SIMBox bypass frauds in cellular networks: Strategies, evolution, and detection survey

Anne Josiane Kouam, Aline Carneiro Viana, Alain Tchana

SIMBox bypass frauds in cellular networks: Strategies, evolution, and detection survey

Anne Josiane Kouam D.
Inria, Ecole Polytechnique - IPP
Palaiseau, France
anne.josiane-kouam.djuigne@inria.fr

Aline Carneiro Viana
Inria
Palaiseau, France
aline.viana@inria.fr

Alain Tchana
Inria, ENS Lyon
Lyon, France
alain.tchana@ens-lyon.fr

***Abstract*—Due to their complexity and opaqueness, cellular networks have been subject to numerous attacks over the past few decades. These attacks are a real problem to telecom operators and cost them about USD 28.3 Billion annually, as reported by the *Communications Fraud Control Association*. *SIMBox* fraud, which is one of the most prevalent of these telephone frauds, is the main focus of this paper. *SIMBox* fraud consists of diverting international calls on the VoIP network and terminating them as local calls using an off-the-shelf device, referred to as *SIMBox*. This paper surveys both the existing literature and the major *SIMBox* manufacturers to provide comprehensive and analytical knowledge on *SIMBox* fraud, fraud strategies, fraud evolution, and fraud detection methods. We provide the necessary background on the telephone ecosystem while extensively exploring the *SIMBox* architecture required to understand fraud strategies. Our goal is to provide a complete introductory guide for research on *SIMBox* fraud and stimulate interest for *SIMBox* fraud detection, which remains little investigated. In this vein, we conclude the paper by presenting insights into tomorrow's *SIMBox* fraud detection challenges.**

***Index Terms*—Telephony networks, *SIMBox* fraud survey, *SIMBox*, fraud detection.**

I. Introduction

Telephone fraud presents a considerable problem for Mobile Network Operators (MNO) around the world. According to the CFCA, the global fraud loss is estimated to USD 28.3 Billion in 2019 [1]. In this context, illegal bypass termination [2] also known as *SIMBox* fraud is by far one of the most prevalent fraud affecting the telecommunication market [3]. In many countries, the international termination rate (ITR) is considerably higher than the local (retail) termination rate (LTR) within the country (e.g., up to $2.8 - 28$ times of difference in Cameroon [4]). This makes it profitable for fraudsters to bypass the regular interconnect operator when terminating calls in the country as they can pay the lower local rate instead of the ITR. *SIMBox* fraud is a major problem in developing countries (e.g. about 78% of African countries and 60% of Middle Eastern countries are fraud destinations [5]). Besides, in some these countries, as much as 70% of incoming international call traffic is terminated fraudulently [6]. This could result in losses of up to USD 39.9 Million as in Cameroon in 2015 [7]. This practice is thus illegal in most countries and mainly, in developing countries.

The simplest way of committing bypass fraud involves setting up a *SIMBox* (VoIP GSM gateway). This is a standard device that can be easily acquired via the internet and equipped with a bundle of SIM cards. The calls are typically routed via an internet flow (VoIP) to the *SIMBox* residing in the terminating country. The *SIMBox* then converts the VoIP call into a local mobile call to the receiving party on the local cellular network.

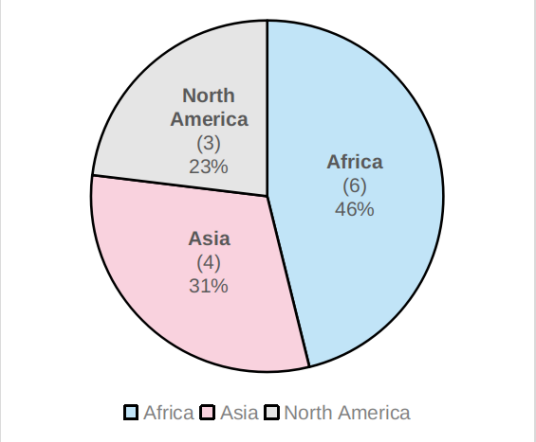

Fig. 1. Distribution of *SIMBox* fraud investigation work by continent.

SIMBox fraud is a significant problem for telecommunication operators and tax authorities of the affected countries, as international traffic taxes cannot be collected. Beyond direct revenue loss, bypass fraud also leads to poor customer experience. Examples of such call quality experience degradation are low voice quality due to latency issues, highly-compressed IP connections, longer call set up time, or still, missing or incorrect Calling Line Identifier (CLI). In particular, this latter results in many call rejections by the called party, while missed calls are not returned. Such degradation impacts the customer experience, which has a direct effect on loyalty, lifetime value, and revenue. All of this results in the *SIMBox* fraud being in the top three types of phone system frauds that cause a significant loss to mobile network operators [8].

Few research works have investigated this issue (a total of 13 publications distributed by continent in Figure 1). We believe this is essentially due to the fact that the fraud mainly affects developing countries, though some papers discussed the issue in USA.

The purpose of this manuscript is to provide all the necessary elements to understand the *SIMBox* fraud problem in its entirety. It is intended to support readers wishing to begin research into *SIMBox* fraud detection and may spark new interest in research in this area which is currently little explored. To the best of our knowledge, this is the first paper in the literature presenting the current state-of-the-art of *SIMBox* fraud. More specifically, we present a comprehensive review of the literature on this topic while providing a better understanding of such fraud and discussing open research issues in the area of *SIMBox* fraud detection. In short, we define our contributions as follows:

- We provide an overview of the full telephony ecosystem with the involved mobile telephony network equipment and stakeholders, as well as discuss the main telephony functionality concerned by *SIMBox* fraud, i.e., call routing (see Section II).
- We describe the *SIMBox* fraud ecosystem while explaining how *SIMBox* fraud works, its financial benefits for fraudsters, and what facilitates its existence (see Section III).
- We deeply explore the system behind the *SIMBox* by commenting on its components and architecture (see Section IV). This work is based only on extensive research into the specifications of different SIMBox models; indeed, no similar work has been done in the past.
- We then narrow our focus and consider newer models of *SIMBox* having the advanced capability of simulating human communication behaviour, which hardens detection. In this vein, we examine the temporal evolution of *SIMBox* fraud strategies related to human behavioural simulation (see Section V).
- After extensively exploring the *SIMBox* fraud ecosystem, we study the related detection methods in literature. Here, a categorization of *SIMBox* detection methods is introduced and a qualitative comparison is presented (see Section VI).
- Our main finding is that *SIMBox* fraud is a very tricky problem due to the following aspects: it involves major stakeholders in the telephony area which could be fraudulent; it is related to economic factors for internal development which it would be difficult for developing countries to do without; it is accentuated by the increasing popularisation among mobile subscribers of VoIP applications for international telephone calls.
- As in most security problems, fraudsters are evolving at the same rate, if not faster, than the research community. The use of AI advances (for instance to mimic human behaviour in frauds) has attracted considerable attention because it hardens the identification of fraudulent calls. In this perspective, we highlight some incoming challenges concerning the evolution of telecommunication technologies (such as 5G and 6G) and take the risk in prospecting the frauds of tomorrow (see Section VII).

Finally, we draw conclusions of our work in Section VIII.

II. The telephony ecosystem

The telephony ecosystem has enormously evolved from a level where calls were made through primitive and pre-traced channels to a more complex environment in which numerous frauds have spread. This environment involves several telephone networks and specialized equipment able to switch from these networks to others, forming a global architecture which is outlined in Figure 2. This section presents the full telephony ecosystem.

A. Mobile telephony networks

1) Cellular networks: Cellular devices are widely used for more than just entertainment and phone calls, and their scale is becoming greater and more significant. The cellular network is an open public network to which end subscribers have direct access through distributed over land areas, named *cells*, each served by at least one fixed-location transceiver, known as *base station*. For a customer, the *Mobile Equipment* is the entry point to the network. It is uniquely identified by its *International Mobile Equipment Identity* (IMEI) and requires a *Subscriber Identity Module* (SIM) card to access the operator's network services. From a more technical perspective, the SIM card is uniquely identified on the network by its *International Mobile Subscriber Identity* (IMSI); it contains a cryptography key assigned by the operator to encrypt communication. The IMSI is used by the mobile equipment (i) to detect the provider's network amongst surrounding emitted beacon frames, and (ii) to get attached to that network through the base station. Each generation of mobile communication uses encryption over the wireless channel and specific equipment to handle communications and customer identification. Yet, at the level of a provider, mobile devices can use different generations of mobile communication to access the network.

Call and SMS routing are services provided by cellular networks and involve several connected types of equipment. Base stations are connected to a *Mobile Switching Center* (MSC) of the core network through the base station controller. The MSC is responsible for keeping and updating subscriber information in the *Home Location Register*, registration and authentication of subscribers with the help of the *Authentication Center*, and call/SMS routing. Some MSCs provide an access gateway to the PSTN network or to an external IP network like the Internet. They are called *Gateway MSCs* (GMSC) and do not manage base station controllers.

Last but not least, operators keep track of all services transiting on their networks and resource use. This is done through network switches, where *Charging Data Records* (CDR[1]) are generated. CDRs are collected and processed through *Charging Gateway Functions* [10] in a central location. There, they can be used for billing purposes or by fraud management units (referred to as *Revenue Assurance and Fraud Management* [11]) for telecommunication fraud prevention or detection.

[1]Once referred to *Call Detail Records* and later to *Charging Data Records* in the 3GPP specification [9]

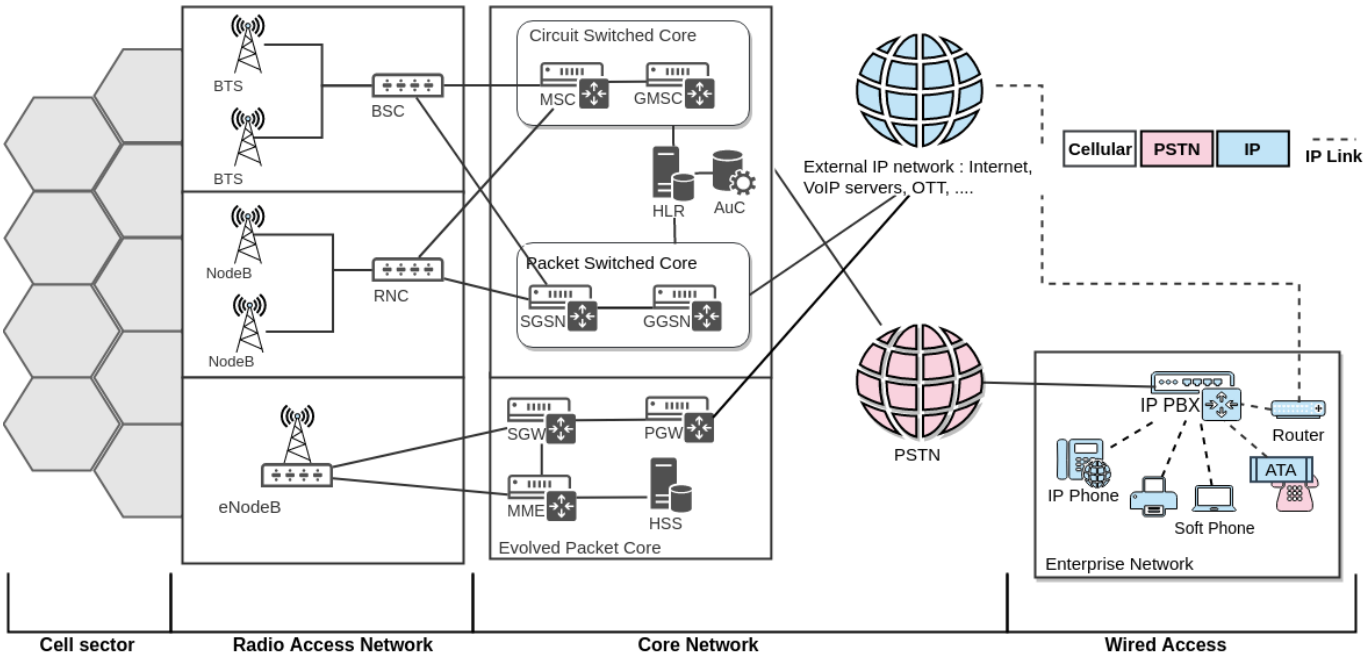

Fig. 2. Telephony networks architecture overview : Cellular network, PSTN network, VoIP network and gateways; from the access to the core.

2) VoIP networks: Voice over IP (VoIP) is a technique that makes it possible to transmit voice over wired (cable/ADSL/optical fiber) or wireless (satellite, Wi-Fi, UMTS or LTE, etc) IP networks, whether private networks or the Internet. VoIP network's whole idea is based on *VoIP servers*, *VoIP clients*, and *VoIP gateways* exchanging data and signaling information according to specific protocols. VoIP servers provide authentication, management, accounting, transport, billing, and routing services to VoIP clients in any IP network. VoIP clients are hard phones (also referred to as IP phones), soft-phones (any computing device with specific software to exchange over VoIP), and even analog phones combined with an *Analog Telephone Adapter*. VoIP gateways are specific network types of equipment, allowing VoIP voice calls and fax to reach the PSTN and the cellular network and vice-versa. Through a media and a signaling components, VoIP gateways compress and packetize voice data, deliver output packets to and from an IP network [12]. Some IP-PBXes (IP Private Branch Exchange)[2], for instance, can ensure a VoIP gateway function from the company's internal IP network to the external PSTN network (see Figure 2).

To ensure the compression and the packetization of analog voice signals for transmission over digital line, while keeping a certain acceptable quality for end-users, specific algorithms (called codecs) are designed. Codecs are used to encode voice traffic picked up by the microphone to digital data and encapsulate that data such that its transmission is faster and the call experience is better. There are several codecs, proprietary, and open-source ones, determining the sound quality and bandwidth usage. The more bandwidth a codec requires, the better the voice quality is. The choice of an appropriate codec is, therefore, strictly relative to the network efficiency. [12] summarizes the most well-known voice codecs. Some codecs are provided on the top of signaling protocols, which ensure codec configuration and other key functions in a call establishment. A signaling protocol enables network components to communicate with each other, establishes a multimedia session between two or more participants of the call, maintains it, and tears it down. *Session Initiation Protocol* (SIP) is a well-know application-layer VoIP signaling protocol. As such, it can allow interoperability between VoIP types of equipment from different manufacturers. More details on signaling protocols covering services are described in [12].

VoIP is also available on cellular networks. Some applications called *Over-The-Top* (OTT) apps (e.g., Skype, Discord) have been developed on the top of VoIP networks and provide for instance cheap call services, attracting more users and seen as a threat by mobile operators [14]. Cellular networks provide mobile connectivity through the Packet Switched Core comprising a *Serving Gateway Support Node* and a *Gateway GPRS Support Node* to fulfill packet-switched data transfers.

VoIP is cheaper than other telephony networks because it relies on an existing service and infrastructure, the Internet. No new connections need to be made whether the call is local or international: the Internet network has already done all. On the other hand, the quality of VoIP calls is generally poor due to bandwidth sharing (for services other than VoIP) and IP network latency. As a result, calls are affected by packet losses, delay, and jitter. Missing packets that are due to packet

[2]Enterprise usually use a PBX to manage their internal and external communication needs. A traditional PBX provides extensions, i.e., an internal phone number to reach each user within the enterprise. A traditional PBX uses phone cables for all internal lines which is expensive to deploy and manage. To the contrary, an IP-PBX can connect IP phones or soft phones over IP [13]

loss or jitter cause gaps in the audio. Such gaps are then filled by default silence audio or generated audio with *Packet Loss Concealment* algorithms.

B. Mobile telephony stakeholders

1) End-users: They are considered as primary actors in telephony because they are the service's purpose: the service exists to satisfy them and is continuously improved according to their needs. As aforementioned, each end-user is identified on the network by a SIM card, which can be prepaid or postpaid. Prepaid cards are easy to get. They can be purchased on platforms not owned by the supplier. For a subscriber using a prepaid card, the contract with the provider is simple: as long as he tops-up his SIM card, he can access the operator's services. Thus, a subscriber using a prepaid SIM has to stop recharging it to unsubscribe from the supplier. Postpaid SIM cards are less straightforward; the mobile operator associates the customer's personal and financial information with the SIM card to ensure payment. A contract between the customer and the provider establishes then, the service to which the customer is entitled, and the charges associated with it. Therefore, the subscriber will be billed for this service whether or not he uses it until he cancels the subscription with the provider.

Some subscribers may have more than one SIM card using multi-SIM devices. Dual-SIM handsets, for instance, have two SIM card slots for the use of two SIM cards from one or multiple providers. Those devices are commonplace in developing markets to always get the best offer from competing networks. In companies, employees can access mobile networks from the internal wired/IP network managed by a PBX/IP-PBX. This is done by the mean of *SIMBoxes* acting as VoIP or *Public Rate Interface* trunks to provide GSM connectivity. A *SIMBox* incorporates several SIM cards; therefore each time an employee calls a subscriber of a mobile network, a SIM card of that mobile network in the *SIMBox* is associated and used to reach the subscriber. The SIM card number can be considered as *a temporary identifier* used by the employee to reach the mobile network.

2) Mobile Operators: They represent service providers in the world of mobile networks. As such they put at the disposal of customers a service portfolio. This portfolio is a packaging of capabilities resulting from the engineering of a set of resources and equipments. They administer a set of equipments and service-specific resources, such as billing means, authentication procedures, and customer profiles databases, which interact for the delivery of value-added services.

Mobile network operators are wireless carriers that own and maintain their towers and all associated equipment. At the opposite, we have *Mobile Virtual Network Operators*, which enter into a business agreement with mobile network operators to obtain bulk access to network services at wholesale rates and then sets retail prices independently.
Mobile operators exchange with other actors to have a broader scope of action for their customers. In this vein, we have *roaming partners* and *international carriers*. With the help of roaming partnerships, a mobile operator can enable its subscribers to use its services in geographical areas that it does not cover, using the same handset and the same telephone number.

3) Regulators: Either *ministries* or *independent regulators* regulate mobile operators' activities and partnerships in some countries. That is because governments in most countries see telecommunications as an essential public service and want to retain a regulatory role to ensure telecommunications services are supplied in a manner consistent with the national perception of the public interest. However, some telecommunication regulation activities are seen as brakes as they are sometimes more damaging than beneficial to the development of national telecommunication services [15].

4) International carriers: The international telecommunications system consists of international carriers interconnecting with each other to exchange international telecommunications traffic. Many of these carriers are also responsible for providing telecommunications facilities and services inside their own countries. Some are government departments, while others may be statutory bodies or even private companies. International carriers have routes to termination or transit countries that they buy and acquire through partnerships and resell to others. Therefore, the route followed by call traffic is similar to carrier-to-carrier hops from the originator's mobile operator to the destination's mobile operator. These routes constitute the service international carriers offer to mobile operators or businesses, besides other ancillary services.

