Freaky Leaky SMS: Extracting User Locations by Analyzing SMS Timings

Evangelos Bitsikas (Northeastern University), Theodor Schnitzler (TU Dortmund), Christina Pöpper (New York University Abu Dhabi), Aanjhan Ranganathan (Northeastern University)

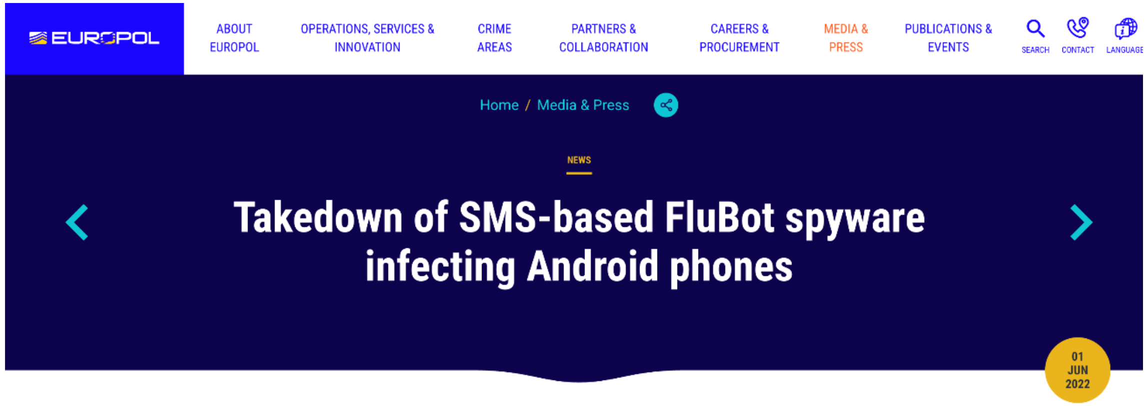
32nd USENIX Security Symposium 2023, Anaheim, CA, USA

Northeastern University
Khoury College of
Computer Sciences

technische universität
dortmund

جامعـة نيويورك ابـوظـبي
NYU ABU DHABI

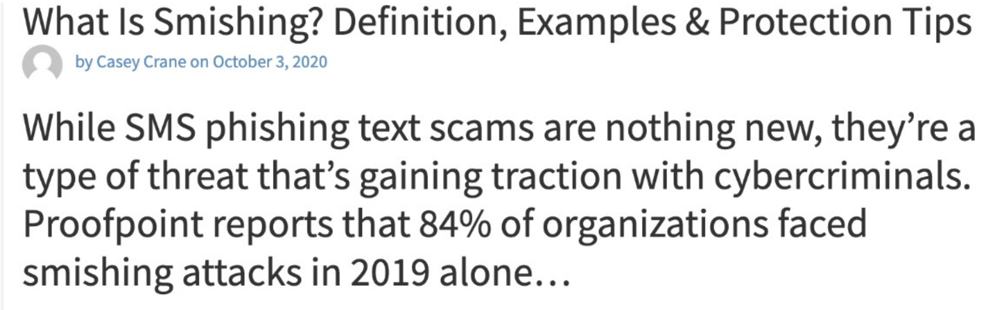
Introduction: SMS Insecurity



NEWS

Takedown of SMS-based FluBot spyware infecting Android phones

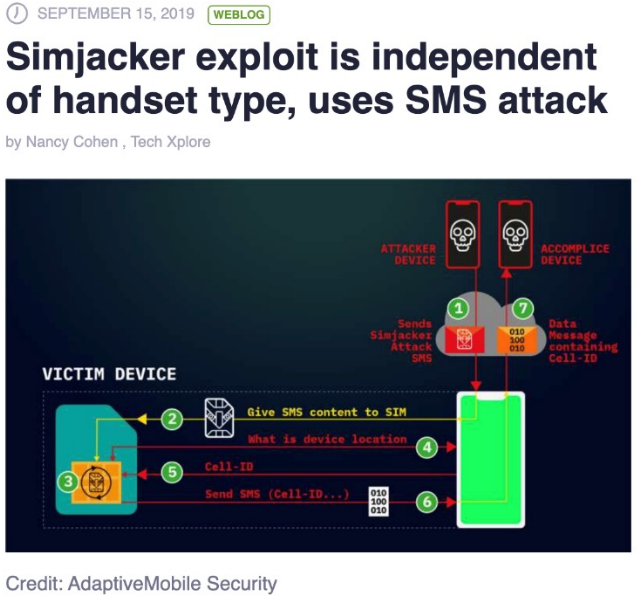
01 JUN 2022



What Is Smishing? Definition, Examples & Protection Tips

by Casey Crane on October 3, 2020

While SMS phishing text scams are nothing new, they're a type of threat that's gaining traction with cybercriminals. Proofpoint reports that 84% of organizations faced smishing attacks in 2019 alone…

