Battle of Wits: To What Extent Can Fraudsters Disguise Their Tracks in International bypass Fraud?

Anne Josiane Kouam, Aline Carneiro Viana, Alain Tchana

Battle of Wits: To What Extent Can Fraudsters Disguise Their Tracks in International bypass Fraud?

Anne Josiane Kouam*
TU Berlin
Germany

Aline Carneiro Viana
INRIA
France

Alain Tchana
Grenoble INP
France

ABSTRACT

International bypass fraud, also known as *SIMBox* fraud, involves diverting international cellular voice traffic from regulated routes and rerouting it as local calls in the destination country. It has significantly affected cellular networks worldwide, generating $3.11 Billion of losses annually and threats to national security. Yet, *SIMBox* fraud remains an ongoing challenge, eluding operators detection due to the continual refinement of fraudulent behavior that is often overlooked in the design and validation of detection methods.

This paper introduces a game-based formalization of the *SIMBox* fraud problem, delineating two key players–the adversary and the investigator–along with their strategies and a set of metrics gauging their efficacy in the game. We develop a practical framework for the empirical evaluation of the fraud, incorporating current adversary and investigator capabilities and accommodating seamless adaptation to the evolving nature of fraud. Our analysis identifies up to 345,600,000 possible adversary strategies from in-market *SIMBox* appliances functionalities. The most sophisticated strategies decisively outperform the most efficient existing detection methods, underscoring the literature's lack of awareness of fraud capabilities. Furthermore, we uncover fraud vulnerabilities and discuss their implications for enhancing future detection strategies in practice. In essence, our work introduces a novel paradigm in *SIMBox* fraud detection that adapts seamlessly to the ever-changing landscape of fraud, treating it as a fundamental aspect of the detection strategy.

CCS CONCEPTS

• **Security and privacy** → **Mobile and wireless security**; • **Computing methodologies** → *Modeling and simulation.*

KEYWORDS

Mobile cellular networks, *SIMBox* fraud modeling, Fraud detection

ACM Reference Format:
Anne Josiane Kouam*, Aline Carneiro Viana, and Alain Tchana. 2024. Battle of Wits: To What Extent Can Fraudsters Disguise Their Tracks in International bypass Fraud?. In *ACM Asia Conference on Computer and Communications Security (ASIA CCS '24), July 1–5, 2024, Singapore, Singapore.* ACM, New York, NY, USA, 17 pages. https://doi.org/10.1145/3634737.3657023

* This work was fully done as part of Anne Josiane Kouam's PhD at Inria.

ASIA CCS '24, July 1–5, 2024, Singapore, Singapore

ACM ISBN 979-8-4007-0482-6/24/07.
https://doi.org/10.1145/3634737.3657023

1 INTRODUCTION

SIMBox fraud is one of the most prevalent scams in cellular networks, being in the top three types of phone system frauds causing a significant loss to network operators [49]. Fraudsters take advantage of the fact that International phone call routing is complex and lacks transparency, to divert the international voice traffic from the regulated routes, through VoIP established links. The diverted traffic is received at the level of a *SIMBox* (VoIP to GSM gateway) in the destination country and re-originated as a national mobile call to its recipient (cf. Fig. 1). Hence, mobile destination operators perceive national termination fees, considerably cheaper than international rates, while fraudsters benefit from the price discrepancy.

The impact of this problem is worldwide, affecting major developed (as the case of the USA [42]) to emerging states, harming operators' revenues, network quality, networking research, users privacy, or national security. Mainly in developing countries, up to 70% of incoming international call traffic is terminated fraudulently, leading to a revenue loss estimated to $3.11 Billion annually [49]. *SIMBox* fraud also degrades consumers' quality of experience due to call initiation delays and network unavailability, which increases churn. Beyond financial aspects, *SIMBox*'s re-originated calls induce bias in operators' network usage records that appear with incorrect origins and locations, impacting multiple analyses and research [39]. More significantly, *SIMBox* appliances allow eavesdropping on international call conversations [16], impeding users' privacy and giving way for international espionage. They can further permit international terrorists to conduct covert activities, masquerading as national subscribers. *The induced possibilities of terrorism attest SIMBox fraud deserves much more attention.*

Due to protocol and regulatory weaknesses rooted in the mobile network organization at an international scale [44], *SIMBox* fraud detection investigations are conducted only at the destination operator experiencing losses. According to existing literature (cf. Table 1), such investigations consist of analyzing network-related datasets (i.e., CDRs, call audio, or signaling data) to distinguish between legitimate users' traffic and *SIMBox*'s one. Here, CDRs (Charging Data Records) refer to time-stamped and geo-referenced events (i.e., data, calls, text) generated from the interactions between mobile devices and operators' cellular network (cf. Table 2). *This paper focuses on CDR-based methods for the mitigation of SIMBox fraud, as being the most widely employed and the most practical within the operational constraints of mobile networks (cf. §2.2).*

Motivation. CDR-based detection involves the supervised classification of network users (identified by an International Mobile Subscriber Identity) as *fraudulent* or *legitimate* based on per-user communication features extracted from CDRs. Despite the high accuracy of the CDR-based detection literature (cf. Table 1), the impact

of *SIMBox* fraud continues to escalate [6, 49]. Indeed, *SIMBox* fraudsters constantly create and refine their strategies to mimic human communication behavior and be indistinguishable [35]. *Therefore, SIMBox fraud mitigation becomes a challenging battle of wits, where ambiguity prevails to the advantage of fraudsters*, as elaborated:

- First, as with supervised classification, *SIMBox* detection relies heavily on operators providing ground truth data about known fraudulent or legitimate users to train, guide, and validate detection methods. Hence, the quality and representativeness of the provided ground truth directly impact the effectiveness of the developed detection methods. In the context of *SIMBox* detection, fraudulent ground truth typically derives from operators' past detection efforts, often involving active methods like test calls [35]. Test calls are intentional calls made by operators to specific destinations to identify potential fraudulent activities. However, *the fraud dynamic nature poses a challenge as fraudulent behavior constantly evolves, potentially rendering the ground truth less representative over time and limiting its ability to offer meaningful insights for detecting evolved SIMBox frauds.* Consequently, *there is a pressing need for continually evolving fraudulent ground truth that accurately reflects the actual fraud capabilities, enabling detection to stay up-to-date on emerging challenges.*
  While one (costly) approach to achieving this is through the regular performance of test calls, some articles [28] shed light on a drawback: fraudsters deliberately allow pools of SIM cards to be detected through test calls to outsmart operators' vigilance. These "sacrificed" SIM cards are chosen for their seemingly naive fraudulent behaviors, rendering them useless for identifying other fraudulent SIM cards. Therefore, we highlight *the importance of exploring alternative strategies for acquiring representative fraudulent ground truth in the dynamic SIMBox fraud landscape.*
- Second, given that detection methods heavily rely on fraudulent ground truth describing a specific set of fraudulent behaviors, interpreting detection validity becomes challenging. Indeed, current literature on *SIMBox* detection has, until now, demonstrated effectiveness within specific and non-characterized contexts provided by the non-public datasets they rely on (e.g., one week's CDRs of an Ethiopian operator [30]). However, as the fraud behavior may vary from one dataset to another, *the accurate interpretation of detection results heavily depends on a deep understanding of the fraud behavior targeted by a detection methodology.* Without this crucial context, assessing the generality and success of a detection approach across different datasets or evolving fraud behaviors becomes complex. Therefore, we emphasize *the need for comprehensive fraud behavior structuring, essential for developing universally applicable and interpretable detection methodologies amidst the shifting landscape of SIMBox fraud.*

Approach. In this paper, we present a pioneering formalization of *SIMBox* fraud detection through CDR-based analysis, acknowledging the fraud dynamic nature. *We conceptualize the problem as a strategic game wherein an adversary systematically introduces fraudulent users among legitimate ones.* The adversary's goal is to replicate the communication behaviors of genuine users using automated functionalities within *SIMBox* appliances. Meanwhile, an investigator aims to identify distinctive communication features that differentiate fraudulent users. Unlike genuine users, who generate traffic and mobility patterns based on their routine-like activities, fraudsters employ computerized methods. Through our formalization, we thoroughly investigate how the fraud dynamic nature, in various computerized human behavior reproduction strategies, impacts detection performance and how to strategically position detection methods in response. This exploration enables us to *(i) assess the actual capabilities of SIMBox fraud detection, previously overestimated in the literature due to an ambiguous fraudulent context, and (ii) gain insights into the fraud strengths and weaknesses for future detection enhancements.* Consequently, this investigation lays the foundation for a new paradigm in fraud detection – a dynamic approach seamlessly adapting to the evolving nature of fraud, integrating it as a fundamental aspect of the detection strategy.

Contribution. This paper achieves the following contributions as sequential steps in implementing the approach described earlier:

(1) First, we introduce a generic methodology to study CDR-based *SIMBox* fraud detection (**§3**). This approach formalizes the process using a game-theoretic approach distinguishing an adversary, who incorporates fraudulent behaviors, based on *SIMBox* functionalities, defined as a "*SIMBox fraud model*" into a group of legitimate users, and an investigator, who suggests a set of communication features to discern the introduced frauds. The game is evaluated using metrics measuring the efficiency of the *SIMBox* fraud model against the investigator's proposal and enabling the assessment of the performance of both parties.
(2) Second, building upon the established formalization, we develop a scalable framework for the empirical examination of *SIMBox* fraud (**§4**). This framework establishes a game environment that generates fraudulent ground truth related to an input *SIMBox* fraud model from which it computes the game metrics for a specified detection strategy. As part of this practical framework, we shed light on the existing *SIMBox* fraud models through an in-depth review of the fraud ecosystem, scrutinizing the functionalities of all 94 *SIMBox* appliances produced by major international manufacturers and used by over 2000 fraudsters in more than 31 countries [13, 17]. Similarly, we shed light on existing detection strategies, enabling us to leverage the gaming environment to comprehensively analyze actual fraud detection capabilities against various *SIMBox* fraud models of different sophistication levels in **§5**.
(3) Our investigations in **§5** reveal that current *SIMBox* capabilities result in elaborated fraud models outperforming existing detection strategies with zero detection precision and recall when the number of fraudulent SIM cards is low (i.e., 50 SIMs). Additionally, we identify *mobility behavior* as a genuine fraud vulnerability, challenging for fraudsters to realistically simulate, especially over an extended *observation period.*
(4) At last, in **§6**, we provide practical recommendations for enhancing future detection based on quantitative and qualitative insights into fraud strengths and weaknesses. The formalization and empirical study framework of this paper give the necessary flexibility to implement these recommendations.

Organization. Besides the above sections, we give in **§2** our work background and discuss the related works. At last in **§7**, we provide our research conclusion.

Table 1: Overview of literature's *SIMBox* fraud detection

Ref.	Date	Data type	Avg. Accur.
[45]	2013	CDRs	98.71%
[38]	2014	CDRs	99.95%
[46]	2014	CDRs	98.8%
[42]	2015	Audio	87%
[37]	2015	CDRs	*No eval*
[32]	2015	CDRs	99.99%
[1]	2016	CDRs	83.34%
[27]	2017	CDRs	*No eval*
[10]	2017	Audio	*No eval*
[30]	2018	CDRs	83.2%
[26]	2019	CDRs	99.9%
[31]	2019	CDRs	99.3%
[53]	2020	CDRs	*No eval*
[51]	2020	CDRs	95.55%
[40]	2023	Signaling	*No eval*

2 SETTING THE *SIMBOX* FRAUD STAGE

This section contextualizes the *SIMBox* fraud problem and its related challenges while discussing the limitations hindering its investigation. Furthermore, it examines existing literature, shedding light on the constraints that inspire our efforts to address the current gaps and enhance the understanding of *SIMBox* fraud detection.

2.1 Fraud ecosystem

***SIMBox* architecture.** The *SIMBox* is a VoIP to GSM gateway: a system configured as a VoIP client to receive VoIP call traffic and terminate it by re-originating cellular mobile calls using numerous SIM cards. The *SIMBox* automatically creates "decoupled" mobile devices by binding SIM cards and GSM modules. In this association, the GSM module provides wireless communication with the cellular network, and the SIM card identifies and authenticates the formed device, referred to as "fraudulent" in this paper. A *SIMBox* architecture comprises three kinds of interacting components:

- The gateway is a rack of GSM modules. It receives incoming VoIP traffic and distributes it to the *SIMBox* GSM modules. E.g, the GoIP324 model [23] is a 2G gateway with 32 GSM modules.
- The SIMBank is a device with numerous SIM slots that remotely holds a bundle of SIM cards. It handles the *SIMBox* SIM cards, i.e., their addition, removal, and their data transfer to other components. E.g., the SMB128 model [24] manages 128 SIM cards.
- The control server is a web server providing the *SIMBox* control functions – i.e., the binding of SIM cards to GSM modules – and the whole architecture's configuration. It can be hosted online to facilitate remote access from a web GUI client.