Interconnection between carriers is a very sensitive point. It follows certain principles governed by an agreement that provides terms and conditions. Those agreements deal with some specific essential issues including the traffic measurement, the *Points of Interconnection* between carriers and the quality of service standards. An interconnection can be set up with satellite links, submarine communication cables, data links, fiber rings, and such, each impacting on the *type*, the *price*, and the *quality* of the route. With regards to the *route types*, we have white, grey, and black routes. A route is considered white when no illegal (black) action is taken all over the interconnection path. On the contrary, grey routes are arrangements where, at one point in the route of a call, illegal action is taken so that even though both sides of the call look legitimate (white), the call is grey. In black routes, both source and destination use unconventional interconnections.

Subsequently, the telecommunication market is very dynamic. A mobile operator can have interconnections with several or even hundreds of carriers from which it has to choose route calls across the world. Moreover, the quality and the price of these routes differ and can change from week to week for the same carrier. In order to keep up with changes, routing reconfigurations are required in shorter timeframes. That is why *Least Cost Routing* algorithms [16] and equipments are set up within a telecommunication carrier to optimize connections between several other carriers as well as

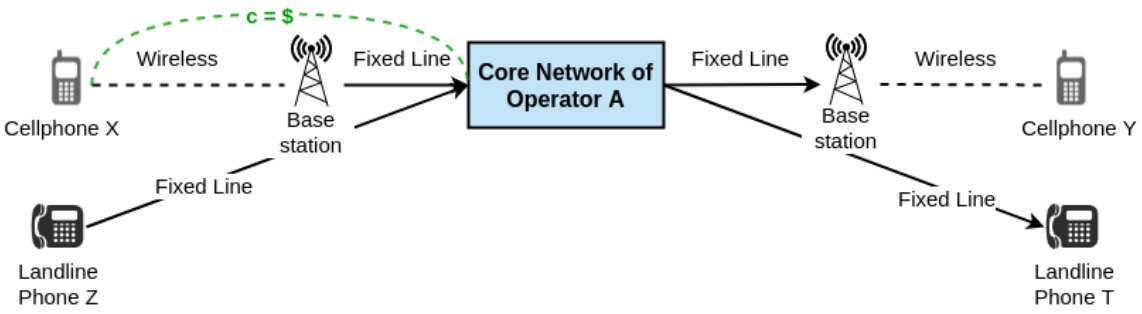

Fig. 3. On-net call scheme

to ensure efficient use of the existing network infrastructure to maximize operators income.

C. Call routing overview

This section presents an overview of a call's routing by distinguishing different carriers that can be involved. At the operator level, we use the terms "On-net" and "Off-net" to distinguish whether a call or another messaging type is made within its network or not. An on-network (on-net) call occurs when the message originates from a customer and terminates at another customer of the same operator. It does not matter if the both ends are using the home network of the operator or another network of a different provider for roaming cases. At the opposite Off-network (off-net) applies when the call or the message is made to a different operator's customer in the same country or abroad.

Figure 3 shows a typical on-net call flow. The call can be from/to a landline phone in the PSTN network or a cellphone connected in the operator's wireless network. In the Figure, cellphone X transmits a call signal to the nearest base station of the operator A. The base station passes the call through the core network of the operator A, to the central switch where the destination party, i.e., the cellphone Y, is recognized as being a customer of operator A as well. The switch sends then the call to the base station where the cellphone Y has made contact, and finally, that base station sends the call to the cellphone of Y. Customer X will get billed for the call. In this example, only the beginning and the end of the call are radio signals (wireless). The in-between steps of the signal are passed through fixed lines, such as glass-fiber. For heterogeneous calls where one end uses wireless technology and the other uses a landline phone, the call is considered on-net if all the installations are made by the same operator. The call transits through the *Gateway Mobile Switching Center* gateway (see Figure 2) to join the landline network.

In Figure 4 an international off-net call flow is depicted. A customer X of the operator A calls a customer Y who has a subscription at the operator B network in another country. Similarly to on-net calls the call can be originated from a mobile or a fixed-line. Cellphone X transmits the call signal to the core network of the operator A through the base station nearest to A. The switch of operator A recognizes the receiving party to be a customer of operator B. Amongst the possible international carriers for call routing, operator A chooses carrier 1 based on the results provided by the Least cost routing team or according to internal policy. The transit arriving at carrier 1 is sent to operator B via carrier 2 according to a routing mechanism similar to that adopted by operator A. Carrier 2 has a direct connection with operator B, it thus sends the traffic directly to the operator who forwards it through the base station where cellphone Y is connected. This example draws a white call routing because all links between international carriers are conventional links (satellite link or submarine cables, for example) established by inter-operator agreements. The example assumes there are only two intermediate hops between operator A and operator B, but this value varies according to existing partnerships between carriers.

D. Money flow

Generally speaking, billing for international calls is higher than for country-wide on-net calls. This is because the routing of international calls very often involves international carriers and, therefore, wholesale billing. An international call travels over multiple intermediate operators before reaching its destination country and operator. Each of these transit operators gets a share from the call revenue, referred to as *settlement rate*, for passing over the call, and the local operator in the destination country receives a *call termination fee* for terminating the international call on its network. In Figure 4, the green dotted line represents an example of money flow. Operator A first gets a *collection charge* $(c_1 - c_2)$ from the subscriber : he receives a charge c_1 from the subscriber and uses c_2 for the termination of the call by Transit Carrier 1. As well, each time the traffic passes by a transit carrier, it keeps some amount of the charge it receives and uses the remainder to ensure the call termination. Finally, what operator B gets (c_4) represents the termination fees. Practically, the routing of a call is very often opaque. Each operator only knows the next hop of the upstream and downstream routes, as well as the originating and the destination number [17]. In some case, the originating number can be absent or incorrect.

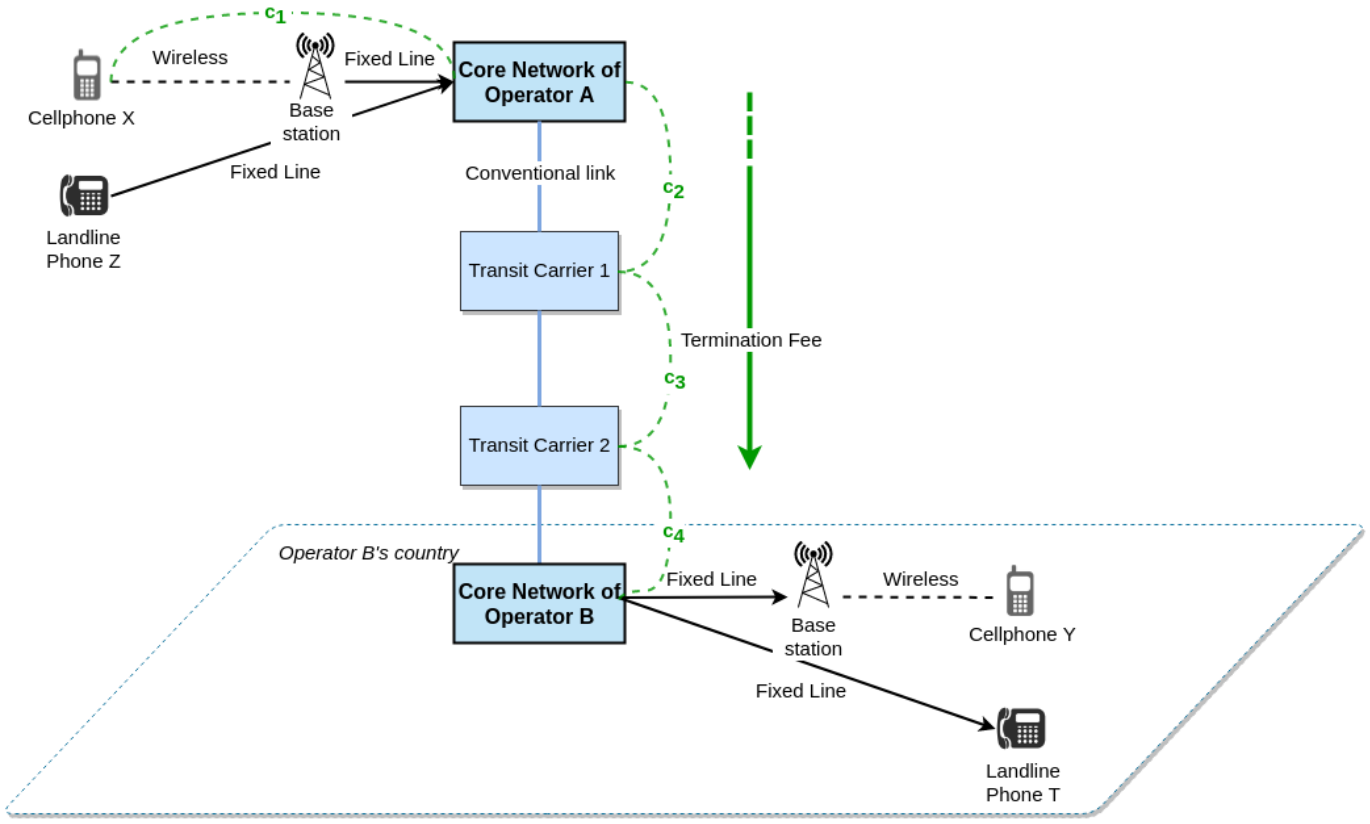

Fig. 4. International off-net call scheme

III. The *SIMBox* fraud

In this section we provide a comprehensive overview of *SIMBox* fraud from an in-depth review of the literature as well as specific research we carried out with *SIMBox* manufacturers. We use the 5-layer taxonomy defined by [17] for telephone frauds as shown in Figure 5. Therefore, we shed light on the different techniques and schemes used to commit fraud, how fraudsters benefit from them, and finally the factors that enable and facilitate the fraud.

A. *SIMBox fraud scheme and techniques*

SIMBox fraud consists of deviating call traffic from the conventional routing routes to the VoIP network using the appropriate gateways. Its scheme can be broken down into four steps, summarized in Figure 6. *(1)* a call (mobile or landline) is emitted from one country to another and transits through regulated routes until a fraudulent carrier. *(2)* The fraudulent carrier uses a VoIP gateway to route the traffic through the VoIP network to a country where fraudsters partners have a *SIMBox*, and the traffic is received at the *SIMBox* level (*manipulation of call routing*). *(3)* The *SIMBox* reconverts the traffic to a mobile call by using a SIM card as the origination of the call. *(4)* The call emitted by the *SIMBox* is routed to the recipient of the call. The call transits through the *local operator* providing the network of the SIM card used by the *SIMBox*.

Two cases should be considered in this scheme.

The first and most recurrent case is when the *SIMBox* is located in the destination country of the call, i.e., operator B's country is equal to fraudsters partner's country in Figure 6. Either a SIM card from operator B will be used to terminate the call or a SIM card from a competing operator if the charges between the two operators are low enough. For this, fraudsters obtain large amounts of prepaid SIM cards with low prices or toll-free from concerned mobile operators and use them to initiate disguised on-net/off-net national calls. In this way they minimize the cost of terminating calls through the *SIMBox* and thus maximize their revenues. SIM cards can as well be obtained illegally through theft or cloning (*superimposed fraud* [18; 19]) or by impersonating existing accounts (*subscription fraud* [20]). Therefore, it is advantageous for fraudsters to have a partner in the country where the call is terminated. The reasons are threefold: *(1)* SIM cards can be easily and cheaply obtained; *(2)* calls can be disguised as on-net calls; and *(3)* operator B does not receive the termination charges that were supposed to be due to it for routing the call to cellphone Y, but rather these charges are diverted to the fraudsters.

The second case to be considered is when the *SIMBox* is located in a country other than the country of destination of the call, i.e., operator B's country is different from fraudsters partner's country in Figure 6. This is hardly mentioned in

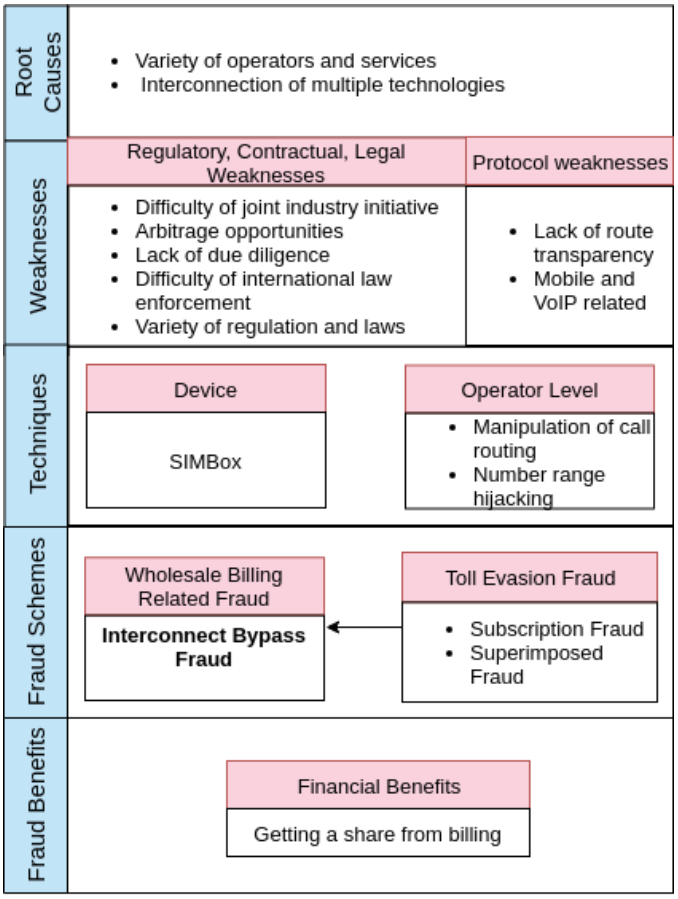

Fig. 5. [17]'s telephone fraud taxonomy that we adapted for *SIMBox* fraud

the literature, but such cases may arise where fraudsters don't have a partner in the country of destination. To better describe it, a fraudulent carrier receives a call to a country with high termination charges. He does not have a partner in that country and is interested in termination charges. Therefore, he goes through a partner he has in another country from which international calls fees to the destination country are lower; passes the traffic to the partner to be converted by the *SIMBox* into an international call to the destination country. Here, although the generated call may be expensive, fraudsters benefit from the difference in termination costs. They leverage *arbitrage*[3] *opportunities* to always make the most money with the different routes they have, based on the deployment of their partners.

Usually, fraudsters form a fraud network with one or a few international carriers and many partners in different countries. With such a network, they set up a system to prevent possible detection methods coming from operators or businesses. They pass voice traffic through VPN tunnels to avoid detection by the Internet service provider and compress the voice traffic using appropriate codecs.

[3]Arbitrage as a concept in economics, is the manipulation of price discrepancies in different markets. Especially in call routing, arbitrage consists of routing traffic via an intermediate country to take advantage of the differences in settlement rates.

B. How do fraudsters benefit?

The main motivation for the *SIMBox* fraud is financial. Fraudsters aim to obtain *a share from international call billing*: the termination fees. The more costly they are the better it is for fraudsters. To obtain money from this activity, fraudsters insert themselves into the voice traffic termination route so they can divert the conventional route and collect the corresponding fees. This can be done either *at the origin of the call* by inducing subscribers not to send the traffic to their operator but to send it directly to them or *somewhere along the route* of the call as fraudulent carriers. In both cases, the typical behaviour of fraudsters is to pretend to be able to route calls at costs low enough to be of interest to their audience.

Based on this principle, we exemplify in Figure 7 different ways of routing international calls. The Figure helps identifying the actors who may be fraudulent (represented in a red circle) and those who may be usurped by fraudsters (represented in a blue circle).
The call is diverted *at its origin* when the caller passes the traffic directly to fraudsters. Subscribers are motivated by the desire to make international calls at lower costs than those offered by operators. They can do this through VoIP using OTT (Over The Top) applications or through international calling services such as *Override providers* and *International calling cards*. Override providers[4] offer low-cost calls to subscribers with an account on their application. They either charge per month or sell call minutes. They are not explicit on the method used to route calls and may be fraudulent. International phone cards are prepaid cards that give access to international calling times directly to end consumers at discounted rates. These cards provide a phone number that the user must dial, and a PIN code requested to access the service. The user is then asked to enter the destination number and is connected to its recipient. International phone cards can be purchased at common commercial platforms.

Through *number range hijacking* fraudulent carriers manage to insert themselves *somewhere along the route* of termination of calls. *Number range hijacking* consists of dishonest carriers advertising very cheap rates for a destination number range by which they attract traffic from other operators. Thus, usurped operators transfer calls to such a number range to those carriers that will be hijacked and routed/terminated fraudulently. This transfer is facilitated by a *lack of due diligence* in partnership agreements between carriers. It is difficult to be detected as there is no mechanism in telephony networks to authenticate the owner of a number range directly or to check if an operator has the connectivity to route a call to a number range (*lack of route transparency*).