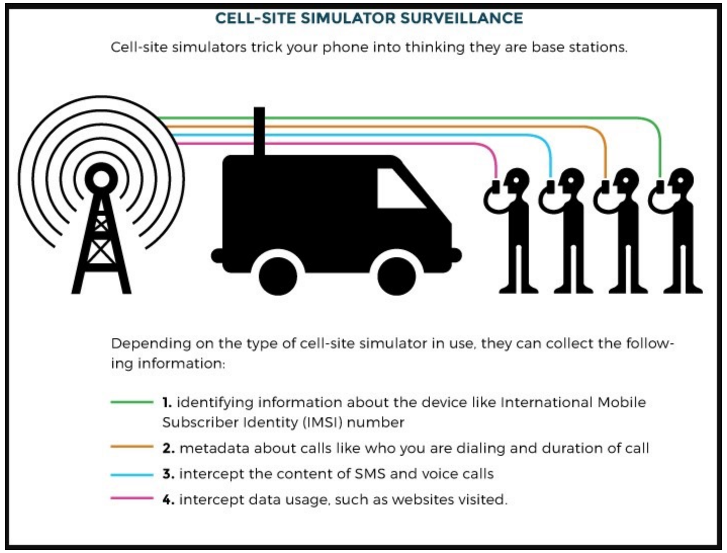


SEPTEMBER 15, 2019 WEBLOG

Simjacker exploit is independent of handset type, uses SMS attack

by Nancy Cohen , Tech Xplore

Credit: AdaptiveMobile Security



CELL-SITE SIMULATOR SURVEILLANCE

Cell-site simulators trick your phone into thinking they are base stations.

Depending on the type of cell-site simulator in use, they can collect the following information:

1. identifying information about the device like International Mobile Subscriber Identity (IMSI) number
2. metadata about calls like who you are dialing and duration of call
3. intercept the content of SMS and voice calls
4. intercept data usage, such as websites visited.

ETSI TS 123 040 V17.3.0 (2023-07)

TECHNICAL SPECIFICATION

**Digital cellular telecommunications system (Phase 2+) (GSM);
Universal Mobile Telecommunications System (UMTS);
LTE;
5G;
Technical realization of the Short Message Service (SMS)
(3GPP TS 23.040 version 17.3.0 Release 17)**

Contributions & Goal

Objective:

Identify the location of the SMS recipient-victim any time worldwide.

High-Level Logic:

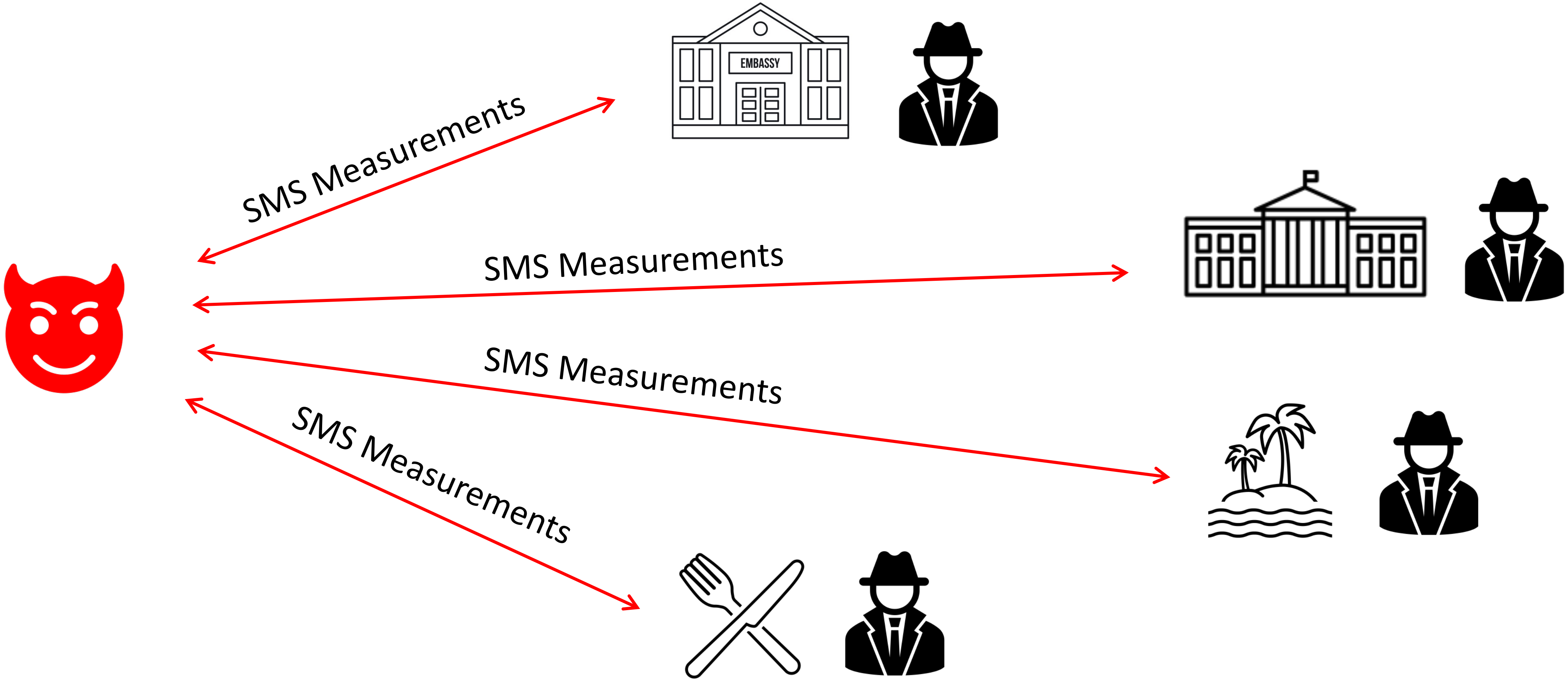
1. Know the routinely locations and mobile number of the victim.
2. Send silent SMSs and receive acknowledgements and delivery reports.
3. Use the SMS timings to generate fingerprints per location.
4. Use the fingerprints to predict the location of the victim using ML techniques.