Fraud mechanism. Fig. 1 illustrates a typical *SIMBox* fraud mechanism. The *flow 1* (F1) of traffic in the figure represents an international call routing scheme *with no fraud*. The call traffic leaves the caller's mobile operator (Operator A) and is routed to the destination country through a set of transit carriers. These carriers facilitate traffic interconnection between countries by buying and reselling international termination routes. Consequently, the callee's operator (Operator B) receives the traffic from a transit carrier and then forwards it to the destination as an international mobile call.

Nevertheless, a transit carrier can be fraudulent. Instead of adhering to legitimate practices, a fraudulent carrier sometimes diverts the traffic it receives through a low-cost VoIP trunk, as in the *flow 2* (F2) on Fig. 1. The diverted traffic is sent to a *SIMBox* (VoIP to GSM gateway) in the destination country and re-originated as *a national mobile call* to its callee. Once in the destination country, there are two possible fraudulent termination scenarios: (i) F2-1 is an on-net termination when the re-originated call is made using a SIM of Operator B, the same operator of the callee and (ii) F2-2 is an off-net termination wherein the fraudster employs a SIM card from a different local operator in the destination country.

2.2 Fraud detection literature

Despite the striking impact of *SIMBox* fraud, this problem remains little tackled due to the challenge of obtaining related data. As reported in Table 1, we identified only 15 *SIMBox* fraud detection approaches since 2011, which is relatively low for a security problem of this importance that continues to plague the world today. The vast majority of literature contributions rely on CDR-based analysis. This method is performed offline, benefiting from operators' historical data for extracting patterns that give a comprehensive view of fraudulent activities over time without impacting the network performance. Conversely, a few contributions (3/15) propose online approaches to detect ongoing fraudulent calls through call audio analysis [10, 42] or to filter network access to *SIMBox* devices through signaling-based fingerprints [40]. Unfortunately, the practical implementation of such online contributions in operators' networks is still challenging due to scalability concerns. Indeed, being real-time, such online solutions should run smoothly across the entire access network surface, constituted of hundreds to thousands of base stations. This would require investigating all local calls audio across the network or maintaining a voluminous database of all network devices' fingerprints at each base station to examine each device's connection.

Hence, given real-world constraints, CDR-based investigations are the primary and practical approach for tackling *SIMBox* fraud in mobile operators' networks. Moreover, operating offline, this method is essential for extending the applicability and reach of other online detection approaches by leveraging historical data.

2.3 Positioning

This paper concentrates on CDR-based mitigation of *SIMBox* fraud, the most widely adopted method in the existing literature. Our goal is not to propose a new CDR-based detection method that outperforms the current state-of-the-art. Instead, we highlight a nuance of CDR-based fraud mitigation often overlooked in prior research: *CDR-based SIMBox fraud detection operates within an environment where fraud consistently evolves to adapt to the target solutions. Consequently, there isn't a singular SIMBox fraud but multiple expressions of SIMBox fraud that can occur in various ways.* This consideration gives rise to the following crucial implications:

First, fraudulent users' behavior may vary significantly from one dataset to another to be more or less elaborate. Therefore, the fundamental question of "*Which precise SIMBox fraud is tackled?*" becomes a crucial starting point for any detection methodology. Unfortunately, this central aspect is frequently omitted in the current CDR-based literature, making it hard to interpret. Indeed, despite excellent classification performances, these techniques often lack a precise characterization of the fraud they aim to address.

Second, this ambiguity surrounding the fraud behavior extends to the design of detection solutions. Unfortunately, *the absence of clarity on the specific fraud being addressed creates a misleading impression that the designed technique universally applies to detecting*

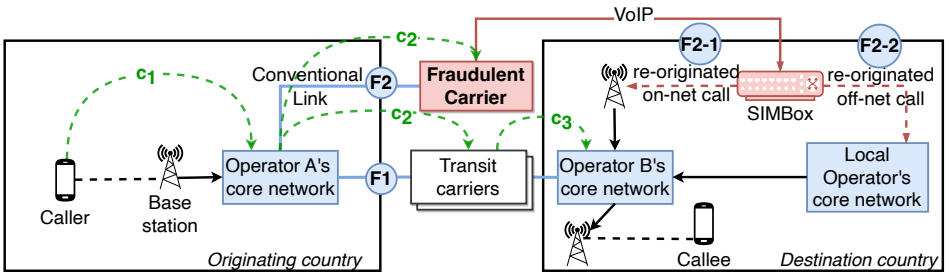


Figure 1: International call routing: (Flow 1, i.e., F1) Legitimate scheme, (Flow 2, i.e., F2) Fraudulent scheme.

Table 2: CDRs format.

	CDR field - r_p
Traffic	Timestamp - t_p Event-type (call/text/data) -et_p Call duration (if call) - em_p Data volume (if data) - em_p
Mobility	Network cell Id - cid_p
Social	Contact's phone number (if call/text) - con_p
Device properties	Phone number - u_p Phone identifier (IMEI) - dev_p

(all forms of) "SIMBox fraud." However, in reality, the solution's efficiency is confined to the specific, often undisclosed, context. Hence, gaining insights from these detection designs to be valuable in contexts with different fraud behaviors becomes challenging. *The critical link between the detection design and the tackled model, which should be the main takeaway, remains entirely concealed.*

The previous discussions underscore the cruciality of modeling *SIMBox* fraud to enable meaningful detection. While not a novel concept, the literature recognizes the importance of attack modeling, often termed "threat model" [5, 43] or "attack model" [4, 8, 52], in addressing evolving security challenges. In the context of *SIMBox* fraud, we position our work as the pioneering attempt in *SIMBox* fraud modeling, constituting a notable contribution to the field.

3 GAME-THEORETIC *SIMBOX* FRAUD

In this section, we frame the *SIMBox* fraud problem through a game-theoretic lens. The scenario involves a mobile operator scrutinizing a CDRs dataset spanning a given time frame T (e.g., one-month CDRs in [38]). The dataset includes legitimate users, representing genuine mobile consumers, and fraudulent users, characterizing SIM cards within *SIMBox* appliances for illicit call termination.

We formalize the problem as a non-cooperative game between an *Adversary* manipulating the *SIMBox* appliances to define fraudulent users' behavior and a *investigator* (i.e., the operator) trying to differentiate fraudulent users from legitimate ones. The game performs in a non-cooperative manner, where stakeholders independently make decisions, and assumes each participant acts without knowledge of the other's choices. Importantly, we posit it as a zero-sum game, where any gain by one stakeholder precisely equates to a loss suffered by the other (e.g., the operator's accuracy in detection is mirrored by the count of blocked SIM cards for fraudsters).

Subsequently we introduce important notations and outline the set of strategies employed by each stakeholder in the game. At last, we establish measures of the game's gain or loss directly tied to the efficiency of each stakeholder's strategy.

3.1 Notation

We denote the set of all users as $U = \{u_1, u_2, u_3, ..., u_N\}$ in the CDRs dataset with $|U| = N$ users. Each user is either legitimate, i.e., $u_i \in U^l$, or fraudulent, i.e., $u_i \in U^f$, with $U = U^l \cup U^f$. We use $r_p = \{u_p \rightarrow (t_p, et_p, em_p, cid_p, con_p, dev_p)\}$ to denote a CDRs record: generated by the user u_p, at the timestamp t_p, for an event type et_p with the related metric em_p (e.g., event duration), interacting contact con_p, while located at the cell id cid_p and using a device identifier dev_p (i.e., IMEI) (cf. Table 2). The event metric, em_p, is the call duration if the event is a call (i.e., $et_p = Call$) and the data volume if the event is a data usage ($et_p = Data$). Similarly,

$$con_p = \begin{cases} u_p & \text{if } et_p = Data \\ u_j, j \in \{1, \ldots, N\} \setminus \{p\} & \text{if } et_p = Call/Text \end{cases}$$

We denote as R^{u_i} the set of CDRs records related to the user u_i, i.e., $R^{u_i} = \{r_p \mid u_p = u_i \text{ or } con_p = u_i\}$. The overall set of users' related records is $R_u = \{R^{u_i} \mid u_i \in U\}$. Similarly, we define $R^{u_i}_{T'} \subseteq R^{u_i}$ as the subset of CDRs records related to the user u_i, generated during a time interval $T' \subseteq T$. Formally, $R^{u_i}_{T'} = \{r_p \mid \forall r_p \in R^{u_i} \text{ where } t_p \in T'\}$. Finally, for a given time interval T', we have $R_{UT'} = \{R^{u_i}_{T'} \mid \forall u_i \in U\}$, representing the set of all users' related records over T'.

3.2 Adversary strategy: *SIMBox* fraud model

While legitimate users' set of records $R^l_u = \{R^{u_i} \mid u_i \in U^l\}$ naturally emerges through genuine human activities, fraudulent users' set of records $R^f_u = \{R^{u_i} \mid u_i \in U^f\}$ is automatically generated through the *SIMBox*. Indeed, the *SIMBox* employs a range of algorithms to control the operations of fraudulent devices autonomously, dictating their behavior across specific CDRs record fields, namely $(t, et, em, con, cid, dev)$. By configuring these algorithms, the adversary shapes the communication behavior of a subset of N_f fraudulent users, where $N_f < N$. The ultimate goal is to devise an optimal configuration that renders these fraudulent users virtually indistinguishable from their legitimate counterparts.

We define the following sets, encompassing the algorithms used by the *SIMBox* for defining fraudulent users' behavior respectively to the CDRs record fields $(t, et, em, con, cid, dev)$:

$$T_a = \{alg^T_1, ..., alg^T_{T_a|}\}, ET_a = \{alg^{ET}_1, ..., alg^{ET}_{|ET_a|}\},$$
$$EM_a = \{alg^{EM}_1, ..., alg^{EM}_{|Dev_a|}\}, CID_a = \{alg^{CID}_1, ..., alg^{CID}_{|CID_a|}\},$$
$$CON_a = \{alg^{CON}_1, ..., alg^{CON}_{|CON_a|}\}, DEV_a = \{alg^{DEV}_1, ..., alg^{DEV}_{|DEV_a|}\}$$

For instance, T_a specifically groups the $|T_a|$ algorithms supported by the *SIMBox* to define the timing of fraudulent users' events generation. An example is, alg^T_1, which could represent a periodic algorithm establishing event generation with a fixed frequency.

Such *SIMBox* algorithms might be parametric. For instance, the periodic timing algorithm is tuned by a single parameter that is the frequency or period. Consequently, algorithms can vary in complexity (the number of required parameters) and efficiency

(their ability to replicate human behavior). Therefore, we define the sets $T_p, ET_p, EM_p, CID_p, CON_p,$ and DEV_p, which consist of parameters associated with each *SIMBox* algorithm. For example, $T_p = \{(P_i)_1^T, (P_j)_2^T, ..., (P_\alpha)_{|T_p|}^T\}$, with $|T_a| = |T_p|$ and $(P_i)_k^T$ representing parameters of the algorithm alg_k^T.

We define a *SIMBox* fraud model fm as the representation of the adversary's strategy for generating the communication behavior of fraudulent users. It involves a combination of algorithm choices drawn from sets, specifically

$$fm \quad = \quad (alg_*^T, alg_*^{ET}, alg_*^{EM}, alg_*^{CID}, alg_*^{CON}, alg_*^{DEV})$$

A *SIMBox* fraud model, denoted as fm, can give rise to multiple *SIMBox* fraud implementations, each characterized by distinct values of the corresponding chosen algorithms' parameters. For a specific instance of a *SIMBox* fraud model, we denote the set of parameter values as pm, and thus, a fraud model instance is represented as the tuple (fm, pm), where

$$pm = (((P_i)_*^T, (P_j)_*^{ET}, (P_k)_*^{EM}, (P_m)_*^{CID}, (P_l)_*^{CON}, P_n)_*^{DEV})$$

We denote as $R_u^{(fm,pm)}$ the set of fraudulent user records generated by a *SIMBox* fraud model instance (fm, pm).

