FraudZen Hackathon

Advancing Mobile Network Security, Privacy, and Fraud Detection

November 18th, 2024
UY1, Yaounde, Cameroon

Anne Josiane Kouam - TU Berlin

Collaborators:
Alain Tchana - Grenoble INP
Aline Carneiro Viana - INRIA
Hippolyte Tapamo - UY1
Konrad Rieck - TU Berlin

Presentation Outline

1. **Context**: FraudZen & En-WDM
2. **Hackathon:** Directives and Goals
3. **Practical** organization

Context: FraudZen & En-WDM

Context: Frauds in cellular networks

Cellular networks are

14.9 Billion in 2021
18.2 Billion in 2025

mobile devices [1]

[1] Forecast number of mobile devices worldwide from 2020 to 2025 (in billions). Statista. 2023
[2] CFCA. CFCA 2021 Fraud Loss Survey. Report. 2021.

Context: Frauds in cellular networks

Cellular networks are

14.9 Billion in 2021
18.2 Billion in 2025

mobile devices [1]

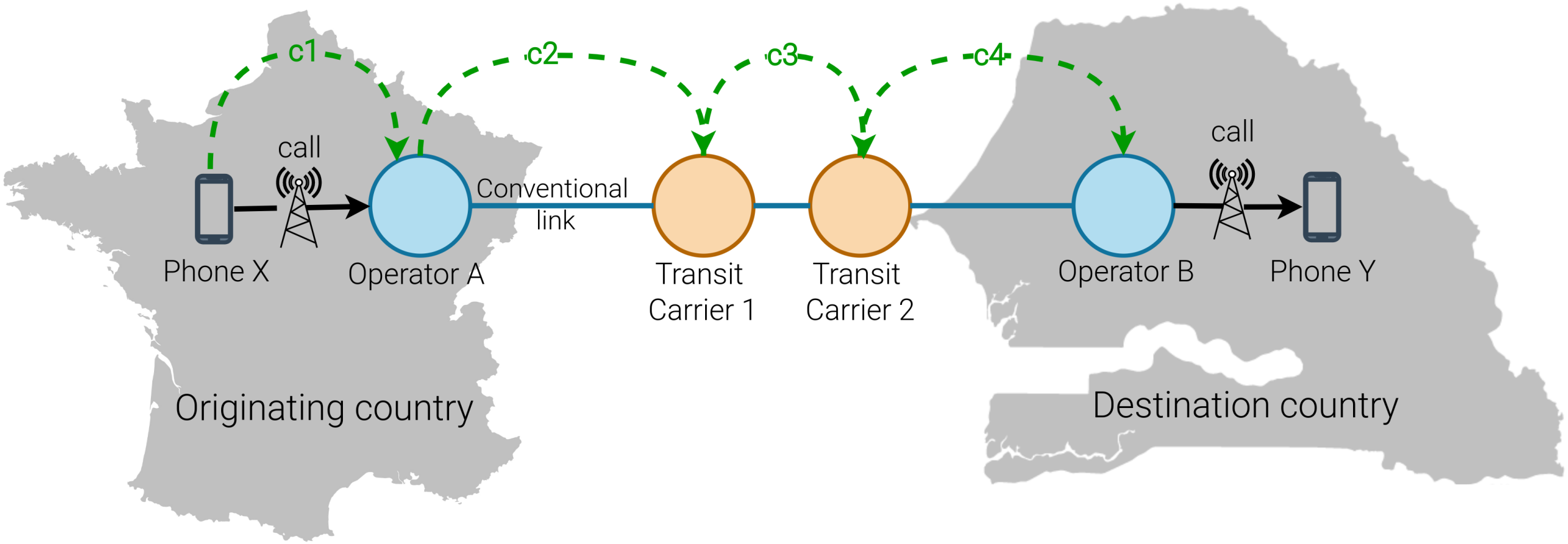
$39.89 billion

mobile operators' annual losses [2]

[1] Forecast number of mobile devices worldwide from 2020 to 2025 (in billions). Statista. 2023
[2] CFCA. CFCA 2021 Fraud Loss Survey. Report. 2021.

