaway: The complexity of the mobile IoT ecosystem is reflected in the signaling patterns that IoT devices generate. In particular, different IoT provider models exhibit different patterns. Both the IoT verticals and the different contract options (such as monthly quotas) affect signaling. Even if domain knowledge helps to explain the complexity and intertwining of the system, an automatic and detailed characterization of each IoT device would allow the network operator to better address the needs of its customers' applications.

V. Feature Engineering

To enable the automatic extraction of actionable information from raw data, we design a machine learning pipeline. Our goal is to automatically group devices that exhibit similar behavior. We obtain these by analyzing the combined MAP and GTP signaling traffic of each device.

In total, we include 143 features encompassing i) overall statistics, ii) message types and signaling patterns, iii) device activity, iv) mobility statistics, and v) longitudinal activity statistics. The features cover different levels and directions to describe devices as holistically as possible. In the following, we outline each feature group and highlight their respective importance in differentiating signaling traffic patterns. Table IV provides a concise overview of the identified feature groups, showing the number of features in each group as



(a) DAT1: IoT Platform. (b) DAT2: Dedicated IoT Operator.

Fig. 4: Empirical CDFs of selected features from overall statistics and message type statistics.

well as feature examples. For all features, we consider the period of time of interest (one month in our case), extract the timeseries of all records for each device and compute the statistical features. Our intent is to characterize the overall device behavior, and avoid sudden changes and anomalies that may occur on short timescales. For the sake of brevity, we provide the full list of features on GitHub.[2] For the interested reader, we offer a detailed listing with discussion of the features, which we will continuously update with the help of community feedback.

TABLE IV: Overview of identified feature groups.

Group	No.	Example Feature(s)
Overall Stats	7	Percentage of SIGTRAN and GTP messages, errors, and successes
Message Types	104	Contribution of individual message types
Device Activity	5	Percentage of time slots with activity
Mobility Stats	15	No. of operator changes, cell changes
Long. Activity	12	Inter-arrival times, mean context duration

A. Overall Statistics

These features aim at capturing the general signaling behavior of IoT devices. We include here i) the total number of signaling dialogues and ii) the percentage of MAP and GTP dialogues, breaking down by the number of successful, erroneous and rejected dialogues. Errors indicate technical inconsistencies such as missing replies, out of order packets, invalid packet contents. Rejects represent actively rejected signaling requests and do not indicate faulty behavior by devices, carriers, or the core network, but problems with the device Service Layer Agreement (SLA, e.g., misconfiguration, data quota violations, attempts to access forbidden MNO).

Intuitively, these features provide general information, and allow for macroscopic differentiation of devices (e.g., those that rely on circuit switched connectivity from those that use data connectivity). They also allow us to identify heavy hitters. Finally, the error and reject features enable us to identify anomalous and/or misconfigured devices that fail to complete successful signaling.

Figure 4 shows the empirical Cummulative Distribution Function (CDF) over all devices for the percentage of GTP, MAP, rejected, erroneous dialogues for each dataset. We generate these metrics over the entire period of analysis. In a nutshell, there is a significant variation of each of these features, with some exhibiting a mostly uniform trend, coupled with a portion of possibly anomalous devices (in the tail of some distributions). Note how the distributions differ between the two datasets, but exhibit similar trends.

[2]https://github.com/lsinfo3/IoT_clustering_features

B. Message Type and Session Statistics

To respect the devices' communication traffic flows, we evaluate what kind of signaling methods IoT devices trigger with their chronological sequence. We generate these features based on the observations of our exploration in Section IV-C. These features indicate what fraction of the signaling traffic corresponds to the MAP dialogues: SAI, UL, UL-GPRS, and CL; and the GTP-related create, update, and delete messages. We also differentiate between successful, erroneous and rejected dialogues, as before. We next assemble messages occurring in proximity to each other to identify *sessions*.

Figure 5 shows two *sessions* consisting of multiple dialogues each. A_S represents the interarrival time between two MAP or one MAP and one GTP dialog. D_{GTP} describes the PDP tunnel duration as determined by the time between a PDP_CREATE and PDP_DELETE.

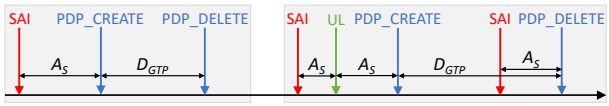


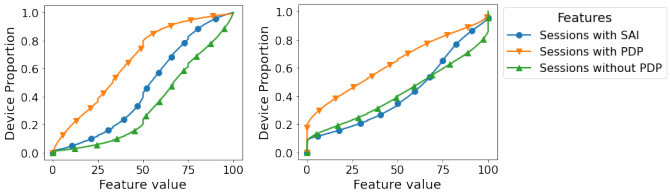
Fig. 5: Two examples of signaling sessions consisting of MAP and GTP dialogues.

The identification of each session requires domain knowledge. We rely on a set of rules developed during data analysis in combination with discussions with experts and the administrators of the dedicated IoT operator:

- MAP dialogues are attached to their successor if the interarrival time A_S is smaller than 30 seconds
- The PDP_CREATE, PDP_UPDATE and PDP_DELETE dialogues of a PDP tunnel always belong to the same session
- A PDP_DELETE dialogue always terminates the current session

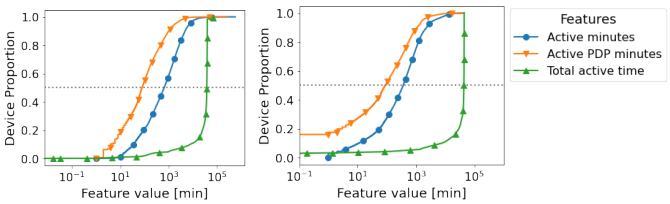
These sessions capture the activity of devices corresponding to specific network activity, e.g. a smart meter uploading hourly measurements to the application server, or a connected car retrieving navigation instructions upon a driver's request. For each IoT device, we compute features that indicate the number of unique dialogue sequences required to describe 90% and 99% of the sessions we derive from its data. These features allow further differentiation regarding both devices' overall signaling behavior with respect to which signaling messages are used, and how variable the signaling patterns of each device are by evaluating the number of uniquely observed signaling sessions.