Main Contributions:

- Unique and stealthy location identification attack based on the SMS infrastructure.
- Large scale evaluation: 3 continents, 9 countries, 10 operators, and 16 devices.
- The attack can currently achieve up to 96% accuracy for international. classifications, and over 70% more for many national/regional classifications.
- Countermeasures against the SMS location inference attack.

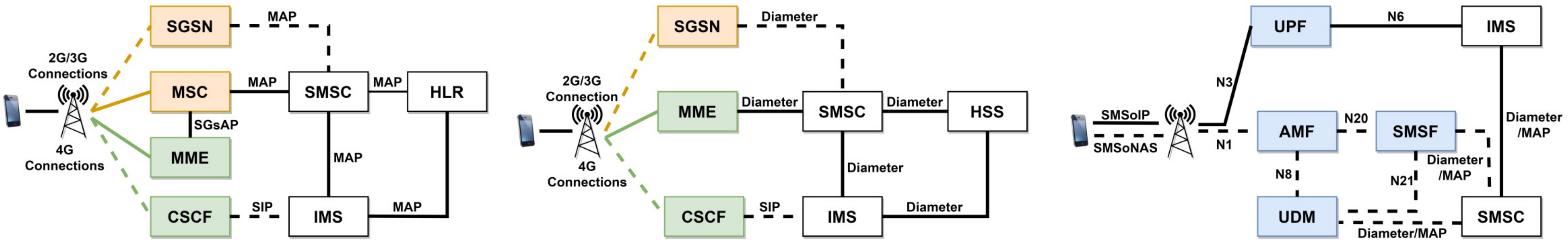
Use Case Example

Tracking the diplomat to routinely locations



Network Architecture

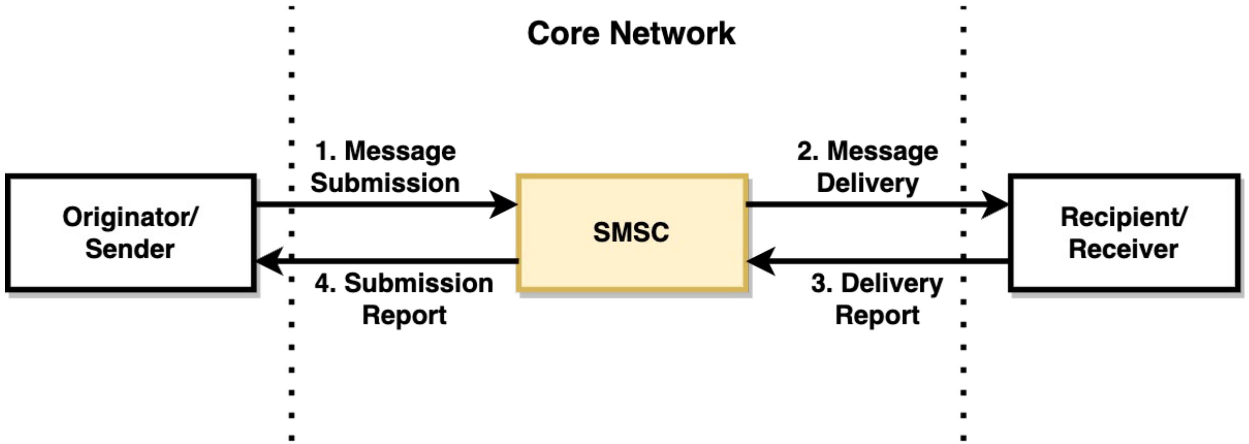
Freaky Leaky SMS: Extracting User Locations by Analyzing SMS Timings
Evangelos Bitsikas, Theodor Schnitzler, Christina Pöpper, Aanjhan Ranganathan
32nd USENIX Security Symposium 2023, Anaheim, CA, USA