Context: International bypass frauds

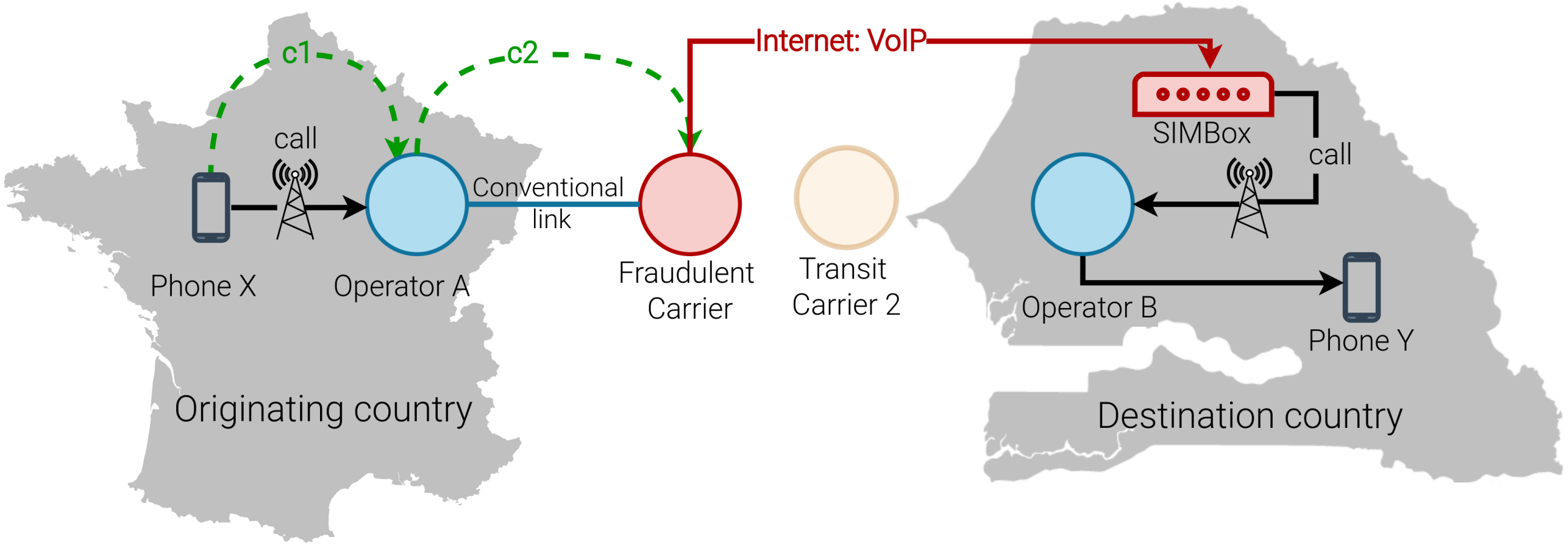
Top 4 most severe phone system frauds [1]: SIMBox fraud



[1] CFCA. CFCA 2021 Fraud Loss Survey. Report. 2021.