Figure 6 shows the cumulative distribution of the percentage of signaling sessions per IoT devices during the period of analysis. We differentiate three types of sessions, namely



(a) DAT1 (b) DAT2

Fig. 6: Empirical CDFs of the percentage of session types per IoT device.



(a) DAT1. (b) DAT2.

Fig. 7: Empirical CDFs of device activity features.

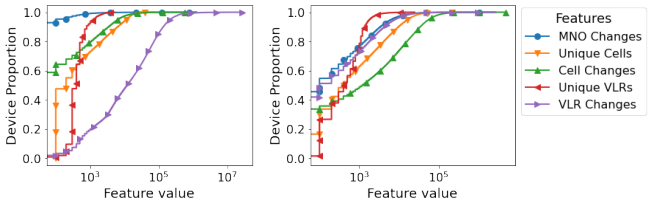
sessions with a SAI dialogue, sessions with PDP and without PDP signaling. We find that in DAT1, devices have on average 70% of sessions with SAI, which is similar in DAT2. However, in DAT2, the percentage of sessions with PDP signaling is larger than for DAT1.

C. Device Activity

We aim at capturing device activity levels, both when idle or in presence of an active data connection. For this, we include two metrics: the overall time in minutes between the first and last signaling dialogue observed in each of the datasets (i.e., the total active time), and the number of minutes with actual signaling activity (i.e., one signaling dialogue occurs in that respective one minute slot), regardless of the type of signaling (i.e., active minutes). For GTP-specific activity (i.e., active PDP minutes), we only count the slots with GTP-specific dialogues. We compute the relative activity as the ratio between the active minutes (i.e., minutes with any signaling activity) and the total active time. Analogously, we compute relative PDP activity.

These features act as an indicator of device verbosity. For instance, they let us differentiate between long-running devices that generate traffic continuously and devices with spurious activity that only trigger signaling traffic occasionally.

Figure 7 shows the CDFs for three activity features generated for each of the two datasets. Note that the total active time per device is consistent with the corresponding entire period of analysis for more than 90% of devices in each dataset that are alive for almost the whole time period (green curve). For both datasets, the active PDP time per device is on average one order of magnitude smaller than the active time. This is expected, since the MAP signaling accounts for a higher volume of traffic than GTP signaling. We also find that the difference on average between the active minutes and the active PDP minutes for DAT1 is higher than for DAT2.



(a) DAT1. (b) DAT2.

Fig. 8: Empirical CDFs of mobility statistics features.

This is a side effect of the way the two groups of devices operate on average (i.e., DAT1 devices have longer-lived and more seldom data connections than DAT2 devices).

D. Mobility Statistics

To characterize the mobility of IoT devices, we introduce features that capture the number of unique visited networks as well as the number of times devices switch between using specific mobile network entities (i.e., radio cells, visited MNOs, SGSNs, and VLRs in the visited MNOs).

These features allow us to differentiate between stationary and mobile devices, i.e., those exhibiting a small or large number of changes in mobility related values, respectively. In addition, we can distinguish devices that alternate between a few operators (e.g., due to coverage issues) from devices that visit a large number of different operators, indicating higher (international) mobility.

Figure 8 shows the empirical CDFs for the mobility features. In this case, the distribution of the number of MNO switches recorded by a single device is very different for the two datasets: In DAT1, more than 90% of devices show zero changes in the visited MNOs (blue curve), i.e., they always rely on the same visited MNO; in DAT2, this percentage is below 50%. In fact, approximately 20% of IoT devices register more than 1,000 MNO changes during the time of one month of analysis. This striking dissimilarity is due to (i) the different approaches of the two providers to operate their platform (i.e., the IoT platform honors their underlying MNO as the preferred roaming partners), and (ii) the mobility profiles of the IoT verticals relying on each provider (e.g., fleet tracking IoT solutions have different mobility patterns than smart meters). In terms of the number of radio cells devices use in DAT1, 60% of devices never change the radio cell to which they initially connect to. This percentage decreases to 35% in DAT2, hinting for a larger proportion of highly mobile devices.

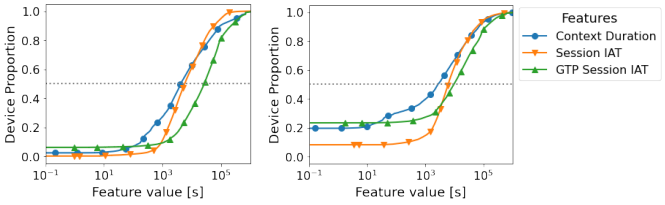
E. Longitudinal Activity Statistics

The last set of quantitative features encompasses information regarding various stochastic processes such as the interarrival times of sessions that we characterize by the mean and standard deviation for each of the following features:

- Interarrival time between sessions: we define separate features for the time between sessions containing GTP dialogs, sessions containing only MAP dialogues, and any two consecutive sessions, irrespective of session content.

- Interarrival time between dialogues within sessions, and separately between GTP update messages.
- Context duration, i.e., the time between PDP tunnel creation and deletion.

These features allow us to identify devices that exhibit low variability (e.g., indicating periodic behavior), or high variability (e.g., alarms or user triggered actions). Furthermore, the context duration enables us to separate the devices that establish many short-lived data tunnels from the ones that establish few long-lasting tunnels.



(a) DAT1 (b) DAT2

Fig. 9: Empirical CDFs of longitudinal features.

Figure 9 shows the empirical CDF of three longitudinal features. We show the CDFs of the average PDP context duration, the interarrival times between sessions (regardless the type of session, MAP or GTP) and, finally, the interarrival time only for GTP sessions. Looking at the two datasets, devices in DAT1 have longer interarrival times between GTP sessions than devices in DAT2. Also, the GTP session duration is longer for DAT1 devices than DAT2 devices. Indeed, in DAT2, GTP dialogues have the highest frequency among dialogue types, explaining the short interarrival times of the GTP sessions, and their short duration. We conjecture that this is a consequence either of the IoT device programming for the specific vertical (some device may require periodic and frequent data communication with the application server), or of the configuration of the IoT provider itself (e.g., the provider has a maximum number of data connection that it can sustain for each gateway).