The payment of fraudsters depends on the means they use to insert themselves into the routing of voice traffic. In the first case studied above, they are paid by selling their international calling cards on commercial platforms or making monthly

[4]RebTel(https://www.rebtel.com/en/) is an example of override provider

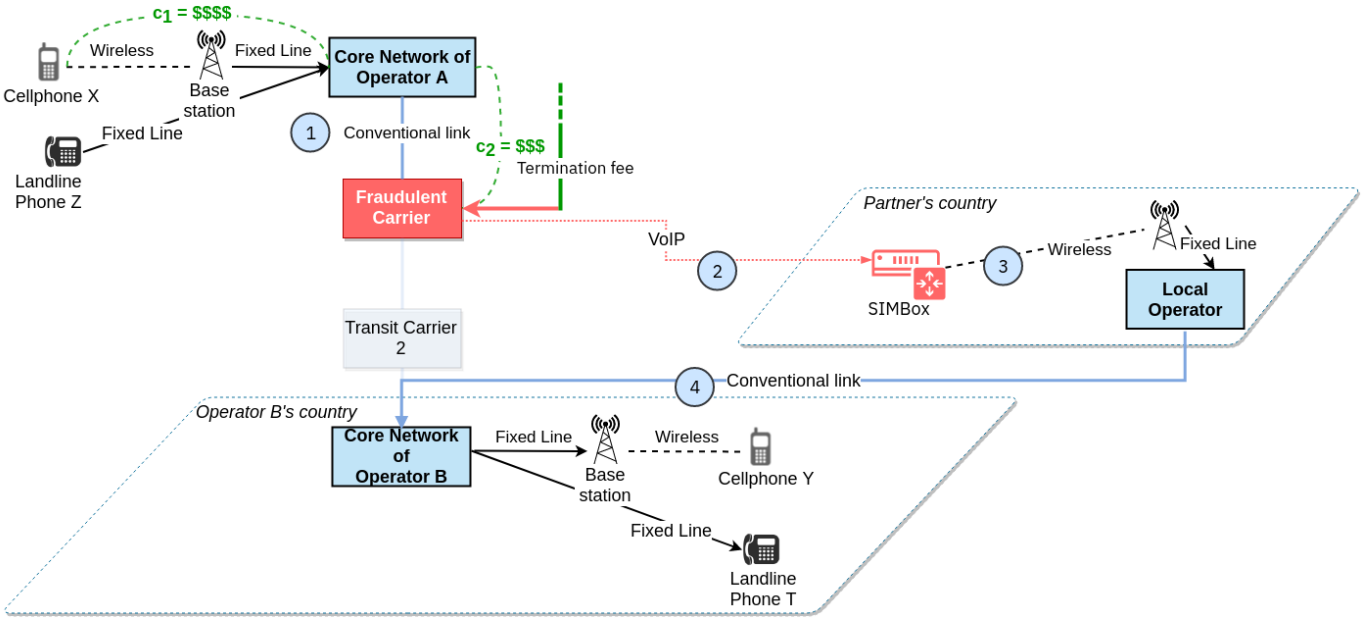

Fig. 6. Generalized scheme of a call flow in case of a *SIMBox* fraud

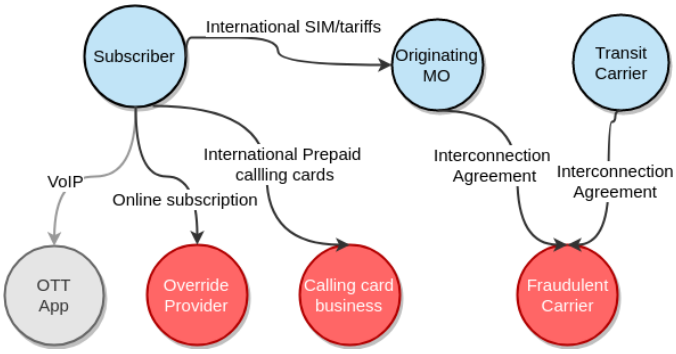

Fig. 7. Diagram of fraud intrusion in an international call routing path. Red circles represent actors who may be fraudulent. Blue circles represent actors who may be usurped by fraudsters.

withdrawals from their subscribers if they act as override providers. On the other hand, if they are fraudulent carriers, as in the second case studied above, they will be paid according to the means and policies specified in the interconnection agreement. Finally, the carrier receiving and routing the voice traffic is responsible for ensuring payment of its collaborators managing the *SIMBox* in destination countries.

C. What motivates and facilitates the fraud?

SIMBox fraud is reported to cause significant losses to telecom operators and its proportion is still considerable to the extent that in 2019, more than 80% of African mobile operators have been victims [21]. Bypass fraud appears in the 2011 Communications Fraud Control Association report as the top 5 most damaging telephone fraud, responsible for 7.2% of telephone fraud losses, moves up then to the second place in the 2015 report with 15.7% of losses and remains there in the 2017 report with 14.62% of losses. Such figures raise questions about the factors in the telephone ecosystem favouring the fraud, despite available detection and prevention solutions. These factors are likely to be related to the environment in which the scam quickly unfolds. Based on the common knowledge of *SIMBox* fraud prevalence in Africa and the Middle East [22], the following factors can be determined.

(Factor 1): The huge difference between International Termination Rates (ITRs) and Local Termination Rates (LTRs). The *SIMBox* fraud's fuel and primary motivation lies in the fact that there is a difference between ITRs and LTRs. Generally, charges associated with international calls are greater than those of on-net calls. Specifically, termination fees or destination country' ITRs have a considerable share in these charges. Indeed, the termination of an international call can be divided into three parts, each of which implies a certain cost: (1) the nationwide transmission of the call from its originating terminal to the originating country international gateway, (2) the routing of the call from the international gateway of the originating country to the destination country, and (3) finally, the access to the destination mobile operator infrastructure for the termination of the call. Depending on the pricing structure of the terminating operator, the last part of the call is the most expensive as it passes through all the switching levels in the destination country [4]. As illustration, in Figure 4, the termination fees c_4 would generally be greater than the settlement rates obtained by transit carriers 2 (i.e., $c_3 - c_4$) and carrier 1 (i.e., $c_2 - c_3$), respectively. Other reasons explained below may justify this.

In their role of protecting the rights and interests of service providers and consumers within their countries, governments and National Regulatory Authorities would instead encourage high ITRs as this could serve as funding for the development

TABLE I
TOP 5 AFRICAN COUNTRIES DESTINATION FOR GSM TERMINATION IN 2020 ACCORDING TO [24]

Country	Population	Mobile penetration	Local call rate
1- Nigeria	203 million	75%	\$0.042- \$0.048
2- Côte d'Ivoire	25.9 million	128%	\$0.1
3- Tanzania	58.9 million	90%	\$0.12
4- Mali	19.9 million	91%	\$0.17
5- Algeria	43.5 million	117%	\$0.02-\$0.22

of the domestic network and thus act as a leverage for the internal economy. Besides, ITRs have no impact on domestic subscribers. Moreover, most voice traffic originates from rich countries; money flows are thus, from developed to developing countries, and high settlement rates favor the recipient countries. All these justify the tendency for developing countries in Africa and the Middle East to increase their termination fees. [23] presents the average price per minute that United States' carriers pay to foreign carriers to terminate traffic in their countries. It reveals that Africa, Caribbean, and Middle East have the highest ITRs over time, increasing in particular, for Africa. This significantly favors fraudulent activity in these areas in close correlation with what [22] mentioned.

On the other hand, to cope with a national competition, telecom companies have aggressive packages and promotions in the form of bundles and unlimited access to calls for a given period, mostly on the same network, aimed at lowering the churn rates to attract new customers. This widens the difference between ITRs and LTRs for countries with high termination fees and particularly attracts fraudsters' attention. A GSM termination business realized a classification of the top 5 *SIMBox* fraud destination countries in 2020 [24]; results are presented in Table I. The more mobile penetration and the lower LCR, the better it is for fraudsters.

(Factor 2): An easy access to prepaid SIM cards. SIMBox fraud requires a large quantity of SIM cards. SIM cards are used to re-originate bypassed international calls as local calls. A huge amount of SIM cards is necessitated for the activity, to handle multiple calls increasing the potential of the fraud and, therefore, revenues, to perform SIM rotations and avoid suspicion, and to continue providing service in case one or more SIMs are blocked. Indeed, the *SIMBox* allows to conduct several calls at the same time and generally rotate the use of SIM cards so that ideally, one SIM card is randomly chosen among a group of SIM cards for each call, limiting the detection of fraudsters (see Section V-A). Besides, at any time, one or more SIM cards may be blocked by anti-fraud activities and should be replaced to maintain the activity.
Moreover, fraudsters are more likely to use prepaid cards to limit their traceability as postpaid SIM cards require precise information on the SIM holder. Such information could facilitate fraudsters' arrestment if their SIM cards are intercepted. From this analysis, it is clear that areas allowing easy access to large quantities of prepaid SIM cards would be areas where fraudulent *SIMBox* activity could quickly spread.

Here again, Africa and the Middle East are the areas most concerned. Indeed, on average, 94% of mobile subscriptions in Africa are prepaid, and there are about 80% in the Middle East [25]. The following facts can justify these high percentages: *(1)* prepaid SIM cards are easy to obtain and give more flexibility to their holders (in fact, 75% of mobile subscriptions in the world are prepaid [25]) and *(2)* the financial inclusion rate in Africa is relatively low[5]; thus it would be difficult for mobile operators to ensure regular and reliable payment of postpaid SIM cards.

(Factor 3): The corruption of the telecom industry.
SIMBox fraud has been a significant fraud issue for at least ten years partly because of the nonchalance of telecom operators who do not use all possible means to overcome it. Although operators are motivated by severe money lost, individualism and the search for personal profit give rise to corruption infiltrated at several levels in the fight against fraud [27]. We identified the following levels. *(1)* In the wholesale telecom market, operators know that low priced routes (20 to 70% off) are more likely to be grey routes. Carrier teams may receive a bribe to buy these routes and tangle them with legitimate ones. In some cases, conflicts may be created within an operator because the carrier team is buying routes that the fraud prevention team is trying to detect. *(2)* Concerning the acquirement of SIM cards in markets where legal identification is required, fraudsters collude with re-seller kiosks. They pay re-sellers to turn a blind eye on fake identity cards (IDs) or to sell SIM cards pre-identified to other subscribers. Indeed, re-sellers can use the IDs of their clients to register SIM cards to their identity without their consent/knowledge and reserve them for fraudsters. This can happen with no suspicion from the mobile operator hiring the re-sellers because operators are satisfied of the purchase of their SIM cards and encourage it. *(3)* Vendors or anti-fraud private companies can also run bypass termination to their profit. They want to be hired by mobile operators and boost *SIMBox* traffic in a mobile network before a proof of concept to inflate the apparent size of the fraud as well as to ensure instant results of their proposed solution. Similarly, within the operators, senior managers could have their own termination business and influence the fraud-prevention team by limiting their activities to avoid being detected. *(4)* At last, even the fraud-prevention team can be corrupted. Fraudsters make enough money to influence some fraud prevention team agents not to block their SIM cards if detected.

(Factor 4): The prevalence of different telecommunications regulatory policies. Telephony ecosystem comprises a variety of regulation policies and law, and the notion of legality can significantly differ depending on the country and the communication medium [17]. That's how the grey routes came into being: on the one hand we have a broad regulation that allows to legally send traffic to a VoIP carrier and on the other

[5]according to the World Bank Group [26] there would be an average of about 40% of adults with a mobile-money account or in a financial institution

hand we have a strict regulation that only considers the traffic of a few legacy operators as legitimate (e.g. USA and India as described in [28]).

Thus, *SIMBox* fraud is fuelled by many actors, legally operating in their country. This is the case of the *SIMBox* manufacturers who produce, advertise, and sell these devices. A *SIMBox* device can be easily ordered online on Amazon or Alibaba by any individual. Some GSM gateways providers such as Sysmaster [29] explicitly provide procedure and methods to counteract detection strategies from operators and authorities. Similarly, some international transit carriers offer their support services to internauts interested in terminating GSM traffic using the *SIMBox* in the context of a partnership with them. This is the case of Antrax [30], a company based in LatVia, which has been operating termination in 74 countries since 2017.

On the other hand the fraud is outlawed and actively combated by many countries whose impact is directly negative for their economies. For example, in Pakistan, the rapid advance of technologies developed by the gateway manufacturers is seen as an "arms race" against which regulators have reacted by introducing additional fraud prevention and detection. Pakistan reportedly went as far as installing deep packet inspection technology that would block all unauthorized virtual private networks in the country [23]. In other countries, bypassing the official termination can be considered a violation of local laws. The National Communications Authority of Ghana regularly arrests offenders, some of whom were sentenced to five years in prison. Some countries ban the usage VoIP usage to protect their revenue from bypass termination [31].

Furthermore *the lack of cooperation* of law enforcement authorities makes identification of fraudsters difficult, even when the fraud is detected [32]. Despite the presence of international organizations, there is a *lack of joint industry initiative* to fight fraud. Due to the privacy concerns and competition, operators are usually not willing to share their pricing terms, routing options or fraud-related findings [17]. Besides, not all operators have the same incentives to fight fraud. Indeed, sometimes a competing operator can profit from the losses and the bad reputation induced by the fraud. In other cases, fighting small scale fraud can be more expensive than the losses due to the fraud itself.

(Factor 5): The success of the SIMBox fraud. The success of the *SIMBox* fraud acts as a factor in its perpetuation. The Communications Fraud Control Association reports the losses caused each year by this fraud as being enormous (in the order of billions of dollars), and it is essential to note that from the fraudsters' point of view, these losses are a profit and a real motivation. With such motivation, it is immediately understandable that the actors involved in this fraudulent activity receive a significant remuneration that would be difficult to drop overnight. In this point, we agree with [33] on the observation that *SIMBox* fraudsters see their activity as a business opportunity: they invest their time and money in it, and they receive a profit.

Fraudsters are rarely arrested, so they are not in any real danger. Anti-fraud teams mostly act by blocking SIM cards [34]. When one or more SIM cards are blocked, all fraudsters have to do is identify the flaw, adjust, and replace the SIM cards to continue the activity. Besides *SIMBoxes* are getting cheaper and more featured; more cheap-international-calling Apps are present on the App Stores (for IoS and Android users) to easily get traffic from users. This motivates other individuals (e.g. unemployed people) to take up the activity in order to benefit from it.

Based on these facts, it can be conjectured that the fight against *SIMBox* fraud is a war that is not yet over. Teams and technical means are being put in place to refine the strategy at the slightest flaw and make it unshakeable because a great reward is behind it. It is more than necessary to realize this to better measure the enemy's strength to adjust the magnitude of the riposte.

IV. *SIMBox* ARCHITECTURE

Figure 8 depicts the functional architecture of the *SIMBox*. It represents a fragmented architecture as all the different components of the *SIMBox* are distinguished. However, smaller architecture variants, performing the same role but with limited functionalities also exist. In this vein, Table II groups the different *SIMBox* architecture variants with some providers. Overall, the *SIMBox* architecture consists of three main components : *the GSM gateway*, *the SIMBank*, and *the Control server*.

A. The GSM gateway

This component allows call termination from the VoIP network to the cellular network and vice versa. It receives VoIP traffic from a softswitch, transcodes it, and ensures its routing on the wireless cellular network. To this end, it supports several VoIP signaling protocols, including SIP and H.323, and several voice codecs.

The GSM gateway includes on-board *GSM modules* each maintaining a *GSM channel* to establish communication with the GSM network. GSM modules can operate at different frequencies corresponding to different cellular network standards, and each has an integrated firmware and an IMEI identifier. They are managed by the *Voice server*, which receives call routing requests from the softswitch, communicates with the GSM board[6]'s hardware to ensure the routing of a call through a specific GSM channel selected according to pre-established routing policies.

The *SIM client* is responsible for assigning SIM cards to GSM channels for call routing. As soon as a GSM channel is released (not bound to any SIM card) the *SIM client* makes a request to the system's *SIM server* for a SIM card. A *voice channel* is the combination of a GSM channel and a SIM card. Once a voice channel is created and registered in the cellular network, it is ready to process calls and perform other activities (send SMS or USSD requests, etc.). The number

[6]Set of GSM modules of the GSM gateway.

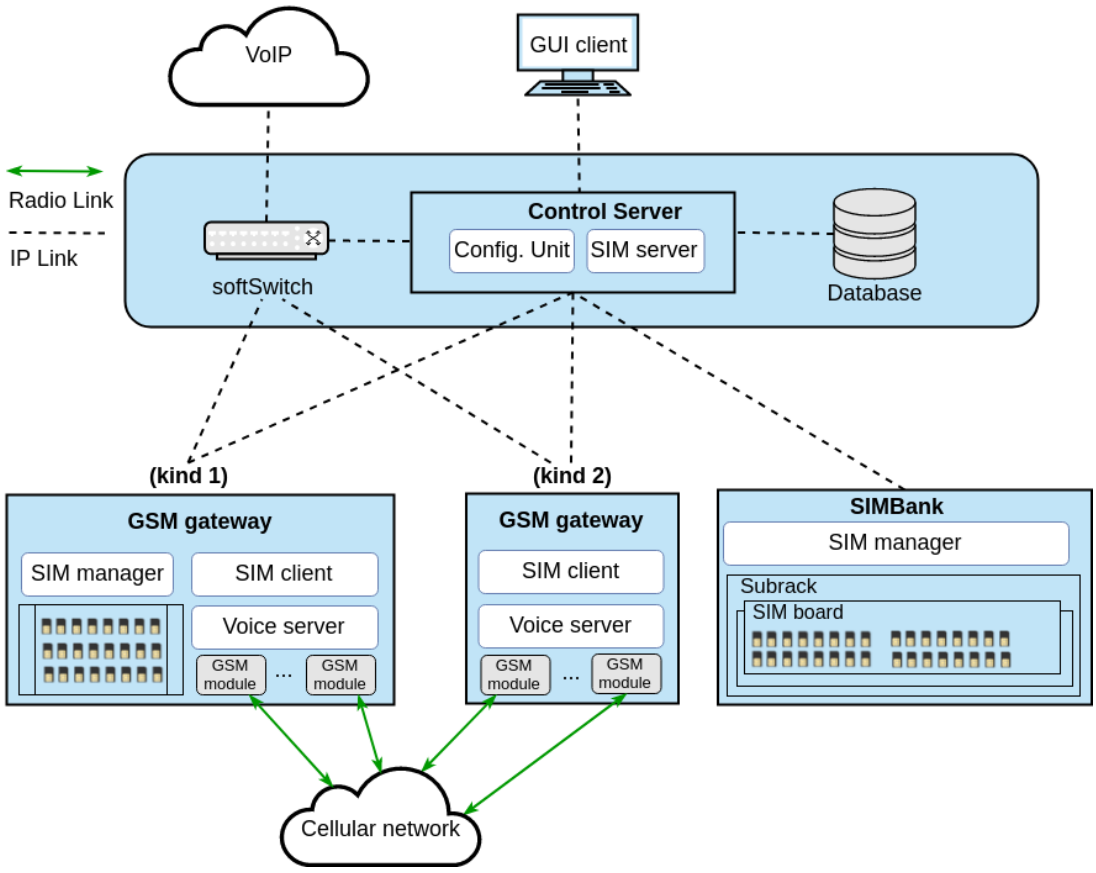

Fig. 8. *SIMBox*/VoIP GSM gateway installation functional architecture

TABLE II
SIMBox ARCHITECTURAL VARIANTS AND SOME PROVIDERS

	GSM gateway		SIMBank		Control server		Some providers
	function(s)	number	function(s)	number	function(s)	number	
1	Voice server SIM client SIM manager (present or not)	1 to n	SIM manager	1 to n	Config. unit SIM server	1	Hybertone Antrax Ejoin 2N VoiceBlue Portech Dinstar
2	Voice server SIM client SIM server Config. Unit	1	none	0	none	0	Hybertone Antrax Ejoin 2N VoiceBlue Portech Dinstar Hypermedia
3	Voice server SIM client SIM manager (present or not)	1 to n	Config. unit SIM server (max. 32 SIM cards)	1	none	0	Hybertone
4	Voice server SIM client SIM manager (at least one)	1 to n	none	0	Config. unit SIM server	1	Hybertone Antrax Ejoin 2N VoiceBlue Portech Dinstar

of GSM channels within the GSM gateway determines the number of simultaneous calls that can be routed through the device. A GSM channel is characterized by:

- An IMEI as aforementioned.
- A status indicating the connection state of the current GSM module to the cellular network, such as: *mobile registered*, *unregistered*, or *no SIM card*, if no SIM card is connected to it.
- The IMSI of the SIM card connected, if any.
- The remaining call duration time of the channel.
- The network provider (mobile operator) of the SIM card connected, if any.
- The current base station ID if mobile registered.
- The signal strength or the *Received Signal Strength Indicator* of the current cell.
- The *Bit Error Rate*, which indicates error rates between the GSM module and the base station.
- The *Answer-Seizure Ratio* of the channel.
- The *Average Call Duration* of the channel.
- The *Post Dial Delay* of the channel.
- The call status of the channel such as: *Idle*, meaning there is no call on the channel; *Processing*, meaning a call is connecting; *Alerting*, meaning the destination is ringing; *Active*, meaning a call is connected; and *Calling Waiting*, meaning the gateway is receiving another call during a running conversation and implement a call waiting service.
- The Idle time indicating the time elapsed since the last call.

