Amplifying Threats: The Role of Multi-Sender Coordination in SMS-Timing-Based Location Inference Attacks

Evangelos Bitsikas, *Northeastern University;* Theodor Schnitzler,
Research Center Trustworthy Data Science and Security and Maastricht University;
Christina Pöpper, *New York University Abu Dhabi;*
Aanjhan Ranganathan, *Northeastern University*

https://www.usenix.org/conference/woot24/presentation/bitsikas

This paper is included in the Proceedings of the
18th USENIX WOOT Conference on Offensive Technologies.

August 12–13, 2024 • Philadelphia, PA, USA

ISBN 978-1-939133-43-4

Open access to the
Proceedings of the 18th USENIX WOOT
Conference on Offensive Technologies
is sponsored by USENIX.

Amplifying Threats: The Role of Multi-Sender Coordination in SMS-Timing-Based Location Inference Attacks

Evangelos Bitsikas[⋆], Theodor Schnitzler[†‡], Christina Pöpper[§], Aanjhan Ranganathan[⋆]
[⋆]*Northeastern University,* [†]*Research Center Trustworthy Data Science and Security,*
[‡]*Maastricht University,* [§]*New York University Abu Dhabi*
bitsikas.e@northeastern.edu, theodor.schnitzler@maastrichtuniversity.nl,
christina.poepper@nyu.edu, aanjhan@northeastern.edu

Abstract

SMS-timing-based location inference attacks leverage timing side channels to ascertain a target's location. Prior work has primarily relied on a single-sender approach, employing only one SMS attacker from a specific location to infer the victim's whereabouts. However, this method exhibits several drawbacks. In this research, we systematically enumerate the limitations of the single-sender approach, which prompted us to explore a multi-sender strategy. Our investigation delves into the feasibility of an attacker employing multiple SMS senders towards a victim to address these limitations and introduces novel features to bolster prediction accuracy. Through exhaustive experimentation, we demonstrate that strategically positioned multiple SMS senders significantly enhance the location-inference accuracy, achieving a 142% improvement for four distinct classes of potential victim locations. This work further highlights the need to develop mitigations against SMS-timing-based location inference attacks.

1 Introduction

SMS (Short Message Service) has emerged as a key vector in numerous cyber-attacks due to its widespread use for purposes such as two-factor authentication [21], identity verification [24, 25], and emergency alerts [24, 25]. Its prevalence, reliability, and global reach have made it a favored medium for malicious activities. Smishing attacks, for example, leverage SMS to distribute links that direct victims to phishing sites, aiming to steal sensitive information [14]. The Flubot virus utilized SMS links to spread trojan apps that compromised banking credentials, personal data, and disabled security features [9]. Beyond these, SMS has been exploited for spamming [8] and to propagate malware such as Simjacker and WIBAttack, which embed malicious commands within binary SMS messages [4, 28].

Most recently, a novel approach to ascertain the location of recipients was demonstrated in [6], utilizing the timing of silent SMS messages in conjunction with machine-learning techniques. This strategy exploits the delivery reports generated upon SMS reception as a timing attack vector for the sender. Rigorous experimentation across various countries, telecommunications operators, and a range of devices demonstrated that an attacker could deduce a recipient's location by analyzing timing data from typical receiver locations. Although this method introduces an innovative side channel for localizing mobile users, it encounters notable limitations. Most importantly, there is a significant probability that the attack originating from a single source/mobile device can be detected and potentially be blocked by the victim's service providers. This is more apparent when the attack requires a substantial amount of SMS transmissions to collect the necessary data. Additionally, as the number of possible victim locations increases, the method's accuracy in predicting locations degrades due to the finite entropy available from single attacker-victim channel timing reports. As a result, there are classifications in which machine learning can perform poorly.

To tackle the above-mentioned limitations associated with single-sender-based SMS location inference attacks, this paper focuses on the following key research questions. The primary question we explore is whether using multiple coordinated SMS senders can improve the accuracy of localization predictions. We hypothesize that using senders from different locations could create unique timing side-channels which, when combined, could lead to more accurate classifications. This multi-sender approach can improve the prediction accuracy, especially as the number of potential victim locations increases. Additionally, using multiple SMS senders spread out geographically could also make the attack more resilient against being blocked, as the victim's service provider now has to identify and block several senders. Optimizing the timing and pattern of SMS

sending could further reduce the likelihood of the attack being detected. Finally, we hypothesize that the attacker can collect a significantly smaller amount of data to conduct this attack efficiently, without compromising the model's accuracy. Consequently, the adversary can save resources, as well as measurement collection and training time.

Motivated by the above hypothesis, in this paper, we make the following contributions:

- We identify limitations of single-sender SMS-timing-based location inference attacks and conceive *multi(ple)-sender SMS-timing-based location inference attack* in cellular networks. To establish a baseline for comparison with our multi-sender approach, we reproduced the single SMS sender-based localization attack described in prior work [6]. Interestingly, our data analysis highlights certain limitations inherent in the single-sender approach which serve as a crucial motivation for the development of our multi-sender approach.

- Through rigorous experimentation, we demonstrate the enhanced capability of multiple SMS senders, strategically placed across different locations, to coordinate and significantly improve the accuracy in determining a victim's location. Our experiments reveal that the multi-sender MMS approach can reach up to 142% accuracy improvement for four classes. This further emphasizes that the effectiveness of the multi-sender attack strategy improves with an increasing number of potential victim locations, thereby overcoming a significant limitation of the single-sender approach.

- We highlight two substantial improvements and insights: (1) From the distinct timing side-channels generated by the multi-sender setup, we identify and introduce *new features* that are instrumental in boosting the prediction accuracy: the statistical mean, median, and standard deviation of the senders' delivery time measurements, allowing us to effectively fuse the timings from multiple senders to improve the accuracy even further. (2) We investigate the required *sample sizes* for location inference attacks and demonstrate that already a few hundred SMS can yield strong results without the need for thousands of collected messages.

2 Background and Motivation

In this section, we provide the technical background for SMS delivery processes and then delve into the concept of *SMS-timing-based Location Inference Attacks*. We subsequently outline its limitations, which serves as the foundation for our research presented in this paper.

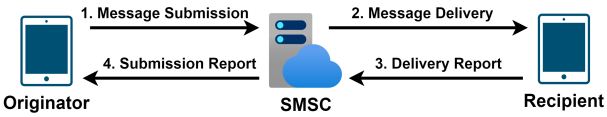


Figure 1: Brief representation of the SMS process, according to GSMA [11].

2.1 Overview on SMS Process

Short Message Service (SMS) is an inherent component of the cellular infrastructure and universally accessible across all network generations from 2G to 5G [1–3, 11]. Figure 1 briefly outlines the SMS delivery process involving the originator (sender), Short Message Service Center (SMSC), and the recipient (receiver).

The process begins with the message submission (Step 1) by the originator, who composes the message and sends it to the SMSC. Upon receiving the SMS, the SMSC performs the necessary network and validation checks and then forwards the SMS to the intended recipient. The SMSC ensures that the message reaches the recipient (Step 2), even if it means storing it temporarily, in case the recipient is unavailable immediately. Additionally, the originators have been informed by now that the submitted message was actually sent.

Next, once the recipient receives the message, the involved device sends the delivery report back to the SMSC. The report confirms that the message has been successfully delivered to the recipient's device (Step 3). Finally, the report is sent to the originator via the SMSC, called the submission report (Step 4). This report ultimately confirms that the message was sent and delivered to the recipient successfully.

2.2 SMS-timing-based Location Inference

In an SMS-timing-based Location Inference attack, an attacker is interested in learning the current physical location of a specific victim by sending them (silent) SMSes. The attack builds upon the time elapsed between sending the SMS and the SMS being delivered to the victim and is conducted in two phases.

In the first phase (fingerprint generation), the attacker repeatedly sends SMSes to the victim while knowing their respective locations and measures the time it takes to deliver the SMS messages. By analyzing the resulting delivery timings and their distributions, the attacker is able to determine a unique fingerprint for each of the locations the victim has visited.