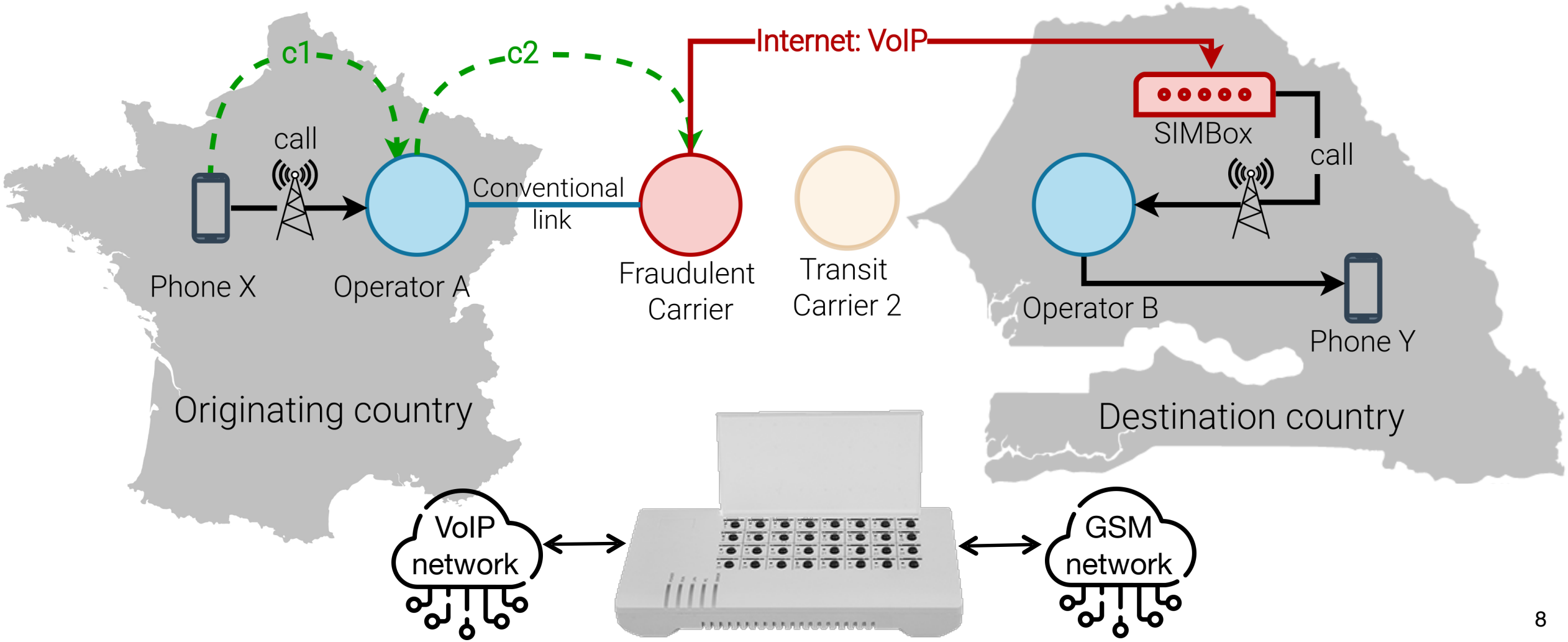
Context: International bypass frauds

Top 4 most severe phone system frauds: SIMBox fraud



Context: International bypass frauds

Top 4 most severe phone system frauds: SIMBox fraud



Context: International bypass frauds

Top 4 most severe phone system frauds, **or even more**

Negative impact:

- **Financial:** $3.11 billion revenue loss annually [1]
- **Network quality:** Poor QoE for network users
- **Privacy:** Phone conversations eavesdropping
- **National security:** International espionage, Facilitating Terrorism operations
- **Research:** Bias to cellular network datasets

[1] CFCA. CFCA 2021 Fraud Loss Survey. Report. 2021.

Bypass fraud mitigation

Where?
At the destination operator

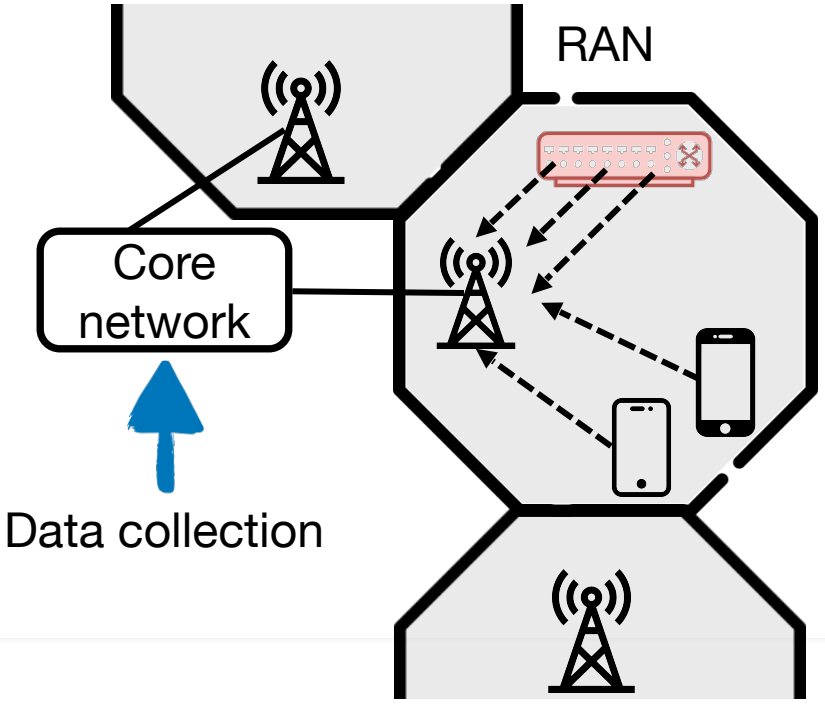
What?
Distinguish legitimate user (IMSI) or device (IMEI) from fraudulent ones