There are two kinds of GSM gateways, as shown in Figure 8. One of them contains slots for inserting SIM cards. GSM gateway models of this kind can be used in a standalone mode as they can carry out the SIMBank (with a *SIM manager* component) and the control server functions (architecture 2 of Table II). However, GSM gateways of both kinds can be used in combination with other components to build substantial architectures with more functionalities (architectures 1, 3 and 4 of Table II).

From an external point of view, the GSM gateway is an off-the-shelf equipment, with one to many plugged GSM antenna(s)[7]. It has one or two Ethernet port(s) intended for intranet or internet connection and network sharing, a USB port to which a modem or flash drive can be connected, and a *Direct Current* port for power supply.

B. The SIMBank

The role of the SIMBank is to hold a bundle of SIM cards used by the system for call routing. The SIMBank is not essential at the *SIMBox* architecture because some GSM gateway models integrate SIM slots (as aforementioned). However, in contrast to GSM gateways, the SIMBank has the advantage of offering remote operation capability, which eases management tasks, minimizes the maintenance expenses, and solves the SIM-blocking issue[8]. A SIMBank allows to install and manage SIM cards of different mobile operators and enhance the working of several GSM gateways placed in different locations. Depending on its model, a SIMBank can comprise between 32 to 256 SIM slots. Nevertheless, several SIMBanks can connect to one system providing the ability to use practically unlimited SIM cards.

SIM slots within the SIMBank are managed by a *SIM manager*, whose central role is to communicate with the SIM board's hardware and to transfer data and status obtained from SIM cards to the *control server*.

Similarly to GSM gateways, some SIMBank models can work in a standalone mode by carrying out the *control server*'s function (architecture 3 of table II). In this mode, the SIMBank's *SIM manager* behaves as a *SIM Server* and provides SIM cards to the GSM gateways. However the SIMBank is limited to 32 SIM cards and only basic configuration functionalities can be performed. Only one SIMBank can be used in this mode and serves several remote GSM gateways.

Finally, to an external point of view, the SIMBank can either be a sub-rack with one or many SIM boards, each SIM board having independent control of many SIM cards, or a lighter unit with multiple SIM slots that is easily transportable for the latest models.

C. The Control server

The control server is the central component of the architecture. It is composed of a Linux-based core with web and database systems. It is often hosted on a *Dedicated Server* or on a *Virtual Private Server* (VPS) on the cloud to be always available and easily accessible. It has three leading roles.

- *Role 1:* It communicates with all the critical components of the architecture to provide centralized remote administration, including visualization of the general state of the system and the configuration of its various components. To this end, the control server persists all necessary data (settings, statistics, logs, etc.) in a database. It communicates with a remotely deployed GUI client through which monitoring features are performed. It also synchronizes all functional units integrated into the system and coordinates the interaction between them.
- *Role 2:* It sets up *voice channels* for call routing by leading the binding of SIM cards to the remote gateways' GSM modules. The establishment of voice channels is a continuous process that stands whether there is a call or not. It is triggered by *SIM clients*' requests to the *SIM Server*. This process is done either manually - by selecting a specific SIM card for a GSM channel - or using SIM and GSM grouping. For the latter method, we identified two ways of binding through groups.

[7]Each GSM module includes a GSM antenna for radio communications.

[8]In an architecture with several GSM gateways at different locations and distant from each other, a SIM card within a GSM gateway that is blocked by anti-fraud activities could require a considerable move to remove it from the system. However, with the SIMBank all SIM cards used by the different GSM gateways are centralized, and therefore the management of the blocking of a SIM card is simplified.

- *(First way)* GSM channels are combined into GSM groups and SIM cards into SIM groups. An administrator creates these groups. A GSM group is a set of GSM channels that can be located on different GSM boards and managed by different *voice servers* but have shared configurations. A GSM group must have a unique name.
  As well, a SIM group is a set of SIM cards, which have shared configurations and a unique name. Similarly to GSM groups, SIM groups can contain SIM cards located on different SIM Boards and managed by different *SIM managers*. SIM groups have a variety of configurations that determine the behavioral pattern of SIM cards.
  Therefore, the administrator makes links between a GSM group and one or more SIM groups. Links indicate that SIM cards from the SIM groups can be bound only to GSM channels from linked GSM group according to SIM groups' rules.
- *(Second way)* Bindings are based on a *scheduling group* which includes several SIM cards and GSM channels. Similarly to *(first way)*, SIM slots can be located on different SIM boards and managed by different *SIM managers*. As well, GSM channels can be located on different GSM boards and served by different *Voice servers*. A scheduling group implies that the SIM Slots and GSM channels within it can be dynamically bound together by the system following the group scheduling rules.

- *Role 3:* It performs routing of VoIP traffic to GSM gateways by interacting with a softswitch. The softswitch interfaces the system with the external VoIP network and receives SIP or H.323 calls traffic, which are transferred to the system according to defined policies. For this purpose, when configuring a GSM gateway, SIP accounts are created and registered on a softswitch (acting as a SIP server) that transmits VoIP traffic per account. Therefore, we have several possibilities:

(i) One account is created for each system's GSM channel, and the traffic will be directly sent to a specific GSM channel eliminating the need for selecting a GSM channel;

(ii) One account is created for all the system's GSM channels or several accounts are created for groups of GSM channels;

(iii) No account is created and instead the gateway can use SIP trunks to receive VoIP calls.

The second option is commonly chosen as it is cheaper than the first one. However it requires a policy to choose the *voice channel* that will terminate the call. To this end, one of the following routing policy is applied :

- *In-Turn:* traffic is routed to the first released voice channel.
- *Balance:* traffic is routed to the fewest historical calls channel.
- *Sequence:* traffic is routed by ascending voice channel.
- *Random:* the voice channel is randomly chosen among available.

Allowed prefixes can be set as well to a GSM channel for intelligent routing. This way, the GSM channel will only accept calls to numbers with a specific prefix, and others will not be routed to it. As an example, if the prefix +237 is set on a GSM channel, it will only terminate calls to Cameroon.

D. Interaction and organization

The communication diagram of Figure 9 summarizes a call termination flow within *SIMBox* components. The following steps can be identified in the Figure. ① A call comes from the external VoIP network to the softswitch. ② The softswitch sends the routing request for the call to the Control server. ③ After processing the request (which includes anti-spam rules, see Section V-I1), the control server responds with the appropriate route (voice channel), if any. If not, the call is dropped. The voice channel is chosen according to the policy defined in the control server's *Role 3* (see Section IV-C). ④ The softswitch routes the call to the *voice server* of the selected voice channel. ⑤ The voice server terminates the call to the cellular network. ⑥ At last, the call connection is completed.

⓪ As a prerequisite step, voice channels are continuously created within the system through *SIM clients*' requests to the *control server* :

- When a *SIM client* requests a SIM card for a released GSM channel, the Control server looks among potential SIM cards (e.g. SIM groups linked to its GSM group).
- These SIM cards are ordered according to a criterion to ensure SIM cards rotation (see Section V-A). According to this criterion, the first SIM card is chosen for the GSM channel, and some checks are made.
- The control server checks the availability of the SIM card according to its activity limitation parameters (see Section V-B). For example, it will check if the SIM card can operate at that time (time limitation) or if the SIM card has not reached the call threshold for the day (parameter limitation per time period). If the SIM card is unavailable, the next SIM card is selected, and the same check is made until a suitable SIM card is found.
- The control server then checks whether the selected SIM card can be connected to the GSM channel according to the SIM card's GSM channel selection configurations for migration (see Section V-C). The control server will choose the first SIM card (according to the criterion) that validates both checks. If no SIM card validates, the voice channel creation procedure is cancelled and can be resumed later.

In practice, installing these different components in the context of a GSM termination business within a country starts with the choice of the different locations to place the GSM gateways. These locations must be crowded places such as

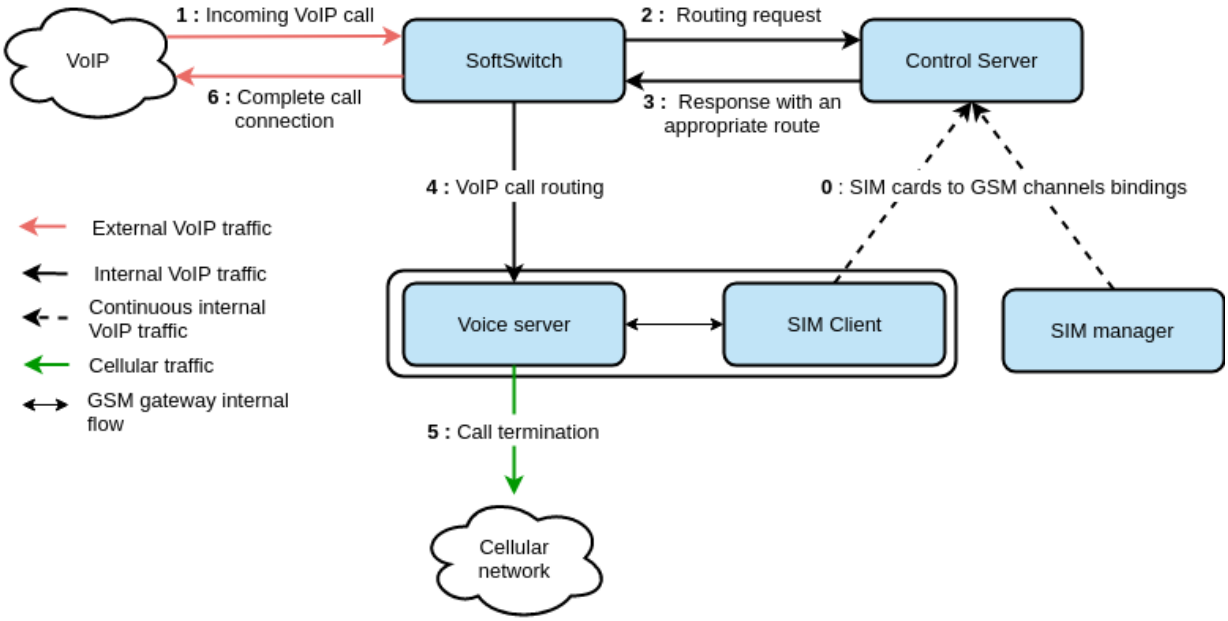

Fig. 9. Communication diagram : main stages of a call flow through *SIMBox* functional components.

city centers, high-rise buildings in the center of the market, station districts, market areas, densely residential districts, or call-center areas. Crowded places enable the camouflage of *SIMBox* calls by the massive flows of calls made in these areas. Once these locations have been chosen, office or apartment locations with stable power and Internet access are rented, and the various GSM gateways are installed. These gateways are then connected to the system by public addresses obtained through NAT. On the other hand, the SIMBank(s) of the system does not emit a GSM signal and can be located anywhere in the country or even abroad. However, the quality of the connection should be good beyond a certain threshold to allow smooth communication with the GSM gateways.

In the Hybertone architecture, for instance, the network environment between the control server, the SIMBanks, and the GSM gateways should meet a packet delay rate of less than 300ms and a packet loss less than 1%. The network bandwidth required, on the other hand, depends on the number of SIM cards used simultaneously. It is at its peak of 11Kbps during a GSM gateway registration. It is preferable to have SIMBanks located where the system administrator resides so that he can easily add or remove a SIM card from the system, when needed. The control server, as mentioned above, is usually hosted on a private server, and its visualization and configuration interface is accessible online. Finally, the softswitches emitting the VoIP traffic to the system are managed by the fraud partners who divert the traffic and send it according to a predefined contract.

V. *SIMBox* FRAUD STRATEGIES

The architecture of the *SIMBox* as described in the previous section depicts the main functionality of call termination and how this is achieved through the *SIMBox*' various components. The newer models of *SIMBoxes* are not limited to this basic functionality, but provide advanced features that help fraudsters in their activities.

Indeed, as mentioned in Section III-C, fraudsters invest enough to obtain large quantities of SIM cards, and it would be a substantial financial blow if their SIM cards were blocked just after some time of activity. To avoid this, they optimize their strategies by simulating user behaviour (e.g. traffic habit) [27]. This is done automatically through various features (strategies) integrated into most new *SIMBox* models. It falls within the framework of what is called in the literature *Human Behavior Simulation (HBS)* [21]. HBS features are thus designed and improved to counteract fraud detection techniques (more details are provided on Section VI where these detection techniques are described).

We present below the different functionalities included in HBS and their current mode of operation in the light of studies we carried out on *SIMBox* models from the most popular suppliers. Table III summarizes such functionalities. Furthermore, we present some features of the *SIMBoxes* not related to HBS, but still relevant to consider, as they help fraudsters avoid detection. Finally, in the last sub-section, we aim to build a history of the temporal evolution of the *SIMBox* from the point of view of these different functionalities.

A. *SIM card rotation*

SIM card rotation allows using multiple SIM cards to operate in one GSM channel. This ensures the distribution of termination traffic among the SIM cards of the system, preventing one SIM card or a small group of cards from being used excessively during working hours or beyond a certain threshold. This will make SIM cards operate in limited hours a day, which simulates the behavior of regular customers. Its

TABLE III
SIMBox HBS FUNCTIONALITIES, PARAMETERS, AND ILLUSTRATIONS

HBS feature	parameter	value	Illustration
SIM rotation	Method for selecting the next SIM card	round-robin method	Hybertone [35], Antrax[36]
		random method	Hybertone [37], Portech [38]
		statistic-based factor	Hybertone [37], Antrax [36], Dinstar [39]
		statistic-based factor per time period	Antrax [36]
	Trigger to switch SIM cards	Threshold method	Hybertone [37], Antrax [40; 41], Portech [38]
		Activity script method	Antrax [40; 41]
SIM activity limitations	Type of limitation	Parameter limitation	Hybertone [37], Dinstar [42], Ejoin [43], Portech [38]
		Parameter limitation per time period	Antrax [40; 41] Dinstar [42]
		Time limitation	Hybertone [37], Antrax [40; 41], Ejoin [43], Dinstar [42], Portech [38]
SIM migration	Method for selecting the next GSM channel	Manually fixed method	Hybertone [37]
		Any except previous	Antrax [44], Hybertone [37]
		Any except previous zone ID	Hybertone [45]
		Any gateway	Antrax [44]
		Specified order	Antrax [44]
Base station switching/locking	Method for selecting a base station	Manually	Portech [38]
		Default	Hybertone [37], Ejoin [43], Portech [38], Dinstar [42]
		Fixed	Hybertone [37], Dinstar [42]
		Random	Dinstar [42]
		Poll	Hybertone [37], Ejoin [43]
		Advanced	Hybertone [37] Dinstar [42]
Changeable IMEI	Method for changes at GSM channel	Manual editing	Hybertone [37]
		IMEI auto change	Hybertone [37]
	Method for changes at SIM slots	Manual editing	Ejoin [43] Antrax [46]
		Random IMEI	Antrax [46], Dinstar [42]
		Prefix IMEI	Ejoin [43], Antrax [46]
		Registry IMEI	Antrax [46]
		IMEI based on TAC	Antrax [46], Dinstar [42]
The usage of other network services	Services used	Internet	Ejoin [43], Dinstar [42], Antrax [47]
		USSD commands	Hybertone [37], Ejoin [43], Dinstar [42], Antrax [48], Portech [38] 2N voiceblue [49]
		SMS	Hybertone [37], Ejoin [43], Dinstar [42], Antrax [50], Portech [38] 2N voiceblue [49]
Family list	Services used	SMS inter-sending	Ejoin [43]
		Inter-calling	Ejoin [43]
Call forwarding	Forwarding conditions	Unconditional	Hybertone [37], Dinstar [42]
		Busy	Hybertone [37], Dinstar [42]
		No reachable	Hybertone [37], Dinstar [42]
		No reply	Hybertone [37], Dinstar [42]

operating principle is based on the assignation of a group of SIM cards to a GSM channel (through *SIM and GSM groups links* or through *scheduling group*). Designated SIM cards are candidates for the formation of a voice channel with the GSM channel to which they are assigned, and will rotate each other in the process of termination. At the switching of an active SIM card, the method for choosing the next SIM card can be one of the following:

- The *round-robin or constant method* puts all available SIM cards in a circular loop and then selects the next SIM card in sequence according to a particular set up step.
- The *random method* selects the next SIM card randomly from the SIM cards available (not including the active SIM card). In a *scheduling group*, for example, SIM cards and GSM channels are randomly related.
- The *statistic based factor* selects the next SIM according to the ascending or descending values of a particular factor that can be edited. Examples of factors are: the remain talk time of SIM cards, the cumulative call duration, the total calls count and the total SMS count.
- The *statistic based factor per time period* method follows the same principle as the previous method, with the exception that factors are aggregated by time period.

The trigger for initiating a SIM card switching is set according to two methods.