In the second phase (location inference), the attacker sends new SMS messages to the victim *without* knowing their current location, measures the time it takes to deliver them, and then *classifies* the collected timings by comparing them to the previously obtained fingerprints. Thus, the attacker can determine and re-identify the victim's location out of a set of known locations.

2.3 Limitations and Motivation

When the SMS-timing-based Location Inference Attack is carried out from a single sender at a fixed location, it has several drawbacks. In particular, the success and performance of the attack depend heavily on the specifics of the chosen location and its mobile network connection, such as the distance to the base station. The quality and reliability of the connection, along with the robustness of the collected data, may also vary depending on circumstances specific to the location, such as fluctuating numbers of people and concurrent mobile network connections throughout the day or week.

Another drawback is that during the initial phase of the attack (fingerprint generation), the attacker engages in non-standard behavior as a mobile network subscriber. Consequently, there is a risk that the adversary may be perceived as suspicious by the network operator and potentially be blocked, particularly if only a single static location is utilized.

From an organizational perspective, the attack outlined in [6] encompasses analyses at various levels of granularity, and a broad range of locations, from regional to worldwide attacks. However, the study lacks a thorough analysis of the sample size impact regarding the classification accuracy. This limitation implies that the attack requires additional evaluation.

Hence, we recognize the necessity for a more systematic evaluation of factors that could impact the SMS-timing-based Location Inference Attack's performance. This entails varying the adversary's location, systematically assessing the attack's performance with different receiving devices at the same locations, conducting repeated evaluations with varying sample sizes, and expanding the attack to encompass attackers operating from multiple vantage points simultaneously.

3 Multi-Sender Location Inference

3.1 Threat Model

We consider an attacker whose primary goal is to determine the presence of a victim's mobile device within a specific geographic area, without the intention to track the victim's exact movements.

The attacker is presumed to possess the victim's mobile number, enabling them to initiate various forms of SMS communications, including personal messages, undirected mass messages such as marketing advertisements, and notably, silent SMSes which the victim's device acknowledges without alerting the user. It is assumed that the attacker has access to an arbitrary number of smartphone devices, SIM cards, mobile numbers, and subscription plans. Furthermore, the attacker can deploy multiple sender devices in different geographical areas to collect data from the victim receivers simultaneously and combine them for location extraction. The adversary is assumed to possess the capability to utilize network services as a *conventional* user: leveraging several SIM cards, having the ability to send messages to any subscriber with a valid number, and maintaining a normal connection for the transmission of text messages and receipt of delivery notifications.

We emphasize that the attacker does *not* require physical access to the victim's mobile device, USIM cards, or any network infrastructure, nor do they seek to obtain or modify sensitive victim data such as cryptographic keys.

3.2 Attack Concept

The foundation of the multi-sender approach rests on the observation that fingerprints generated from the SMS exchanges between a single sender (attacker) location and a receiver can be limited in their effectiveness for accurate location classification. This limitation becomes particularly pronounced in complex environments, such as certain German locations in [6], where the variability and granularity of the urban landscape can dilute the distinctiveness of timing fingerprints.

To address these challenges, this work pioneers the integration of multiple attacker locations into the analysis framework. By orchestrating SMS exchanges from various (unique) attacker positions to the receiver, a richer and more nuanced dataset emerges. Each unique pairing of attacker and receiver locations contributes a distinct timing fingerprint to the dataset. These timing fingerprints, when aggregated, undergo further processing to distill additional dataset features, thereby forging more robust and comprehensive fingerprints. This enriched dataset plays a crucial role in enhancing the efficacy of machine-learning models during both the training and prediction phases.

For conducting a multi-sender location inference attack, we essentially replicate the attack methodology presented in [6] and simultaneously execute it from multiple locations. Consistent with previous work, the attack comprises two phases: fingerprint generation and location inference, but both are conducted from multiple sender locations. Basically, multiple instances of the

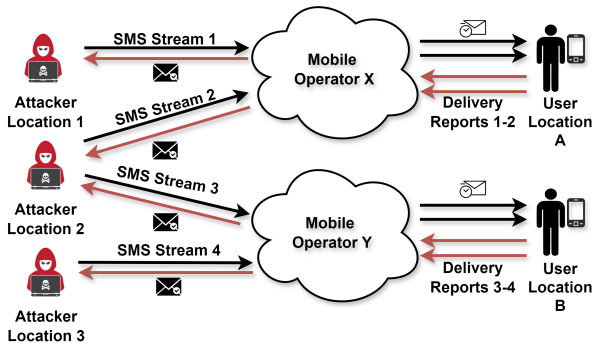


Figure 2: Multiple attackers in different locations establish SMS streams to send silent messages to the victim in various locations and receive delivery reports. This is possible even with distinct network providers.

single-sender location inference attack are executed in parallel.

Multi-Sender Setup. To gather data from multiple vantage locations and eventually enhance the accuracy of the location identification attack, the attacker deploys the setup at various geographical locations. Intuitively, by employing more attacking locations that are diverse, an adversary could generate more precise receiver location fingerprints. This distributed approach allows the attacker to collect measurements of the victim's location from different "angles", increasing the robustness and reliability of the subsequent analysis.

Attacking Process. The attacker, situated in multiple locations, initiates the process by sending a barrage of silent SMS messages to the victim. The victim, unknowingly participating in this scheme, moves across different locations at different times. The silent nature of these messages means that the receiver's device does not notify the victim of the incoming SMS, thus keeping the process clandestine. Each time a message is received, the victim's device automatically generates and sends back delivery reports as part of its standard operating procedure. These reports, unbeknownst to the victim, reveal valuable information for the attacker, notably the sent and delivered times. By analyzing the time discrepancies between when a message was sent and when the delivery report was received, the attacker can infer certain aspects of the victim's location.

Since this procedure is repeated multiple times in the multi-sender attack, it accumulates a substantial dataset of measurements. The attacker categorizes the measurements based on the victim's known locations during the attack, forming distinct datasets for each location. These datasets are then aggregated and analyzed to predict the victim's location in the future. According to Figure 2, the attacker creates several SMS streams, which could be established with different operators since the attacker can operate from different countries. The victim may also move to different countries and sends back the delivery reports to the corresponding SMS.

In the prediction stage, the attacker collects fresh measurements from the current location of the victim in the same fashion. These measurements serve as input for a machine-learning model that has been trained on the previously collected data, representing potential locations of the victim. Then, the model processes this input and outputs a prediction of the victim's current location.

4 Experimental Validation

In this section, we detail our experimental validation of the SMS-timing-based location inference attack with multiple senders and report on our setup for data collection, processing, and evaluation.

4.1 Data Collection Setup

At the core of the attacker's setup is the use of typical computer devices equipped with a smartphone running Android Debug Bridge (ADB). ADB allows for a wide range of communication with a connected device, in this case, to transmit silent SMS messages and record the sent and delivered timestamps. As in [6], the SMS transmission and recording of the timing metrics is conducted by an Android application, which also stores results for further processing. Controlling the application via ADB allows us to automate this process since it should be repeated multiple times to collect a sufficient number of timing metrics. This process also happens stealthily, without altering the victim, since the attacker utilizes silent SMSs which are accepted by the network operator. Moreover, the attacker's equipment includes a SIM card, granting access to the cellular network.

Adhering to the aforementioned attacking concepts, over a period of 12 weeks, we repeatedly send SMS messages between smartphones in different locations in Germany and the Netherlands. We do not consider locations that are very far apart, as they are easier for an attacker to identify [6]. We use three smartphones, each placed in a fixed location that remains unchanged during the experiments, to send messages to four phones whose positions are periodically rotated. For sending SMS messages, we use two locations in Germany and one in the Netherlands. The receiving phones are placed in five different locations in Germany and three in the Netherlands (including the locations of the sending devices). Table 1 lists the devices we used for sending and receiving SMS messages, and Table 2 provides an overview of the locations used during our measurements and the amounts of data collected.