How?
Analysis of cellular network traces

Call Audio Traces (2/15) - Signaling Data (1/15) - Charging Data Records (12/15)

Bypass fraud mitigation

Call Audio Traces (2/15) - Signaling Data (1/15) - **Charging Data Records (12/15)**



Time	Event	Caller ID	Called ID	Call Duration	Data Volume	CellID
12:00:05	CALL	09879	09997	10 min	0	lat=12.5 lon=72.8
12:15:12	DATA	00779	/	0	20 MB	lat=12.4 lon=73.5
12:30:00	SMS	09875	09879	0	0	lat=12.0 lon=73.5

Classification: Human vs Automated behavior

Limitation:

Human Behavior Simulation

Bypass fraud mitigation

SIMBox fraud evolution

- No unique SIMBox fraud
- Missing: "Which type of SIMBox fraud is tackled?"
- Detection efficiency is restrained to an undisclosed context
- No takeaway: relationship btw detection design and tackled fraud

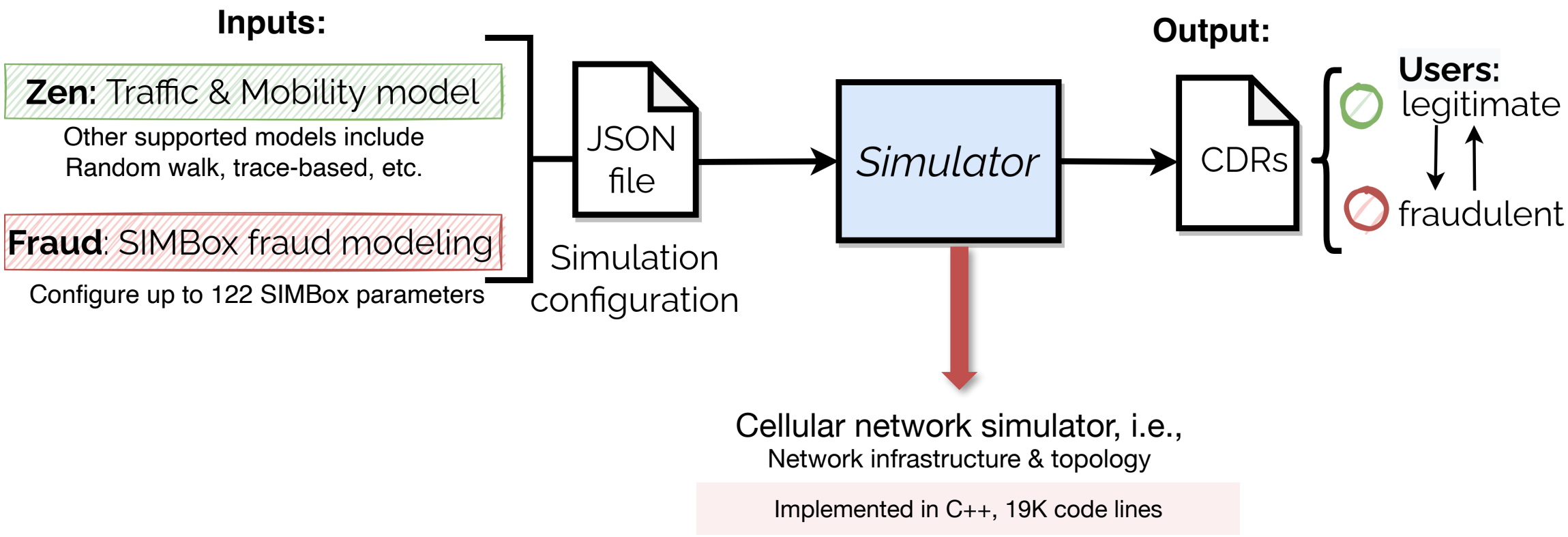
Positioning

Acknowledgment of the evolutive nature of the fraud through bypass frauds modeling