For DAT1, we note that there is an order of magnitude difference between the interarrival time of all sessions (GTP or MAP sessions) and the interarrival time for GTP sessions, in particular. This is consistent with the fact that MAP dialogues have higher frequency than GTP dialogues for devices in DAT1. Again, this might be a direct consequence of the application these devices serve (e.g., energy smart meters only trigger one data connectivity per day to upload their measurements, while vehicle tracking devices require much more frequent connections).

Takeaway. The characterization of IoT devices requires multiple features that project each device in a multidimensional space. We design 143 features that cover different aspects and dimensions to describe devices as holistically as possible. We expect similar devices to exhibit similar features. This calls for an automatic approach to highlight clusters of devices exhibiting similar patters, which can explain major signaling patterns in the data and let operators understand the footprint of each group of devices.

VI. Device Classification

In Section V, we have shown that the two deployments present similar but different characteristics. This is due to a multitude of factors, e.g., different data connectivity, different IoT verticals the two providers serve, and different business models and systems the operators run. We now investigate if it is possible to identify common behaviors, e.g., groups of devices that exhibit similar signaling traffic patterns. Next, we are interested if there exist generic patterns that emerge in both deployments.

For this, we rely on unsupervised machine learning approaches. First, we generate the full set of features for each device in each dataset. Then we run dimensionality reduction with Principal Component Analysis (PCA) on the feature set. This significantly reduces the dimensionality of the feature space to a number of principal components that retain at least 80% of the variability of the initial input data. We next perform k-means clustering on this reduced feature set. We identify $k = 8$ as the optimal number of clusters based on the Dunn index [33] in both cases. Similarly, the Silhouette-Score suggests to use k between 7 and 8. For consistency, we opt to set $k = 8$. All operations are performed using the R statistical language (machine learning) and Apache Spark (feature extraction).[3]

A. Clustering results

Following the same methodology on DAT1 and DAT2, we separately extract eight main clusters that represent the eight main signaling profiles. Interestingly, we verify that these eight profiles are consistent for both providers, and that they group devices with similar communication patterns (e.g., high mobility devices, stationary devices such as smart sensors). In the following, we analyze the results.

1) DAT1 Profiles – IoT Platform: We provide an overview of the clustering results for the IoT platform in Table V. We show the mean values for the subset of features with the highest feature importance to distinguish and characterize clusters from one another. For this, we consider four groups of characteristics, which we map to the sets of features we previously introduced (Section V). The first group includes overall statistics that allow us to distinguish between the use of pure phone connections or phone plus data connections (i.e., percentage of GTP/MAP Dialogues). The second group includes overall statistics regarding the errors and rejects in the signaling procedures. The third group includes features related to the mobility of devices. The fourth group includes message type statistics, which breaks down the use of specific signaling procedures. For example, some clusters only contain devices that are characterized by the low use of GTP messages, thus little signaling related to data traffic.

All percentage values we include in Table V indicate the ratio between the one feature value to the total number of dialogues triggered by a device. We show the average values of the features over all the devices that are included in a given cluster. To help the reader, we color cells according to the

[3] 3 nodes with 2xAMD EPYC 7513 32-Core Processor, 1TB memory

baseline range of the entire dataset: blue – if the values are above the 80%, and red – if the values are below the 20% quantile. This highlights outliers in both directions.

Cluster 1 has a high percentage of UL_GRPS messages, with an average of 15.3% of the total messages per device. This indicates a high signaling volume originating from the SGSN, which we further corroborate with the occurrence of SAI messages originating at the SGSNs (not included in the table). Mobility and error features show no irregularities. Thus, we attribute cluster 1 to a device profile with regular behavior, but a high volume of signaling originating from SGSNs.

TABLE V: Mean values for features that distinguish clusters in DAT1. Blue and red indicate if values are globally above the 80% or below the 20% quantile.

Device Cluster	1	2	3	4	5	6	7	8
Overall Statistics: Phone & Data Usage								
MAP Dialogues [%]	83.4	77.8	87.2	89.2	66.9	92.2	25.0	77.7
GTP Dialogues [%]	16.6	22.2	12.8	10.8	33.1	7.8	75.0	22.3
Overall Statistics: Errors & Rejects								
Errors [%]	0.46	0.20	0.46	0.08	0.14	0.36	0.67	0.16
Rejects [%]	0.43	0.68	55.71	0.46	1.15	1.14	0.21	0.58
Selected Mobility Features								
Cell ID Changes	21.96	0.08	0.23	128.13	9.20	5.46	54.53	51.34
Operator Changes	3.08	0.01	0.03	77.58	0.38	0.28	1.49	1.14
Message Type Statistics: Protocol Usage								
SAI [%]	30.6	21.8	15.4	32.5	29.7	46.6	10.1	39.5
UL [%]	13.0	11.1	5.5	19.8	15.2	20.0	5.4	14.9
UL_GPRS [%]	15.3	4.0	4.0	19.1	5.9	6.6	1.6	6.0
PDP_CREATE [%]	9.2	11.3	6.4	6.6	16.9	4.2	38.2	11.8
PDP_DELETE [%]	7.4	10.7	6.3	4.1	16.2	3.6	36.2	10.4
PDP_UPDATE [%]	8.3	11.1	6.4	4.8	16.6	3.9	37.5	11.1

Moving on, cluster 2 captures devices that are likely stationary, with an average of 0.08 radio cell changes and 0.01 operator changes over the 30-day period. This is also consistent with the low ratio of UL_GPRS messages, indicating the lack of mobility-related HLR updates. Devices in cluster 2 are also characterized by an error rate of only 0.2%, and low reject rate of 0.68%. We conjecture that, due to their lack of mobility, these devices have good signal strength to successfully complete signaling procedures most of the time. Furthermore, we find that devices in this cluster exhibit long GTP tunnel durations, indicating long-lived data connections.

Focusing on cluster 3, we see it is mainly characterized by a high reject rate (55.71% per device), indicating it contains devices that tend to fail to complete the signaling procedures. These may include, for example, devices trying to establish connections to invalid roaming partners, devices that are using deactivated SIM cards, or devices that attempt to connect in poor coverage areas.