Table 1: Device Specifications

ID	Device	Chipset	OS	Model	Release
Sending Devices					
D	**Samsung Galaxy A53**	Samsung Exynos 1280	Android 12	SM-A536E/DS	2022
V	**Nokia 5.3**	Qualcomm Snapdragon 665	Android 11	TA-1234	2020
B	**Huawei P8 Lite 2017**	HiSilicon Kirin 655	Android 8	PRA-LX1	2017
Receiving Devices					
px6a	**Google Pixel 6a**	Google Tensor	Android 12	G1AZG	2022
a53	**Samsung Galaxy A53**	Samsung Exynos 1280	Android 12	SM-A536E/DS	2022
op7	**OnePlus 7 Pro**	Qualcomm Snapdragon 855	Android 11	GM1910	2019
p8l	**Huawei P8 Lite 2017**	HiSilicon Kirin 655	Android 8	PRA-LX1	2017

Table 2: Data Collection Summary

	Number of SMS per Receiving Device				**Distances [km] to Sender**		
	px6a	p8l	op7	a53	Sender B	Sender D	Sender V
Receiver Locations in Germany							
DE-1	3160	3280	420	–	11	0	140
DE-2	1540	1560	–	–	2	11	130
DE-3	4960	4540	8920	6900	0	11	129
DE-4	420	460	–	–	4	14	126
DE-5	1220	320	–	–	5	11	140
Receiver Locations in the Netherlands							
NL-1	7140	5500	0	1440	125	135	4
NL-2	5820	5280	10300	8700	129	140	0
NL-3	2020	960	1680	1120	125	136	7

Locations (Cities): *DE-1,5*: Dortmund, *DE-2,3,4*: Bochum, *NL-1*: Eindhoven, *NL-2*: Veldhoven, *NL-3*: Valkenswaard

Locations in the same country are chosen to be relatively close to each other. The distance from a receiving location to the *closest* sending device is 11 km at maximum, which also corresponds to the distance between the two sending devices in Germany.

4.2 Data Collection Procedure

We replicated the attack to use an Android app that sends one silent SMS at a time to a designated target phone number. Additionally, the app waits for the *Sent* and *Delivered* notifications and collects and stores the timestamps for the SMS transmission and both notifications. In line with previous work, we schedule 20 consecutive SMS transmissions on an hourly basis. We automate SMS transmissions by controlling the app remotely via a Python script issuing ADB commands to the smartphone. We simultaneously send SMS messages from all senders to the same receiver by scheduling the script to start once per hour at the same time for a specific receiver (i. e., :00 for the first receiver, :15 for the second receiver, ...) across all senders. While this does not guarantee perfect sender synchronization due to potential offsets in their individual system clocks, we consider this a best-effort approach to approximate the behavior of an adversary simultaneously probing a specific target from multiple locations.

Our data collection tooling builds upon the code released by Bitsikas et al. [6] available on GitHub[1] and is extended to fit with the phones we use for sending messages. We also follow the guidelines provided along with the framework to implement the missing code handling the actual SMS transmission and timestamp collection procedures.

4.3 Feature Set Generation & Multi-Sender Fusion

To generate the timing features for each SMS transmission and combine the multi-sender datasets, we take the following steps:

[1] https://github.com/vaggelis-sudo/SMS-Location-Identification-Attack

Step 1: Calculating the initial metrics. Following [6], we calculate the initial metrics for each SMS transmission in the collected dataset: the real sent duration T_{sent}, the real delivery duration T_{del}, the total delivery duration T_{tot}, and the delivery ratio P.

$$T_{sent} = t_{sent} - t_{tx} \tag{1}$$

$$T_{del} = t_{del} - t_{sent} \tag{2}$$

$$T_{tot} = T_{del} + T_{sent} \tag{3}$$

$$P = \frac{T_{del}}{T_{tot}} = \frac{t_{del} - t_{sent}}{t_{del} - t_{tx}} \tag{4}$$

Then, for every two consecutive SMS transmissions ($j-1$ and j), we calculate the differences in sent duration $T_{\Delta sent}$ and delivery duration $T_{\Delta del}$, respectively:

$$T_{\Delta sent} = (T^{j}_{sent} - T^{j-1}_{sent})/T^{j-1}_{sent} \tag{5}$$

$$T_{\Delta del} = (T^{j}_{del} - T^{j-1}_{del})/T^{j-1}_{del} \tag{6}$$

Moving beyond [6], the fingerprint does not conclude with this calculation, as we do not consider only one but multiple senders.

Step 2: Combining the sender datasets. Let D_i represent the dataset for sender i, where $i = 1, 2, \ldots, m$, with n receiver locations. Additionally, let $t_{\text{del},i,r,j}$ denote the delivery time of the j-th SMS transmission from sender i to receiver r. Finally, let $S_{i,r,j}$ represent the data associated with the j-th SMS transmission from sender i to receiver r, including $t_{\text{del},i,r,j}$. Then, D_{concat} is the dataset resulting from the concatenation process, where each element is derived by matching $S_{i,r,j}$ from all senders based on the closest matching $t_{\text{del},i,r,j}$.

For each $S_{i,r,j}$ in D_i, we seek to find $S_{k,r,l}$ in D_k ($k \neq i$) such that the difference in delivery times $|t_{\text{del},i,r,j} - t_{\text{del},k,r,l}|$ is minimal or zero, indicating the closest matching timestamps across different senders. This process occurs for every receiver separately and every available sender, until the new D_{concat} dataset contains per row the data of each sender to the same receiver, but synchronized. Algorithm 1 shows briefly the process.

Step 3: Fusing the sender datasets statistically. Given m senders, the number of unique combinations of two senders is given by the binomial coefficient:

$$\binom{m}{2} = \frac{m!}{2!(m-2)!} \tag{7}$$

For each pair of senders and for every z consecutive SMS transmissions (in this study, $z = 5$ [2]), we calculate the `Mean`, `Median`, and `Standard Deviation` of the delivery times. Let $t^{(s,r)}_{\text{del},i}$ denote the delivery time of the i-th SMS in a sequence of z consecutive messages from sender s to receiver r. The statistics are calculated as follows:

$$\mu^{(s,r)} = \frac{1}{z}\sum_{i=1}^{z} t^{(s,r)}_{\text{del},i} \tag{8}$$

$$\text{Median}^{(s,r)} = \text{Median}\{t^{(s,r)}_{\text{del},1}, t^{(s,r)}_{\text{del},2}, \ldots, t^{(s,r)}_{\text{del},z}\} \tag{9}$$

$$\sigma^{(s,r)} = \sqrt{\frac{1}{z-1}\sum_{i=1}^{z}(t^{(s,r)}_{\text{del},i} - \mu^{(s,r)})^2} \tag{10}$$

Differences in these statistics for the delivery time between pairs of senders are calculated as their actual differences. For example, for means between sender pair (s_1, r) and (s_2, r):

$$\Delta\mu^{(s_1,s_2,r)} = \mu^{(s_1,r)} - \mu^{(s_2,r)} \tag{11}$$

These differences, $\Delta\mu^{(s_1,s_2,r)}$, $\Delta\text{Median}^{(s_1,s_2,r)}$, and $\Delta\sigma^{(s_1,s_2,r)}$, are incorporated into the dataset for each sender pair accordingly, as additional features.

Algorithm 1 Match and Concatenate SMS Transmissions based on Timestamps

```
Initialize D_concat = ∅ as empty dataset
for each receiver location r from 1 to n do
    for each S_{i,r,j} in D_i for all i do
        Initialize a list L_{i,r} to hold data for concatenation
        for each D_k where k ≠ i do
            Find S_{k,r,l} in D_k such that |t_{del,i,r,j} − t_{del,k,r,l}| is minimized
            Add S_{k,r,l} to L_{i,r}
        end for
        NewRecord_{i,r} ← Concatenate(L_{i,r})
        D_concat ← D_concat ∪ {NewRecord_{i,r}}
        Clear L_{i,r}
    end for
end for
```

4.4 Multi-Sender Techniques

Simple Integration of Senders. In this method, the initial features are generated based on the timing data from individual sender-receiver pairings (Step 1). Subsequently, datasets corresponding to multiple senders are amalgamated (Step 2) *without* the application of sophisticated statistical fusion techniques (Step 3). Thus, we create datasets that are matched and concatenated based on the timestamps, but without incorporating unique feature types.