(Method 1) The threshold method allows to set certain limit values for specific factors beyond which the SIM card will be replaced. These parameters are indicated in the column *tresholding for switching* of Table IV. In the same table, the *number of outbound SMS* refers to the total number of text messages sent via SMS during an active SIM session; The *SIM in use duration* refers to the duration of an active SIM session; the *no cell service duration* refers to the duration of not be able to access mobile service, potentially caused by an invalid SIM or weak RF signal; The *number of call*

TABLE IV
RECORDED FEATURES FOR *SIMBox* CONFIGURATION

Parameter	Tresholding for switching	Script-lets composition for switching	Parameter Limitation	Parameter limitation per time period	SMS and call inter-sending	Base station balancing
Number of outbound SMS	x	x		day		
SIM in use duration	x				x	
No cell service duration	x					
Number of call attempts	x	x		day	x	x
Total call duration	x	x	x	day, month	x	x
Total credit used	x		x	day	x	
Number of successful calls		x		day		
Consecutive successful calls					x	
Consecutive failed calls			x		x	x
Consecutive non-answered calls			x			
Consecutive no carrier calls			x			
Consecutive short duration calls			x			
Consecutive fast alerting calls			x			
Consecutive fast answered calls			x			
Consecutive SMS count			x			
Consecutive failed SMS			x			
Total failed to send SMS			x			
Total received SMS			x			
Consecutive registration failures			x			
Consecutive SIP release			x			
Consecutive GSM release			x			
Consecutive module release			x			
Consecutive outbound calls with no ringback tone			x			
Minimum ASR			x			
ASR			x			
ACD			x			
SIM Balance			x			
PDD			x			
Failed to send USSD commands count			x			

attempts refers to the total number of outgoing calls during an active SIM session with the option of including or excluding unanswered calls; And the *total call duration* refers to the total talk time of all calls during an active SIM session. When at least one of the above parameters exceeds the defined value, the SIM switching process begins when there is no active call in progress. If there is, the action to be taken should be defined, either hang up the call or wait for its end.

(Method 2) The activity script method represents a set of scrip-lets (simple scripts) combining several limit parameters. Herein parameters are linked by the logical operand *and*, meaning that both conditions should be performed to end a SIM session and the logical operator *or*, meaning that at least one condition must be performed to end the SIM session. These parameters are indicated in the column *script-lets composition for switching* of Table IV.

B. SIM card activity limitations

Most *SIMBoxes* incorporate usage restrictions as well, to distribute the termination traffic between the different SIM cards of the system and thus, prevent them from being used beyond a certain threshold. Configured limitations allow locking SIM cards of the system automatically in such a way that the unlocking is only done either manually or automatically after a certain set period. Limitations are classified as follows:

- *Class 1 – Parameter limitation:* Numerous factors related to SIM cards' call and SMS behaving are used to limit their activity. They are deducted in terms of consecutive occurrence, total number or total duration. These factors are listed in the column *parameter limitation* of the Table IV.
- *Class 2 – Parameter limitation per time unit:* Some parameters are aggregated and measured per time unit (day, week and month). They are listed in the column *parameter limitation per time period* of the Table IV. For each parameter, the reported aggregation time periods have been filled in.
- *Class 3 – Time limitation:* The system makes it possible to set for each SIM card or for a group of SIM cards working periods of time, break times per day or delays between each use. For some systems, the user can select the days of the week on which a SIM card or a group of SIM cards will be able to operate. A *scheduling group* for example, is featured by two properties named *re-allocation interval* referring to the working time and *sleep time* relating to a break time after each work session. When the working time (the r*e-allocation interval*) ends, SIM cards and GSM channels cancel the binding and turn into an hibernation state. The duration of the hibernation state is the *sleep time*.

In some systems, these different limitation factors can be combined using the *"and"* and *"or"* operands to form more efficient control scripts.

C. SIM card migration

The purpose of SIM card migration is to simulate human mobility. Indeed, *SIMBox* equipment is static (as aforementioned, its installation requires the setting up of a favourable environment), which does not correspond at all to the behaviour of a regular customer as people are moving around and may make calls at many different places. To avoid being detected *SIMBoxes* allow to simulate the movement of SIM cards by migrating their bindings with GSM channels from GSM gateway to GSM gateway (the system's GSM gateways being located at different places in the city).

For a SIM card, the choice of the GSM gateway to which it will be connected is made according to the following method:

- *Manually fixed method:* The administrator can manually specify a binding between a specific SIM card and a specific GSM channel. Therefore, the binding will be stopped either at the disabling of one of the two elements (GSM channel or SIM slot) or manually by the administrator.
- *Any except previous:* A SIM card can register on any GSM channel, except a stated quantity of the previous ones to which it was bound. The quantity of previous GSM channels is edited as the *previous gateway depth*.
- *Any except previous zone ID:* A SIM card should not select a GSM channel with the same zone ID than the previous GSM channel. This way, SIM cards automatically switch GSM channel and location at each re-allocation.
- *Any gateway:* SIM cards register on any gateway.
- *Specified order:* SIM cards register according to a stated sequence list of GSM channels. The list includes GSM channels selected from the available ones and ordered according to a given progression defined by the administrator.

D. Base station switching/locking

Some *SIMBoxes* can simulate the smaller movements of mobile phones, which are conveniently reflected by connecting one surrounding base station to another depending on the signal strength emitted by them. Therefore the base station switching and locking functionality allow to configure the selection and connection to a base station for a GSM channel.

First, the system provides a list of all surrounding base stations with for each base station the values of the parameters *Mobile Country Code*, *Local Area Code*, Cell Identifier, *Base Station Identity Code*, *Broadcast Control CHannel*, and *Received Signal Level*. It is as well possible to have a spatial arrangement of these base stations around the GSM channel. From this list, the selection of a specific base station is made according to the following modes:

- *Mode 1 – Manually:* The system administrator manually selects the base station to which the GSM channel should be connected;
- *Mode 2 – Default:* This mode uses the default GSM base station selection mechanism;
- *Mode 3 – Fixed:* This mode locks the GSM channel to be registered to a specified base station or to switch between up to 3 fixed base stations;
- *Mode 4 – Random:* The base station is randomly chosen by the system in accordance to some conditions, including minimum signal strength, the frequency for base station switch, and whether the system can make a switch during a running call;
- *Mode 5 – Poll:* This mode enables the device to switch to the next base station of an ordered list at a specified frequency. The list is called base station *polling list*. It gathers all the surrounding base stations in the descending order of their signal strength, as recorded by the GSM module. The *maximum polling channel* defines the maximum number of base stations in the polling list. The *channel switching interval* establishes the frequency of base station switching occurrence. The frequency value can be randomly chosen at each switching, between a range set by the administrator. A white-list and a blacklist of base stations can be edited to define base stations that are going to be used in the polling and base stations that will not be part of the polling list respectively;
- *Mode 6 – Advanced:* This mode triggers base station switching when the GSM channel reaches the threshold value set on different parameters in the column *base station balancing* of the Table IV. A minimum allowed signal strength can be fixed as well.

E. Changeable IMEI

A GSM gateway counts one IMEI per GSM channel. Therefore, all SIM cards used by a GSM channel match the IMEI of this GSM channel in the CDRs of mobile operators. It is an obvious way to detect *SIMBoxes* as using a large number of SIM cards in a single mobile device is quite unusual. The *SIMBox* provides the ability to set an IMEI to any used SIM card to overcome this weakness and simulate the behaviour of a regular customer. Depending on the manufacturer of the *SIMBox*, IMEI changes can be made either to GSM channels or to SIM cards.

Changes related to the GSM channel can be done manually by the *SIMBox* administrator or automatically by the device itself. In the latter case, the IMEI is modified according to one of the following: a specified frequency (by default /hour and not more than /10 minutes), a threshold on the number of calls made by the channel (not less than 10) and each time a SIM card changes.

Changes at the SIM card level are either manual or automatically applied to a pre-formed group of SIM cards according to one of the following patterns: randomly, prefix-based (similar to random, but with a prefix), *Type Allocation Code*[9]-based, or registry-based (the full IMEI code comes from a registry).

[9]The Type Allocation Code (TAC) is the initial eight-digit portion of the 15-digit IMEI and 16-digit IMEISV codes used to uniquely identify wireless devices.

They are made according to a *generation rule* which is either *null* – meaning that IMEI codes will be generated only once –, *periodic* – meaning that IMEI codes will be generated every period of time (hour, day, week or month) – or *registration count* – meaning that the IMEI codes will be generated after a certain amount of registrations for the SIM card.

Finally, it is relevant to mention that *changeable IMEI* can, in some cases, become a weakness because, with an intensive GSM termination, a SIM-card can be assigned different IMEI numbers at high frequencies, which is suspicious for mobile operators.

F. The usage of other network services

The primary use of *SIMBox*'s SIM cards is to terminate voice traffic through phone calls. This is a weakness as regular users use other services such as SMS, USSD commands, and data. Some *SIMBoxes* allow to use SMS/MMS services or the Internet, and to send USSD commands, to simulate this human behaviour.

- *Internet:* An amount of data can be used for a group of SIM cards. For this purpose, the administrator has to specify a time interval and a consumption flow in MB (generally less than 2048 MB), which will be distributed among the SIM cards within the specified time interval. Finally, he must specify some URLs to websites where the SIM cards will surf to consume the data and the Access Point Names (APN) for the connection.
- *USSD commands:* USSD[10] commands are sent at the GSM channel level to get the number of the SIM cards they use, their balances or to make top-ups. For this purpose, the administrator selects one or more GSM channel(s) or all GSM channels[11] from the interface, enters the USSD command in the appropriate field and sends it. It is also possible to automatically send USSD commands under certain conditions set by schedule, call duration or by GSM channel connection time.
- *SMS:* SMS are sent in a similar way as USSD commands, i.e., by GSM channels. They can be used as well to make top-ups, get SIM numbers or SIM balances. Thus, one, several, or all GSM channels are selected for sending an SMS. A list of recipients as well as the message content are configurable. The encryption mode of the message can be chosen between ASC7/8 (ASCII 7/8 bit) and UCS2 (Unicode 16 bit), and the maximum length of the SMS varies according to the chosen mode. Finally, it is possible to consult the messages received by the GSM channel and the history of the messages sent.

[10]Unstructured Supplementary Service Data (USSD) is a service that is provided by telecom operators and allows GSM/WCDMA mobile phones to interact with the telecom operator's computers. USSD messages travel over GSM/WCDMA signalling channels and are used to query information and trigger services. Unlike similar services (SMS and MMS), which are stored and forwarded, USSD is a session-based service. It establishes a real-time session between mobile phones and telecom operators' computers or other devices.

[11]With the option "select all" one can select all the GSM channels available on the GSM gateway.

G. List of family

The family list consists of building a virtual family of network consumers for each SIM card used in the *SIMBox*. Indeed, SIM cards used by the *SIMBox* make several outgoing calls to a large number of non-related network consumers as part of the termination of voice traffic. Such is an abnormal behaviour because most regular consumers only call and receive calls from a restricted group of network consumers, named *family list*, who, in some cases, also make calls to each other. Therefore, some *SIMBoxes* offer the possibility to exchange SMS and voice traffic between the different SIM cards of the system as follows, so that these SIM cards constitute a family list for each of them.

- *SMS inter-sending:* A *SIMBox* administrator can manually schedule the sending of an SMS. For this, he selects GSM number(s) of one or several SIM cards of the system to set them as recipients of the SMS of which he has to write the content. It is also possible to send a message to a local SIM card for a GSM channel (that means that the SIM card belongs to a GSM group assigned to the GSM channel). In this way, by default, a SIM card is randomly selected from the SIM card group. Furthermore, it is possible to select the recipient SIM cards according to factors listed in the column *SMS and call inter-sending* of the Table IV. This way, as soon as a factor exceeds a defined threshold for a SIM card, a predefined message is sent to the SIM card.
- *Inter-calling:* Calls between SIM cards are featured by a *minimum/maximum duration* and a *message sending* option indicating whether an SMS should be sent by the called SIM to the caller one, just before the call. The inter-calling function is activated at a GSM channel. Calls are triggered from an idle SIM card to the other SIM cards of a GSM group. This is done randomly or according to threshold conditions on the parameters listed in column *SMS and call inter-sending* of the Table IV. If the *message sending* option is activated, the callee will, before each call, select a message from a predefined message list and send it to the caller.

In addition to the here-above traffic exchange mechanism between SIM cards, some providers offer call routing based on the *SIMBox* activity history. As a result, a SIM used by the *SIMBox* to route a call from a specific number will be preferred to route future calls from that number, if they occur while the SIM is available. In this way, the fraudulent SIM inserts itself into the family list of the number which it routes calls.

H. Call forwarding

The call forwarding feature allows a call intended for a SIM card used in the *SIMBox* to be forwarded to a specific number, so that a human agent can answer the call. It can be edited according to the three following conditions :

- *Unconditional:* it allows to forward all incoming calls unconditionally;

- *Busy:* it allows to forward incoming calls only when the called number is busy;
- *Not reachable:* it allows to forward incoming calls when the called number is not reachable or cannot register to the mobile operator network;
- *No reply:* it allows to forward incoming calls when there is no reply from the called number.

I. Other relevant SIMBox features

Most *SIMBox* models incorporate features that are not part of the simulation of human behaviour but are still relevant to explore as they help to achieve and maintain fraudulent activity. At the following, we discuss some of these features.

1) Anti-spam: Anti-Spam consists of setting up lists of SIM numbers from which calls can be filtered. Fraudsters use this feature to filter test calls coming from detection teams as explained in Section VI-A1. For this purpose, we distinguish *white*, *gray*, and *black* lists of numbers.

The *white list* represents the phone numbers authorized to make outgoing calls when the outgoing call authentication mode is set to "white list". A number is added to this list either manually or automatically, according to certain conditions essentially related to the number of calls and the duration of each call over a certain period.

The *grey list* represents the phone numbers listed for a period to be reviewed by the system. Phone numbers are automatically added to the grey list based on filtration algorithms. Indeed, if over a certain period, the amount of calls made by a number exceeds a specific value established for the grey list, the phone number is added to the grey list. Similarly, if over a defined period, the number of calls of short duration exceeds a maximum allowed, the phone number is added to the grey list. A phone number remains in the grey list for a defined time and, depending on its behaviour, the blocking will be extended or not.

The *black list* represents the list of numbers that are not allowed to make outgoing calls through the system. Phone numbers are added to the blacklist either manually or based on filtration algorithms. In this latter, thresholds are set for phone numbers in the grey list. Such limits are related to the number of calls made and their duration over a given period. If a phone number exceeds the allowed values, it is transferred from the grey list to the blacklist.

2) Voice and codec configuration: Configurations to voice and codecs used by the system, can improve the call quality. This is useful for fraudsters as some detection methods are based on audio call pitfalls (packet losses and jitter) identification. The *SIMBox* has a variety of codecs. The user can activate and order them according to his preferences. Some codecs are: G.711 a-law, G.711 μ-law, G.723, G.723.1, G.729, G.729-16, G.729-24, G.729-32, G.729-40, G.729A and G.729AB. Besides, the user can make voice configurations. He can define a minimum length for each VoIP packet received and activate the jitter buffer. The jitter buffer is designed to remove the effects of jitter from the decoded voice stream, buffering each arriving packet for a short interval (called *jitter buffer delay*) before playing it out. The jitter buffer substitutes additional delay and packet loss (discarded late packets). If a jitter buffer is too small, an excessive number of packets may be discarded, which can lead to call quality degradation. If a jitter buffer is too large, then the additional delay can lead to conversational difficulty. A fixed jitter buffer maintains a constant size, whereas an adaptive jitter buffer has the capability of adjusting its size dynamically to optimize the delay/discard trade-off. Three modes of jitter buffer are supported:

- *Fixed:* The fixed mode, which is the default mode, is a simple *First In First Out*(FIFO) mode for arriving packets, with a fixed jitter buffer delay.
- *Sequential:* The sequential mode is a fixed jitter buffer delay mode in which the jitter buffer function looks at packets' timestamps for dropped or out of sequence packet problems. The data packets are sorted based on the packets' timestamps.
- *Adaptive:* The adaptive mode optimizes the size of the jitter buffer delay and depth in response to network conditions, in addition to the sequential mode functions.

Finally, some *SIMBoxes* allow the activation of audio silence suppression through *Voice Activity Detection* (VAD) in combination with the *Comfort Noise Generator* (CNG). The purpose of VAD and CNG is to maintain an acceptable perceived quality of service while simultaneously keeping transmission costs and bandwidth usage as low as possible. CNG, in conjunction with VAD algorithms, quickly determines when periods of silence occur and inserts artificial noise until voice activity resumes. The insertion of artificial noise gives the illusion of a constant transmission stream, so that background sound is consistent throughout the call, and the listener does not think the line has gone dead. Mechanisms of echo cancellation can be supported as well.

3) CDR management: The *SIMBox* provides CDR generated by its activity for traffic and accounting management.A line of CDRs as collected by the *SIMBox* can contain items reported in Table V.
CDRs are saved either on an external disk or on a server to which queries can be made. Queries aim to obtain CDR records that meet certain conditions on call duration, caller and callee identifiers, call start time, call end time, and call type. They enable fraudsters to identify SIM cards/GSM channels that may be behaving suspiciously and thereby refine their activity.

J. SIMBox temporal evolution

The *SIMBox* has a hardware architecture and various features that have evolved to adapt to needs and be more efficient. We try in this section to build a chronological sequence to this evolution as it gives better visibility on the pace at which the *SIMBox* (and therefore, fraud) has evolved, provides insights on the motivations for this evolution, and allows to get a global

TABLE V
SIMBox VOICE CDR FIELDS

CDR Field	Description
Port	GSM channel of the *SIMBox* which terminated the call
Slot	Identifier of the SIM slot used by the GSM channel for the call termination
Source number, Original destination number	Dialing and dialed numbers respectively
Filtered destination number	dialed number after the appliance of a filter (where relevant). Otherwise identical to the Original destination number.
Billing and collection	When and how to collect traffic data, to exchange bills and to make payment
Dial date and time	The moment the gateway received the call request
Alert date and time	The moment the received party has begun ringing (if reported by the remote party and supported by the outgoing resource
Hangup date and time	The moment any of the parties had ended the call
Call type	Either mobile originated or terminated call
Source hangup direction	0 = source party had ended the call 1 = call was ended by LCR (due to termination of the destination party, no route to destination, unavailable destination resource, etc.)
Destination hangup direction	0 = source party had ended the call 1 = call was ended by LCR (due to termination by the source party, etc)
Hangup reason cause	Code as reported by the party that had first ended the call
Call duration in second	Time measured between answer time and hangup time

view of the potential of the *SIMBox* (and therefore, fraud) positioned in time. We cover here two aspects of the evolution of the *SIMBox*: the material evolution and the functional evolution centred on HBS features and other relevant features we mentioned in V-I.