Cluster 4 clearly contains devices that exhibit a very high degree of mobility. They frequently change the cell and the operator, and have a high percentage of UL and UL_GPRS messages that support this observation. Devices in this cluster also exhibit low GTP signaling ratios, indicating limited use of data connectivity.

Clusters 5 and 8 do not exhibit extreme characteristics aside from their low error and reject rates. The difference between the two clusters is the higher mobility indicator (cell changes) of cluster 8. Hence, we categorize both clusters 5 and 8 as containing devices with no exceptional properties.

Clusters 6 and 7, on the other hand, exhibit mainly opposing properties regarding the type of signaling procedures they trigger. Cluster 6 devices trigger on average a high percentage of MAP signaling messages (92.2% of signaling dialogues are MAP), whereas cluster 7 devices trigger mostly GTP signaling (with 75.0% GTP-related signaling per device). The per-message-type breakdown ratios further support this separation of the two clusters. Cluster 6 devices use then circuit-switched technologies, while cluster 7 devices use mostly data connectivity.

TABLE VI: Mean value of relevant features in each cluster of DAT2. Blue and red indicate if values are globally above the 80% or below the 20% quantile.

Device Cluster	1	2	3	4	5	6	7	8
Overall Statistics: Phone & Data Usage								
MAP Dialogues [%]	39	99.4	63	86	81	81	54	99.1
GTP Dialogues [%]	61	0.7	37	14	19	19	46	0.9
Overall Statistics: Errors & Rejects								
Errors [%]	5.4	1.9	5.1	2.6	2.6	28	2	0.34
Rejects [%]	0.51	0.24	1.5	0.72	1.7	0.76	0.67	97
Unclassified [%]	1.7	35	3.2	3.1	2.6	2.6	2	0.27
Select Mobility Features								
Cell ID Changes	129.42	0.19	567.43	18.03	43.38	6.04	317.93	8.30
Operator Changes	4.41	4.20	114.51	15.29	5.76	0.63	45.30	0.51
Message Type Statistics: Protocol Usage								
SAI [%]	27.2	5.3	41.8	40.4	53.7	44.3	32.8	94.5
UL [%]	6.0	38.4	5.7	12.7	19.7	19.8	9.6	4.0
UL_GPRS [%]	2.5	0.6	4.0	18.5	2.3	6.7	3.3	0.1
PDP_CREATE [%]	32.5	0.3	15.3	6.5	9.2	10.3	15.1	0.5
PDP_DELETE [%]	27.0	0.1	11.0	4.1	7.3	3.8	13.6	0.3
PDP_UPDATE [%]	1.7	0.2	10.4	2.9	2.2	5.2	16.9	0.1

2) DAT2 Profiles – Dedicated IoT Operator: Following the same clustering methodology, we extract signaling profiles from DAT2. In Table VI we show the results.

Here, cluster 1 exhibits two major characteristics: devices show a high ratio of data-related signaling, which both the high percentage of GTP dialogues and the high percentages of PDP_CREATE and PDP_DELETE messages confirm. The low ratio of PDP_UPDATE messages indicates that devices use short-lived GTP tunnels, instead of relying on the update mechanism. We infer that this might be an artifact of the manufacturer settings (e.g., due to quota limits). Devices also exhibit a significantly high number of radio cell changes, but a low number of network operator changes. This indicates that these are mobile devices that largely remain attached to the same operator. We infer that these devices have regional mobility, over a relatively small geographical area.

Cluster 2 aggregates devices that mainly trigger MAP signaling. Devices in this cluster also exhibit a large ratio of unclassified dialogues indicating the use of non-data related communication, like SMS[4]. The mobility features show that devices are mostly stationary. These devices may thus reflect very old IoT devices that connect via SMS.

Clusters 3 and 7 both contain devices that exhibit a large degree of mobility, as indicated by the high numbers for both cell changes and operator changes. This occurs if devices move over large geographical distances, traversing regions covered by multiple operators, or for devices that frequently alternate between two operators in areas with coverage of more than

[4]Here we also include the percentage of unclassified dialogues. This group of dialogues contains every type of MAP signaling that is not related to authentication or mobility handling. These include mainly signaling dialogues related to SMS transmission, which are beyond the scope of the analysis we expose in this work.

TABLE VII: Signaling traffic profile mapping for both datasets and all four identified traffic profiles. ▲ cluster exhibits high values; ▼ low values; - medium values.

Cluster	Mobility	Data Usage	Reject Behavior	Protocol Usage
DAT1 (IoT Platform)				
1	-	-	▼	▲ UL_GPRS
2	▼	-	▼	▼ SAI ▼ UL_GPRS
3	▼	-	▲	▼ SAI
4	▲	▼	▼	▲ UL ▲ UL_GPRS
5	-	-	▼	-
6	▼	▼	▼	▲ SAI ▲ UL
7	-	▲	▼	▲ GTP ▼ MAP
8	-	-	▼	-
DAT2 (Dedicated IoT Operator)				
1	-	▲	▼	▼ PDP_UPDATE
2	▼	▼	▼	▲ UL
3	▲	-	▼	▲ PDP_UPDATE
4	-	-	▼	▲ UL_GPRS
5	-	-	▼	▼ PDP_UPDATE
6	▼	-	▼	▲ Error Rate
7	▲	-	▼	▲ PDP_UPDATE
8	▼	▼	▲	▲ SAI

one operator. Both types of mobility result in the same pattern when it comes to the observed signaling traffic.

Clusters 4 and 5 exhibit largely similar characteristics and contain mainly nondescript devices. The major difference between the two clusters is the degree of mobility these devices exhibit. While cluster 4 shows a low average number of cell changes, the number of operator changes is almost three times higher compared to cluster 5. We assume that this is an artifact of the radio deployments available to the devices in the areas where they operate (i.e., per country or region differences in radio deployment).

Cluster 6 aggregates devices that exhibit a very high error rate. This means that observed signaling interactions violate the signaling standardization through unexpected data, out of order or missing messages, or invalid signaling attempts.

Finally, cluster 8 contains devices that fail to perform successful signaling, with a very high ratio of rejected dialogues (i.e., 97%). Similar to DAT1, this occurs if devices try to roam with operators without valid roaming agreements or if deployed sim cards are either not yet or no longer activated.