Specifically, we consider double- and triple-sender datasets as distinct (simple) approaches. For the double-sender cases, we create the BV, VD, and BD datasets,

[2] We determined that the number should be less than 10 in *our* dataset to accommodate small sample sizes while not covering too many transmissions at a time. A middle value of 5 was chosen as a result.

while for the triple-sender cases, we create BDV, based on Table 2. The total number of features for double-senders is 12, and for triple-senders is 18, according to Algorithms 1- 6 from Step 1. This exploratory step seeks to discern whether straightforward sender concatenation can bolster the machine-learning model's predictive accuracy compared to single senders and to statistically combined datasets.

Statistical Fusion of Senders. Advancing beyond the simple approach, the statistical combination of sender datasets represents a more refined approach to dataset enhancement. This technique encompasses a comprehensive process involving the generation of initial features (Step 1), the combination of sender measurements (Step 2) followed by the fusion of datasets from multiple senders through the statistical metrics (Step 3). Unlike the simple method, this approach enriches the combined dataset with additional features derived from the statistical analysis of delivery times: using the means, medians, and standard deviations between the sender measurements. For this approach, we use all three senders with their maximum sample size available for each receiver location.

In this work, we explore the following two strategies:

1. `Enhanced Mean Datasets.` Datasets statistically enhanced by the *mean* of the delivery time. A total number of 21 features is used, corresponding to the 18 combined features for the three senders and the 3 additional ones generated by the differences between the sender means.

2. `Enhanced MMS Datasets.` Datasets statistically enhanced by the *mean*, *median* and *standard deviation* of the delivery time. A total number of 27 features is utilized, correlated with the 18 combined features for the three senders and the 9 extra ones engendered by the differences between the sender means, medians, and standard deviations.

This dual-strategy approach aims to demonstrate the superiority of statistically enhanced datasets over both single-sender datasets and those trivially combined. The hypothesis posits that the inclusion of a broader array of statistical features not only increases the accuracy of location predictions beyond that achievable with simpler dataset combinations but also highlights the comparative advantage of the "Enhanced MMS" over the "Enhanced Mean" approach. This distinction underscores the principle that the depth and complexity of features within the dataset are pivotal to the refinement of model accuracy.

4.5 Attack Training & Prediction

In this study, we employ a Multilayer Perceptron (MLP) Classifier, a type of feedforward artificial neural network, as the core predictive model to analyze the relationship between the features derived from SMS transmission data and the target outcomes. The MLP Classifier is instantiated with a specific configuration of hyperparameters to optimize its performance for the given dataset. The architecture of the neural network is defined by hidden layer sizes = (10, 40, 10), indicating a three-layered structure where the input data is first processed by a layer of 10 neurons, followed by a denser layer of 40 neurons, and finally, the information is aggregated through a layer of 10 neurons before reaching the output layer. This configuration is designed to capture the nonlinear relationships between the input features.

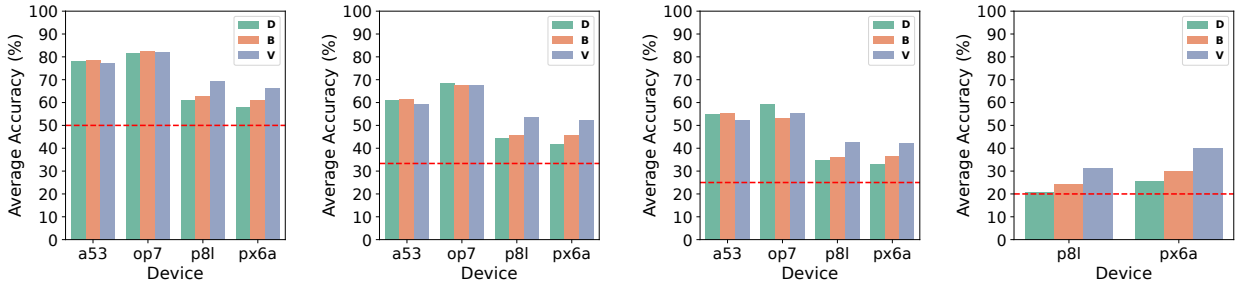
The model utilizes the stochastic gradient descent (SGD) algorithm for optimizing the network's weights. This choice is motivated by SGD's efficiency in handling large datasets and its capability to escape local minima during training. The regularization term, alpha= 0.0001, is set to a low value to prevent overfitting while allowing the model to learn complex patterns in the data. With learning rate='constant' and a max iteration of 5000, the learning rate is kept fixed across all epochs of training, and the model is allowed a substantial number of iterations to converge towards an optimal set of weights. Batch processing is employed with a size of 32 to leverage computational efficiency and stability in gradient descent updates. Model evaluation is conducted through a 10-fold cross-validation process providing a robust estimate of the model's predictive accuracy on various random data. Finally, the `Accuracy` metric is calculated to quantify the model's performance, offering a measure of how often the model predictions match the true labels. In our experimentation, we repeatedly run the model prediction with increasing numbers of samples per class, (i. e., 100, 200, 300, 500, 1000, 5000, and 10000), to analyze differences in the classification accuracy.

5 Experimental Evaluation and Results

We next describe the exact experimental setup we used in our experiments and then delve into our results with the multiple-sender approaches.

5.1 Single Senders: Baseline

We ran the classifications for single senders (D, B, and V) to establish the baseline for the subsequent improvement. Figure 3 illustrates the results of all classifications for all sample sizes. Generally, the lowest accuracy is observed for sender D on the device p8l with 5 classes (21%), while the highest accuracy is observed for sender B on the device op7 with 2 classes (82%). In fact, we make similar observations for the single sender classifications with [6], regarding the average accuracy scores



(a) Scores for 2 Classes (b) Scores for 3 Classes (c) Scores for 4 Classes (d) Scores for 5 Classes

Figure 3: Average single-sender accuracy scores across devices and classes. These scores are considered the established baseline for which we provide improvement. The presented results take into account all possible sample sizes. The red dashed line indicates random guessing.

and the decline across the increasing number of classes.

Specifically, for each device examined ranging from a53 to px6a, the data showcases a nuanced relationship between the number of classes involved in the classification task and the single sender accuracy scores. Notably, as the number of classes increases, a general trend of decreasing accuracy is observed, which is consistent across all devices. This trend is particularly evident when comparing results from 2-class configurations to those with 4 or 5 classes, where the average accuracy scores tend to diminish, highlighting the increased complexity and challenges associated with classifying a larger number of classes. Moreover, some devices and senders exhibit a more graceful degradation in accuracy as more classes are added. For example, V on px6a degrades from 66% with 2 classes to 40% with 5 classes, a relatively modest decline compared to D on p8l, which plummets from 61% with 2 classes to 21% with 5 classes.

In the comparative analysis of device performance, the op7 and a53 models significantly outperform the p8l and px6a devices across all metrics. In particular, the p8l and px6a devices achieve a maximum accuracy of 69% and 66%, respectively, when tested with sender V. Furthermore, sender V consistently surpasses senders B and D in performance on the p8l and px6a devices, highlighting a notable disparity in efficacy. Conversely, when evaluating the performance on the op7 and a53 devices, the results among senders B, D, and V demonstrate a remarkable uniformity, with only minimal variations in accuracy. The most significant discrepancy observed is a 6% difference between senders B and D when assessed with four classes on the op7 device. This suggests that while op7 and a53 provide more consistent and higher performance across different senders, p8l and px6a exhibit limitations, particularly in terms of accuracy and sender variability. Consequently, sender V not only shows higher accuracies across the board but also appears to be more resistant to accuracy drops as the number of classes increases. This suggests that V's data might be inherently more separable or that V employs more consistent patterns in location-related behavior. Overall, the presence of differences in performance between the senders within the same device and class configuration underscores the variability in sender effectiveness.