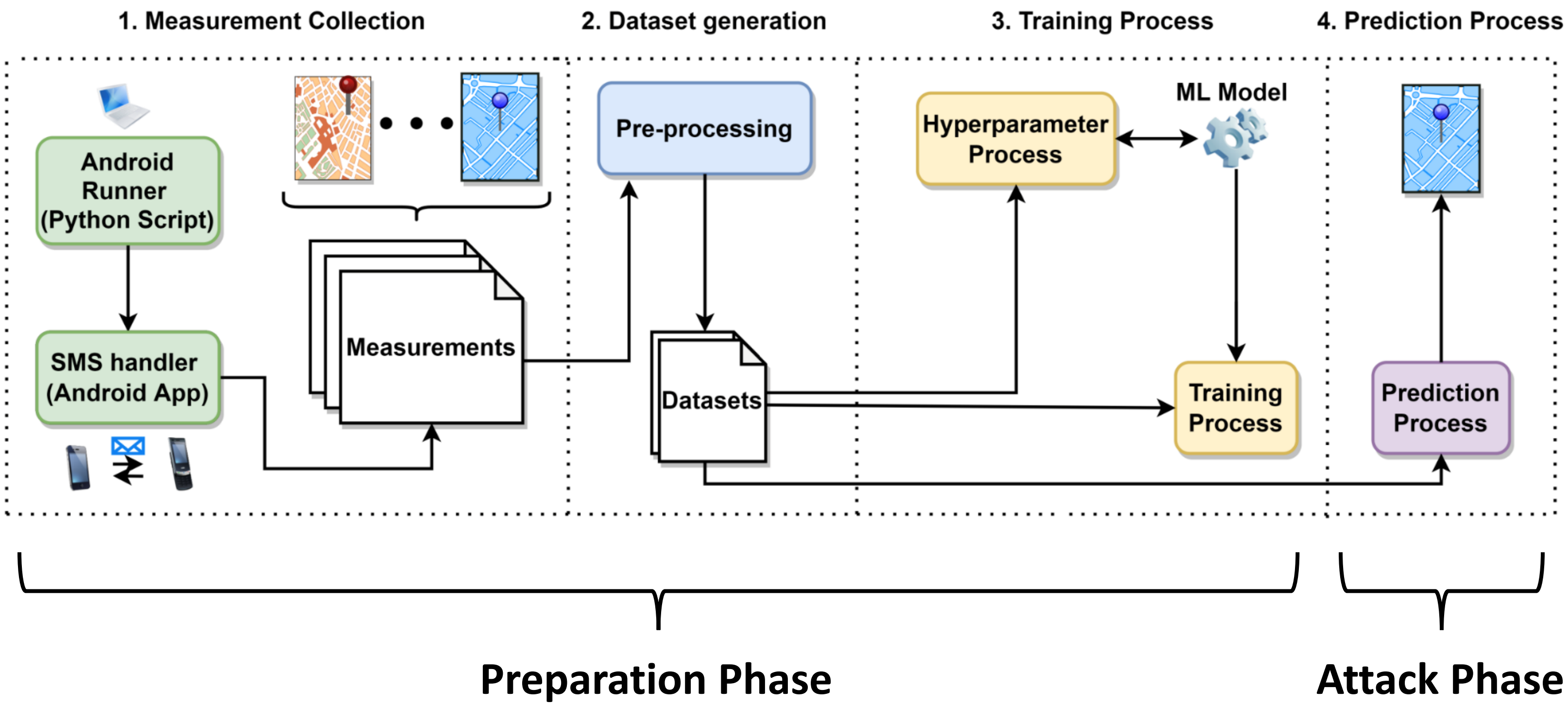
(a) 2G/3G/4G with MAP and IMS

(b) 2G/3G/4G with Diameter and IMS

(c) 5G Standalone with IMS and NAS



Attack Process & Setup



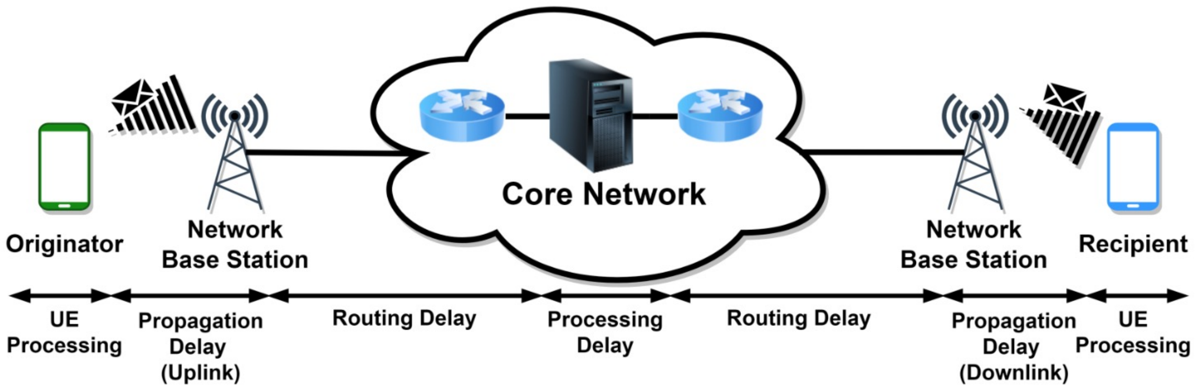
Device Types:

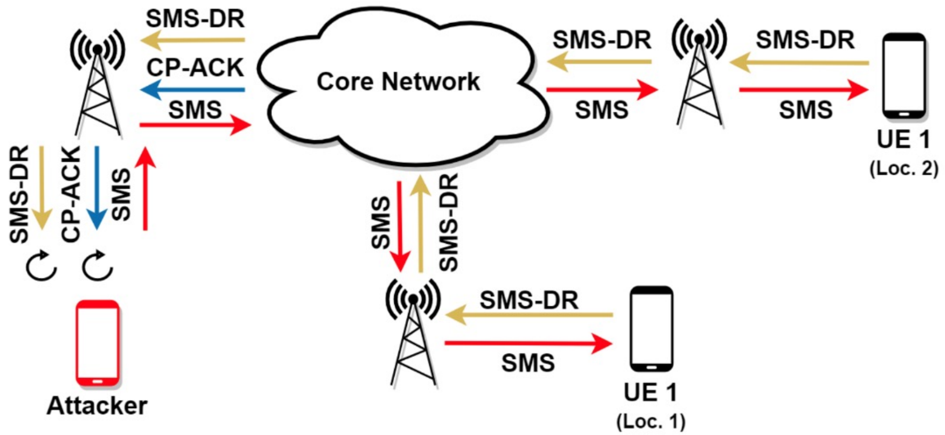
- *Active* (used for disseminating messages)
- *Passive* (receiving messages only at various locations)