Takeaway: The analysis of device signaling characteristics allows identifying macroscopic patterns that are explainable. The unsupervised learning approach greatly simplifies the analysis of the administrators, who can further use their domain knowledge to interpret the results. In addition, some clusters clearly pinpoint groups of devices that have problems, e.g., generating large amounts of errors, or are using old technology. This information is instrumental for the network operator for both management and business actions.

B. Comparison of Clustering Results

While the operations and customer bases of these two providers are different, we found that the datasets contain similar profiles, but at different granularity. We provide a direct comparison of the results for the two independent datasets.

We condense the identified descriptions into four profiles and assign qualitative indicators according to the values in Table V and Table VI. This simplifies the comparison of signaling patterns. Table VII presents this mapping between signaling profiles and clusters in each of the datasets.

Mobility. In both datasets, we identify clusters (cluster 4 in DAT1, and clusters 3 and 7 in DAT2) that exhibit significant mobility (high number of operator and cell changes). When looking at the specific feature values for these clusters, we observe that devices – while exhibiting strong mobility in both datasets – differ in the way they use data connectivity. On average, only around 11% of signaling traffic consists of GTP dialogues in DAT1, while devices in DAT2 exhibit 37% and 46% GTP dialogues on average, with DAT1 having more IoT devices that do not rely on data connectivity (e.g., shipment tracking) than DAT2.

Data Usage. Data usage characteristics represent our second behavioral profile. Differentiation by data usage shows that our clustering approach on both datasets is able to isolate devices that either exhibit excessive data usage (cluster 7 in DAT1, cluster 1 in DAT2), or very limited data usage (clusters 4 and 6 in DAT1, cluster 2 in DAT2). Note that cluster 8 in DAT2 also shows very low data-related signaling. However, this is due to the fact that devices in this cluster largely fail to perform successful signaling and hence never get the chance to establish data connectivity. This is again due to the different configurations of the two providers (e.g., quota limits). These clusters may suggest the IoT operators to adapt their plans.

Reject Behavior. The reject behavior of devices maps to cluster 3 for DAT1 and cluster 8 for DAT2. With a mean reject rate of 55% and 97%, respectively, both exhibit significantly higher reject rates than all other clusters. This profile points to the fact that these devices largely fail to perform any meaningful signaling, likely due to deactivated SIMs, or buggy implementations. They bring unnecessary cost for operators and carriers, and justify the need for network management decisions to alleviate the issues they pose.

Protocol Usage. Finally, we were able to identify clusters showing distinct protocol usage patterns in both datasets. Namely, clusters 1 and 4 in DAT1 as well as cluster 4 in DAT2 show high amounts of signaling messages originating from the SGSN (UL_GPRS) while at the same time only showing low to moderate data usage, indicating irregular signaling behavior. Conversely, clusters 1 and 7 in DAT2 exhibit low and high usage of PDP_UPDATE dialogues despite showing low and moderate data usage, respectively. This indicates irregular device behavior, such as excessive creation and deletion of data tunnels for cluster 1 or abnormally high data tunnel duration in the case of cluster 7. This suggests possible implementation issues.

Takeaway. The profiles we identify through clustering generalize to both deployments for our datasets. We expect this to be applicable to other networks and datasets as well. We see this as a step towards creating common profiles for IoT devices based on mobile signaling that researchers can use as a starting point, and network operators can leverage to simplify their actions. We achieve this thanks to the generality of the features that characterize device behavior, thus allowing common and macroscopic patterns to emerge.

VII. Discussion

Understanding the signaling behavior of IoT devices is crucial for network optimization in modern mobile networks, specifically when it comes to the current complex roaming landscape. Administrators benefit from detailed traffic analyses, gaining insights into peak usage times, similarities, and differences between regions, operators, and device types, and a deeper understanding of how devices react to issues in the network, such as congestion. Specifically, this knowledge can directly be applied to many aspects of the current mobile network landscape, as outlined in the following.

Network Optimization and Resource Usage. By improving the usage of the available resources, reducing congestion, and improving overall network performance in the signaling plane, administrators are able to improve the Quality of Service (QoS) for devices. Specifically, administrators can implement effective load balancing strategies based on signaling patterns, preventing congestion in specific network areas, and optimize resource utilization. By identifying and minimizing unnecessary signaling overhead, network efficiency can be improved, leading to a more sustainable and cost-effective operation. Dynamic resource allocation, guided by real-time signaling behavior analysis, enables the network to adapt to changing conditions, while the identification and mitigation of latency issues contribute to a responsive and reliable network.

Security and Diagnostics. Establishing a general baseline for the signaling patterns observed in mobile networks is a critical requirement in both security related and troubleshooting tasks. Anomalies, specifically deviations from the previously established baseline, in signaling behavior can indicate potential security threats or attacks on the network. Monitoring and analyzing signaling patterns enable the early detection of unusual activities, allowing for prompt security measures to be implemented. Similarly, signaling insights are invaluable for troubleshooting network issues. By analyzing signaling data, administrators can quickly identify and resolve problems, reducing downtime and improving the overall reliability of the network.

Planning and Migration. A deep understanding of the signaling behavior across the entire network is crucial for effective network planning and expansion. As IoT deployments grow, understanding how devices communicate, and thereby generate load for the network, helps administrators design scalable and resilient networks. This ensures that the infrastructure can accommodate the increasing number of devices while maintaining optimal performance and reliability. This is especially relevant as networks progress from the current 2G/3G/LTE landscape towards 5G and 6G deployments. Understanding the signaling behavior is essential to successfully transition towards novel technologies, as it is important to understand the workload generated by devices after the transition towards a novel network paradigm. The insights obtained in this work facilitates this migration processes, ensuring that existing IoT devices remain connected and operational during network upgrades.

Sustainability. Finally, in the recent advent of sustainable networking, understanding the behavior of devices is crucial when it comes to optimizing networks regarding energy consumption and sustainability. IoT devices often operate on limited battery power, making efficient energy management crucial. Analyzing signaling behavior allows administrators to implement strategies that minimize the frequency and duration of signaling events. This proactive approach can be leveraged to extend the battery life of connected devices, a critical factor for many IoT applications and a crucial Key Performance Indicator (KPI) for operators going forward.