5.2 Multiple Senders: Simple Combination

In this subsection, we start by comparing the double- and triple-sender accuracy scores with the single-sender scores. In Figure 4, we show all classification accuracy scores with the worst (minimum) and best (maximum) performances of the single- and double-sender data, across all devices and sample sizes. The aim here is to show the minimum and maximum improvement of the multi-senders with simple combinations, based on this collected dataset.

Reflecting on previous discussions, sender V consistently emerges as the top performer across all metrics, capturing both the lowest and highest scores. However, this trend does not uniformly extend to scenarios involving double- and triple-sender configurations. Initially, all multi-sender combinations yield superior accuracy rates compared to individual efforts by senders B and D, underscoring the premise that pooling sender data can enhance overall performance. Notably, in binary classification tasks, sender V is marginally eclipsed by combinations such as DV, BV, BD, and BDV, and similarly by DV, BV, and BDV in contexts involving three and four classes. On the contrary, the BD pairing underperforms for three and four classes, highlighting that sender D's contributions do not bolster the collective accuracy to the same extent as other senders in these specific instances.

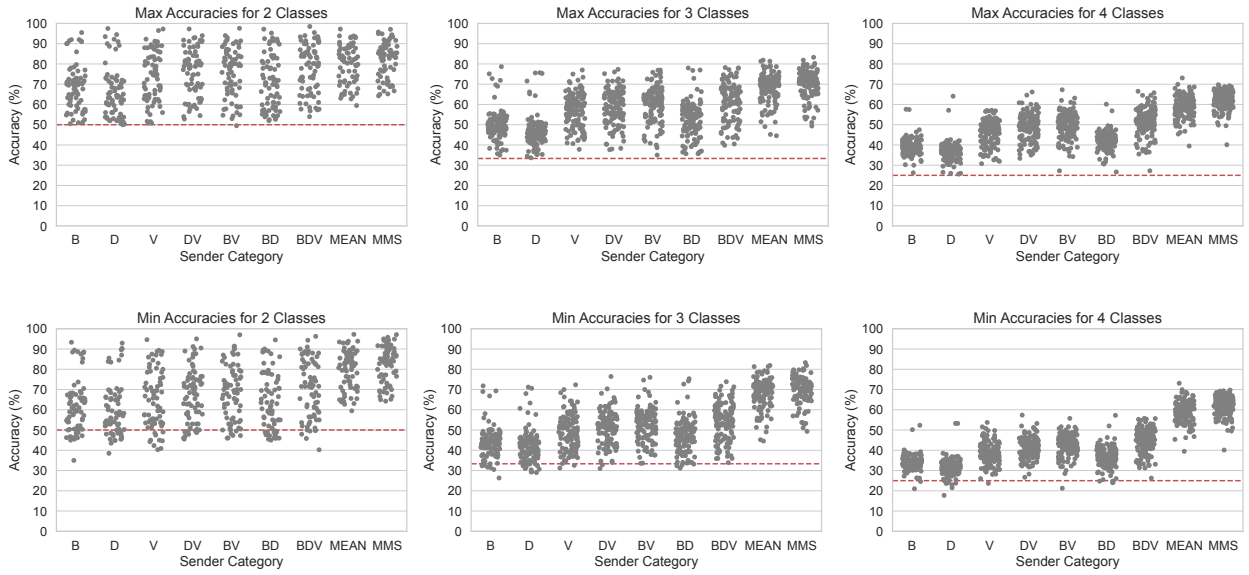


Figure 4: The scatter plots illustrate the accuracy points between different sender types and classes. All devices and sample sizes are considered. The plots with the **minimum** accuracy scores take into account the worst performance of the single- and double-sender data, while the **maximum** accuracy scores focus on the best possible (in this setup).

This phenomenon underscores a critical insight: a sender with generally lower performance can, in certain conditions, detrimentally impact the collective accuracy of multi-sender configurations.

To illustrate the enhancements in accuracy we achieved by integrating multi-sender data over single-sender benchmarks, we included Figure 5. This figure highlights the maximal accuracy improvements realized in our study for configurations involving two and three senders combined. It provides a detailed examination of the specific devices engaged in our experiments and quantifies the average accuracy enhancement across different class numbers. For each classification category, we pinpointed the lowest accuracy scores from single-sender scenarios and juxtaposed these with the highest-performing scores from multi-sender configurations across all sample sizes. This approach was designed to showcase the performance improvements achievable with multi-sender strategies within our dataset. The underlying principle is that the attacker can always adapt the classifications by choosing the best-performing multi-sender combination.

The analysis reveals that for devices a53 and op7, enhancements from multi-sender configurations are relatively modest for binary classifications. This is attributed to the already high performance of single-sender setups in these instances (as detailed in Figure 3). However, the narrative shifts significantly for classifications involving three and four classes, where we observe improvements of approximately 20%. The scenario is even more pronounced for the p8l and px6a devices, which exhibit progressively larger gains in accuracy with an increase in the number of classes. Notably, the peak improvement recorded is an impressive 120% for the px6a device within four-class scenarios using three senders (namely, the BDV combination).

This data suggests a clear trend: *Classifications that initially present lower accuracy in single-sender formats tend to benefit substantially from the incorporation of multi-senders, particularly in multi-class classifications.*

5.3 Multiple Senders: Statistical Combination

In this subsection, we delve into a comparative analysis between the performance of individual senders and the aggregated results from multiple senders, specifically focusing on the statistically enhanced Mean and MMS datasets. These datasets incorporate data from all three senders at their largest sample sizes, representing the best dataset advancements explored in this study.

By observing Figure 4 once more, it becomes apparent that the Mean and MMS datasets exhibit superior performance for binary classifications compared to other methodologies. This is particularly noticeable in their minimum accuracy scores, which significantly exceed those achieved by alternative approaches. The gap between the Mean and MMS datasets is relatively narrow,

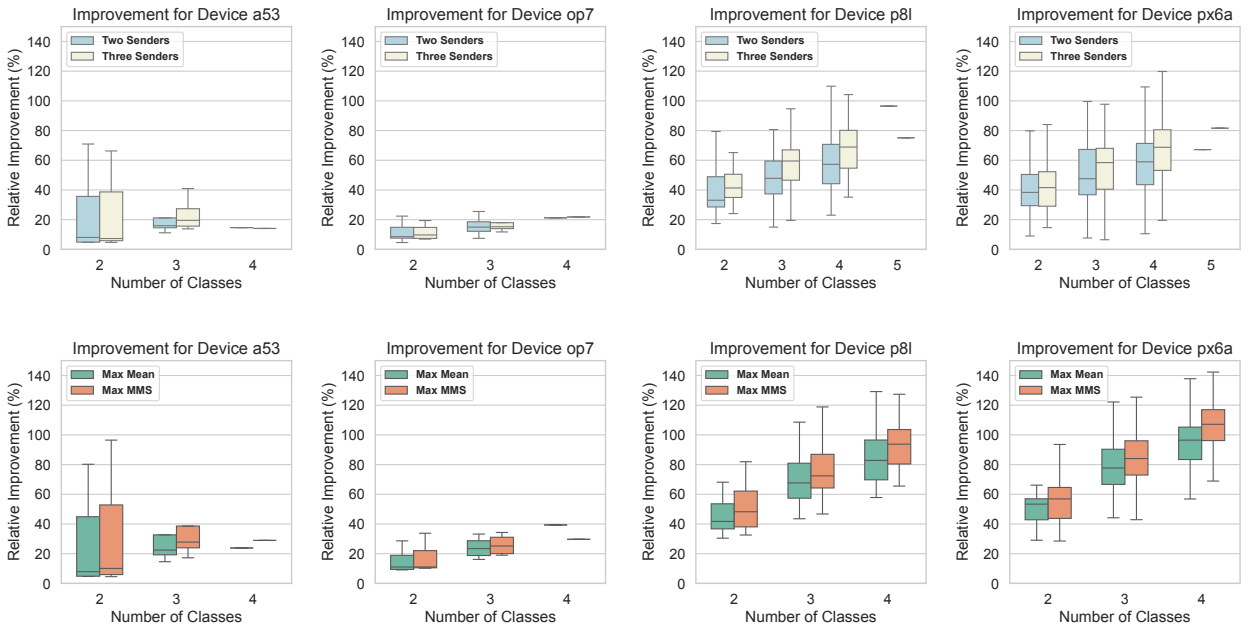


Figure 5: Best accuracy improvement of all multi-sender techniques from the single-sender baseline (not globally optimal), across all sample sizes. Lines in 4 and 5 classes indicate that there was only one classification, meaning one accuracy outcome.

with the MMS dataset showing a marginal enhancement in accuracy. However, the distinction in performance between these advanced datasets and other techniques becomes starkly apparent in the analyses for three and four classes. For these more complex classifications, the MMS dataset demonstrates a better performance than the Mean dataset, unlike the improvement observed in binary classifications. The results indicate that the MMS is currently the best-performing method for location identification, especially for multi-class classifications.