Very little information is available on the evolution of *SIMBoxes*. *We have been able to collect the data presented in Figure 10 by reporting news about the addition or update of components on the websites of the leading SIMBox suppliers.*

On the image, the updates in pink represent hardware evolution, and those with blue titles represent functional evolution. The graph gathers the upgrades of four different manufacturers (Hybertone, Dinstar, Ejoin and Antrax). For each update, we indicate the name of the supplier who realized it as a header.

From a hardware point of view, the SIMBank appears as the first update for each supplier. Its first occurrence is in 2012 at Hybertone with the capacity of only 32 SIM slots. Over time its size, i.e. the number of simultaneously SIM cards evolves very quickly, to the maximum of 256 SIM slots from 2015 at Ejoin. The presence of the SIMBank implies the possibility of managing SIM cards remotely from the gateways as well as the realization of the SIM card migration functionality studied in section V-C, as long as there are at least 2 GSM gateways in the architecture.
Concerning gateways, their evolution is seen in terms of the number of channels, the number of SIM slots per device, and the cellular network technology handled by the device. Similarly to SIMBank, the number of GSM channels evolves from 8 at Hybertone in January 2012 and has a maximum of 32 in 2012 with the Dinstar provider. This means that since 2012, it is possible to make 32 calls simultaneously using a GSM gateway. We notice that despite the presence of the SIMBank, throughout the evolution, manufacturers are designing GSM gateways incorporating SIM Slots. The number of ports is always a multiple of the number of channels and gradually evolves to a maximum of 512 ports, in March 2017 with Ejoin. This shows that these models remain popular in the market because they allow easy management to have reduced architecture to a single device. We think that for this type of equipment, the physical evolution has been made in terms of volume: they have gone from sub-rack with many GSM boards and SIM boards to lighter and easily transportable units. Moreover, only Antrax provides GSM gateways without SIM slots, making it clear that it promotes a distributed operation architecture.

The cellular technologies handled by GSM gateways have evolved with the advent of 3G and 4G networks. In 2013, we had GSM and CDMA gateways, WCDMA in 2015, and LTE in 2016. The more networks a gateway serves, the more efficient the gateway is, offering more significant possibilities. Hence, over time, Dinstar and Ejoin manufacture gateways that combine several cellular technologies.

Information on functional evolution is more challenging to obtain. We were only able to collect them for two manufacturers: Hybertone and Antrax. Both suppliers make a point of honour of *changeable IMEI* feature and *SIM card activity limitations*. Indeed, from 2011 with Hybertone, the *changeable IMEI* feature was already available as well as the SIM card call activity limitation. This is understandable as the IMEI, if not changed, makes it easy to determine all SIM cards operating in a *SIMBox*' GSM channel and block them. As well, if the call traffic is not regulated and distributed to the different SIM cards of the architecture, the SIMs will be easily identifiable. Antrax, which starts its activity later in time (in 2015), added in 2016 the *Anti-Spam* feature to register and block numbers used by operators for test calls (see below) and developed many other features above-mentioned. It is worth noting that from 2017, it is possible to use the Antrax *SIMBox* equipment to record and obtain audio tracks of calls routed by the *SIMBox*. This represents a real intrusion because fraudsters can eavesdrop on bypassed calls without the permission of the sender and recipient of the call.

The evolution of the *SIMBox*, both from a hardware and functional point of view, is significant. With the limited information collected, we notice that the *SIMBox* adapts to different challenges to stay up to date technologically. Over time, it extends its hardware limitations to provide higher termination capacity in a single device and new architectures. On the functional level, we notice that over time the existing functionalities are refined to resemble more and more human behaviour, but also new features can be born and, in some cases, pose real security problems.

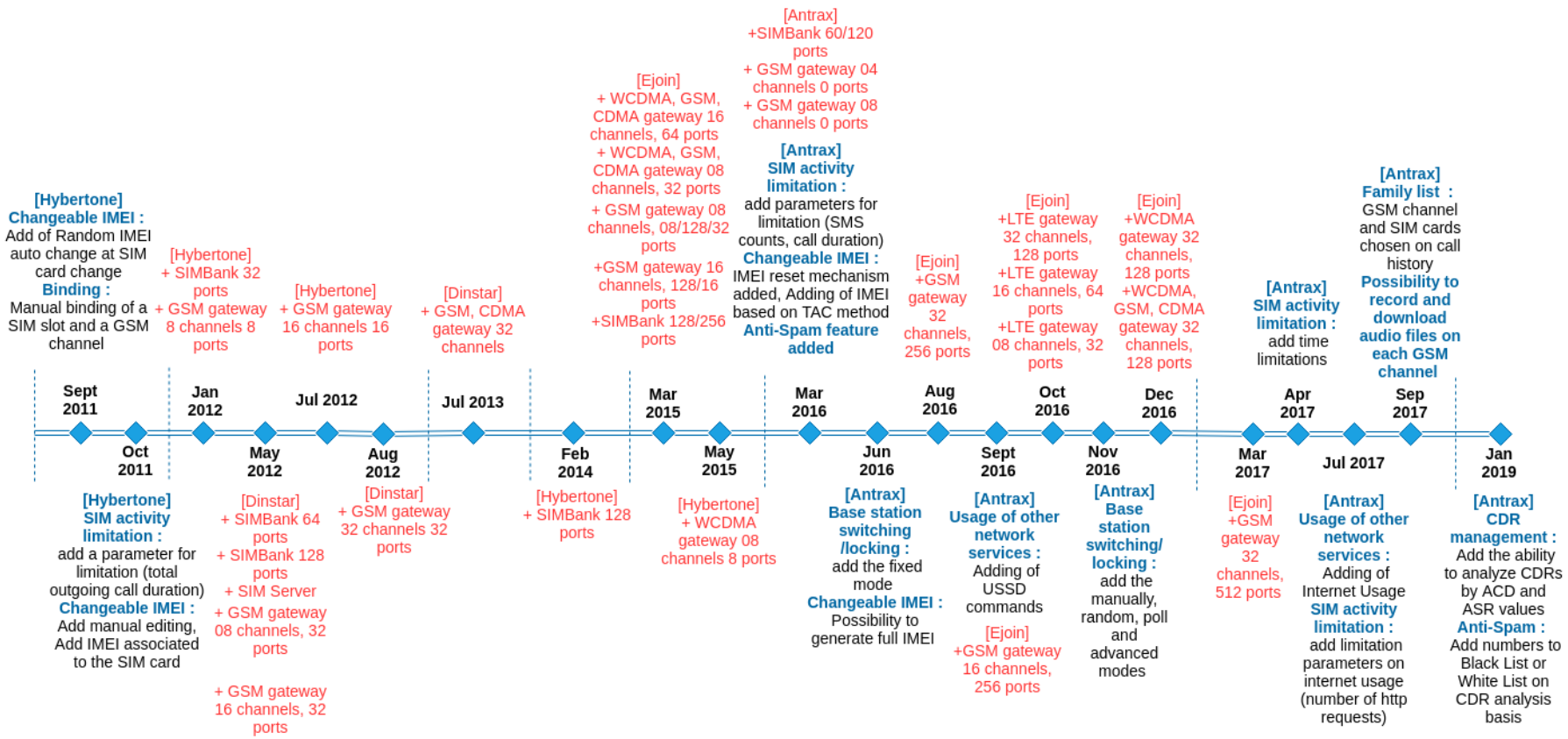

Fig. 10. Timeline with the evolution from material and functional points of view.

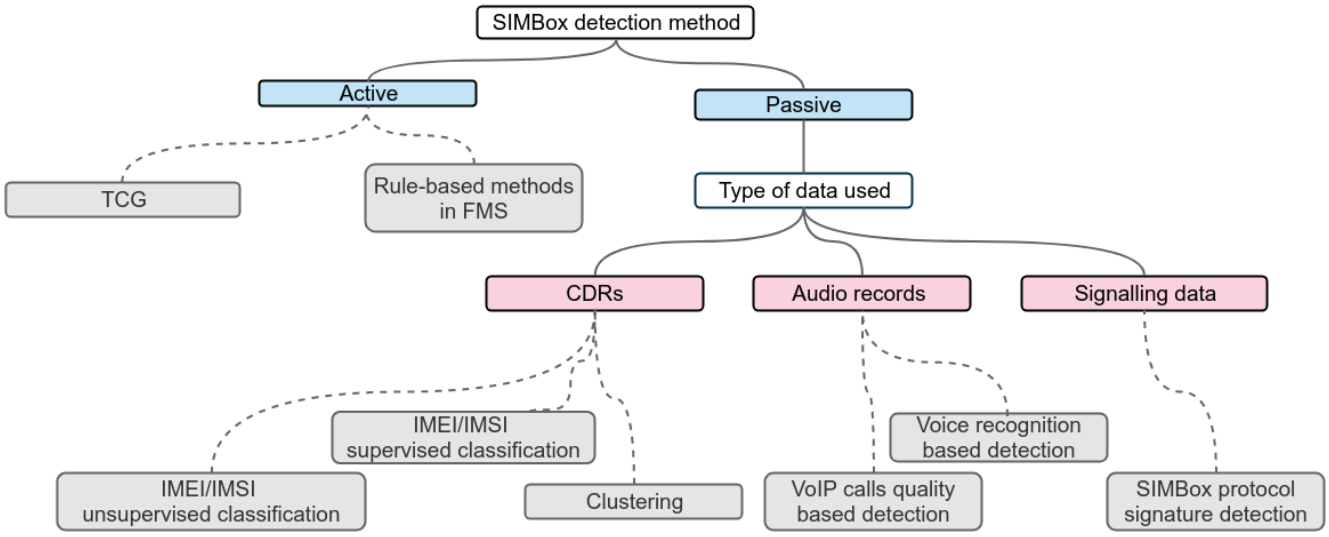

Fig. 11. Categorization of existing *SIMBox* fraud detection methods.

VI. *SIMBox* FRAUD-PREVENTION STRATEGIES

Despite the negative impact of *SIMBox* fraud, there are a few amounts of studies on this topic. This can be explained by the lack of publicly available data on which to conduct research combined with the absence of a simulation/emulation environment to test solutions. This section presents all existing solutions for detecting and preventing *SIMBox* fraud in the literature, which are summarized in Figure 11. Concerning their operation mode, these solutions can be organized into two categories: active and passive solutions. We organize passive solutions into three sub-categories concerning the studied data type.

A. Active methods

Active methods require permanent action by one or more entities. These methods are commonly considered as *classical methods* because they represent the first response of telecommunication companies to *SIMBox* fraud. They require significant material resources to be implemented.

1) Test Call Generation (TCG): The principle of TCG consists of setting up test phone numbers in a target mobile network and make calls to those test numbers from different countries, through many different interconnect voice routes around the world. This way, a local *Calling Line Identification* (CLI) indicates a *SIMBox* number and can be acted upon accordingly. Once routes having a high volume of *SIMBox* terminations are detected, the call campaign focus on them in order to maximize detection as much as possible. A large number of test calls are generated in a very short time and may use anti-white list services [51] and specialized SIM cards with *CLI Restriction Override* (CLIRO) to overcome the hiding of the CLI in calls routed through the *SIMBox*.

As can be seen, TCG is all about probability; that is, the more test calls cover routes, the more likely *SIMBox* fraud cases are to be detected. TCG is known for not making false positives, which explains its wide adoption by anti-fraud services [52–56]. Yet, it is expensive because it requires to make several calls and every test call is associated with cost/network resource consumption.

TCG method worked very successfully for many years. Yet, around 2012 and 2013 its effectiveness dropped off significantly because of the following reasons. First, *SIMBox* fraudsters figured out how to avoid detection by test calls. For instance, they perform analysis on the voice call traffic coming toward their *SIMBoxes*. Based on usage patterns (e.g., as discussed in Section V-I1), they can differentiate calls related to real subscribers from those originating from a TCG campaign. They can then either block test calls and prevent them from reaching the *SIMBox* or reroute them to a legitimate route. Second, fraudsters can allocate pools of SIM cards to be sacrificed. They therefore allow these SIM cards to be detected by TCG in order to make the mobile operators feel confident of their results. This creates a diversion for other SIM cards to conduct the bypass activity.

2) Rule-based methods in Fraud Management System (FMS): Rule-based methods [11] in FMS consist of establishing basic rules for subscriber profiling in order to identify fraudulent SIM cards. This involves the analysis and monitoring of call patterns (outgoing call count, distinct destinations ratio, cell sites used, incoming to outgoing call ratio, SMS originating/terminating counts, etc.) of a set of subscribers by experts looking for an abnormal behaviour originating from an operator's SIM card or terminating over it. Any case identified and validated (through a call or a similar action) can then be used to profile and uncover other similar SIM cards.

This approach is less costly and has a better coverage than TCG because once a profile is established, it can be extended over all available subscribers for a wide detection range.

TABLE VI
CDR FIELDS USED IN THE LITERATURE FOR *SIMBox* FRAUD DETECTION GROUPED BY FEATURE TYPE

Feature type	CDR Field	Description	Usage	Illustrative papers
All features type	Originating number	Phone number of the caller	Gives the number of calls made per user during a period of time	[20],[57],[58], [59],[11],[60]
	Terminating number	Phone number of the called party	Allows to obtain the number of calls received per user during a period as well as to calculate ratios in relation to the number of calls sent.	[20],[57],[58], [59],[11],[60]
	Originating IMSI	Calling subscriber unique SIM card identifier	Gives the number of unique subscribers calling a given subscriber during a period of time.	[20],[57],[58], [59],[60]
	Terminating IMSI	Called subscriber unique SIM card identifier	Gives the number of unique subscribers called by a subscriber during a period of time.	[20],[57],[58], [59],[60]
Call Behavior	Call type	Mobile originated/terminated call	Distinguishes calls made to the mobile network from calls received by the network	[20],[57],[60]
	Originating country code	Country code of the caller	Detects calls from international destinations	[20], [60]
	Terminating country code	Country code of the called party	Detects calls destined for international destinations	[20],[60]
	Event type	Local or international destination	Similarly to originating and terminating country codes, it allows the detection of calls made or intended for international calls.	[11]
	Time	Date and time of the call	Allows to make districts, sorting and selection according to the period of time in which the events occurred. The time period can be weekday, daytime, peak hours during the day, time period during the night etc.	[20],[57], [58], [59],[11],[60]
	Duration	Call duration	Gives the average or cumulative call time over a period of time.	[20],[57], [58], [59],[11],[60]
Mobility	LAC-CID at the origination of the call	Local Area Code and Cell Id (base station location identifier) at the start of the call	Allows the study of a subscriber's mobility for a given period of time	[20],[57],[60]
	LAC-CID at the termination of the call	Local Area Code and Cell Id (base station location identifier) at the end of the call	Allows to study of a subscriber's mobility during a call	[57]
Mobile services Usage	Service type	Call, SMS, MMS or mobile data	Allows to distinguish between the different types of service used by a subscriber and to study the frequency of use	[57],[11]
Entity properties	IMEI	Device identifier	Allows to identify the mobile devices acting on the network and to study the number of SIM cards per device	[20],[57],[11]
	Account age	Time since account activation	Allows you to study the average duration of SIM cards and SIM cards associated with a given device	[20],[60]
	Customer segment	Prepaid/postpaid/corporate account	Serves as an indicator as to which accounts may or may not be fraudulent	[20],[60]

However, it has several limitations. First, it has a fairly large percentage of false positives and thus, requires continuous monitoring and field expert intervention. Second, through time, the whole process of analyzing data gets more complex as rules are added in the system. This increases the detection latency, allowing fraudsters to make enough profit before being blocked. Finally, this method can not scale as it requires human intervention.

FMS has been effective in detecting *SIMBox* fraud prior the integration of *Human Behavior Simulation* (HBS) into *SIMBoxes* (presented in Section V). HBS significantly increases the false positive rate to the point of paralyzing the decision of blocking a detected SIM card.

B. Passive methods

Passive methods don't require a permanent human action: they are deployed in a system to detect fraudulent entities automatically. We categorized in Figure 11 passive methods based on the type of data analysed throughout the detection process. This classification allows us to distinguish 3 subgroups discussed in the following: CDR analysis-based approaches, audio analysis-based approaches and signalling data analysis-based approaches.

1) CDR analysis-based approaches: This is a passive method consisting of analysing both the content and the occurrences of CDRs, unlike Rule-based methods that only focuses on the latter.

CDRs gather all traces of events carried out on the network, whether it is voice, text message, or Internet data, as listed in Table VI. As a result, the large majority of *SIMBox* fraud detection solutions leverage Machine Learning to analyse CDR content data in order to shed light on anomalies of all kinds [61]. More specifically, CDR-based *SIMBox* fraud detection can be abstracted as a *classification problem*. The entity to be classified here is either a *SIM card* (identified by its IMSI) or a *mobile device* (identified by its IMEI).

A common methodology[12] is applied in each CDR-based *SIMBox* fraud solution. It can be summarised as data preparation step followed by model building and evaluation.

Data preparation: Data preparation includes data under-

[12]Similar to the six typical steps of the CRISP-DM process [62].

standing, feature selection as well as any form of data pre-processing. A CDR have several fields, some of which may not be meaningful for *SIMBox* fraud detection. The Table VI summarizes all the fields used in the literature for *SIMBox* detection, their common usage, and all the works, to the best of our knowledge, where they are exploited. We organize commonly used features (built by aggregated fields) into four types.

(Feature type 1) Call behavior. They highlight how calls are generated within the *SIMBox*. Because of its role in traffic termination, the *SIMBox* is known to generate tremendous outgoing calls than incoming calls. Therefore, a feature such as the ratio of the number of outgoing calls to the number of incoming calls over a given period of time is usually considered to detect this pattern. In addition, SIMs within the *SIMBox* are considered to be more or less heavily used and may be at irregular hours. This usage behavior can be evaluated by features such as: the total and average number of calls, and the cumulative or average duration of calls made during a period of time (day, week, etc) and at irregular hours. Finally, the *SIMBox* is known to make outgoing calls to a large number of different individuals; thus, the total number of individuals called is generally considered as well. Early fraud detection works [58; 59] are solely based on features of this category.

(Feature type 2) Mobility. As far as mobility is concerned, the *SIMBox* is known to be a not very mobile device that makes calls through the few surrounding base stations to its location. As a result, movement during a call will be practically non-existent and also a significant amount of traffic will be generated in these cells to the point of qualifying them as hot cells. To detect this, the total number of different cell IDs where a SIM or device is located over a given period of time, the event load (number of calls, SMS, etc. per unit of time) on a cell or a set of geographically close cells, the number of subscribers making voice calls within the same location, the average or cumulative distance travelled by an individual (SIM) during a call, and the ratio of calls made without displacement etc. are examples of attributes that can be considered. Murynets [20] exploited this category of attributes.