VIII. Conclusions

In this work, we performed a first-of-its-kind characterization of IoT signaling traffic in mobile networks. We showed how measuring the signaling plane of IoT providers can illuminate the characteristics of the network, the device footprint, and how unsupervised machine learning allows the identification of groups of devices with similar behavior. Our work provides the basis for network operations teams to improve their management and general service delivery, by highlighting groups of devices with critical signaling characteristics that need to be taken into account for the correct management of the IoT ecosystem in general. In addition, we lay the groundwork for future research activities in the areas of anomaly detection, device classification, and traffic modeling in general. Our results showed the need for device-level characterization to improve both our understanding of specific IoT verticals and the network load generated by deployments. In a privacy-aware system, where the network operator is blind to the end-user application, this knowledge allows the IoT provider to deploy a proactive approach when running their service.

Interestingly, after separately processing two independent, large-scale datasets, we found that similar signaling characteristics are well-suited to identify groups of devices across both datasets. Here, domain knowledge is paramount in building a meaningful set of features, while enabling a generalizable approach. We believe this is key to enabling interpretability of results, that we prove with our dataset for two IoT providers.

Further, the data work builds the base for future research and applications, building more complex machine learning models to apply in production. We presented features that were crucial for distinguishing clusters of devices, as highlighted in Tables V and VI. Specifically, we identified device mobility, errors, and the ratio of GTP and MAP dialogues to be the most relevant factors when explaining the clusters. Intuitively, these groups of features explain three key characteristics of devices: mobility, anomalous behavior, and data usage.

To enable researchers and practitioners to benefit from our findings, to compare future datasets to our results and to validate and reproduce our findings, we provide the aggregated feature values for the dedicated IoT operator (DAT2). The included dataset contains all feature values for 270,000 unique IoT devices. This, in addition to the provided insights presented in this paper and the engineered feature set, allows operators to reproduce our findings in their respective deployments. By comparing aggregated results across device clusters or detailed device characteristics within specific clusters, e.g., highly mobile devices, network technicians can leverage

our findings to identify anomalous devices. By applying our proposed methodology and resulting models, operators can identify and optimize a wide range of aspects of their deployments, by tailoring system configurations to the behavioral characteristics observed in real world data.

As global connectivity becomes more important, and over-the-top providers (e.g., Twilio, EMnify, Truphone, Google Fi) penetrate deeper into the market, we argue that scrutinizing mobile signaling traffic between different entities involved in the end-to-end communication solution is crucial for monitoring the health of the ecosystem.

Acknowledgement

This work was partially supported by the Smart Networks and Services Joint Undertaking (SNS JU) under the European Union's Horizon Europe research and innovation program under grant agreement no.101139270 ("ORIGAMI"), and by the European Union's Horizon 2020 research and innovation program under grant agreement no.101017109 ("DAEMON").

Glossary

CDF Cummulative Distribution Function. 7
CL Cancel Location. 3, 6, 7
GGSN Gateway General Packet Radio Service (GPRS) Support Node. 3
GRX General Packet Radio Service (GPRS) Roaming eXchange. 3
GTP GPRS Tunneling Protocol. 3, 4, 9
HLR Home Location Registry. 3, 4, 10
IMSI International Mobile Subscriber Identity. 3
IoT Internet of Things. 1–4, 9
IPX IP Packet Exchange. 2, 3
KPI Key Performance Indicator. 12
MAP Mobile Application Part. 3–6, 9
MNO Mobile Network Operator. 1–3, 7, 8
MSC Mobile-services Switching Centre. 4
MVNO Mobile Virtual Network Operator. 1
PCA Principal Component Analysis. 9
PDP Packet Data Protocol. 4, 6, 8
PoP Point of Presence. 4
QoS Quality of Service. 12
RAN Radio Access Network. 2, 4, 5
SAI Send Authentication Information. 3, 6–8
SCCP Signaling Connection Control Part. 3, 4
SGSN Serving General Packet Radio Service (GPRS) Support Node. 3, 8, 10
UL Update Location. 3, 6, 7, 10
UL-GPRS Update General Packet Radio Service (GPRS) Location. 6, 7
VLR Visitor Location Registry. 3, 4, 8

References

[1] A. Lutu, B. Jun, A. Finamore, F. E. Bustamante, and D. Perino, "Where Things Roam: Uncovering Cellular IoT/M2M Connectivity," in *Proceedings of the ACM Internet Measurement Conference*. New York, NY, USA: Association for Computing Machinery, 2020, p. 147–161.

[2] P. Arriandiaga, E. Goodness, L.-O. Wallin, and K. Chao, "Gartner Magic Quadrant for Managed IoT Connectivity Services, Worldwide," Gartner Research, White Paper G00749396, 2022.

[3] P. Schmitt, M. Vigil, and E. Belding, "A study of MVNO data paths and performance," in *International Conference on Passive and Active Network Measurement*. Springer, 2016, pp. 83–94.

[4] A. Xiao, Y. Liu, Y. Li, F. Qian, Z. Li, S. Bai, Y. Liu, T. Xu, and X. Xin, "An in-depth study of commercial MVNO: Measurement and optimization," in *Proceedings of the 17th Annual International Conference on Mobile Systems, Applications, and Services*, 2019, pp. 457–468.

[5] A. Lutu, D. Perino, M. Bagnulo, and F. E. Bustamante, "Insights from Operating an IP Exchange Provider," in *Proceedings of the 2021 ACM SIGCOMM 2021 Conference*. Association for Computing Machinery, 2021, p. 718–730.

[6] A. Lutu, B. Jun, F. E. Bustamante, D. Perino, M. Bagnulo, and C. G. Bontje, "A First Look at the IP EXchange Ecosystem," vol. 50, no. 4, p. 25–34, 2020. [Online]. Available: https://doi.org/10.1145/3431832.3431836

[7] S. Geissler, F. Wamser, W. Bauer, M. Krolikowski, S. Gebert, and T. Hoßfeld, "Signaling Traffic in Internet-of-Things Mobile Networks," in *2021 IFIP/IEEE International Symposium on Integrated Network Management (IM)*. IEEE, 2021, pp. 452–458.