To further investigate the improvement of the Mean and MMS datasets per device, we study the corresponding boxplots of Figure 5 which illustrate the improvement percentages for the enhanced datasets for the four distinct devices. These plots reveal the percentage improvements of the advanced datasets across four distinct devices. For devices a53 and op7, the increments between the Mean and MMS methods are relatively modest. However, as we shift our focus to devices p8l and px6a, especially with an increasing number of classes, the distinction becomes more significant. The MMS dataset showcases the maximum improvement, reaching up to 142% for a four-class scenario on the px6a device. Furthermore, when juxtaposing the performance of the Mean and MMS datasets against results from two or three senders, the superiority of the MMS strategy becomes more evident. Particularly, the MMS dataset demonstrates considerable superiority over the conventional multi-sender combinations, highlighting its effectiveness not just in enhancing accuracy, but also in providing a more consistent and reliable performance across varying class complexities and devices. This comparative analysis not only underscores the value of the MMS approach but also positions it as a notably advanced methodology within the scope of our investigation, significantly outpacing traditional techniques in terms of performance improvement. Still, Figure 5 displays our best improvements, but they are not considered as global optimal, since there might be ways to enhance these techniques even further. Finally, Figure 6 provides additional information comparing the Mean and MMS results to all single senders with all sample sizes.

5.4 Sample Size Comparisons

In machine learning, the sample size is a significant factor that influences the model's performance. A sufficient sample size ensures that the model can capture the diversity of the entire population within the data. Typically, larger sample sizes provide more data points for the mode to learn from, which can lead to higher accuracy and reliability. In our work, we explore the connection between the model's performance and the sample size. Our goal is to determine whether the accuracy increases as the sample size increases. To this end, we

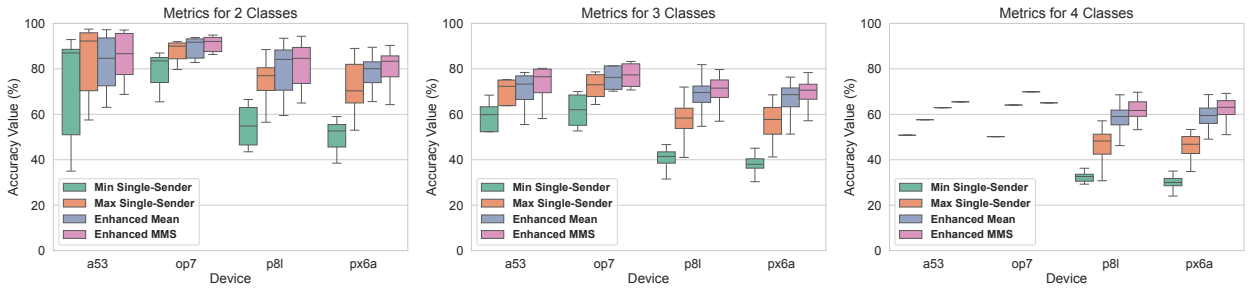


Figure 6: Accuracy boxplots between the single-senders and the enhanced multi-sender approaches for all classifications. The plots consider the **worst** and **best** performing accuracy scores for single senders. These distributions show that MMS achieves the best improvement (not global optimal).

meticulously analyze the performance metrics of single-, double-, and triple-sender results across a sample size range from 100 to 1000.

For single-senders B, D, and V, Figure 7 shows the average accuracy for all number of classes in each device, unveils a trend where accuracy generally stabilizes with an increase in sample size across various device contexts. For double-senders BD, DV, and BV, Figure 8 reveals a consistent pattern of steady or small improved accuracy with larger sample sizes, across all class numbers. This pattern persists into Figure 8, representing triple-sender configurations, where the trends once again affirm the model's steady performance with increased data volume for each class number per device.

Regarding the classification, the trends give us the insight that the model might be well-tuned to the complexity of the task at hand, effectively capturing the patterns within the available data. In addition, this means that the key features and patterns necessary for making accurate predictions are already captured within the smaller dataset. Steadiness after a certain sample size also shows that the model's structure is robust enough to perform reliably under varying dataset conditions.

Consequently, for the attacker, these are promising results as it is not necessary to collect large amounts of data, corresponding to the SMS transmissions, in order to conduct the location identification attack. This can be beneficial in reducing the measurement collection time, computational costs, and training time, making the model more efficient to develop and deploy, where acquiring large volumes of data is challenging or impractical. Additionally, this can also make the adversary less susceptible to detection, since the attacker can adapt to the least amount of SMS transmission and senders for the desired accuracy.

6 Discussion

In this section, we discuss the distribution of the sender locations in our study. Then, we provide our insights on the countermeasures against multi-sender SMS location inference attacks and explain their potential limitations.

6.1 Geographical Distribution of Senders

The strategic placement of sender locations, adhering to the principle of distancing them by several Kilometers, aims to capture diverse timing characteristics (e. g., via different routing), since the networks are black-box to the attacker based on our threat model. In our study, we utilize the most suitable locations from our options, for which we can collect a sufficient amount of data continuously and for a long time. We confine our options to two adjacent countries since it is more challenging to conduct the location inference attack in lower granularity levels. Expanding the number of senders and diversifying locations internationally as well can potentially improve the accuracy of attack even further.

6.2 Countermeasures

Ways to mitigate this attack can span from the elimination of silent SMSes and delivery reports to the implementation of more rigorous SMS filtering mechanisms for spam and flooding, which represents one of the most direct and practical countermeasures against location identification attacks [6]. Enhancing the core concept of resilient spamming/flooding filters, networks are encouraged to integrate advanced anomaly detection systems in order to accurately distinguish between normal and anomalous patterns of SMS traffic. However, it's important to acknowledge that these systems primarily operate based on predefined rules and thresholds for anomaly detection, thereby limiting their efficacy to

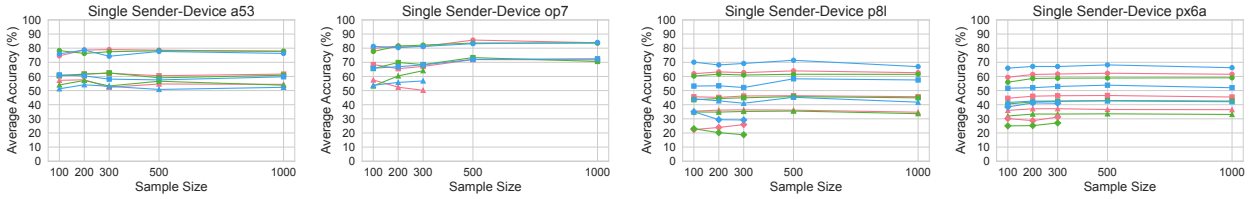


Figure 7: Single-Sender accuracy trend plots for each device, per number of classes. The trends behave steadily and continuously in most cases, as the sample sizes expand. We have included B, D, and V for 2 (○), 3 (□), 4 (△), and 5 (◇) classes.