(Feature type 3) Mobile services usage. SIMBoxes are known to be specialised in terminating calls. As a result the SIMs used in *SIMBoxes* make little or no use of other mobile services such as SMS, MMS, GPRS, or other mobile internet services. The ratio of the number of voice calls to the total number of other services can be studied globally and per individual for a SIM card in order to detect this usage behavior. Further analysis can be carried out on a service-specific study (SMS, data upload and download traffic etc.).

(Feature type 4) Entity properties. Finally, there are features used to perform a detection based on the account information and properties, and not on the entity (SIM card or Mobile equipment) activity. In the literature, for example, we count features such as the number of SIM cards per IMEI, which is normally high for *SIMBoxes* because they are designed to hold quantities of up to hundreds of SIM cards. The age of a customer account is also considered by [20], as fraudulent SIMs operate for less time than regular SIMs because they are usually blocked by operators as soon as they are detected and it can be used for classification purposes. The customer-segment also helps to identify fraudulent entities since prepaid accounts are more likely to be fraudulent than postpaid or corporate accounts (as discussed in Section III-C).

Data preparation often includes data sampling to reduce the size of the input data in order to ensure a proper proportionality between fraudulent and non-fraudulent entities, leading to better results [42]. In this vein, the proportionality of 66% normal cases versus 34% fraudulent cases is commonly adopted as in [58], [59], and [20].

Model building and evaluation: Several classification methods have been used in the literature as described in the following.

Artificial Neural Network (ANN). ANN [63] is the first classifier used for *SIMBox* fraud detection. ANN is composed of several neuron layers: an input layer, zero of several hidden layers, and an output layer. Each layer forwards its outputs as input to the next layer until a final result is delivered by the output layer. The optimal weights associated with the nodes are calculated during the neural network training. This is usually done according to an optimization algorithm which aims to minimize training errors. The back propagation algorithm is commonly used.

In the context of *SIMBox* fraud, each neuron of the input layer represents a feature of the entity to be detected and the output layer comprises two nodes, one indicating a fraudulent entity when it is activated and the other indicating a normal entity. It is as well possible to have a single node whose output indicates one or the other of the two possibilities depending on its value.

Sallehudin et al. [58] tested 240 NN models for the detection of *SIMBox* fraud considering the Sigmoid function as activation function. The authors selected 9 nodes on his input layers and varied the values of 4 parameters to choose the most optimal architecture: the number of hidden layers, the number of hidden nodes per hidden layer, the learning rate and the momentum[13]. This work showed that very high learning and momentum rates significantly degrade the classification accuracy of the models. The best results (accuracy of 98.7% and RMSE of 0.10380) are obtained when a low value of momentum is used with relatively high value of learning rate.

Kashir et al. [64] proposed a similar NN architecture for fraud detection but with the Sign function as activation function. The authors selected 25 attributes for its input layer and built its model on the basis of a dataset with 8695 normal subscribers and 50 fraudulent subscribers divided into 3 groups: training, testing, and validation. They tested 5 variants of NN by varying the optimization algorithms of the model and obtained very good performances: an accuracy of

[13]The momentum in an ANN helps in the early stages of the algorithms, by increasing the rate at which the weights approach the neighbourhood of optimality

99.87% with the *Bayesian regularization algorithm* and an RMSE of 0.01654.

The Support Vector Machine (SVM). SVM classifies cases by finding a hyperplane separator in the feature space between two classes in such a way that the distance between the hyperplane and the closest data points of each class (referred to as support vectors) is maximized. When the dataset is not linearly separable, the training samples are mapped to a higher dimensional space by applying a kernel. The most common used are linear kernel, polynomial kernel with a p parameter as the polynomial degree, Radial Basis Function (RBF) kernel with a ∂ parameter and Sigmoid kernel.

Sallehudin et al. [59] tested out 40 SVM models considering 3 kernel functions and varying for each kernel the values of parameters. The authors used the same features as in their previous work [58] (based on ANN, described above) and obtained a maximum accuracy of 98.9% and minimum RMSE of 0.105. The authors also provided an extensive comparison between the ANN model and the SVM model for fraud detection. They showed evidence that SVM model is more efficient in classifying fraudulent subscribers with a lower false negative rate overall than ANN model as the training data percentage increases.The ANN model would be better suited to classify normal subscribers, however it presents dramatically increased outliers for certain percentages of training data considered. As far as classification accuracy is concerned, SVM model has the highest value regardless of the percentage of data used for training. Finally, the evaluation of the time taken to build the model showed that the SVM model is about three times faster than the ANN model and therefore requires less computational power.

Kashir et al. [64] tested out 5 SVM kernels compared by classification accuracy and regression. The authors reported a poor performance compared to their previous work (presented above) based on ANN. This is in contradiction with Sallehudin et al. [59], questioning the validity of these results.

Fuzzy logic. Unlike SVM, a binary and non-probabilistic classifier, Fuzzy logic deals with approximate reasoning rather than fixed and exact. Therefore, an element belongs to a fuzzy set according to a *Membership Function* (MF), whose value is between 0 and 1. If this value is 0 for an element x for a fuzzy set A then, it has no membership to this set, and if it is 1 then, there is a full membership. From this logic, fuzzy rules are defined as conditional statements based on elements belonging to fuzzy sets. A fuzzy rule is therefore, true at a certain rate, which results from calculations on MF values of the different fuzzy sets included in the conditional statement. The collection made up of fuzzy sets with their MFs and fuzzy rules constitutes a fuzzy system. Fuzzy logic brings the reasoning closer to that of a human with linguistic expressions that refer to fuzzy sets.

Marah [57] proposed a fuzzy system for *SIMBox* fraud detection based on five fuzzy sets. Each fuzzy set is established based on a fraud detection pattern related to mobility, call behaviour, or the use of mobile services by subscribers. For each input SIM card, the authors find to what extent it conforms to each fraudulent pattern: This means to determine the MF value of each fuzzy set for that SIM card. The MFs are calculated in a triangular way by identifying for each pattern the maximum and minimum values of a dataset and applying the ratio $\frac{Value-Min}{Max-Min}$ for an input value. The SIM card detection process is based on the average of the MF values for the 5 patterns. If it is above a certain threshold the SIM can be considered fraudulent or to be watched and a decision will be taken by the operator depending on the case. In Marah et al. [57], evaluations have not been conducted to validate the model because the dataset was not provided with ground truth data.

We note that contrary to the algorithms previously presented, this model is not very adjustable; only the value of the threshold determines the detection, which could present some limitations.

Random Forest (RF). It is a classification method consisting of many decisions trees. It uses bagging and feature randomness when building each individual tree to try to create an uncorrelated forest of trees whose prediction by committee is more accurate than that of any individual tree. The number of trees to build is a parameter which can be selected during the training phase. The prediction made by random forest is determined by majority rule of the generated decision trees.

Using a 10 fold cross validation and a 64-36 percentage split validation, Hagos et al. [65] compared the RF model with ANN and SVM for three datasets obtained by aggregation of features on a 4-hour, daily and monthly basis. Its results suggest that all models have comparable performance measures. However, the RF model attains a little bit better than the two others in accuracy precision and build time for the 4-hour dataset. Similarly, outcomes of the confusion matrices verify that the RF model have a little bit lesser false positive rate compared to the others. However, for the daily and the monthly datasets the RF performance is slightly lower in terms of accuracy. We also note that generally the false negative rate obtained with the RF model is higher than that of the other models.

Murynets et al. [20] makes a similar observation regarding false negative and false positive rates. The authors showed that this issue can be mitigated by combining RF with an *Alternating Decision Tree* (ADTree) model [66; 67] and a Functional Tree model [68]. For each record to be classified, a linear combination of the predictions of these three classifiers (in the form 0 and 1) is carried out on the basis of coefficients ($\in [0.1]$ and whose sum is equal to 1) chosen to minimize the classification error. This value is compared to a threshold α (chosen by minimizing the classification error) for decision making. The resulting model has a lower false negative rate and a better accuracy.

Some remarks: First, the choice of the entity (IMSI vs IMEI) has a meaningful impact on the results as it'll lead fea-

tures determination (and therefore, detection patterns). Using HBS features, fraudsters work to ensure that only SIM cards (and not devices i.e. GSM gateways) are similar to regular ones. Therefore detection based on IMSIs is more hampered and therefore, more difficult than detection based on IMEIs.

Second, the *period of operation* of the entity is also relevant. It refers to the entity's features collection time prior to a classification. The longer it takes to detect a fraudulent SIM card or *SIMBox*, the more revenue it can generate to the detriment of mobile operators. It is worth noting that most of the contributions that we studied do not focus on optimizing the *period of operation*. In some cases, that information is conveyed in a disguised way and the methodologies are presented as being effectively applicable regardless of the period of data collection, which remains to be validated.

2) Audio analysis-based approaches: Audio records hold valuable information to obtain attributes such as the origin of a call, the types of telephone networks a call has traversed and to perform analysis and profiling on packet loss and noise as in [69]. From far of our knowledge, two research works used audio record analysis to detect *SIMBox* fraud.

[14] leveraged the fact that calls performed over the VoIP network suffer from audio degradation in terms of packet losses and jitters (as discussed in Section II-A). Therefore, the authors try to detect calls with such degradations whether they are concealed by a *Packet Loss Concealment* (PLC) algorithm or not; which would indicate that they have been bypassed through the VoIP network by means of a *SIMBox*. The system takes as input a stream of GSM audio frames and a vector indicating which audio frames were erased by the GSM air correcting mechanism applied at the base station level. This vector is intended to ignore erased frames caused by the GSM network so that they are not misinterpreted.
For evaluations the authors used audio recordings from the TIMIT Acoustic-Phonetic continuous Speech corpus [70] to generate 1960 calls from a set of 98 randomly chosen speakers. By considering a SIM to be fraudulent when at least 25% of the calls it makes are considered fraudulent by the system, 100% of fraudulent SIMs are identified with a 5% VoIP packet losses rate for the three simulated codecs. Fewer SIM cards are identified as loss rates decrease and in the case of 1% VoIP packet losses, 43% of G.711 SIMs and 28% of GSM-FR SIMs could be identified. Finally, experiments conducted with a real *SIMBox* showed that this system can detect 87% of fraudulent SIMs with no false positives.

Elrajubi [71] proposed a voice-recognition-based approach to fraud detection. His system is based on the fact that fraudulent SIMs are used to terminate traffic that may originate from several different callers to local numbers. However, these calls appear in the CDR traces as coming from a single SIM. This single SIM actually used by different subscribers could be detected by analyzing the voices of the different speakers using it to differentiate between them. The methodology adopted is therefore to identify from audio samples extracted from the calls each period of time, the subscriber speaking, for each SIM card using feature extraction algorithms specific to the field of speech recognition. This speaker would either be added as a new speaker for this SIM card if not already existing, or used to increment the number of calls made by the corresponding speaker if already existing in the database. From this information three variables are defined for each SIM card : M the greatest number of calls made by an unique speaker using this SIM card, T the total number of calls made with this SIM card and F a threshold value between 0 and 1 that should be experimentally chosen so that if M is less than $F \times T$ the SIM is considered fraudulent and else the SIM is considered normal. Although the idea is very promising the system was not implemented due to privacy issue in telephone calls which raises a number of questions about its effectiveness. Will the quality of the audio signal in a real case allow to recognize the voice of the speakers? How efficient the system will be and can it not be circumvented by fraudsters ? Is it possible to alter the voice of the users at the level of the *SIMBox*? If so, is it within the reach of fraudsters and how can the system be extended to prevent this ?

3) Signalling data analysis approaches: The analysis of signalling data for the detection of *SIMBox* fraud is a recent and not very exploited technique. It was mentioned by LATRO Services [72] in 2015 and is described as highly effective.
Indeed, signalling messages are exchanged between *User Equipments* (UEs) and the core mobile network according to well-defined protocols. They aim at controlling the UEs, managing access to the network, monitoring terminals in case of mobility and access to the service. The protocols are distributed according to the *Access Stratum* (AS) and the *Non Access Stratum* (NAS). The AS (RRC, PDCP, RLC, MAC and PHY) manages signalling between UEs and base stations for radio resource management, handover, and data encryption/compression. The Non-Access Stratum (EMM and EPS in LTE) manages the signalling between UEs and the core network, including the establishment of data or call sessions and mobility.

Network Attachment for example, is a procedure that involves the AS and the NAS. It is carried out when the UE is switched on, after a loss of network coverage or a change of *Mobile Management Entity* (MME)[14]. It consists of the authentication of the SIM card to the network, the authentication of the network to the SIM card[15], the identification of the UE and the updating of the mobile subscriber's location on the core network, through specific information exchanges (IMSI, IMEI, authentication vector, etc.). Similarly, signalling information is exchanged when a call service is set up or for the transfer of SMS.

Authors in [72] argue that *SIMBox* components (GSM gateways, SIMBanks and Control servers) generate a specific set of these signalling messages which constitutes a fingerprint, allowing the fraudulent devices to be distinguished from other

[14]The MME in LTE replaces the *Visited Location Register*(VLR) in 2G and 3G mobile network standards.

[15]From 3G, mutual authentication is realised. The SIM card verifies that the terminal is connected to a legitimate serving network to prevent 'fake base station' attacks.

devices on the mobile network.

The analysis of these messages' data and parameters can be performed in real-time. As an example, the *SIMBox* signature can be detected when it connects to the network to block the use of the SIM cards it owns. This technology, therefore, has several advantages, including the fact that it is passive (no permanent human action) but also stops fraud before any revenue is lost. High motivation is thus attached to the exploration of its possibilities. However, there are some difficulties related to the accessibility to this type of data from mobile operators. Their processing as well might be challenging because data volume is much higher than with CDRs.

4) Concluding remarks: First, detection methods which are based on audio analysis are more efficient than those based on CDRs. Indeed, the effectiveness of the former does not depend on prior detection work and they can be deployed in a telecommunication environment for real-time fraud detection. However, they remain limited because they are dependent on the voice codec used by the *SIMBox*: for each codec an algorithm must be adapted for optimal detection. Several voice codecs can be dynamically deployed into the *SIMBox*. This makes the audio analysis-based approach not scalable.

Second, almost no CDR analysis-based solution takes into account the presence of HBS in current *SIMBoxes*, making them inefficient today.

Third, signalling data analysis-based solutions still virtually unexplored, promise more efficient and accurate *SIMBox* fraud detection regardless of the strategy used. Although challenging, this provides a great incentive to investigate solutions based on this type of data.

VII. *SIMBox* FRAUD IN A COUPLE OF YEARS

SIMBox temporal evolution discussed in Section V-J shows that *SIMBox* fraud evolves over time. As a result, the challenges encountered today in fraud detection will not be the same in a few years.

In this section, we identify the different factors that may influence *SIMBox* fraud, and on this basis, we forecast what tomorrow's fraud may be. The purpose of this exercise is to allow readers interested in fraud detection to propose detection solutions that will not be limited to today's challenges and quickly become outdated, but that will be able to adapt and face tomorrow's possible challenges.

We distinguish three categories of factors that can influence *SIMBox* fraud : *(1)* Technological advances of fraud ecosystem elements; *(2)* Economic variations in the billing of calls; *(3)* And finally, improvements to the *SIMBox* for more efficient and accurate fraud.

A. Technological advances of fraud ecosystem elements

Cellular networks in which the *SIMBox* fraud is deployed are rapidly evolving. Indeed, since 2014 several studies [73–75] focus on 5G, whose deployment is on-going, while scientists are already working on beyond 5G specifications, with the definition of the next-generation 6G wireless system [76; 77]. Through technologies such as high-speed connectivity, *Internet of Things* (IoT), augmented virtual reality and so on, these new standards will considerably increase the quality of VoIP communications by allowing high data rates and low latency [73]. Hence, phone calls routed through the *SIMBox* will be indiscernible from a quality point of view and may even be of better quality than cellular-only voice calls. As a result, the efficiency of audio quality-based detection methods [14] will be significantly reduced, if not nullified because VoIP current pitfalls (packet losses and jitters) will be practically indistinguishable.

Similarly, with a massive connection of devices to the cellular network as planned with the IoT, the amount of remote communications will increase and thus, voice traffic. With the massive number of devices connected (1 million per km^2 [78]) to the cellular network, we could envisage the birth of a new form of *SIMBox* fraud applied to *Machine Type Communication* [79]. We thus claim that technological advances will significantly impact *SIMBox* fraud.

Furthermore, the subject of virtual SIM cards [80] is topical [81–84] and a good option for mobile operators to handle the emergent massive mobile connectivity. Virtual SIM cards technology could revolutionize the way fraud is currently carried out and give rise to pro-fraud (e.g. easily obtaining large quantities of SIM cards, advanced *Human Behavior Simulation*-capabilities architecture) or counter-fraud (e.g. more precise control of SIM card distribution) possibilities depending on how it will be implemented by mobile operators.

B. Economic variations in the billing of calls

The use of OTT applications for voice, audio, video or other media delivery services is expanding. This tendency will be accentuated with the democratization of 5G, which will give rise to high bandwidth allowing the birth of new types of OTT applications (e.g. tactile internet applications [85]). This growing trend is seen as a credible and measurable threat to mobile operators [86] because OTT apps provide services (voice calls and messaging) that can substitute their own relatively more expensive ones. Besides, OTT apps use mobile operators' infrastructure and network to deliver services without directly[16] contributing to mobile operators' revenue. To balance this, efforts are being made to regulate and tax OTT and VoIP services in some countries [87; 88], while in others VoIP usage is banned [31]. The former remains challenging because of the difficulty of finding a consensus on what digital content is and how tax should be applied [89]. However, this will likely come into effect in some countries in a few years, and consequently, VoIP calls will be billed. We believe that this could increase *SIMBox* fraud as current users of OTT apps for international calls might instead use *cheap-international-calling apps* to get a better quality/cost compromise. As discussed in Section III-B, these *cheap-international calling apps* are a way for *SIMBox* fraudsters to get mobile subscribers' voice traffic.

[16]We still have an indirect contribution because OTT service usage requires a subscription of Data pack thus driving the data revenue.

C. SIMBox's improvements

Fraudsters create/refine their fraud strategies to (1) adapt to existing detection solutions and be able to evade them or (2) to have higher traffic termination capacity (more GSM channels for simultaneous calls, more SIM slots on a single device, support for CDMA, LTE or other new unpredictable functionalities).