[8] A. Sivanathan, H. H. Gharakheili, F. Loi, A. Radford, C. Wijenayake, A. Vishwanath, and V. Sivaraman, "Classifying IoT Devices in Smart Environments Using Network Traffic Characteristics," *IEEE Transactions on Mobile Computing*, vol. 18, pp. 1745–1759, 2019.

[9] H. Kolamunna, I. Leontiadis, D. Perino, S. Seneviratne, K. Thilakarathna, and A. Seneviratne, "A First Look at SIM-Enabled Wearables in the Wild," 2018.

[10] C. E. Andrade, S. D. Byers, V. Gopalakrishnan, E. Halepovic, D. J. Poole, L. K. Tran, and C. T. Volinsky, "Connected cars in cellular network: a measurement study," in *Proceedings of the 2017 Internet Measurement Conference*. ACM, 2017, pp. 235–241.

[11] 3GPP, "3rd Generation Partnership Project; Technical Specification Group Core Network and Terminals; Location management procedures (3GPP TS 23.012)," *Technical specification (TS)*, 2022.

[12] D. Kumar, K. Shen, B. Case, D. Garg, G. Alperovich, D. Kuznetsov, D. Kuznetsov, R. Gupta, and Z. Durumeric, "All Things Considered: An Analysis of IoT Devices on Home Networks," in *Proceedings of the 28th USENIX Conference on Security Symposium*. USA: USENIX Association, 2019, p. 1169–1185.

[13] J. Martin, E. Rye, and R. Beverly, "Decomposition of MAC Address Structure for Granular Device Inference," in *Proceedings of the 32nd Annual Conference on Computer Security Applications*. New York, NY, USA: Association for Computing Machinery, 2016, p. 78–88. [Online]. Available: https://doi.org/10.1145/2991079.2991098

[14] S. Marchal, M. Miettinen, T. D. Nguyen, A.-R. Sadeghi, and N. Asokan, "AuDI: Toward Autonomous IoT Device-Type Identification Using Periodic Communication," *IEEE Journal on Selected Areas in Communications*, vol. 37, no. 6, pp. 1402–1412, 2019.

[15] J. Ren, D. J. Dubois, D. Choffnes, A. M. Mandalari, R. Kolcun, and H. Haddadi, "Information Exposure From Consumer IoT Devices: A Multidimensional, Network-Informed Measurement Approach," in *Proceedings of the Internet Measurement Conference*. New York, NY, USA: Association for Computing Machinery, 2019, p. 267–279. [Online]. Available: https://doi.org/10.1145/3355369.3355577

[16] P. Torkamandi, L. Kärkkäinen, and J. Ott, "An Online Method for Estimating the Wireless Device Count via Privacy-Preserving Wi-Fi Fingerprinting," in *Passive and Active Measurement*, O. Hohlfeld, A. Lutu, and D. Levin, Eds. Cham: Springer International Publishing, 2021, pp. 406–423.

[17] H. Guo and J. Heidemann, "Detecting IoT Devices in the Internet," *IEEE/ACM Transactions on Networking*, vol. 28, no. 5, pp. 2323–2336, 2020.

[18] R. Perdisci, T. Papastergiou, O. Alrawi, and M. Antonakakis, "IoTFinder: Efficient Large-Scale Identification of IoT Devices via Passive DNS Traffic Analysis," *2020 IEEE European Symposium on Security and Privacy (EuroS&P)*, pp. 474–489, 2020.

[19] S. J. Saidi, A. M. Mandalari, R. Kolcun, H. Haddadi, D. J. Dubois, D. Choffnes, G. Smaragdakis, and A. Feldmann, "A Haystack Full of Needles: Scalable Detection of IoT Devices in the Wild," in *Proceedings of the ACM Internet Measurement Conference*. New York, NY, USA: Association for Computing Machinery, 2020, p. 87–100. [Online]. Available: https://doi.org/10.1145/3419394.3423650

[20] ETSI, "Machine-to-Machine Communications (M2M), Functional Architecture," ETSI, Tech. Rep. TS 102 690 V2.1.1, 2013, work in Progress. [Online]. Available: https://www.etsi.org/deliver/etsi_ts/102600_102699/102690/02.01.01_60/ts_102690v020101p.pdf

[21] V. Gazis, "A Survey of Standards for Machine-to-Machine and the Internet of Things," *IEEE Communications Surveys Tutorials*, vol. 19, no. 1, pp. 482–511, 2017.

[22] P. K. Verma, R. Verma, A. Prakash, A. Agrawal, K. Naik, R. Tripathi, M. Alsabaan, T. Khalifa, T. Abdelkader, and A. Abogharaf, "Machine-to-Machine (M2M) Communications," *J. Netw. Comput. Appl.*, vol. 66, no. C, p. 83–105, 2016. [Online]. Available: https://doi.org/10.1016/j.jnca.2016.02.016

[23] M. Zhao, A. Kumar, T. Ristaniemi, and P. H. Chong, "Machine-to-Machine Communication and Research Challenges: A Survey," *Wirel. Pers. Commun.*, vol. 97, no. 3, p. 3569–3585, 2017. [Online]. Available: https://doi.org/10.1007/s11277-017-4686-1

[24] A. Biral, M. Centenaro, A. Zanella, L. Vangelista, and M. Zorzi, "The challenges of M2M massive access in wireless cellular networks," *Digital Communications and Networks*, vol. 1, no. 1, pp. 1–19, 2015. [Online]. Available: http://www.sciencedirect.com/science/article/pii/S235286481500005X

[25] Y. Mehmood, N. Haider, M. Imran, A. Timm-Giel, and M. Guizani, "M2M Communications in 5G: State-of-the-Art Architecture, Recent Advances, and Research Challenges," *IEEE Communications Magazine*, vol. 55, no. 9, pp. 194–201, 2017.

[26] M. Klymash, H. Beshley, O. Panchenko, and M. Beshley, "Method for optimal use of 4G/5G heterogeneous network resources under M2M/IoT traffic growth conditions," in *2017 International Conference on Information and Telecommunication Technologies and Radio Electronics (UkrMiCo)*, 2017, pp. 1–5.