3.3 Investigator strategy: detection model

The investigator is tasked with determining the legitimacy of users based on their respective sets of records, denoted as $R_u = \{R^{u_i} \mid u_i \in U\}$. Typically, investigators employ ML classifiers (e.g., Support Vector Machine [30, 46, 51] or Random Forest [30, 38, 51]) to make these determinations. Here, we delve deeper into the strategic choices the investigator precisely employs for this detection.

Within a specified time interval $T' \subseteq T$, we define a *feature* as a function $f^{T'} : R_{UT'} \rightarrow \mathbb{R}$. This function takes a user's set of records $R_{T'}^{u_i}$ during the time interval T' as input and computes a mathematical function that aggregates these records into a unique value. For example, a feature might calculate the number of records in $R_{T'}^{u_i}$ or the average value of the *em* field across all the user's records. From a set of features $F^{T'} = \{f_1^{T'}, f_2^{T'}, ..., f_{|F^{T'}|}^{T'}\}$, we derive a matrix $V^{T'}$ of size $|U| \times |F^{T'}|$, where $V(u_i, f_j^{T'}) = f_j^{T'}(R_{T'}^{u_i})$. The vector $V^{T'}(u_i)$, representing the i^{th} row of the matrix $V^{T'}$, serves as a representation of the user's records $R_{T'}^{u_i}$, allowing the investigator to assess the user's communication behavior.

We model the investigator's strategy, or *detection model*, i.e., dm, as a composite structure comprising:

(1) an *observation period* $T_O \subseteq T$,
(2) a *set of features* $F^{T_O} = \{f_1^{T_O}, f_2^{T_O}, ..., f_{|F^{T_O}|}^{T_O}\}$, representing users' communication behavior within the observation period, and
(3) a binary *classifier* clf, categorizing users, based on such representation, as fraudulent or legitimate.

The investigator has the flexibility to define multiple features of varying complexities in terms of the number of fields involved in the calculation. The ultimate goal is to propose a representation that maximizes, in the vector space, the distance between legitimate users, i.e., $V_l^{T_O} = \{V^{T_O}(u_i) \mid u_i \in U^l\}$, and fraudulent users resulting from the fraud model instance (fm, pm), i.e., $V_{(fm,pm)}^{T_O} = \{V^{T_O}(u_i) \mid R^{u_i} \in R_u^{(fm,pm)}\}$, to accurately distinguish between fraudulent and legitimate users. By adjusting the observation period T_O, the investigator includes more or fewer user records in the feature computation, thereby enhancing differentiation.

Each classifier, clf, uses a specific algorithm to analyze the vector space for the categorization process. When applied to the user's vector representation V^{T_O}, it produces a binary vector $\hat{Y}$ (0 for legitimate, 1 for fraudulent) of size N, aiding the investigator in making informed decisions about user legitimacy.

3.4 Game metrics

Given our game-based definition, we assess the fraud detection in terms of the (i) the adversary's proficiency in concealing fraudulent users and (ii) how well it influences the investigator's classification performances. To this end, we define two sets of metrics.

3.4.1 The in-crowd blending capability. This metric assesses how well the adversary's generated fraudulent users blend into the crowd of legitimate users based on their communication behavior. Indeed, the more a fraud model yields fraudulent users' behaviors close to human ones, the more efficient it is and harder to detect.

To quantify this capability, we apply a multi-variate unsupervised clustering algorithm (e.g., DBSCAN) to user vectors given by the detection model. The clustering algorithm identifies clusters in areas of high density in the vector space, indicating regions with a minimum number of samples within a given distance. Isolated samples beyond a certain distance to a cluster are considered outliers. The algorithm, therefore, groups users with similar cellular communication behavior as specified by the investigator's strategy.

We classify fraudulent users into three categories based on the clustering results: (i) outlier users, referred to as the *outlier group* (*OG*), (ii) users in the same clusters as legitimate users, known as the *hybrid group* (*HG*), and (iii) users in clusters comprising only fraudulent users, labeled the *fraudulent-only group* (*FG*). The distribution of fraudulent users across these three categories provides insights into how effectively the *SIMBox* fraud model instance (fm, pm) camouflages into the legitimate crowd. We formalize the *in-crowd-blending capability*, denoted as $ICB_{(fm,pm)}^{dm}$, of the fraud model instance (fm, pm) in the detection model dm's space as follows:

$$ICB_{(fm,pm)}^{dm} = \frac{|HG|}{|HG| + |FG| + |OG|} \tag{1}$$

The *in-crowd-blending capability* is measured on a scale from 0 to 1. A value closer to 1 indicates the effectiveness of the adversary's strategy in making more fraudulent users behave similarly to legitimate users, putting the investigator at a disadvantage. Conversely, a value approaching 0 implies that the investigator has gained an advantage over the adversary, by proposing a set of features that effectively discriminates fraudulent users.

3.4.2 Classification capability. We consider three classification metrics assessing the effectiveness of the investigator strategy in determining which users are fraudulent given the fraud model:

(1) The *balanced accuracy (BA)* – rather than the classical one, as legitimate users outnumber fraudulent ones – is defined as the average of good classification obtained on each class.

(2) The *recall* indicates the investigator strategy's performance to find all the introduced fraudulent users, giving its capacity to resolve the fraud.
(3) The *precision* measures the ability of the investigator's strategy to not label as fraudulent a user that is not one, giving the confidence level in blocking a detected fraudulent user.

4 *SIMBOX* FRAUD PRACTICAL STUDY

Having formalized the *SIMBox* fraud problem with a game-theoretic approach, this section focuses on establishing a practical framework for its empirical study. Our objective is to assess the problem within this formalized structure, considering the current strategies employed by both the adversary (i.e., *SIMBox* fraudsters) and the investigator (i.e., existing detection methodologies). To this end, we use the following methodology, as illustrated in Fig. 2:

(1) First, with the aim of identifying real-world adversary strategies, we conduct an in-depth review of the *SIMBox* market in §4.1. This review enables us to uncover and classify the capabilities of current *SIMBox* appliances, shedding light on the *SIMBox* fraud models currently accessible to adversaries.
(2) Moving to the investigator side in §4.2, we undertake a parallel examination, identifying and classifying investigator's strategies proposed in the existing literature.
(3) Finally, in §4.3, we establish a controlled environment for the game process. This environment takes inputs of adversary and investigator strategies and generates the corresponding mobile interactions, yielding to a CDRs dataset of fraudulent and legitimate users. The output of this controlled environment are the in-crowd-blending and classification capabilities, measuring the game's overall effectiveness.

4.1 Adversary strategies in the *SIMBox* market

The adversary's strategy, or *SIMBox* fraud model, involves a thoughtful selection of the available *SIMBox* functionalities to generate fraudulent communication behaviors of varying effectiveness. This section presents our in-depth survey of in-market *SIMBox* appliances functionalities, giving the basis for *SIMBox* fraud modeling. We first introduce our employed methodology followed by a classification of the reported functionalities.

4.1.1 Methodology. To ensure a comprehensive analysis, we meticulously reviewed the functionalities of all 94 appliances from the *SIMBox* manufacturers accessible online. This assessment involved (i) analyzing user manuals, (ii) reviewing additional guiding resources in 1212 blog posts [18] and 66 video tutorials [12, 19], and (iii) manipulating five acquired *SIMBox* appliances.

Our study first considered the top 5 *SIMBox* manufacturers as reported by GoAntiFraud [17], a cloud-based service assisting *SIMBox* termination businesses. Since 2013, GoAntiFraud has helped over 2000 fraudsters in more than 31 countries, making the statistics fairly representative [13]. In addition, we included all manufacturers providing "VoIP GSM gateways" on commercial platforms such as Alibaba and Amazon, resulting in 12 *SIMBox* manufacturers outlined in Table 7 in the appendix. Subsequently, we collected and analyzed user manuals of all *SIMBox* component models on such manufacturers' websites. We also considered blog posts on the related functionalities, video tutorials, and public codes when available [11]. At last, we refined our study by manipulating five *SIMBox* appliances from the Hybertone manufacturer (top in the market) in an operator network test-bed inside a Faraday shield.

4.1.2 SIMBox fraud modeling. Thanks to the previous in-depth *SIMBox* market review, we uncover a full set of adversary strategies that we report in Table 3. To ensure clarity in presenting these strategies, we organize the uncovered *SIMBox* algorithms (col. 5) and their related parameters (col. 6) according to the intended motive (col. 3) in human behavior mimicking. All these are categorized based on the human communication behavior they try to reproduce (col. 1) and the corresponding impacted CDRs field (col. 2).

From Table 3, the construction of a *SIMBox* fraud model involves deciding, for each CDR field, whether to pursue a specific motive (M_i) followed by the selection of an algorithm ($M_i.j$) to achieve the desired motive. For example, consider the fraud model

$$fm = (alg_*^T = M_1.1,\ alg_*^{ET} = null,$$
$$alg_*^{EM} = M_1.1 \times M_2.1,\ alg_*^{CID} = M_1.1 \times M_5.1 \times M_6.1 \times M_7.1;$$
$$alg_*^{CON} = M_1.2,\ alg_*^{DEV} = M_1.3 \times M_2.1)$$

which illustrates an adversary's strategy where algo. $M_1.1$ is chosen to fulfill only the motive M_1 for the CDR t_p field. Regarding the CDR et_p field, no sophistication is sought. Concerning the CDR em_p field, both motives M_1 and M_2 are pursued with their respective algo. $M_1.1$ and $M_2.1$, and so forth.

The organization and classification process above provides an overview of the SIMBox fraud models obtainable from the combination of existing SIMBox algorithms in the market (col. 4) and their related motives (col. 3). This results in approximately 345,600,000 fraud models with varying levels of similarity based on their shared characteristics. Furthermore, the ability to parameterize these fraud models provides a limitless spectrum of instances. For a comprehensive understanding of the construction of these *SIMBox* fraud models, please refer to the discussion in §A.

4.2 Investigator strategies in existing detection

Here, we extract existing investigator strategies from state-of-the-art detection features. We conduct a systematic analysis of all contributions presented in Table 1, providing insights into their proposed observation periods (T_O), the utilized detection feature sets (F^{T_O}) and ML classifier (clf), where available.

Detection feature set, F^{T_O}. The final column of Table 3 outlines the identified detection features organized by CDRs field. The table references only literature that lists the employed detection features, ensuring reproducibility, which corresponds to 58.3% of related works. Notably, each detection study introduces a distinct set of features. The table indicates a prevalence of methodologies proposing features for detecting fraudulent traffic, followed by social and mobility behavior. The consideration of the device properties is infrequent, observed only in [38].

Observation period, T_O. The considered observation period is less transparent in the literature, often omitted or challenging to determine. Most methodologies are presented as universally applicable, irrespective of the observation period, which requires validation.

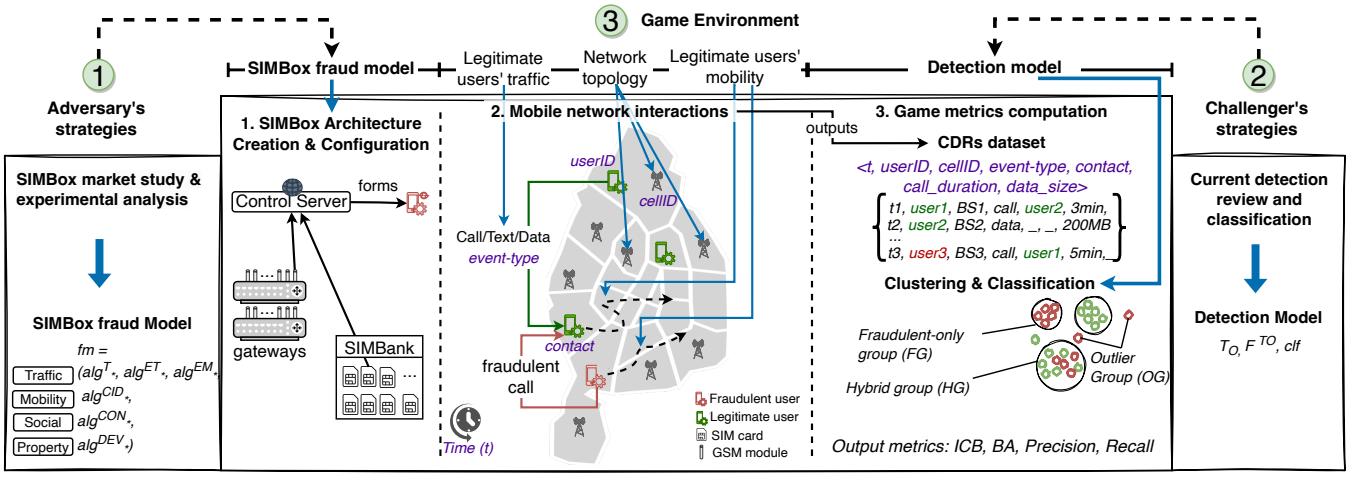


Figure 2: Framework for *SIMBox* fraud practical study.