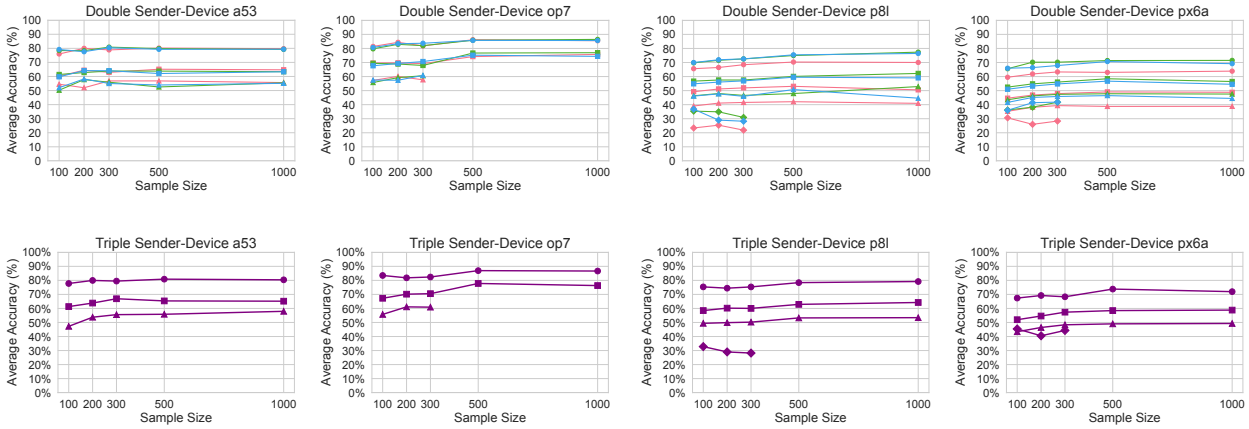


Figure 8: Accuracy trend plots per number of classes for two and three senders. The trend is rather steady and continuous as the sample sizes expand. We include BD, BV, DV, & BDV for 2 (○), 3 (□), 4 (△), & 5 (◇) classes.

merely delaying, rather than outright preventing, the execution of such attacks.

To further complicate the attacker's efforts in utilizing timing information, the implementation of adaptive jitter mechanisms introduces a more nuanced counterstrategy. These mechanisms, capable of introducing variable delays in SMS processing, adjust dynamically in response to fluctuating network conditions and traffic patterns. This adaptability ensures that networks can impede side-channel analysis through effective timing obfuscation. Nevertheless, considering the sophisticated strategy of attackers deploying multiple senders across different geographical locations and leveraging various networks, the effectiveness of previously mentioned countermeasures could be compromised. To address this, networks could adopt a multi-layered defense strategy that also considers the following methods:

1. **Geographic Analysis of Source:** Implement anomaly detection systems that not only monitor the frequency and pattern of messages but also analyze the geographic origins of SMS traffic. By identifying unusual patterns of messages coming from multiple locations (also through roaming) targeting a single number, the system can flag potential coordinated attacks.

2. **Adaptive Routing:** Dynamically alter the routing of messages based on real-time analysis to disrupt the timing measurements of attackers. This could involve randomizing the path messages take through the network or introducing variable delays for messages from identified suspicious sources and roaming.

3. **Joint Defense Initiatives:** Since the attacks can happen internationally from any location, it is imperative to establish shared intelligence on known attack patterns, including the use of multiple senders, across networks. Networks that work together can implement joint defense measures, such as coordinated blocking of attack sources and unified response strategies to emerging threats.

6.3 Limitations

In this work, we alleviated the problems of some limitations present in the location identification attack. First, the attacker is not constrained by one location only and can combine multiple sender measurements to significantly improve the model's accuracy. In addition, our sample size study showed that the attacker is not constrained by the data size in most cases, making the attack more efficient. The adversary has also the flexibility to choose the best-performing multi-sender technique per classification and is not restricted by one method only.

Despite the initial success of our experimentation, several challenges remain in multi-sender attacks. Firstly, while our study did not directly encounter coordination or resource challenges, expanding the attack to incorporate multiple senders may necessitate significant resources. This includes not only hardware but also logistical efforts to strategically position devices across various locations. Such expansion could substantially increase the complexity, cost, and effort required, potentially making the attack viable only for adversaries with substantial resources. Secondly, even though our experiments did not face any issues with anomaly detection systems, attacks conducted by multiple senders are more likely to be identified as anomalous, resembling patterns of spam or malicious activity more closely than those conducted by single senders. Lastly, our focus has largely been on closed-world scenarios, where the attacker has predefined knowledge of the victim's potential locations. The efficacy of multi-sender attacks in open-world scenarios, where the victim's location is unknown, remains less explored. We are planning to investigate these aspects of the attack in the future.

7 Related Work

Recent studies have increasingly focused on the exploitation of timing side-channel analysis for various security and privacy implications. Schnitzler et al. [27] explored the feasibility of distinguishing the location of message recipients in messenger applications using a technique based on timing differences, focusing on Internet infrastructure, similar to the concept examined by Bitsikas et al. [6] which was centered on cellular networks. This line of inquiry is part of a broader spectrum of research into timing side-channel analysis even across different web aspects, as evidenced by works such as Rasmussen et al. [23], Kohlbrenner et al. [15], Brumley et al. [7], and Goethem et al. [10], highlighting the versatility and risk of timing attacks in various online environments.

In the domain of cellular networks, a rich body of literature has methodically explored both active and passive techniques to localize cellular network users. Studies range from capturing specific identifiers to leveraging vulnerabilities within the network's paging messages and Radio Link Failure reports [12, 13, 17, 18, 29, 30]. The MAC layer and timing advance values have been investigated for their potential in enhancing localization accuracy [22, 26]. Notably, LTrack [16] demonstrated an improvement in localization accuracy to as precise as 20 meters, significantly enhancing tracking capabilities with minimal adversary involvement. Furthermore, Lakshmanan et al. [18] showed that by collecting data from the public scheduling channel and finding unique identifiers, one could trace a target's path with an accuracy of less than 1 kilometer.

Various SMS attacks have been demonstrated, exploiting vulnerabilities to extract sensitive user information or execute commands, as seen in the case of Simjacking [4] and studies on spamming, spoofing, DoS, and silent SMS in LTE networks [31]. Mulliner et al. [19] introduced a vulnerability analysis framework for monitoring unexpected smartphone behaviors leading to large-scale DoS attacks. Furthermore, audio call features have been explored for security applications, such as fingerprinting and anomaly detection to combat call redirection/hijacking. Techniques leveraging audio latency and network characteristics have been investigated, with notable examples including Sonar [20] and PinDr0p [5].

8 Conclusion

In this work, we explored various multi-sender techniques of the SMS location inference attack, which provide a substantial accuracy improvement compared to the single-sender approaches. Our results showed that the best-performing method for all devices, sample sizes, and number of classes was the multi-sender MMS method. Additionally, we performed an analysis on the effects of the sample size on the model's accuracy for single- and multi-sender attacks, which revealed that the attacker can leverage smaller sample sizes to conduct the attack saving measurement collection time, resources and reducing the possibility for detection. Finally, we re-examined the potential countermeasures with extra suggestions.

Acknowledgements

This work was supported by NSF grant 2144914, by UA Ruhr under the Research Alliance Ruhr program, and by the Center for Cyber Security at New York University Abu Dhabi (NYUAD). The authors would like to thank Michel Lang, Philipp Markert, and Lena Schnitzler for their help with data collection.

References

[1] 3GPP. Digital cellular telecommunications system (Phase 2+)(GSM); Universal Mobile Telecommunications System (UMTS); LTE; 5G; Support of SMS over IP networks; Stage 3. Technical Specification (TS) 24.341, 3rd Generation Partnership Project (3GPP), 05 2022. Version 17.1.0.

[2] 3GPP. Digital cellular telecommunications system (Phase 2+)(GSM); Universal Mobile Telecommunications System (UMTS); LTE; IP Multimedia Subsystem (IMS); Stage 2 . Technical Specification (TS) 23.228, 3rd Generation Partnership Project (3GPP), 05 2022. Version 17.3.0.