[27] T. de Andrade, C. Astudillo, and N. da Fonseca, "Impact of M2M traffic on human-type communication users on the LTE uplink channel," 2015.

[28] M. Z. Shafiq, L. Ji, A. X. Liu, J. Pang, and J. Wang, "A first look at cellular machine-to-machine traffic: large scale measurement and characterization," *ACM SIGMETRICS Performance Evaluation Review*, vol. 40, no. 1, pp. 65–76, 2012.

[29] M. Z. Shafiq, L. Ji, A. Liu, J. Pang, and J. Wang, "Large-Scale Measurement and Characterization of Cellular Machine-to-Machine Traffic," vol. 21, no. 6, 2013.

[30] N. Vallina-Rodriguez, S. Sundaresan, C. Kreibich, N. Weaver, and V. Paxson, "Beyond the radio: Illuminating the higher layers of mobile networks," in *Proceedings of the 13th Annual International Conference on Mobile Systems, Applications, and Services*. ACM, 2015, pp. 375–387.

[31] A. M. Mandalari, A. Lutu, A. Custura, Ali, Ö. Alay, M. Bagnulo, V. Bajpai, A. Brunstrom, J. Ott, M. Mellia, and G. Fairhurst, "Experience: implications of roaming in Europe," 2018.

[32] F. Michelinakis, H. Doroud, A. Razaghpanah, A. Lutu, N. Vallina-Rodriguez, P. Gill, and J. Widmer, "The Cloud that Runs the Mobile Internet: A Measurement Study of Mobile Cloud Services," 2018.

[33] J. C. Dunn, "Well-separated clusters and optimal fuzzy partitions," *Journal of cybernetics*, vol. 4, no. 1, pp. 95–104, 1974.

Stefan Geißler received the PhD degree for his thesis "Performance Evaluation of Next-Generation Data Plane Architectures and their Components" from the University of Würzburg, Würzburg, Germany, in 2022. He is a Research Associate with the Chair of Communication Networks, University of Würzburg, where he leads the Cloud Applications and Networks Research Group. His current research is focused on the analytical and simulative performance evaluation and optimization of next-generation communication technologies, including TSN, 5G, and Internet of Things.

Andra Lutu is a Senior Researcher at Telefónica Innovación Digital, in Madrid, Spain. Her main research interests lie in the areas of mobile network architecture design, internet measurements and (network) data representation. Andra's work translated into impactful industrial innovation projects, and was published in top venues, including ACM MobiCom, ACM SIGCOMM or IEEE INFOCOM.

Florian Wamser is Lecturer and Researcher for Cyber Security and communication networks with an affinity for monitoring, data science and data engineering. He received his PhD in 2015 from the University of Wuerzburg, Germany. His thesis is recognized in GI's 2015 Lecture Notes in Informatics for excellent computer science dissertations. He is currently employed at the Lucerne University of Applied Sciences and Arts in the field of Cyber Security.

Thomas Favale got his Master Degree in Computer Engineering in 2019 and his Ph.D in 2023 at Politecnico di Torino under the supervision of Professor Marco Mellia, joining the Interdepartmental Center for SmartData@PoliTo. His research activity was developed in collaboration with Huawei Paris Research Center with the vision of a novel environment to enhance Cybersecurity. He addressed the problem of anonymisation of network traffic and the cyber attackers behavior. Thomas is now involved in a challenge to strengthen security in underwater wireless communications within WSense.

Viktoria Vomhoff is a Research Assistant at the Chair of Communication Networks at the University of Würzburg, where she is currently pursuing her PhD. She earned her Master's degree in Computer Science from the University of Würzburg in 2021. Her research interests include traffic measurements and modeling, performance evaluation, and the modeling of communication systems, with a particular focus on mobile communications and the Internet of Things (IoT).

Michael Krolikowski is the Chapter Lead - Applications at emnify, where he has been instrumental in leading cross-company software and cloud architecture initiatives since 2018. Before joining emnify, he worked at Syniverse Technologies Messaging GmbH, focusing on software development and architecture. With a Bachelor of Engineering in Computer Science from Technische Hochschule Würzburg-Schweinfurt, Michael has been active in the telecommunications industry since 2009. Throughout his career, he has been involved in developing multiple mobile network components, with a strong emphasis on scalability and high availability. His professional interests include telecommunication and large-scale online and offline data processing.

Marco Mellia (F'21), Ph.D., is a full professor at Politecnico di Torino, Italy. His research interests are in the areas of Internet monitoring, users' characterisation, cyber security, and machine learning applied to different sectors. He has co-authored over 300 papers published in international journals and presented at leading conferences. He won the IRTF ANR Prize at IETF-88, and the best paper awards at IEEE P2P'12, ACM CoNEXT'13, IEEE ICDCS'15, ACM CCR'16, ITC'18. He is the Editor in Chief of the Proceedings of the ACM on Networking.

Diego Perino is an organization manager, technical leader and scientist with passion to work in cutting edge projects with industrial impact. He currently is an engineering manager at Meta, but he has been working for different companies in the ITC sector (Telefonica, Bell Labs, Orange Labs) and covered different technologies and research areas (above all AI, networks, systems). Apart from his industrial experience, he has also been very active in the scientific community with several publications, participation in conference committees, and editorial board contributions. He holds a Ph.D. from the Paris Diderot-Paris 7 University, MSc. from Politecnico di Torino, Eurecom Institute and Université de Nice-Sophia Antipolis.

Tobias Hoßfeld is professor at the Chair of Communication Networks at the University of Würzburg, Germany, since 2018. From 2014 to 2018, he was head of the Chair "Modeling of Adaptive Systems" at the University of Duisburg-Essen, Germany. He has published more than 100 research papers in major conferences and journals, receiving several best conference paper awards, 3 awards for his PhD thesis, and the Fred W. Ellersick Prize 2013 (IEEE Communications Society) for one of his articles on QoE. He is member of the editorial board of IEEE Communications Surveys & Tutorials, Springer Quality and User Experience, ACM SIGMM Records and elected chairperson of the ITG/VDE expert group "Communication Networks and Systems" within the German society of Information Technology (ITG).