This period directly influences the information gathered to distinguish between fraudulent and legitimate users. Our research reveals that detection features are typically computed on a *daily* basis [30, 46, 51], with occasional *weekly* computations [38]. In some cases, a shorter observation period of *four hours* is proposed, as seen in [30, 51], utilizing either fixed or one-hour sliding windows.

ML classifier, clf. Our review reports literature usage of the following ML classifiers for *SIMBox* fraud detection: Artificial Neural Network (ANN) and Support Vector Machine (SVM) [30, 45, 51], Random Forest (RF) [30, 38, 51], and Alternating Decision Tree [22, 38].

4.3 Game environment

We present the conceptualization of a controlled environment tailored for orchestrating the strategic game between the adversary and the investigator. As depicted in Figure 2, the game environment inputs include the adversary's *SIMBox* fraud model and the investigator's detection model, along with essential parameters defining the mobile network context, such as network topology, legitimate users' traffic, and mobility behaviors. The environment subsequently reproduces (1) the formation of fraudulent users by the *SIMBox* architecture, (2) the interactions between fraudulent and legitimate users in the destination country resulting in a CDRs dataset, and (3) the exploitation of these CDRs to compute the game's metric, which serves as the environment's output.

In the following, due to space constraints, we provide a brief overview of "*FraudZen*," the mobile network simulator handling stages 1 to 2 in Fig. 2 to produce a CDRs dataset. Stage 3 is then externally computed and implemented in Python.

We build *FraudZen* to meet the following properties: (i) *flexibility* to generate multiple *SIMBox* fraud models and instances, (ii) *modularity* to facilitate extension with new fraud functionalities, (iii) *efficiency* to generate, in a short time, CDRs covering a significant period and (iv) *ease of use* to facilitate the configuration of simulation scenarios. Hence, *FraudZen* has been written in C++, using the object-oriented paradigm as an event-driven simulator. It comprises 90 classes, 247 files, and approximately 19,000 code lines. In addition, it provides great flexibility through a configuration file of 122 parameters allowing for the simulation of countless *SIMBox* fraud model instances. Yet, to ease usage, default values allow an inexperienced user to modify a maximum of 20 essential parameters related to (i) the *SIMBox* architecture creation (cf. Fig. 10) and (ii) configuration (cf. Figs. 11 and 12), and to (iii) legitimate users' mobility and traffic (cf. Fig. 10).

FraudZen is constituted by four modules : the *SimulationManager*, the *NetworkManager*, the *TrafficManager*, and the *MobilityManager*. Each module performs a key role summarized in Table 4. For additional descriptions of these modules, please refer to §B.

5 EMPIRICAL STUDY AND INSIGHTS

This section conducts an empirical investigation into the *SIMBox* fraud problem, building upon the the practical framework established in §4. Our objective is to enrich the understanding of CDR-based *SIMBox* fraud detection by analyzing the problem outcome, i.e., the game metrics, in the context of dynamic strategies employed by both the adversary and the investigator. In essence, that is to provide insights into two pivotal questions: first "*How do adjustments to the adversary strategy impede detectability?*" and second, "*How can the investigator adjust her strategy to increase detectability*?"

To this end, we first design, in §5.1, an experimental setup that involves purposeful selections of adversary strategies (i.e., *SIMBox* fraud models) and investigator's strategies (i.e., detection models) while considering a real-world CDRs dataset for legitimate users' behavior. In §5.2, we present the results obtained from the experimental setup, deriving key insights that shed light on the existing strengths and weaknesses in fraud and detection design.

5.1 Experimental setup

We constitute 1280 scenarios encompassing heterogeneous adversary and investigator's strategies, as shown in Table 5.

5.1.1 Adversary configuration. It mainly involves the adversary's *SIMBox* fraud model, along with practical parameters such as the amount of diverted incoming calls and fraudulent SIM card count.

Table 3: Real-world *SIMBox* fraud and detection models organized by communication behavior

Com. Behavior	CDR field	Adversary's strategy			Investigator's strategy
		Motive (M_i)	**Algorithms ($M_i.j$)**	**Parameters $(P_i)^X$**	**Feature set (F^{I_o})**
Traffic	time (t_p)	M1- SIM activity limitation (time-related)	1- Working period per day	Day period	- nb. of calls at night (0-5AM) [37, 46] - night call duration [46] - nb. of unique contacts at night [46] - average inter-call time [30]
			2- Working period per week	Week period	
		M2- Network activity generation (data/text/calls)	1- Fixed inter-event-time (IET)	IET value	
			2- Random IET inside a fixed interval	IET interval	
			3- Multiple IET per day period	IET values per day periods	
			4- Triggered by a metric threshold (#call, call duration, allocation time)	- Metric choice - Metric value	
	Event Type (et_p)	M1- Network activity generation (Data)	1- Activate	/	- ratio of in. to out. calls [30, 37, 38] - nb. of out. calls for users with no other events (out_calls_no_evt) [37] - nb. of out. calls [30, 32, 38, 46, 51] - nb. of in. calls [30, 32, 38, 46, 51] - nb. of out.: intl. calls [38], texts [30] - nb. of in. intl. calls [38] - ratio of intl. calls [38]
		M2- Network activity generation (Text)	1- Activate	/	
		M3- Network activity generation (Calls)	1- Activate	/	
	Event metrics (em_p)	M1- SIM activity limitation (metric-related)	1- Metric (call duration/ #calls) threshold per period (day/week/month)	- Metric choice - Threshold value - Time period	- total call duration [46, 51] - avg. call duration [46] - max. call duration [32]
		M2- Incoming traffic routing	1- Balance (to the SIM with fewest historical calls)		
Mobility	Cell ID (cid_p)	M1- SIM to module allocation i.e., choice of the next location	1- Manually fixed, i.e., no change	SIM' locations i.e., cell ids	- nb. of unique visited cells [30, 32] - ratio of the nb. of cell Ids to the nb. of calls [30]
			2- Any except previous	/	
			3- Any except previous zone ID	/	
			4- Specified order	The sequence of location for each SIM	
		M2- Short Base Station (BS) movements, i.e., choice of the next location	1- Random	/	
			2- Default	/	
			3- Specified order	Sequence of BS	
			4- Manually fixed	The BS selection	
		M3- SIM to module allocation, i.e., choice of when to move	1- Periodic	Period	- nb. of calls without mobility [37] - nb. of calls in the most recurrent cell Id [32]
			2- Metric threshold (call duration, #calls)	Threshold value	
			3- Specified duration	The sequence of duration	
		M4- Short movements in the surroundings, i.e., choice of when to move	1- Fixed duration	Duration value	
			2- Threshold of metric (call duration, #calls)	Thresholds value	
			3- Specified duration	The sequence of duration	
		M5- Practical gateway deployment	1- Most visited locations	- Number of gateways - Locations' geo. positions	/
		M6- Displacement mode	1- Automatic (handled by the SIMBox)	/	/
			1- Physical (with a car/motobike)	/	
		M7- Mobility uniqueness	1- *SIMBox* architectural organization	- Number and size of gateways - Distribution of SIM cards in gateways (i.e., SIM and GSM groups and binding)	- avg. nb. of users making calls in the same cell Id [37]
Social	Contact (con_p)	M1- Incoming traffic routing	1- History	- max. #contacts per SIM card	- nb. of unique called [30, 32, 46] - nb. unique callers [32] - ratio of the nb. of unique contacts called to the nb. of calls [46]
			2- Random	/	
			3- In-turn, i.e., first available SIM card	/	
			4- Sequence	/	
			5- Balance	/	
		M2- Network activity generation (contact ctrl.)	1- Activate	- max. #contacts per SIM card	
Device Property	Device ID (dev_p)	M1- IMEI modification generation rule	1- None, i.e., no IMEI modification	/	- nb. of SIM per device [38]
			2- One generation (SIM)	/	
			3- Periodic generation (SIM/GSM mod.)	- period	
			4- Metric-based generation (GSM mod.)	- metric choice (#calls, #movement) - metric threshold	
		M2- IMEI value setting	1- Random	/	
			2- Tac-based IMEI	/	
			3- Prefix-based IMEI	Prefix	
			4- Registry-based IMEI	Registry, i.e., list of IMEI values	

SIMBox fraud models: A multiplicity of fraud models, utilizing existing *SIMBox* functionalities, can be formulated within the structure outlined in Table 3. Building these models involves tweaking several parameters (cf. cols 4 to 6), yielding fraud model instances of varying realisticness (i.e., closeness to human behavior) regarding traffic, mobility, and social aspects. We deliberately exclude consideration of the least-explored device property for space constraints.

In line with fraudsters' pursuit of high realisticness, we introduce the *SIMBox* fraud cube illustrated in Fig. 3. This conceptual framework designates each communication behavior as an axis.

Table 4: *FraudZen* modules.

Module	Functions
Simulation Manager	- Creates/Handles/Ends events - Manages simulation time
Network Manager	- Creates the network topology and devices: Operators, Cells, *SIMBox* architecture, UEs, etc.
Traffic Manager	- Handles legitimate and fraudulent traffic generation
Mobility Manager	- Handles legitimate and fraudulent mobility

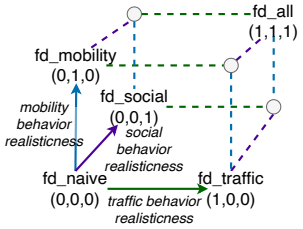


Figure 3: *SIMBox* fraud cube.

Moving along a specific axis from 0 to 1 indicates an increasing degree of realisticness of the corresponding communication behavior in the fraud model. The value of 1 denotes the highest achievable realisticness, leveraging *SIMBox* functionalities. Hence the *SIMBox* cube effectively encapsulates the spectrum of feasible fraud models.

Our study focuses on the five *SIMBox* fraud models, represented in Fig. 3. These models are strategically chosen to encompass the axis of fraud effectiveness, providing a holistic understanding of the potential scope of *SIMBox* fraud models. To ensure the acquisition of the most realistic parameters for constructing these models, we extract values from the distributions of the corresponding communication behavior features in the CDR dataset of legitimate users. For instance, in the case of the time algorithm "*working period per day*," we select from the dataset of legitimate users the period with the highest traffic across all days. This process results in the models outlined in Table 5 and discussed hereafter:

- ***fd_naive*** performs fraud with no effort to mimic human behavior and serves as a baseline. The *SIMBox* assumes its simplest function of routing the received international calls regardless of the day time and contact. Moreover, the fraudulent SIM cards remain static, being located in crowded city areas.
- ***fd_traffic*** focuses on mimicking human's traffic behavior, while keeping the other behavioral features naive. The fraudulent SIM cards generate all event types, while respecting legitimate Inter-Event-Time distributions during an active day period. They also comply with a daily threshold of call count.
- ***fd_mobility*** applies human-like mobility to each *SIMBox* gateway using the Working Day Mobility model [9], while keeping other features naive. Gateways commute by car between home and workplaces based on the time of day. Limited movements between base stations are introduced in the evening. To ensure uniqueness, only two SIM cards are allocated per gateway.
- ***fd_social*** solely modifies the social behavior of *SIMBox* users to imitate human's one. Hence, an history-based call routing is used to limit the number of fraudulent call destinations, i.e., contacts. Similarly, the number of contacts making inter-calls is restricted from legitimate users' contacts statistics.
- ***fd_all***, at last, includes all advanced configurations of each behavioral feature. *As such, this fraud model highlights how the most advanced in-market SIMBox functionalities can make fraud resilient to detection in all behavioral features.*

The percentage of incoming international traffic varies from operator to operator and impacts the number of calls diverted to the *SIMBox*. We consider two values (i.e., 3% and 12%) inspired by a statistical report [7] of a victim mobile operator. Correspondingly, the fixed frequency of international calls directed to the *SIMBox* is 7min and 2min, respectively.