[3] 3GPP. Digital cellular telecommunications system (Phase 2+)(GSM); Universal Mobile Telecommunications System (UMTS); LTE; Use of Data Terminal Equipment - Data Circuit terminating Equipment (DTE - DCE) interface for Short Message Service (SMS) and Cell Broadcast Service (CBS). Technical Specification (TS) 27.005, 3rd Generation Partnership Project (3GPP), 04 2022. Version 17.0.0.

[4] Adaptive Mobile Security Limited. Simjacking. https://f.hubspotusercontent10.net/hubfs/8487362/Reports/AdaptiveMobile_Security_Simjacker_Technical_Paper_v1.01.pdf.

[5] Vijay A. Balasubramaniyan, Aamir Poonawalla, Mustaque Ahamad, Michael T. Hunter, and Patrick Traynor. Pindr0p: Using single-ended audio features to determine call provenance. In *Proceedings of the 17th ACM Conference on Computer and Communications Security*, CCS '10, page 109–120, New York, NY, USA, 2010. Association for Computing Machinery.

[6] Evangelos Bitsikas, Theodor Schnitzler, Christina Pöpper, and Aanjhan Ranganathan. Freaky leaky SMS: Extracting user locations by analyzing SMS timings. In *32nd USENIX Security Symposium (USENIX Security 23)*, pages 2151–2168, Anaheim, CA, August 2023. USENIX Association.

[7] David Brumley and Dan Boneh. Remote timing attacks are practical. In *12th USENIX Security Symposium (USENIX Security 03)*, Washington, D.C., August 2003. USENIX Association.

[8] Cloudmark. SMS spam overview — preserving the value of SMS texting. https://www.cloudmark.com/en/resources/white-papers/sms-spam-overview-preserving-value-sms-texting.

[9] Europol. Takedown of sms-based flubot spyware infecting android phones. https://www.europol.europa.eu/media-press/newsroom/news/takedown-of-sms-based-flubot-spyware-infecting-android-phones.

[10] Tom Van Goethem, Christina Pöpper, Wouter Joosen, and Mathy Vanhoef. Timeless timing attacks: Exploiting concurrency to leak secrets over remote connections. In *29th USENIX Security Symposium (USENIX Security 20)*, pages 1985–2002. USENIX Association, August 2020.

[11] GSM Association. Official Document NG.111 - SMS Evolution. Technical Specification (TS) 111-v2.0, GSM Association, 11 2020. Version 2.0.

[12] Byeongdo Hong, Sangwook Bae, and Yongdae Kim. GUTI reallocation demystified: Cellular location tracking with changing temporary identifier. In *25th Annual Network and Distributed System Security Symposium, NDSS 2018, San Diego, California, USA, February 18-21, 2018*. The Internet Society, 2018.

[13] Syed Rafiul Hussain, Mitziu Echeverria, Omar Chowdhury, Ninghui Li, and Elisa Bertino. Privacy attacks to the 4G and 5G cellular paging protocols using side channel information. In *26th Annual Network and Distributed System Security Symposium, NDSS 2019, San Diego, California, USA, February 24-27, 2019*. The Internet Society, 2019.

[14] Kaspersky. What is smishing and how to defend against it. https://www.kaspersky.com/resource-center/threats/what-is-smishing-and-how-to-defend-against-it.

[15] David Kohlbrenner and Hovav Shacham. On the effectiveness of mitigations against floating-point timing channels. In *USENIX Security Symposium*, 2017.

[16] Martin Kotuliak, Simon Erni, Patrick Leu, Marc Röschlin, and Srdjan Capkun. LTrack: Stealthy tracking of mobile phones in LTE. In *31st USENIX Security Symposium (USENIX Security 22)*, pages 1291–1306, Boston, MA, August 2022. USENIX Association.

[17] Denis Foo Kune, John Kölndorfer, Nicholas Hopper, and Yongdae Kim. Location leaks over the GSM air interface. In *19th Annual Network and Distributed System Security Symposium, NDSS 2012, San Diego, California, USA, February 5-8, 2012*. The Internet Society, 2012.

[18] Nitya Lakshmanan, Nishant Budhdev, Min Suk Kang, Mun Choon Chan, and Jun Han. A stealthy location identification attack exploiting carrier aggregation in cellular networks. In *30th USENIX Security Symposium (USENIX Security 21)*, pages 3899–3916. USENIX Association, August 2021.

[19] Collin Mulliner, Nico Golde, and Jean-Pierre Seifert. Sms of death: From analyzing to attacking mobile phones on a large scale. In *USENIX Security Symposium*, 2011.

[20] Christian Peeters, Hadi Abdullah, Nolen Scaife, Jasmine Bowers, Patrick Traynor, Bradley Reaves, and Kevin Butler. Sonar: Detecting ss7 redirection attacks with audio-based distance bounding. In *2018 IEEE Symposium on Security and Privacy (SP)*, pages 567–582. IEEE Computer Society, 05 2018.

[21] Christian Peeters, Christopher Patton, Imani N. S. Munyaka, Daniel Olszewski, Thomas Shrimpton, and Patrick Traynor. SMS OTP security (SOS): hardening SMS-based two factor authentication. In *ASIA CCS'22: ACM Asia Conference on Computer and Communications Security, Nagasaki, Japan, 30 May 2022 - 3 June 2022*, pages 2–16. ACM, 2022.

[22] Benjamin A Pimentel. *Passive Geolocation in a 4G WIMAX Single Base Station Scenario*. Phd thesis, Naval Postgraduate School, Monterey California, 2013.

[23] Kasper Bonne Rasmussen and Srdjan Capkun. Location privacy of distance bounding protocols. *Proceedings of the 15th ACM conference on Computer and communications security*, 2008.

[24] Bradley Reaves, Nolen Scaife, Dave Tian, Logan Blue, Patrick Traynor, and Kevin R. B. Butler. Sending out an SMS: characterizing the security of the SMS ecosystem with public gateways. In *IEEE Symposium on Security and Privacy, SP 2016, San Jose, CA, USA, May 22-26, 2016*, pages 339–356. IEEE Computer Society, 2016.

[25] Bradley Reaves, Luis Vargas, Nolen Scaife, Dave Tian, Logan Blue, Patrick Traynor, and Kevin R. B. Butler. Characterizing the security of the SMS ecosystem with public gateways. *ACM Trans. Priv. Secur.*, 22(1):2:1–2:31, 2019.

[26] John D. Roth, Murali Tummala, John C. Mceachen, and James W. Scrofani. On location privacy in LTE networks. *IEEE Transactions on Information Forensics and Security*, 12:1358–1368, 2017.

[27] Theodor Schnitzler, Katharina Kohls, Evangelos Bitsikas, and Christina Pöpper. Hope of Delivery: Extracting User Locations From Mobile Instant Messengers. In *Network and Distributed System Security Symposium*, NDSS '23, San Diego, CA, USA, February 2023. The Internet Society.

[28] Security Affairs. After simjacker, wibattack hacking technique disclosed. billions of users at risk. https://securityaffairs.co/wordpress/91800/hacking/wibattack-sim-attack.html.

[29] Altaf Shaik, Jean-Pierre Seifert, Ravishankar Borgaonkar, N. Asokan, and Valtteri Niemi. Practical attacks against privacy and availability in 4G/LTE mobile communication systems. *ArXiv*, abs/1510.07563, 2016.

[30] Ankush Singla, Syed Rafiul Hussain, Omar Chowdhury, Elisa Bertino, and Ninghui Li. Protecting the 4G and 5G cellular paging protocols against security and privacy attacks. *Proc. Priv. Enhancing Technol.*, 2020(1):126–142, 2020.

[31] Guan-Hua Tu, Chi-Yu Li, Chunyi Peng, Yuanjie Li, and Songwu Lu. New security threats caused by IMS-based SMS service in 4G LTE networks. In *Proceedings of the 2016 ACM SIGSAC Conference on Computer and Communications Security*, CCS '16, page 1118–1130, New York, NY, USA, 2016. Association for Computing Machinery.