Location Types:

- Fixed Position
- Area (includes fixed positions)

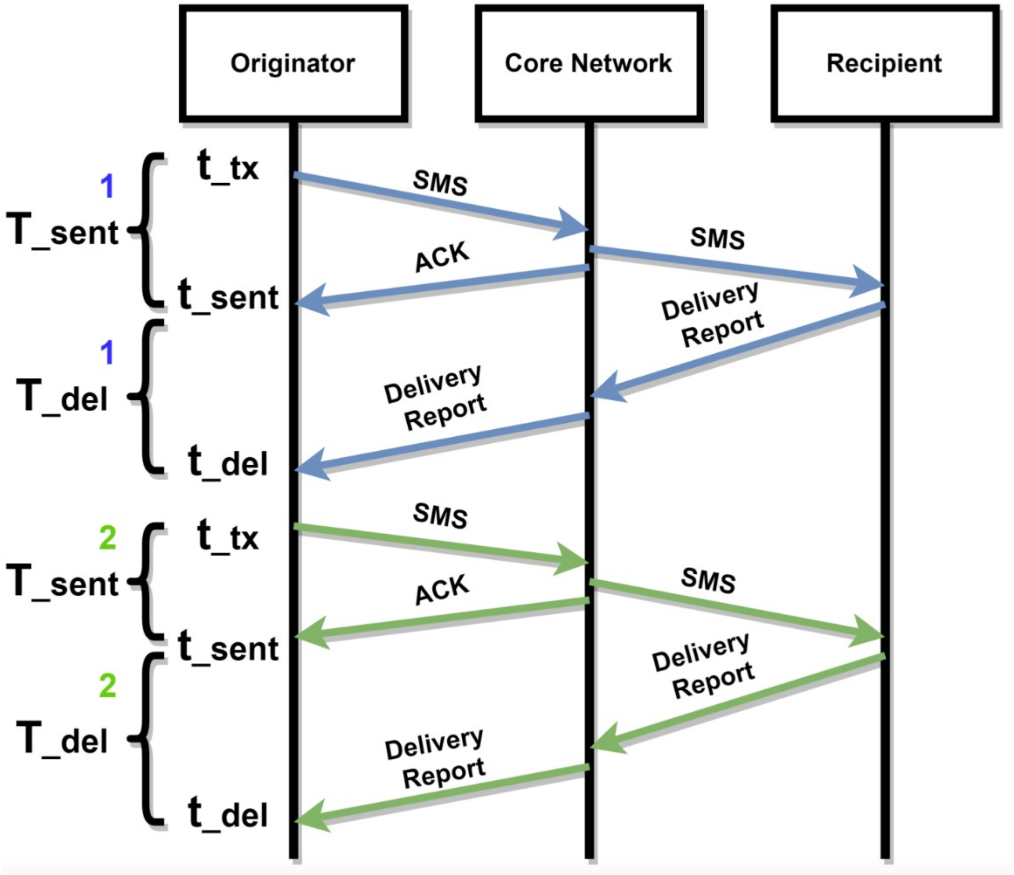
Measurement Collection





- *SMS burst:* 20 silent SMSs per hour (continuously).
- Various times of the day, network configurations, and levels of network loads.
- Locations in GR, DE, DK, UK, US, AE, NL, BE, LU.
- Connection Types: LTE, LTE+, 5G NSA/SA
- Routing Modes: SMSoIP, SGsAP/Diameter
- Approximately 155,512 SMSs in total.

SMS Timings Features



Timing Features

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

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

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

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

SMS specific

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

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

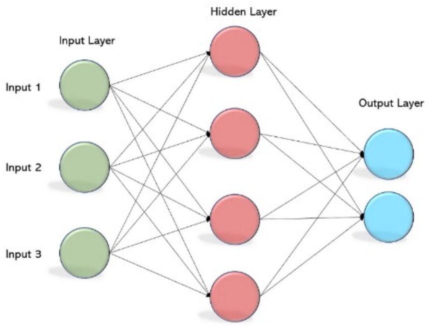
Pattern specific

The ***location signature/fingerprint*** is a combination of these six features: (T_{sent}, T_{del}, T_{tot}, P, $T_{\Delta sent}$, $T_{\Delta del}$)