To respond to CDR analysis-based detection solutions, we believe that *SIMBox* fraudsters can leverage ML algorithms to control the behaviour of SIM cards. Indeed, fraudsters have CDRs generated by *SIMBox* activity and have access to the publicly available ML-based detection algorithms developed by researchers (discussed in Section VI-B1). By replicating these algorithms, for instance, they could check in real-time if a SIM card is detectable to limit its usage. Concerning SIM cards migration, [23] mentions a type of GSM gateway that can be mounted in a vehicle and powered by a car battery to simulate a moving traffic. This suggests that fraudsters may develop more of this type of model to limit detection based on mobility behaviour. They could go further and design mobile GSM gateways with integrated batteries to power themselves and allow any movement.

To respond to audio analysis-based detection solutions, fraudsters could incorporate a *SIMBox* feature to modify call audio characteristics (which they can already eavesdrop as discussed in Section VI-B1) uniquely during each call. Many methodologies exist to do such modification [90–92] and this will considerably limit the efficiency of methods based on voice and audio features recognition [71].

VIII. Conclusion

SIMBox fraud is a genuine problem that causes enormous losses. It is complex and delicate because it is based on economic, technical, and even character factors (people's mentality). Moreover, it evolves by adapting to existing detection solutions and is, therefore, a real challenge.

In this document, we surveyed both the *SIMBox* manufacturer's community to highlight the fraud scheme and different strategies used by fraudsters and the scientific literature in terms of fraud detection. We identified the limitations of current solutions and why fraud continues to be rampant. Besides, we discussed in detail all the necessary elements to understand *SIMBox* frauds and their ecosystem, focusing on optimal detection solutions.

Our review provides key elements to tackle this not-recent but not-enough-studied security issue, which may become more challenging in the future with the forthcoming technological developments and possible economic variations.

References

[1] CFCA, "Communications fraud control association announces results of 2019 global telecom fraud survey," 2019. [Online]. Available: https://cfca.org/sites/default/files/Fraud%20Loss%20Survey_2019_Press%20Release.pdf

[2] K. Baskar, "A study on internet bypass fraud: national security threat," 09 2019.

[3] CFCA, "2017 global fraud loss survey," 2017.

[4] C. NCC Policy and E. A. Department, "An assessment of international voice traffic termination rates," 07 2015.

[5] "Articles," https://goantifraud.com/en/blog/categories/article, accessed: 2020-09-03.

[6] Revector, "Simbox fraud and ott bypass biggest threats to mobile operator revenues."

[7] D. B. Africa, "Cameroun : 22,2 milliards fcfa de pertes en 2015 sur les appels téléphoniques frauduleux par simbox."

[8] M. Yelland, "Fraud in mobile networks," *Computer Fraud & Security*, vol. 2013, no. 3, pp. 5–9, 2013.

[9] "Specifications-3gpp." [Online]. Available: http://www.3gpp.org/specifications

[10] D. Naboulsi, M. Fiore, S. Ribot, and R. Stanica, "Large-scale mobile traffic analysis: A survey," *IEEE Communications Surveys Tutorials*, vol. 18, no. 1, pp. 124–161, Firstquarter 2016.

[11] M. R. AlBougha, "Comparing data mining classification algorithms in detection of simbox fraud," 2016.

[12] S. Karapantazis and F.-N. Pavlidou, "Voip: A comprehensive survey on a promising technology," *Computer Networks*, vol. 53, no. 12, pp. 2050 – 2090, 2009. [Online]. Available: http://www.sciencedirect.com/science/article/pii/S1389128609001200

[13] M. Sahin, A. Francillon, P. Gupta, and M. Ahamad, "Sok: Fraud in telephony networks," in *2017 IEEE European Symposium on Security and Privacy (EuroS P)*, April 2017, pp. 235–250.

[14] B. Reaves, E. Shernan, A. Bates, H. Carter, and P. Traynor, "Boxed out: Blocking cellular interconnect bypass fraud at the network edge," in *24th USENIX Security Symposium (USENIX Security 15)*. Washington, D.C.: USENIX Association, Aug. 2015, pp. 833–848. [Online]. Available: https://www.usenix.org/conference/usenixsecurity15/technical-sessions/presentation/reaves-boxed

[15] E. S. Hank Intven, Jeremy Oliver, *Telecommunications Regulation Handbook*, ser. World Bank Publications. The World Bank, 2000, no. 15249. [Online]. Available: https://ideas.repec.org/b/wbk/wbpubs/15249.html

[16] P. Cholda, M. Kantor, A. Jajszczyk, and K. Wajda, "Least cost routing in inter-carrier context." 11 2006.

[17] M. SAHIN, "Understanding telephony fraud as an essential step to better fight it," Ph.D. dissertation, TELECOM ParisTech, 9 2017.

[18] U. Murad and G. Pinkas, "Unsupervised profiling for identifying superimposed fraud," in *Principles of Data Mining and Knowledge Discovery*, J. M. Żytkow and J. Rauch, Eds. Berlin, Heidelberg: Springer Berlin Heidelberg, 1999, pp. 251–261.

[19] S. Rosset, U. Murad, E. Neumann, Y. Idan, and G. Pinkas, "Discovery of fraud rules for telecommunications—challenges and solutions." New York, NY, USA:

Association for Computing Machinery, 1999. [Online]. Available: https://doi.org/10.1145/312129.312303

[20] I. Murynets, M. Zabarankin, R. P. Jover, and A. Panagia, "Analysis and detection of simbox fraud in mobility networks," in *IEEE INFOCOM 2014 - IEEE Conference on Computer Communications*, April 2014, pp. 1519–1526.

[21] F. Okumbor N. Anthony and A. A. J. Olokunde, "Grappling with the challenges of interconnect bypass fraud," *IOSR Journal of Mobile Computing and Application (IOSR-JMCA)*, vol. 6, pp. 35 – 41, Jan - Feb 2019.

[22] Subex, "White paper bypass fraud - are you getting it right?"

[23] OECD, "Working party on communication infrastructures and services policy international traffic termination," 01 2015.

[24] "Goantifraud gsm termination in africa: Top 5 destinations in 2020," https://goantifraud.com/en/blog/1186-gsm-termination-in-africa-top-5-destinations-in-2020.html, accessed: 2020-02-20.

[25] OECD, "Access to mobile services and proof of identity 2019: Assessing the impact on digital and financial inclusion," 2019.

[26] W. B. Group, "The global findex database 2017 measuring financial inclusion and the fintech revolution," 2019.

[27] D. Morrow, "Telco corruption fuels simbox frauds," *Comms Risk*.

[28] H. Kumar, "Technical note on illegal international long distance telephone exchange in india," 08 2012.

[29] Sysmaster, http://www.sysmaster.com/products/gsm_termination.php.

[30] Antrax, https://en.antrax.mobi/.

[31] S. Guerraoui, "Morocco banned skype, viber, whatsapp and facebook messenger. it didn't go down well," *Middle East Eye*.

[32] S. Limited., "Subex wholesale fraud management survey 2013 is the industry ready to tackle a growing issue?" 2013.

[33] I. Ighneiwa and H. Mohamed, "Bypass fraud detection: Artificial intelligence approach," *ArXiv*, vol. abs/1711.04627, 11 2017.

[34] "10 misconceptions about gsm termination," https://goantifraud.com/en/blog/368-10-misconceptions-about-gsm-termination.html, accessed: 2020-04-24.

[35] Hybertone, *GoIP32-X4 Quick Setup Manual*, Hybertone. [Online]. Available: http://www.hybertone.com/uploadfile/download/20180913163352145.pdf

[36] Antrax, *Sim Server Factor Script*, Antrax. [Online]. Available: https://gitlab.com/flamesgroup/antrax/-/blob/master/doc/manual/scripts/sim-server-factor-script.md#sim-server-factor-script

[37] Hybertone, *GoIP User Manual*, Hybertone. [Online]. Available: http://www.hybertone.com/uploadfile/download/20140304125509964.pdf

[38] Portech, *SIM Server User Manual*, Antrax. [Online]. Available: https://www.portech.com.tw/data/SIM%20Server%20User%20Manual%20V2.pdf

[39] Dinstar, *Instructions for Using Multi-SIM Function of DWG*, Dinstar. [Online]. Available: https://www.dinstar.com/WEB/files/48826/2019-05-22/Multi-SIM%20of%20DWG%20Instruction.pdf

[40] Antrax, *Session Period Script*, Antrax. [Online]. Available: https://gitlab.com/flamesgroup/antrax/-/blob/master/doc/manual/scripts/session-period-script.md

[41] ——, *Activity Period Script*, Antrax. [Online]. Available: https://gitlab.com/flamesgroup/antrax/-/blob/master/doc/manual/scripts/activity-period-script.md

[42] Dinstar, *UC2000-VE/F/G GSM/CDMA/WCDMA VoIP Gateway User Manual*, Dinstar. [Online]. Available: https://www.dinstar.com/WEB/files/47154/2019-04-30/UC2000-VE&VF&VG%20GSM&CDMA&WCDMA%20VoIP%20Gateway%20User%20Manual.pdf

[43] Ejoin, *EJOIN ACOM5xx VoIP Gateway User Manual*, Ejoin.

[44] Antrax, *Gateway Selector Script*, Antrax. [Online]. Available: https://gitlab.com/flamesgroup/antrax/-/blob/master/doc/manual/scripts/gateway-selector-script.md

[45] Hybertone, *SIM Bank Scheduler Server User Manual*, Hybertone. [Online]. Available: http://www.hybertone.com/uploadfile/download/20171222180904804.pdf

[46] Antrax, *IMEI Generator Script*, Antrax. [Online]. Available: https://gitlab.com/flamesgroup/antrax/-/blob/master/doc/manual/scripts/imei_gen_script.md

[47] ——, *HTTP request*, Antrax. [Online]. Available: https://gitlab.com/flamesgroup/antrax/-/blob/master/doc/manual/scripts/business-activity-scripts/http-request.md

[48] ——, *USSD*, Antrax. [Online]. Available: https://gitlab.com/flamesgroup/antrax/-/blob/master/doc/manual/scripts/business-activity-scripts/ussd.md

[49] N. VoiceBlue, *2N VoiceBlue Enterprise User Manual*, 2N VoiceBlue.

[50] Antrax, *SMS*, Antrax. [Online]. Available: https://gitlab.com/flamesgroup/antrax/-/blob/master/doc/manual/scripts/business-activity-scripts/sms.md

[51] B. S. T. Journal, "Araxxe on the art of deception and analysis in sim box fraud warfare," 2019.

[52] CSGi, https://www.csgi.com/portfolio/digital-wholesale/assure/assure-sim-box-detection/.

[53] Araxxe, https://www.araxxe.com/p/our-services/global-transaction-verification/global-transaction-verification/test-call-generator-outsourcing.html.

[54] Pixip, https://www.pixip.net/index.php/solutions/test-call-generation.html.

[55] Calltic, https://www.calltic.com/.

[56] MediaFon, https://www.mediafonts.lt/.

[57] H. M. Marah, O. M. Elrajubi, and A. A. Abouda, "Fraud detection in international calls using fuzzy logic," in *International Conference on Computer Vision and Image Analysis Applications*, 2015, pp. 1–6.

[58] R. Sallehuddin, S. Ibrahim, A. Zain, and A. Elmi,

"Detecting sim box fraud using neural network," in *IT Convergence and Security 2012*, K. J. Kim and K.-Y. Chung, Eds. Dordrecht: Springer Netherlands, 2013, pp. 575–582.

[59] ——, "Detecting sim box fraud by using support vector machine and artificial neural network," vol. 74, 04 2015, pp. 137–149.

[60] N. A. Ibrahem Soliman Alsadi, "Study to use neo4j to analysis and detection sim-box fraud," vol. 17, Jan. 2019. [Online]. Available: http://sebhau.edu.ly/journal/index.php/jopas/article/view/422

[61] P. Ferreira, R. Alves, O. Belo, and L. Cortesão, "Establishing fraud detection patterns based on signatures," in *Advances in Data Mining. Applications in Medicine, Web Mining, Marketing, Image and Signal Mining*, P. Perner, Ed. Berlin, Heidelberg: Springer Berlin Heidelberg, 2006, pp. 526–538.

[62] C. Shearer, "The crisp-dm model: The new blueprint for data mining," *Journal of Data Warehousing*, vol. 5, no. 4, 2000.

[63] S. Haykin, *Neural Networks: A Comprehensive Foundation (3rd Edition)*. USA: Prentice-Hall, Inc., 2007.

[64] M. Kashir and S. Bashir, "Machine learning techniques for sim box fraud detection," in *2019 International Conference on Communication Technologies (ComTech)*, 2019, pp. 4–8.

[65] H. Kahsu, "Sim-box fraud detection using data mining techniques: The case of ethio telecom," Ph.D. dissertation, School of Electrical and Computer Engineering Addis Ababa Institute of Technology, 11 2018.

[66] Y. Freund and L. Mason, "The alternating decision tree learning algorithm," in *Proceedings of the Sixteenth International Conference on Machine Learning*, ser. ICML '99. San Francisco, CA, USA: Morgan Kaufmann Publishers Inc., 1999, p. 124–133.

[67] G. Holmes, B. Pfahringer, R. Kirkby, E. Frank, and M. Hall, "Multiclass alternating decision trees," in *Machine Learning: ECML 2002*, T. Elomaa, H. Mannila, and H. Toivonen, Eds. Berlin, Heidelberg: Springer Berlin Heidelberg, 2002, pp. 161–172.

[68] J. a. Gama, "Functional trees," *Mach. Learn.*, vol. 55, no. 3, p. 219–250, Jun. 2004. [Online]. Available: https://doi.org/10.1023/B:MACH.0000027782.67192.13

[69] V. A. Balasubramaniyan, A. Poonawalla, M. Ahamad, M. T. Hunter, and P. Traynor, "Pindr0p: Using single-ended audio features to determine call provenance," in *Proceedings of the 17th ACM Conference on Computer and Communications Security*, ser. CCS '10. New York, NY, USA: Association for Computing Machinery, 2010, p. 109–120. [Online]. Available: https://doi.org/10.1145/1866307.1866320

[70] J. S. Garofolo, L. F. Lamel, W. M. Fisher, J. G. Fiscus, D. S. Pallett, N. L. Dahlgren, and V. Zue, "Timit acoustic-phonetic continuous speech corpus," *Philadelphia: Linguistic Data Consortium*, 1993.

[71] O. M. Elrajubi, A. M. Elshawesh, and M. A. Abuzaraida, "Detection of bypass fraud based on speaker recognition," in *2017 8th International Conference on Information Technology (ICIT)*, 2017, pp. 50–54.

[72] L. S. Technology Research Institute (TRI), "White paper: Network protocol analysis: A new tool for blocking international bypass fraud before revenue is lost," Tech. Rep., 2015.

[73] M. Shafi, A. F. Molisch, P. J. Smith, T. Haustein, P. Zhu, P. De Silva, F. Tufvesson, A. Benjebbour, and G. Wunder, "5g: A tutorial overview of standards, trials, challenges, deployment, and practice," *IEEE Journal on Selected Areas in Communications*, vol. 35, no. 6, pp. 1201–1221, 2017.

[74] J. G. Andrews, S. Buzzi, W. Choi, S. V. Hanly, A. Lozano, A. C. K. Soong, and J. C. Zhang, "What will 5g be?" *IEEE Journal on Selected Areas in Communications*, vol. 32, no. 6, pp. 1065–1082, 2014.

[75] A. Gupta and R. K. Jha, "A survey of 5g network: Architecture and emerging technologies," *IEEE Access*, vol. 3, pp. 1206–1232, 2015.

[76] K. Letaief, W. Chen, Y. Shi, J. Zhang, and Y.-J. Zhang, "The roadmap to 6g: Ai empowered wireless networks," *IEEE Communications Magazine*, vol. 57, pp. 84–90, 08 2019.

[77] W. Saad, M. Bennis, and M. Chen, "A vision of 6g wireless systems: Applications, trends, technologies, and open research problems," *IEEE Network*, vol. 34, no. 3, pp. 134–142, 2020.

[78] document ITU-R M.[IMT-2020.TECH PERF REQ], "Minimum requirements related to technical performance for imt-2020 radio interface(s)," October 2016.

[79] N. A. Mohammed, A. M. Mansoor, and R. B. Ahmad, "Mission-critical machine-type communication: An overview and perspectives towards 5g," *IEEE Access*, vol. 7, pp. 127 198–127 216, 2019.

[80] M. Richarme, "The virtual SIM - a feasibility study," Richard Petersens Plads, Building 324, DK-2800 Kgs. Lyngby, Denmark, compute@compute.dtu.dk, 2008, in collaboration with DTU and Nokia Denmark A/S. [Online]. Available: http://www.compute.dtu.dk/English.aspx

[81] S. Zhao and B. Smeets, "Virtual sim card cloud platform," Jul. 9 2019, uS Patent 10,349,272.

[82] D. L. Polehn, P. R. Chang, F. Weisbrod, and C. J. Christopherson, "System and method for virtual sim card," Nov. 6 2018, uS Patent 10,123,202.

[83] J. Liu, X. Qin, and B. Du, "Method and system for international roaming using virtual sim card," Jan. 24 2008, uS Patent App. 11/746,493.

[84] G. Shi, V. Tangirala, T.-Y. Siu, J. Durand, and S. A. Sprigg, "Virtual sim card for mobile handsets," Aug. 19 2014, uS Patent 8,811,969.

[85] M. Simsek, A. Aijaz, M. Dohler, J. Sachs, and G. Fettweis, "5g-enabled tactile internet," *IEEE Journal on Selected Areas in Communications*, vol. 34, no. 3, pp. 460–473, 2016.

[86] J. Sujata, S. Sohag, D. Tanu, D. Chintan, P. Shubham, and G. Sumit, "Impact of over the top (ott) services on telecom service providers," *Indian Journal of Science and Technology*, vol. 8, no. S4, pp. 145–160, 2015.

[87] M. D. Bhawan and J. L. N. Marg, "Regulatory framework for over-the-top (ott) services," *Telecom Regulatory Authority of India*, pp. 1–118, 2015.

[88] N. Wasmi, "Telecoms regulator says viber is' unlicensed'in the uae," 2014.

[89] ITU, "The impact of taxation on the digital economy," 2015.

[90] A. H. Zhou, T. T. Zhou, and D. T. Zhou, "Systems and methods for digital multimedia capture using haptic control, cloud voice changer, and protecting digital multimedia privacy," Mar. 3 2015, uS Patent 8,968,103.

[91] F. Horikawa, "Telephone with voice changer and control method and control program for the telephone," Nov. 30 2006, uS Patent App. 11/438,834.

[92] P. Bonnard, I. Bourmeyster, X. Fourquin, and P. Ladouce, "Telecommunication terminal able to modify the voice transmitted during a telephone call," Sep. 14 2010, uS Patent 7,796,748.