The number of SIM cards in the *SIMBox* architecture impacts the efficiency of the implemented fraud. Hence, we consider four values of SIM number in our analyzes: 50, 100, 150, and 200.

5.1.2 Investigator configuration. It consists of defining the detection strategies in the set of selected features, the observation period and the ML classifier for classification implementation.

Detection features set. As explored in §4.2, existing literature considered diverse feature sets (cf. Table 3). In line with our approach to adversary strategies, we center our investigation on how the communication behavior (i.e., traffic, mobility, or social) influences the effectiveness of the investigator's detection strategy. Consequently, we categorize all detection features by behavior, resulting in four distinct sets (cf. Table 5): traffic features only ($F^{T_o}(tra)$), traffic with mobility features ($F^{T_o}(tra+mob)$), traffic with social features ($F^{T_o}(tra+soc)$), and all features, i.e., traffic with social and mobility ($F^{T_o}(all) = F^{T_o}(tra+mob+soc)$). Hence, each feature set targets the detection of the communication behaviors it considers.

Observation period. In alignment with the existing detection (cf. §4.2), we consider two periods: the day and the week.

ML classifier. Drawing from the existing detection (cf. §4.2), we consider the ANN, SVM and RF classifiers. Instead of the no longer utilized Alternative Decision Tree found in literature, we adopt its evolved counterpart—the Gradient Boosting Decision Tree (GBDT) model. GBDT combines multiple decision trees using the boosting method. We conduct hyper-parameter tuning to ascertain the best model in each scenario.

5.1.3 Legitimate behavior. At the generation of legitimate cellular traffic, we leverage the traffic behavior as described in real-world, non-public, and fully anonymized CDRs from a major telecom operator. It describes one month of per-user traffic (local and international outgoing calls and SMS, data) in about 3 million time-stamped events generated by 28K users from the provider operator. We filter out users interacting with other operators' phone numbers (i.e., 7000 users), and build our simulation scenarios with a unique operator and 21K legitimate users.

Because the leveraged CDRs lack daily spatio-temporal mobility information of users (i.e., users' cell ID positions), we assign realistic trajectories to every 21K users using the *Working Day Mobility Model* (WDM) [9] operating in the ONE simulator [33]. Our motivation to use WDM is twofold. First, contrary to related mobility models [47, 54], WDM originality comes from its representation of various mobility aspects present in people's daily life (e.g., home and workplaces, day periods). Second, WDM closely reproduces wireless interactions (i.e., inter-contact and contact time) distributions found in two real-world measurement experiments (i.e., iMote [48] and Dartmouth [34]), asserting modeling generality.

We further enhance WDM to consider real-world parameters and behaviors in human mobility. In this vein, we configure WDM with the Helsinki city infrastructure (i.e., the geographical area where users move) and its corresponding public transportation information [29]. Then, leveraging literature investigations on laws dictating human mobility [3, 20], we assign to users: (i) trajectories

Table 5: Experimental setup parameters

	Parameters	Values						#Scenarios	
Adversary configuration	*SIMBox fraud models*	t_p	et_p	em_p	cid_p	con_p	dev_p	5	Total= 1280
	fd_naive	*null*	*null*	*null*	$P_1.1$	$P_1.2$	$P_1.1$		
	fd_traffic	$P_1.1 \times P_2.3$	$P_1.1 \times P_2.1 \times P_3.1$	$P_1.1$	$P_1.1$	$P_1.2$	$P_1.1$		
	fd_mobility	*null*	*null*	*null*	$P_1.1 \times P_2.2 \times P_4.3 \times P_5.1 \times P_6.2 \times P_7.1$	$P_1.2$	$P_1.1$		
	fd_social	*null*	*null*	*null*	$P_1.1$	$P_1.1 \times P_2.1$	$P_1.1$		
	fd_all	$P_1.1 \times P_2.3$	$P_1.1 \times P_2.1 \times P_3.1$	$P_1.1$	$P_1.1 \times P_2.2 \times P_4.3 \times P_5.1 \times P_6.2 \times P_7.1$	$P_1.1 \times P_2.1$	$P_1.2$		
	% of incoming calls defrauded	3% =>one call/7min, 12% =>one call/2min						2	
	#fraudulent users	50, 100, 150, 200						4	
Investigator configuration	*Detection features*	$F^{T_O}(tra)$, $F^{T_O}(tra+mob)$, $F^{T_O}(tra+soc)$, $F^{T_O}(all) = F^{T_O}(tra+mob+soc)$						4	
	Observation period, i.e., T_O	a day, a week						2	
	ML classifier	*Hyper-parameter values used for tuning*						4	
	ANN	- #hidden layers: 1, 2, 3 - #nodes in hidden layers: 5, 9, 18 - learning rate: 0.1, 0.3, 0.6, 0.9 - optimizer: RMSProp, SGD, Adam							
	SVM	- kernel: RBF, polynomial - gamma: 0.125, scaled - degree: 2, 3 - C: 1, 10, 100, 1000							
	RF	#trees: 1, 2, 5, 10, 20, 50, 100, 200, 500							
	GBDT	- #trees: 1, 2, 5, 10, 20, 50, 100, 200, 500 - learning rate: loguniform(0.01, 1)							
Legitimate Behavior	*Users' traffic*	Trace-based from real-world traffic CDRs						1	
	Incoming intl. traffic								
	Users' mobility	Trace-based from WDM realistic simulation							

containing routine- and exploration-based locations, (ii) displacement profiles, as well as preferential neighborhoods (e.g., residential zones, business districts). Finally, we extract from OpenCellID [41] the network topology (i.e., cell tower distribution) of operators actuating in the emulated city. We end up with complete CDRs describing real-world users' traffic, mobility, and social behaviors.

Note that the lack of mobility information in the used raw CDRs and its enrichment with *as-realistic-as-possible* spatiotemporal trajectories do not impact the numerical assessment of fraud models presented next. *Indeed, the configuration offered by WDM brings the flexibility to add physical organization and daily behaviors of an actual city and its inhabitants, which constitutes the underlying structure required to play the designed fraud models, (most importantly) leveraging in-market SIMBox functionalities.*

5.2 Numerical assessment and Key insights

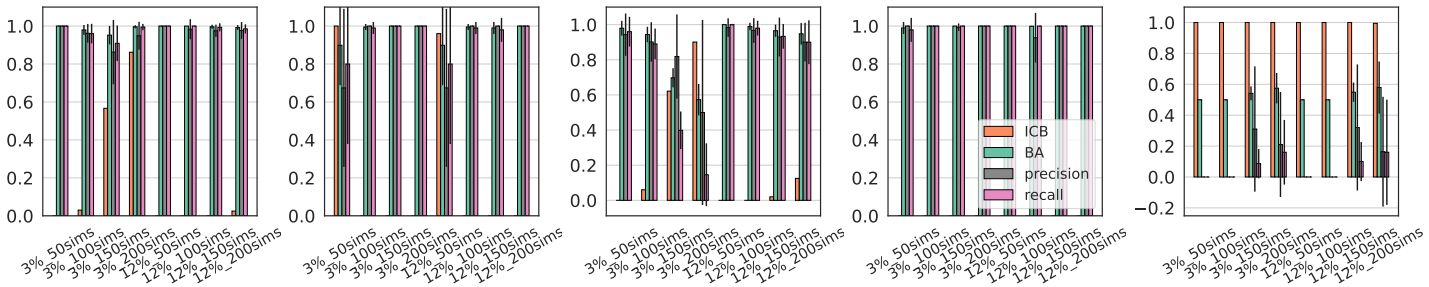
Here, we report the game metrics under the scenarios outlined in Table 5, encompassing diverse adversary and investigator strategy configurations. For a structured presentation of the results, we vary a focal parameter while keeping all other variables constant, aiming to clarify its impact on the game outcome and derive related insights. While the analysis of the *adversary configuration* allows assessing how detection performs across different fraud scenarios, revealing the fraud strengths and weaknesses, *investigator configuration* guides the selection of optimal options in designing an efficient fraud detection. Therefore, we first discuss the impact of the *investigator configuration* – the *SIMBox* fraud model, the amount of defrauded traffic, and the number of fraudulent SIMs – in §5.2.1. Then, we explore the effects of the *investigator configuration*: the chosen set of features in §5.2.2, the observation period in §5.2.3, and finally, the ML classifier in §5.2.4.

5.2.1 Adversary configuration. Fig. 4 shows the game metrics per fraud model for a fixed investigator configuration: $F^{T_O}(all)$ feature set, T_O=day, ANN classifier.

We calculate the in-crowd-blending capability through a two-step process: (i) we apply a clustering based on $F^{T_O}(all)$ features on only legitimate users and filter out legitimate users that are outliers, (ii) we add fraudulent users to the remaining legitimate users and infer the in-crowd-blending capability per fraud model. Besides, for results explainability, Fig. 5 illustrates the average impact of each *feature* $f \in F^{T_O}(all)$ on the generation of groups (i.e., Hybrid, Fraudulent-only, and Outliers Groups, cf. §3.4.1) for in-crowd-blending computation across all implemented scenarios, organized by fraud model. The drawn insights are discussed hereafter.

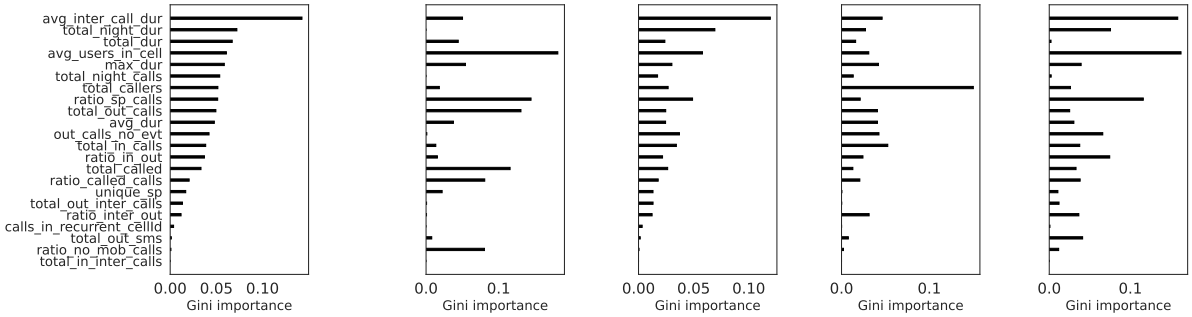
fd_naive (Fig. 4a) makes no human mimicking effort. Nevertheless, we observe that with 3% of incoming international traffic, the *ICB* value rises, and classification metrics (BA, precision, recall) decline as the number of fraudulent SIMs grow. This is because the same amount of diverted traffic is distributed among more fraudulent users, resulting in reduced traffic per user. At 12%, despite the surge in users, the *ICB* remains low, and classification metrics stay high (with a minimum precision of 0.96) due to a substantial increase in generated traffic (1 call/2 min). Confirming this interpretation, Fig. 5a show the most impacting features include the average inter-call duration (*avg_inter_call_dur*) and total call duration at night (*total_night_dur*). These feature values strictly ascend with an increasing amount of traffic.

fd_traffic (Fig. 4b) exhibits counter-intuitive performance compared to the *fd_naive* fraud model, where no effort is made to defraud. When the count of fraudulent users is low, at 3% and 12% of incoming international calls, the fraud model simulates human-like behavior, reflected in the highest *ICB* and lowest classification metrics (BA, precision, and recall). However, with an increasing number of users, the *ICB* becomes null, making fraudulent users easily detectable. Fig. 5b elucidates that this phenomenon primarily stems from naive mobility behavior. Notably, average users in cells *avg_users_in_cell*) and the ratio of unique stay points to the number of outgoing calls (*ratio_sp_calls*) remain exceptionally high for all fraudulent users compared to legitimate ones. This is due to their allocation to a single gateway location and a absence of mobility.