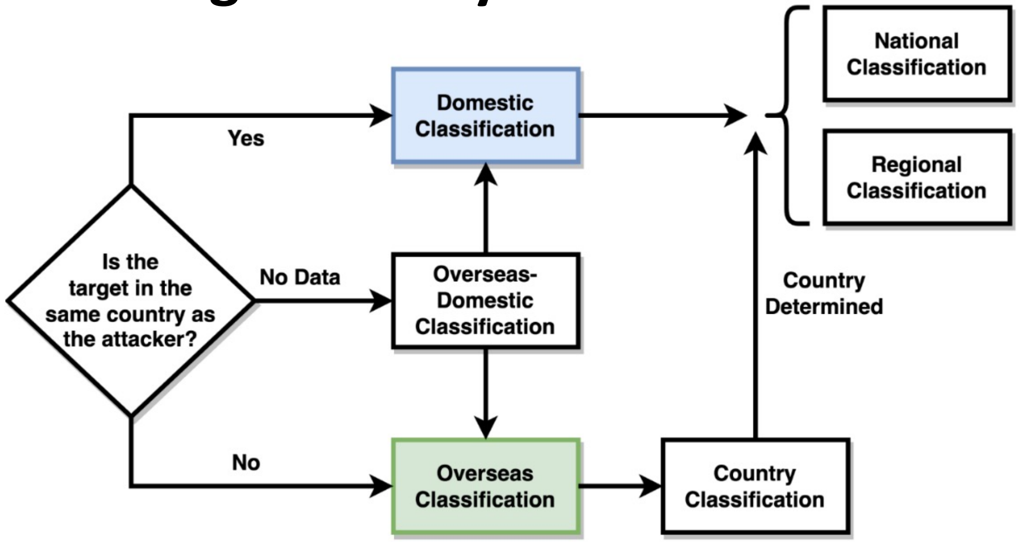
ML Training & Prediction

Multi Layer Perceptron (MLP) Neural Network

- Manual & Automatic Hyperparameter tuning
- Stochastic gradient descent solver
- SoftMax and Sigmoid activations
- Three layers of 10, 40, 10
- Maximum iterations: 5000
- Constant learning rate
- Batch size: 32
- Alpha: 0.0001



What about location granularity? 🤔



What's next?

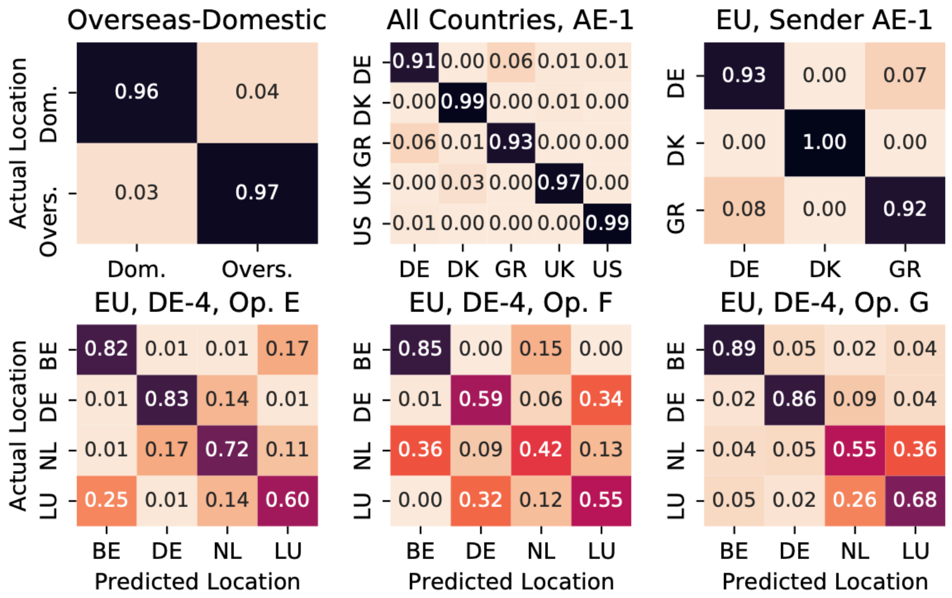
The model is trained based on fingerprints of each location.

The attacker sends new SMS messages and generates the timing features.

The new timing features are fed into the model to predict the current location.

Results: International

Classification	Size/Class	Operators	Receiver Locations	Sender Location	Accuracy
Overseas-vs.-Domestic	1200	A, C, E, H, I, J	AE-X, Int-X	AE-1	96%
All Country-based	280	C, E, H, I, J	Int-X	AE-1	96%
EU Country-based	280	C, E, I	Int-GR, Int-DE, Int-DK	AE-1	95%
EU Country-based	257	G	DE-4, NL-4, BE-1, LU-1	DE-4	75%
EU Country-based	319	E	DE-4, NL-4, BE-1, LU-1	DE-4	74%
EU Country-based	313	F	DE-4, NL-4, BE-1, LU-1	DE-4	62%



Overseas-Domestic

	Dom.	Overs.
Dom.	0.96	0.04
Overs.	0.03	0.97

All Countries, AE-1

	DE	DK	GR	UK	US
DE	0.91	0.00	0.06	0.01	0.01
DK	0.00	0.99	0.00	0.01	0.00
GR	0.06	0.01	0.93	0.00	0.00
UK	0.00	0.03	0.00	0.97	0.00
US	0.01	0.00	0.00	0.00	0.99

EU, Sender AE-1

	DE	DK	GR
DE	0.93	0.00	0.07
DK	0.00	1.00	0.00
GR	0.08	0.00	0.92

EU, DE-4, Op. E

	BE	DE	NL	LU
BE	0.82	0.01	0.01	0.17
DE	0.01	0.83	0.14	0.01
NL	0.01	0.17	0.72	0.11
LU	0.25	0.01	0.14	0.60

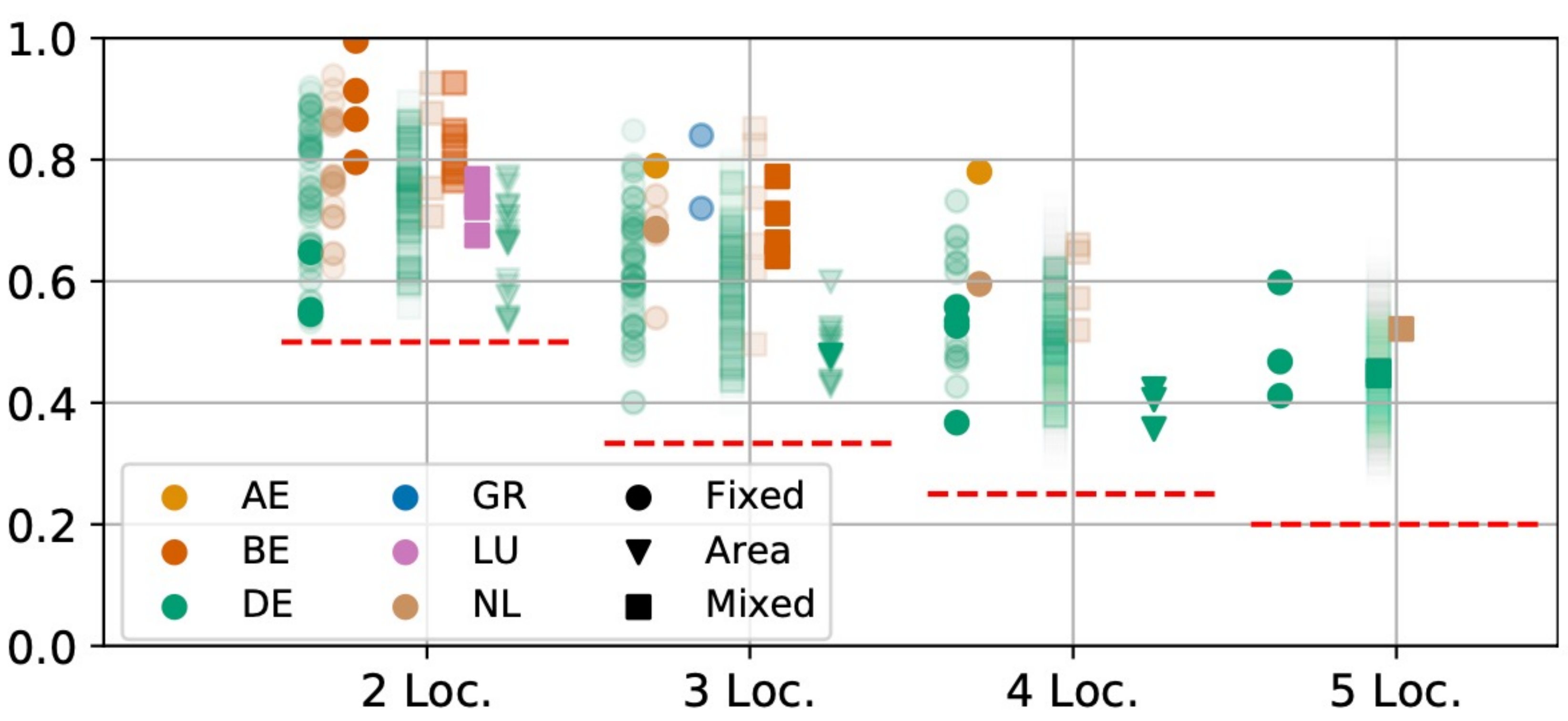
EU, DE-4, Op. F

	BE	DE	NL	LU
BE	0.85	0.00	0.15	0.00
DE	0.01	0.59	0.06	0.34
NL	0.36	0.09	0.42	0.13
LU	0.00	0.32	0.12	0.55

EU, DE-4, Op. G

	BE	DE	NL	LU
BE	0.89	0.05	0.02	0.04
DE	0.02	0.86	0.09	0.04
NL	0.04	0.05	0.55	0.36
LU	0.05	0.02	0.26	0.68

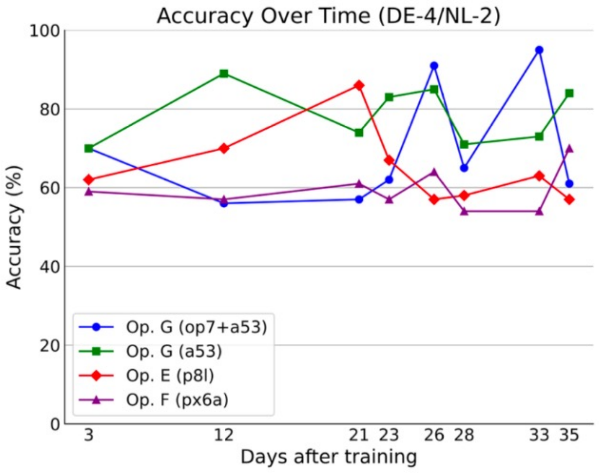
Results: National/Regional



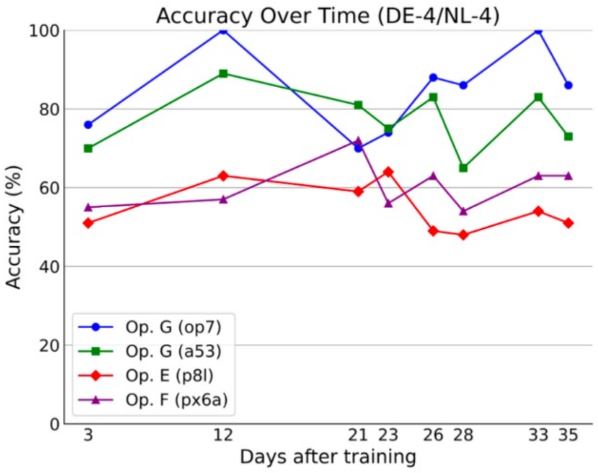
Receiver Locations	Accuracy
Sender Location: DE-4, Operator E	
BE-1, BE-2	83 %
BE-1, BE-3	80 %
BE-2, BE-3	74 %
LU-1, LU-3	64 %
Sender Location: DE-4, Operator F	
BE-1, BE-2	95 %
BE-1, BE-3	72 %
BE-2, BE-3	80 %
LU-1, LU-3	66 %
Sender Location: DE-4, Operator G	
BE-1, BE-2	86 %
BE-1, BE-3	84 %
BE-2, BE-3	84 %
LU-1, LU-3	72 %

Additional Insights

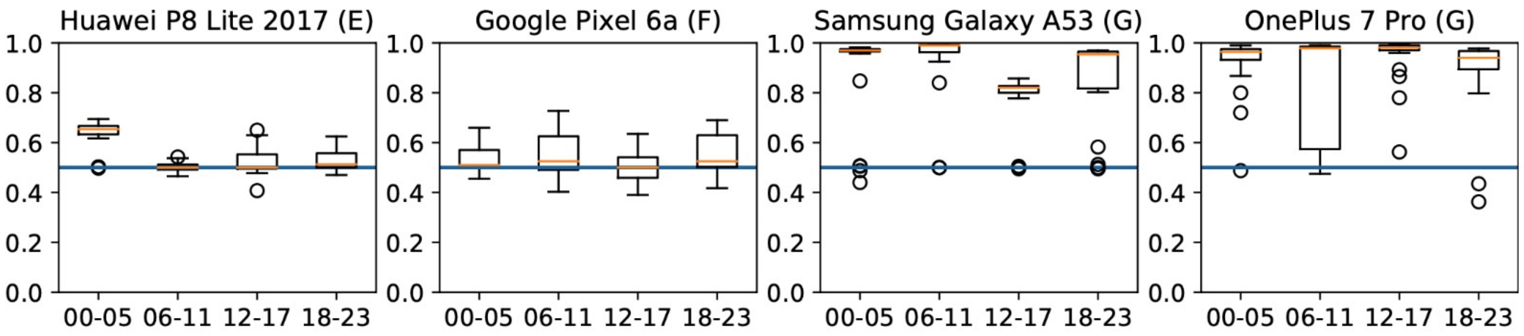
Stability

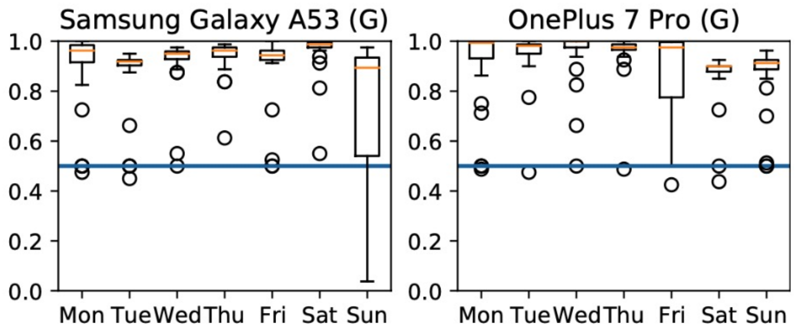


(a) DE4-NL2



(b) DE4-NL4





The attacker does ***not*** need to know the device manufacturer!

Advantages & Limitations

Advantages

- Low equipment requirements and cost (e.g., no false base stations)
- Exploiting the existing and ubiquitous SMS infrastructure
- No internet access is needed, only a mobile number
- System automation, low manual effort
- Stealthy by using silent SMS
- High accuracy in many cases

Limitations

- Less accurate for location granularity below 1-2 Km
- Adaptation to open-world scenarios might be limited
- ML techniques cannot perform completely correct in all cases

Countermeasures

- ✓ **Rejecting/Dropping Silent SMS at the Core Network** 🛡
- ✓ **More Robust Spamming/Flooding Filters** 🛡
- ✓ **Artificial Random Delays for the Delivery Report** 🛡🛡
- ✓ **Total Elimination of the Delivery Reports** 🛡🛡🛡

Takeaway Points

- SMS location identification is possible, but it is a complex problem (with network and human aspects).
- It applies to various devices, networks and location granularities.
- It can have worldwide application and be stealthy.
- More resources, manpower and ML experience means more impactful attacks.

GSMA Mobile Security Research Acknowledgements* under *CVD-2023-0072

GitHub

Longer Version

Thank You!
Questions?

Evangelos Bitsikas
bitsikas.e@northeastern.edu

Northeastern University
Khoury College of Computer Sciences

technische universität
dortmund

جامعة نيويورك ابوظبي
NYU ABU DHABI