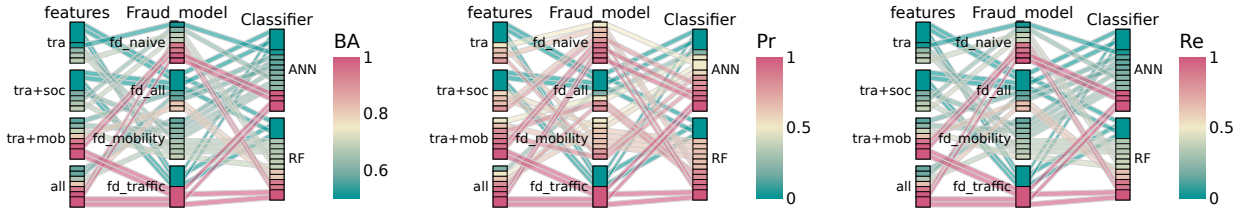
(a) *fd_naive* (b) *fd_traffic* (c) *fd_mobility* (d) *fd_social* (e) *fd_all*

Figure 4: Game metrics per fraud model, $F^{T_o}(all)$ feature set, T_o=day, ANN classifier.



(a) *fd_naive* (b) *fd_traffic* (c) *fd_mobility* (d) *fd_social* (e) *fd_all*

Figure 5: Average relative importance of detection features per fraud model, $F^{T_o}(all)$ feature set, T_o=day.



(a) balanced accuracy (b) precision (c) recall

Figure 6: Classification metrics w.r.t. the feature set F^{T_o}, T_o=day, ANN and RF classifiers, 3% scenario and 200 fraudulent users.

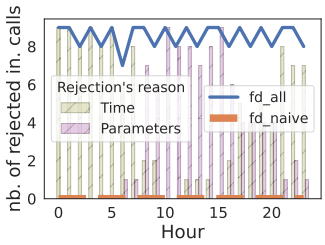


Figure 7: Rejected calls per hour

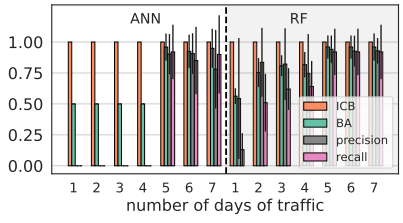


Figure 8: Game metrics *w.r.t. the observation period* T_o, $F^{T_o}(all)$ feature set, ANN and RF classifiers, 3% scenario, 200 fraudulent users.

Table 6: Classification metrics w.r.t. the ML classifier, $F^{T_o}(all)$ feature set, T_o=day, 3% of intl. calls.

Model	#fraudulent users	*fd_naive* balanced acc	*fd_naive* precision	*fd_naive* recall	*fd_all* balanced acc	*fd_all* precision	*fd_all* recall
ANN	200	0.99	0.95	0.99	0.57	0.21	0.16
	50	1	1	1	0.5	0	0
SVM	200	0.98	0.82	0.97	0.56	0.37	0.14
	50	0.97	1	0.94	0.49	0	0
RF	200	0.94	0.93	0.88	0.80	0.82	0.64
	50	0.97	1	0.94	0.49	0	0
GBDT	200	0.99	0.94	0.99	0.70	0.88	0.41
	50	0.98	0.92	0.96	0.5	0	0

fd_mobility (Fig. 4c) demonstrates a performance trend similar to that of *fd_naive* but with an overall enhancement, characterized by higher *ICB* and lower values in classification metrics (BA, precision, and recall). This implies that prioritizing improvements in mobility rather than traffic has a more favorable impact on the effectiveness of the fraud model. Indeed, the traffic behavior is naturally more realistic with an increasing number of fraudulent SIMs. Backing this finding, Fig. 5c illustrates that, when compared with *fd_naive*, mobility-related features (i.e., *ratio_sp_calls* and *avg_users_in_cell*) exert a more. substantial influence in elevating the *ICB*.

fd_social (Fig. 4b) demonstrates less favorable results in reproducing human behavior, featuring a null *ICB* and nearly perfect classification metrics (BA, precision, recall) regardless of the scenario. As depicted in Fig. 5d, the *fd_social* fraud model primarily improves the count of callers of fraudulent users (i.e., *total_callers*), while leaving other features in a naive state, easily detectable. This limited efficiency reveals enhancing social behavior does not contribute significantly to the overall effectiveness of the fraud.

fd_all (Fig. 4e) yields an *ICB* of 1 and poor values for classification metrics, with the max. precision reaching 0.3 and the max. recall around 0.15, regardless of the scenario. This highlights that *currently in-market SIMBox functionalities can generate fraud models closely mimicking human behavior, effectively evading current detection strategies.* Nevertheless, the settings associated with the *fd_all* fraud model lead to the *SIMBox* architecture refusing to route most diverted international calls (cf. Fig. 7 in appendix). This rejection stems from the fraud model's constraints on event timing (i.e., no operations during nighttime) and parameters (i.e., maximum number of calls per day). As a result, *while the fd_all fraud model proves effective, it enables fraudsters to achieve only long-term financial gains due to the rejection of calls.*

Insight 1: *Current fraud capabilities undoubtedly counter current detection strategies, confirming that literature's detection performances are overestimated. Nevertheless, our results underscore the high cost associated with implementing sophisticated strategies for fraudsters. A more advantageous approach for them is to limit the traffic generated by individual fraudulent cards, thereby hardening the extraction of conclusive detection insights. Achieving this can involve manipulating a massive amount of SIM cards.*

5.2.2 Investigator configuration: features set. Fig. 6 shows the impact of the feature set on the detection performance of both ANN and RF classifiers, for 3% of incoming intl. traffic and 200 users. All fraud models are taken into account, excluding *fd_social* due to its minimal contribution to the overall outcome (cf. § 5.2.1).

In summary, the results indicate that the inclusion of social features, as seen in $F^{T_o}(tra)$ and $F^{T_o}(tra + soc)$ feature sets, does not alter the values of balanced accuracy, precision, and recall significantly. On the other hand, the inclusion of mobility detection features has a discernible impact on detection performance, varying across fraud models. Notably, *fd_traffic* exhibits the most substantial performance increase, followed by *fd_naive*, *fd_all*, and, to a minimal extent, *fd_mobility*. It is noteworthy that the influence of mobility detection features is more pronounced on the precision metric than on balanced accuracy and recall. Lastly, evaluations based on $F^{T_o}(all)$ features slightly improve the detection metrics compared to those derived from $F^{T_o}(tra + mob)$ features.

Insight 2: *Compared to traffic and social behaviors, the mobility behavior emerges as the best facet to distinguish between fraudulent and legitimate users. Indeed, unlike traffic and social aspects, mobility requires costly hardware resources to be realistically simulated by fraudsters. To leverage this fraud vulnerability, there should be a greater emphasis on detection features within this category, currently under-explored in the existing literature (cf. Table 3).*

5.2.3 Investigator configuration: observation period. As depicted in Fig. 8, extending the observation period from 1 to 7 days notably enhances the detection performance for both the ANN and RF ML classifiers. These findings are based on the *fd_all* fraud model, considering 200 fraudulent users and 3% of incoming intl. traffic. In particular with the ANN classifier, a substantial performance improvement is evident after five days.

Insight 3: *The observation period enlargement proves highly effective in uncovering even the most sophisticated fraud strategies. Indeed, while SIMBox functionalities may approximate human behavior over a day or a week, sustaining this deception becomes increasingly challenging over more extended duration.*

5.2.4 Investigator configuration: ML classifier. Considering the feature set $F^{T_o}(all)$, Table 6 presents the classification metrics for the *fd_naive* and *fd_all* fraud models, with 50 and 200 fraudulent users, and accounting for 3% of incoming intl. traffic.

With *fd_naive*, all ML classifiers exhibit comparable performance, achieving at least 97% BA and 82% precision, irrespective of whether the number of fraudulent users. Conversely, with *fd_all*, all ML classifiers exhibit poor performance with 50 fraudulent users, displaying 0 recall and precision. This indicates that all fraudulent users are misclassified as legitimate. However, as the number of fraudulent users increases, RF and GBDT outperform ANN and SVM. While GBDT shows higher precision, RF exhibits superior balanced accuracy and recall.

Insight 4: *As in [2, 30, 51], a combination of decision rules (as in RF and GBDT) related to singular behavioral features is more efficient than classifying users considering their global behavior (as in SVM). Our analysis confirms the persistence of this pattern across various fraud models, indicating its robustness to fraud evolution.*

6 PRACTICAL IMPLICATIONS

This section discusses how this paper's contributions and findings should be considered in practice to enhance fraud mitigation.

The formalization and the empirical study framework presented in this work serve as practical tools capable of capturing current fraud capabilities while enabling the seamless incorporation of the fraud evolution. Therefore, operators (and other *investigators*) can use these tools to obtain fraudulent ground truth up-to-date regarding the fraud advancements. This enables them to access the most effective fraud strategies and to challenge the detection models they develop, which a first-of-a-kind progress in the field.

Furthermore, although not exhaustive, the empirical fraud study of §5 leads to the following practical recommendations:

- As noted in Insight 1, detecting fraud becomes arduous when the adversary possesses a large number of SIM cards. Therefore, strengthening SIM cards access policies plays a crucial role in *SIMBox* fraud mitigation, and fraud detection investigations should pay particular attention to new SIM cards.
- Building on Insight 2, we advocate for a more comprehensive exploration of mobility-based features, specifically focusing on the *positioning*, *displacement modes*, and *mobility uniqueness* of fraudulent users. Also, such features should be more elaborated, drawing from the literature richness on human mobility behavior, which demonstrates adherence to difficult-to-mimic laws such as diversity [50], confinement given by the radius of gyration [21], and unpredictability due to the exploration phenomenon [3].
- With Insight 3, we recommend that detection techniques systematically expand the observation period, including features spanning a day, a week, or even longer, to improve detection robustness. Hence, the detection time will depend on the fraudulent user's profile (naive or sophisticated). Moreover, accumulating evidences daily helps in identifying advanced frauds while also expediting the detection of new SIM cards used by fraudsters.
- Finally, based on Insight 4, we recommend utilizing tree ensemble ML classifiers or a linear combination thereof for efficient fraud detection. Combining a set of models, each tailored to specific fraud profiles and trained with corresponding fraudulent ground truth, appears to be a more effective approach.

7 CONCLUSION

Despite its significant impact on operators' revenue and national security, *SIMBox* bypass fraud remains an open issue in cellular networks. This paper highlights the *SIMBox fraud evolution* as a critical factor hindering substantial progress in fraud mitigation. Our game-based approach clarifies adversary and investigator strategies in CDR-based *SIMBox* fraud investigations, presenting a pioneering contribution. The practical framework we introduce enables scalable empirical studies of the fraud, providing a comprehensive evaluation of detection methods aware of real-world fraud capabilities. We believe these contributions, the first of their kind in the literature to the best of our knowledge, establish a robust foundation for in-depth investigations into *SIMBox* fraud's evolution and offer valuable insights for future fraud detection research.

PUBLIC CODE AND DATA

To encourage further research on *SIMBox* fraud detection, we make publicly available the fraud simulation environment's code, i.e., *FraudZen*, corresponding to stages 1 and 2 of Fig. 2. We further release the generated fraudulent CDRs datasets associated with the fraud strategies investigated in this study:
https://gitlab.inria.fr/simbox-fraud-mitigation/fraudzen

ACKNOWLEDGMENTS

We express our gratitude to Inria and the MLNS2 associate research team of Inria and IRP-CNRS for their valuable support. Special thanks go to Serge Fonkou and Jean-Marie Zambo for their assistance and guidance throughout the project.

REFERENCES

[1] Mhd Redwan AlBougha. 2016. *Comparing Data Mining Classification Algorithms in Detection of Simbox Fraud.* Master's thesis. St. Cloud State University.
[2] Mohammad Almseidin, Maen Alzubi, Szilveszter Kovacs, and Mouhammd Alkasassbeh. 2017. Evaluation of machine learning algorithms for intrusion detection system. In *2017 IEEE 15th International Symposium on Intelligent Systems and Informatics (SISY)*. IEEE, Subotica, Serbia, 000277–000282. https://doi.org/10.1109/SISY.2017.8080566
[3] Licia Amichi, Aline Carneiro Viana, Mark Crovella, and Antonio A.F. Loureiro. 2020. Understanding Individuals' Proclivity for Novelty Seeking. In *Proceedings of the 28th International Conference on Advances in Geographic Information Systems* (Seattle, WA, USA) *(SIGSPATIAL '20)*. ACM, New York, NY, USA, 314–324. https://doi.org/10.1145/3397536.3422248
[4] Abderrahim Benslimane and Huong Nguyen-Minh. 2017. Jamming Attack Model and Detection Method for Beacons Under Multichannel Operation in Vehicular Networks. *IEEE Transactions on Vehicular Technology* 66, 7 (2017), 6475–6488. https://doi.org/10.1109/TVT.2016.2645478
[5] Rémi Canillas, Rania Talbi, Sara Bouchenak, Omar Hasan, Lionel Brunie, and Laurent Sarrat. 2018. Exploratory Study of Privacy Preserving Fraud Detection. In *Proceedings of the 19th International Middleware Conference Industry* (Rennes, France) *(Middleware '18)*. ACM, New York, NY, USA, 25–31. https://doi.org/10.1145/3284028.3284032
[6] CFCA. 2019. CFCA 2019 Fraud Loss Survey. Report. https://cfca.org/document/cfca-2019-fraud-loss-survey-pdf/
[7] Agence de Régulation des Télécommunications Cameroun. 2018. *INFORMATIONS STATISTIQUES.* Technical Report. ART Cameroon, Cameroon.
[8] Xuewen Dong, Feng Wu, Anter Faree, Deke Guo, Yulong Shen, and Jianfeng Ma. 2019. Selfholding: A combined attack model using selfish mining with block withholding attack. *Computers & Security* 87 (2019), 101584. https://doi.org/10.1016/j.cose.2019.101584
[9] Frans Ekman, Ari Keränen, Jouni Karvo, and Jörg Ott. 2008. Working Day Movement Model. In *Proceedings of the 1st ACM SIGMOBILE Workshop on Mobility Models* (Hong Kong, Hong Kong, China) *(MobilityModels '08)*. ACM, New York, NY, USA, 33–40. https://doi.org/10.1145/1374688.1374695
[10] Osama Mohamed Elrajubi, Ali Mustafa Elshawesh, and Mustafa Ali Abuzaraida. 2017. Detection of bypass fraud based on speaker recognition. In *2017 8th International Conference on Information Technology (ICIT)*. IEEE, Amman, Jordan, 50–54. https://doi.org/10.1109/ICITECH.2017.8079914
[11] Antrax FlamesGroup. 2021. Antrax. https://gitlab.com/flamesgroup/antrax.
[12] Antrax FlamesGroup. n.d.. Antrax FlamesGroup playlists. YouTube video. https://www.youtube.com/@FlamesGroup/playlists
[13] GoAntiFraud. 2016. GoAntiFraud is a cloud service for efficient GSM Termination. Article. https://goantifraud.com/ Accessed: 2020-04-24.
[14] GoAntiFraud. 2019. GSM Termination for "Dummies": What are the "Favorite" Numbers? https://goantifraud.com/en/blog/1125-gsm-termination-for-dummies-what-are-the-favorite-numbers.html.
[15] GoAntiFraud. 2019. GSM Termination for Dummies: What is IMEI? https://goantifraud.com/en/blog/1146-gsm-termination-for-dummies-what-is-imei.html.
[16] GoAntiFraud. accessed 2023. Call Recording. https://goantifraud.com/en/ejointech-skyline-gsm-termination-solution#call-recording.
[17] GoAntiFraud. accessed 2023. Top 5 Popular GSM Gateway Manufacturers. https://goantifraud.com/en/blog/818-top-5-popular-gsm-gateway-manufacturers.html.
[18] GoAntiFraud. n.d.. GoAntiFraud Blog. https://goantifraud.com/en/blog
[19] GoAntiFraud. n.d.. GoAntiFraud playlists. YouTube video. https://www.youtube.com/@GoAntiFraud/playlists
[20] Marta C. González, César A. Hidalgo, and Albert-László Barabási. 2008. Understanding individual human mobility patterns. *Nature* 453, 7196 (jun 2008), 779–782. https://doi.org/10.1038/nature06958
[21] Marta González, César Hidalgo, and Albert-László Barabási. 2008. Understanding individual human mobility patterns. *Nature* 453 (2008), 779–782. https://doi.org/10.1038/nature06958
[22] Geoffrey Holmes, Bernhard Pfahringer, Richard Kirkby, Eibe Frank, and Mark Hall. 2002. Multiclass Alternating Decision Trees. In *Machine Learning: ECML 2002*, Tapio Elomaa, Heikki Mannila, and Hannu Toivonen (Eds.). Springer Berlin Heidelberg, Berlin, Heidelberg, 161–172.
[23] Hybertone. 2022. GSM VoIP Gateway Model GoIP324. http://www.hybertone.com/en/pro_detail.asp?proid=63.
[24] Hybertone. 2022. Remote SIM Bank. http://www.hybertone.com/en/pro_detail.asp?proid=57.
[25] Hybertone. 2022. Remote SIM Bank. http://www.hybertone.com/en/pro_detail.asp?proid=43.
[26] Nassir Abuhamoud Ibrahem Soliman Alsadi. 2019. Study to use NEO4J to analysis and detection SIM-BOX fraud. *Journal of Pure & Applied Sciences* 17, 4 (Jan. 2019), 31–35.

[27] Ibrahim Ighneiwa and Hussamedin Mohamed. 2017. Bypass Fraud Detection: Artificial Intelligence Approach. arXiv:1711.04627 [cs.CY]

[28] Technology Research Institute. 2015. *Network Protocol Analysis: A New Tool for Blocking International Bypass Fraud Before Revenue is Lost.* Technical Report. LATRO Services.

[29] Ari Jaakola, Teemu Vass, Solja Saarto, and Lotta Haglund. 2019. *Helsinki facts and figures 2019.* Technical Report. City Executive Office, Urban Research and Statistics, Helsinki, Finland.

[30] Hagos Kahsu. 2018. *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.

[31] M. Kashir and S. Bashir. 2019. Machine Learning Techniques for SIM Box Fraud Detection. In *2019 International Conference on Communication Technologies (ComTech).* IEEE, Rawalpindi, Pakistan, 4–8. https://doi.org/10.1109/COMTECH.2019.8737828

[32] KGDC Kehelwala, HMND Bandara, RA Yasaratne, P De Almeida, IKKS Ilesinghe, and PDKE Wickramasinghe. 2015. *REAL-TIME GREY CALL DETECTION SYSTEM USING COMPLEX EVENT PROCESSING.* Technical Report. IET, Sri Lanka. http://theiet.lk/wp-content/uploads/2017/10/22-p7.pdf

[33] Ari Keränen, Jörg Ott, and Teemu Kärkkäinen. 2009. The ONE Simulator for DTN Protocol Evaluation. In *Proceedings of the 2nd International Conference on Simulation Tools and Techniques* (Rome, Italy) *(Simutools '09).* ICST (Institute for Computer Sciences, Social-Informatics and Telecommunications Engineering), Brussels, BEL, Article 55, 10 pages. https://doi.org/10.4108/ICST.SIMUTOOLS2009.5674

[34] David Kotz, Tristan Henderson, Ilya Abyzov, and Jihwang Yeo. 2009. CRAWDAD dataset dartmouth/campus (v. 2009-09-09). Downloaded from https://crawdad.org/dartmouth/campus/20090909. https://doi.org/10.15783/C7F59T

[35] Anne Josiane Kouam, Aline Carneiro Viana, and Alain Tchana. 2021. SIMBox Bypass Frauds in Cellular Networks: Strategies, Evolution, Detection, and Future Directions. *IEEE Communications Surveys & Tutorials* 23, 4 (2021), 2295–2323. https://doi.org/10.1109/COMST.2021.3100916

[36] Siyuan Liu, Lei Li, Christos Faloutsos, and Lionel M. Ni. 2011. Mobile Phone Graph Evolution: Findings, Model and Interpretation. In *2011 IEEE 11th International Conference on Data Mining Workshops.* IEEE, Vancouver, BC, Canada, 323–330. https://doi.org/10.1109/ICDMW.2011.123

[37] Hussein M. Marah, Osama Mohamed Elrajubi, and Abdulla A. Abouda. 2015. Fraud detection in international calls using fuzzy logic. In *International Conference on Computer Vision and Image Analysis Applications.* IEEE, Sousse, Tunisia, 1–6. https://doi.org/10.1109/ICCVIA.2015.7351891

[38] I. Murynets, M. Zabarankin, R. P. Jover, and A. Panagia. 2014. Analysis and detection of SIMbox fraud in mobility networks. In *IEEE INFOCOM 2014 - IEEE Conference on Computer Communications.* IEEE, Toronto, ON, Canada, 1519–1526. https://doi.org/10.1109/INFOCOM.2014.6848087

[39] Diala Naboulsi, Marco Fiore, Stephane Ribot, and Razvan Stanica. 2016. Large-Scale Mobile Traffic Analysis: A Survey. *IEEE Communications Surveys & Tutorials* 18, 1 (2016), 124–161. https://doi.org/10.1109/COMST.2015.2491361

[40] Beomseok Oh, Junho Ahn, Sangwook Bae, Mincheol Son, Yonghwa Lee, Minsuk Kang, and Yongdae Kim. 2023. Preventing SIM Box Fraud Using Device Model Fingerprinting. In *Network and Distributed Systems Security (NDSS) Symposium.* The Internet Society 2023, San Diego, CA, USA.

[41] OpenCellID. 2022. The world's largest Open Database of Cell Towers. https://www.opencellid.org/

[42] Bradley Reaves, Ethan Shernan, Adam Bates, Henry Carter, and Patrick Traynor. 2015. Boxed Out: Blocking Cellular Interconnect Bypass Fraud at the Network Edge. In *Proceedings of the 24th USENIX Conference on Security Symposium* (Washington, D.C.) *(SEC'15).* USENIX Association, USA, 833–848. https://doi.org/10.5555/2831143.2831196

[43] Syed Rizvi, Ryan Pipetti, Nicholas McIntyre, Jonathan Todd, and Iyonna Williams. 2020. Threat model for securing internet of things (IoT) network at device-level. *Internet of Things* 11 (2020), 100240. https://doi.org/10.1016/j.iot.2020.100240

[44] Merve Sahin. 2017. *Understanding Telephony Fraud as an Essential Step to Better Fight It.* Ph. D. Dissertation. TELECOM ParisTech.

[45] Roselina Sallehuddin, Subariah Ibrahim, Azlan Zain, and Abdikarim Elmi. 2013. Detecting SIM Box Fraud Using Neural Network. In *IT Convergence and Security 2012*, Kuinam J. Kim and Kyung-Yong Chung (Eds.). Springer Netherlands, Dordrecht, 575–582. https://doi.org/10.1007/978-94-007-5860-5_69

[46] Roselina Sallehuddin, Subariah Ibrahim, Azlan Zain, and Abdikarim Elmi. 2015. Detecting SIM Box Fraud by Using Support Vector Machine and Artificial Neural Network. *Jurnal Teknologi* 74, 1 (Apr. 2015). https://doi.org/10.11113/jt.v74.2649

[47] Matthias Schwamborn and Nils Aschenbruck. 2013. Introducing Geographic Restrictions to the SLAW Human Mobility Model. In *2013 IEEE 21st International Symposium on Modelling, Analysis and Simulation of Computer and Telecommunication Systems.* IEEE, San Francisco, CA, USA, 264–272. https://doi.org/10.1109/MASCOTS.2013.34

[48] James Scott, Richard Gass, Jon Crowcroft, Pan Hui, Christophe Diot, and Augustin Chaintreau. 2009. CRAWDAD dataset cambridge/haggle (v. 2009-05-29). Downloaded from https://crawdad.org/cambridge/haggle/20090529/imote. https://doi.org/10.15783/C70011 traceset: imote.

[49] Luke Taylor. 2021. *Communications Fraud Control Association 2021 Fraud Loss Survey.* Technical Report. CFCA. https://cfca.org/document/2021-fraud-loss-survey/

[50] Douglas Do Couto Teixeira, Aline Carneiro Viana, Jussara M. Almeida, and Mrio S. Alvim. 2021. The Impact of Stationarity, Regularity, and Context on the Predictability of Individual Human Mobility. *ACM Trans. Spatial Algorithms Syst.* 7, 4, Article 19 (jun 2021), 24 pages. https://doi.org/10.1145/3459625

[51] Fitsum Tesfaye. 2020. *Near-Real Time SIM-box Fraud Detection Using Machine Learning in the case of ethio telecom.* Ph. D. Dissertation. School of Electrical and Computer Engineering Addis Ababa Institute of Technology.

[52] Haicheng Tu, Yongxiang Xia, Chi K. Tse, and Xi Chen. 2020. A Hybrid Cyber Attack Model for Cyber-Physical Power Systems. *IEEE Access* 8 (2020), 114876–114883. https://doi.org/10.1109/ACCESS.2020.3003323

[53] Bruno Veloso, Shazia Tabassum, Carlos Martins, Raphael Espanha, Raul Azevedo, and João Gama. 2020. Interconnect bypass fraud detection: a case study. *Annals of Telecommunications* 75 (Oct. 2020), 583–596. https://doi.org/10.1007/s12243-020-00808-w

[54] Vladimir Vukadinovic, Ólafur Ragnar Helgason, and Gunnar Karlsson. 2013. An analytical model for pedestrian content distribution in a grid of streets. *Math. Comput. Model.* 57 (2013), 2933–2944.

Table 7: Reviewed market *SIMBox* appliances.

Manufacturer	*# gateways*	*# SIMBank*	*# Control server*
Hybertone	10	2	1
Dinstar	6	1	1
Antrax	2	2	1
Ejoin	8	2	1
Portech	6	2	3
2N VoiceBlue	2	0	0
SunComm	8	0	0
Yeastar	5	0	0
Synway	1	1	1
OpenVox	6	3	1
Skyline	9	5	1
NICEUC	3	0	0
Acquired appliances			
Hybertone	3, i.e., [23]	1, i.e., [25]	1

```
{
    "simulationName": "name",
    "simulatedDuration": 30,
    "start_hour": 0,
    "date": "2021-03-17",
    "mcc": "244",
    "nbOp": 1,
    "mnCodes": [
        "005"
    ],
```

Figure 9: General *FraudZen* simulation parameters

A ADVERSARY'S STRATEGY DESCRIPTION

The following discusses insights derived from Table 3 for each identified communication behavior (i.e., traffic, mobility, social, device property), providing a clearer understanding of the implications associated with each adversary's strategy.

(1) Traffic behavior. A user traffic behavior relates to the type of generated network events (et_p, i.e., incoming or outgoing, national or international calls and text, data), the temporal occurrence of these events (t_p), and the associated metrics (em_p, i.e., call duration or data session size). The *SIMBox* default behavior is to generate only outgoing calls (i.e., re-originated diverted intl. calls) regardless of the time of the day, according to the traffic it receives from

```
11    "interCallStrategy": "trace-based",
12    "regularMobilityStrategy": "trace-based",
13    "regularTrafficStrategy": "trace-based",
14    "simbox_fraud": true,
15    "frauded_call_frequency": 3,
16    "simbox_architecture": {
17
18        "gateways1": [8, 8],
19        "gateways2": [],
20        "gateways2_sims": [],
21        "simbanks": [54],
22        "fillingPercentages":[1],
23        "controlServer": true,
24
25        "fraudulentMobilityStrategy": "trace-based:static",
26        "groupSize": 0,
27
28        "routingPolicy": "inTurn",
29
30        "probabilisticCalls": true,
```

Figure 10: *FraudZen* simulation parameters related to (Lines 11-13) Legitimate users' mobility and traffic and (Lines 14-30) *SIMBox* architecture creation

```
32        "nbSimGroup": 1,
33        "simGroup0":
34        {
35            "location": [[0, 54,1]],
36            "unblockFrequency": 20,
37            "parameterLimitation":{
38                "state": false,
39                "parameters" : ["callCount", "totalCallDur"],
40                "values" : [-1, -1, 61]
41            },
42            "timeParameterLimitation":
43            {…
68            },
69            "timeLimitation":{…
85            },
86            "rotationTrigger":{…
90            },
91            "rotationPolicy":
92            {…
96            },
97            "migrationPolicy":
98            {…
102           },
103           "imeiGenerationRule":
104           {…
114           },
115           "networkActivityGeneration":
116           {…
145           }
146       },
```

Figure 11: *FraudZen* simulation parameters related to *SIMBox* SIM card groups creation and configuration

multiple countries. To counter this default tendency, fraudsters rely on *SIMBox* functionalities with three distinct motives:

- *SIM activity limitation* seeks to regulate the excessive use of fraudulent SIM cards, either by restricting their operation to specific time intervals within a day or week (time-related) or by imposing thresholds on metrics tied to the SIM cards' traffic behavior (metric-related, e.g., number of calls, total/average call duration, etc.).
- *Network activity generation* allows fraudsters to generate diverse network traffic (i.e., inter-call, inter-texts, and mobile data) beyond the default *SIMBox* outgoing calls and reproduce the human behavior. For instance, fraudulent SIM cards can automatically engage in web browsing or exchange calls and texts with each other. This involves deciding whether fraudulent users should generate a specific event type (event-type-related) and determining the frequency of such generation (time-related).
- *Incoming traffic routing* control empowers fraudsters to dictate which specific fraudulent user within the *SIMBox* architecture will re-originate an incoming diverted call. This decision-making process can be guided by reducing the number of calls fraudulent users make (metric-related).

```
148       "nbGSMGroup": 2,
149       "GSMGroup0":
150       {
151           "location": [[0, 8, 1]],
152           "simCheckFrequency": 20,
153           "imeiGenerationRule":
154           {…
163           },
164           "baseStationSelection":
165           {…
174           },
175           "copy": false,
176           "copyFromGroupId": 0
177       },
178
179       "GSMGroup1":
180       {…
206       },
207       "linkage": [[0, 0], [1, 0]]
208   },
```

Figure 12: *FraudZen* simulation parameters related to *SIMBox*' GSM module groups creation and configuration

(2) Mobility behavior. User mobility is identified in operators' traces by the position of the base station relaying a generated event (cid_p) and the trajectory formed by a sequence of these positions. Such elements must be meaningful in their alignment with the typical daily patterns governing human mobility, encompassing activities like commuting, shopping, and socializing. As *SIMBox* appliances are generally bulky and demand a fixed wired internet connection, fraudulent users inherently remain stationary and by default confined to an individual cell or a few nearby cells for prolonged periods. In response to this easily detectable behavior, fraudsters employ various strategies:

- *SIM to module allocation* regulates the binding between SIM cards and GSM modules. By binding a SIM card to GSM modules in various locations, fraudsters simulate the mobility of their fraudulent devices, bringing them closer to human behavior. This entails the formulation of algorithms to decide when and where to simulate movement based on past trajectories.
- *Short base station movements* allow connecting GSM modules to surrounding base stations with accessible signals. The choice of algorithms determines when to initiate such movements and the connecting base station.
- The *strategic gateways placement* influences the number and locations of *SIMBox* gateways in the architecture. Fraudsters often opt for crowded zones (such as city centers, densely populated residential areas, or marketplaces) to camouflage *SIMBox* traffic amid the substantial call volumes in these areas. Moreover, these locations are chosen to yield realistic trajectories as fraudulent SIM move association from one gateway to another.

- The *displacement mode* of fraudulent SIMs is automated through *SIM to module allocation* and *short base station movements*. However, though costly, fraudsters might manually transport gateways by car or motorcycle, equipped with the necessary tools and internet connection, to achieve a more realistic trajectory.
- Lastly, fraudsters strive to ensure the "*mobility uniqueness*" of their fraudulent users' trajectories. They aim to reproduce the realistic pattern of predominantly individual movements instead of grouped ones, which are the default scenario with a *SIMBox* (being a box of SIMs). They achieve this by adjusting the number of SIM cards assigned to each *SIMBox* gateway; the higher the number, the more the *SIMBox* generates group mobility.

(3) Social behavior. Calls and text events are direct indicators of inter-user interactions and social dynamics. In particular, a user's contacts (con_p), the importance, and the direction (in/out) of his interactions with these contacts provide insights into his social behavior. These interactions, commonly represented as mobile call graphs in the literature [36], often encapsulate significant relationships, such as familial ties or connections within a friend group. In the context of *SIMBox* fraudulent SIMs, which frequently re-originate multiple calls with a diverse set of contacts, there is a notable absence of these meaningful relationships. Fraudsters employ a two-pronged approach to counteract this inherent pattern [14]:

- First, by handling *incoming traffic routing*, specific *SIMBox* users can be prioritized for terminating calls from international users based on the corresponding contact. This strategic control serves to limit fraudulent users' number of contacts.
- Fraudulent users leverage *network activity generation* functionalities, enabling them to exchange calls and text amongst themselves. This tactic simulates the appearance of close-knit groups, effectively mimicking genuine user relationships.

(4) Device property. The device property relates to the number of SIM cards a user's device operates with, typically limited to one or two for legitimate users. However, in the case of a *SIMBox*, designed to handle multiple SIM cards per GSM module, a single fraudulent device identifier (dev_p, i.e., IMEI code) may be associated with several SIM cards. Additionally, a SIM card connecting to multiple GSM modules accumulates multiple recorded device identifiers (IMEIs). To obfuscate this noticeable behavior, fraudsters have devised methods to simulate a distinct mobile device identifier for each fraudulent SIM card [15]. They achieve this by employing *IMEI modification* or spoofing that links the IMEI code to the SIM card rather than the GSM module. Therefore, fraudsters can automatically alter the IMEI code by specifying an IMEI generation rule and the corresponding new IMEI value(s).

B *FRAUDZEN* DESCRIPTION

B.1 The *SimulationManager*

Being event-driven, it chronologically executes timestamped events issued by simulation objects throughout the simulation time. An event execution runs a function into the related simulation object, which can schedule new events.

B.2 The *NetworkManager*

This module is responsible for creating the cellular network infrastructure, the user devices, and the fraud architecture (cf. Fig. 2), all described in the following.

Network infrastructure. The mobile cellular network, as designed in *FraudZen*, is multi-operator. Each operator provides its subscribers with communication services through voice, text messages, and mobile data to or from external networks (i.e., local or international operators). Each operator's network architecture is based on the standards, comprising a Radio Access Network (RAN) and a core network. Each RAN's base station transmits service initiation requests to the core network and performs paging. The core network provides, in addition to network communication services, control operations, namely authentication and mobility management. It also continually records in a file timestamped network service usage events constituting each operator's CDRs.

Network devices. A network device is formed by adding a SIM to a cell phone. Such a partition between SIMs and user appliances allows for distinguishing legitimate devices from fraudulent ones, created in the *SIMBox*. Each legitimate device has a mobility component and a traffic component. The mobility component keeps up-to-date devices' connected base station by scheduling handovers, i.e., movements between the network cells, according to the chosen mobility model. Similarly, the traffic manager schedules each device's traffic generation according to the defined traffic model.

SIMBox architecture. From inputs such as the number of SIMBank, gateways, and SIM/GSM configuration groups, *FraudZen* builds a *SIMBox* architecture and related configuration units. The *SIMBox* operates with SIM cards provided by the simulation operators, from which it forms *SIMBox* SIM cards (i.e., *SimbSIMCard*). Each *SimbSIMCard* has a state (e.g., Free/Blocked) and a set of parameters related to its traffic (e.g., call count, total call duration), allowing the control of its activity. Each SIM and GSM group works according to a set of functionality configurations given by the *SIMBox* fraud model. Once, the *SIMBox* architecture is formed and configured, it continually receives, with a fixed frequency, international traffic to be routed as local calls to legitimate users.

B.3 The *TrafficManager*

This module handles traffic generation for legitimate and fraudulent devices. For legitimate devices, traffic events are repeatedly generated according to the input traffic model. Each traffic event has the attributes: timestamp, event type (call, international call, text, or data), metric (i.e., call duration or data size), and contact in case of a call or text. *FraudZen* currently includes a trace-based traffic model, reading next events from an input file. Based on the returned event type, legitimate devices transmit a call/text/data service request to their connected base station. Fraudulent devices make use of these same requests to generate traffic; however, this is coordinated by *SIMBox* architecture's *network activity generation* and *incoming traffic routing* algorithms.

B.4 The *MobilityManager*

It handles legitimate and fraudulent users cell-granularity displacements during the simulation. Legitimate users have a mobility attribute implementing the input specified mobility model. *FraudZen*

currently includes mobility models based on existing mobility traces, which can be easily extended. On the other hand, the mobility of a fraudulent device matches the movements of its belonging gateway. Such movements are governed by a strategy defined at the level of the whole *SIMBox* architecture and given by the input *SIMBox* fraud model.