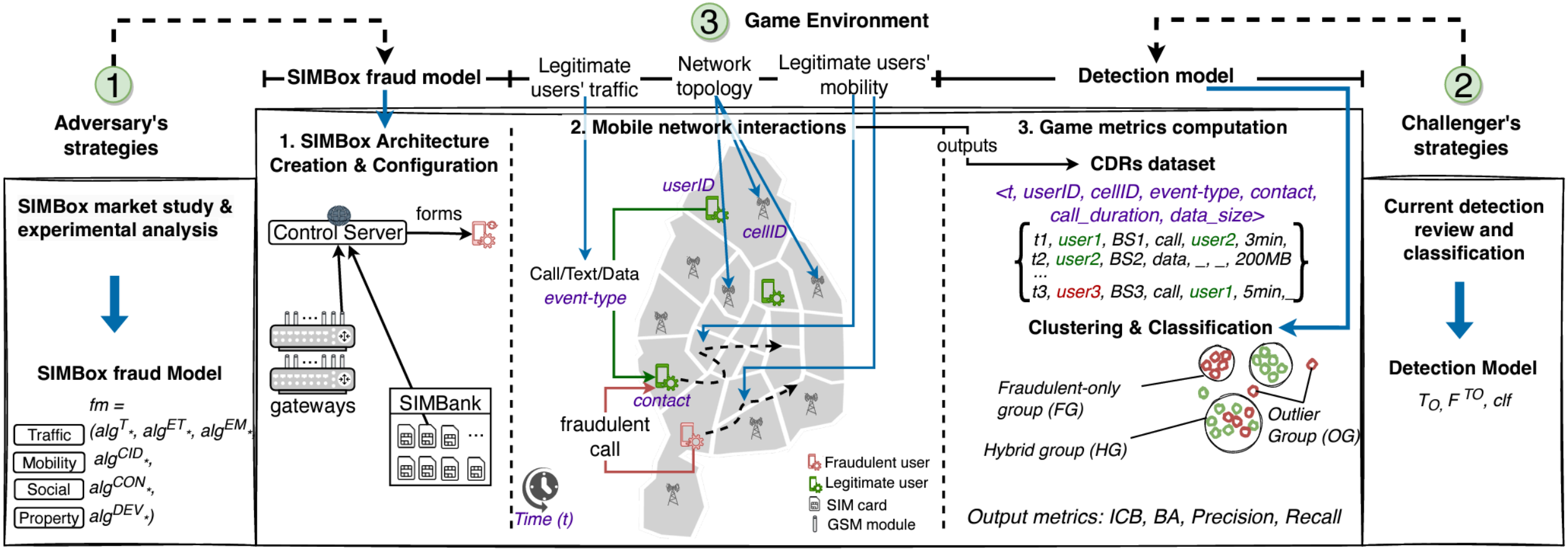
Charging Data Records (12/15)

Venue, Year	Reported Avg. Accuracy
PhD Thesis, 2020	95.55%
Annals of Telecommunications, 2020	
ComTech, 2019	99.3%
Journal of Pure & Applied Science, 2019	99.9%
PhD Thesis, 2018	83.2%
ArXiv, 2017	83.34%
Master Thesis, 2016	No evaluation
Tech. Rep., 2015	99.99%
ICCVIAA, 2015	No evaluation
Jurnal Teknologi, 2014	98.8%
INFOCOM, 2014	99.95%
ITCS, 2013	98.71%

Modeling bypass frauds mitigation



FraudZen architecture

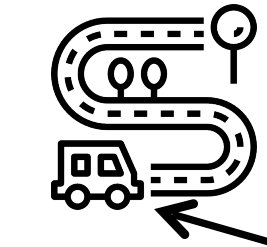


Cellular network simulator, i.e.,
Network infrastructure & topology

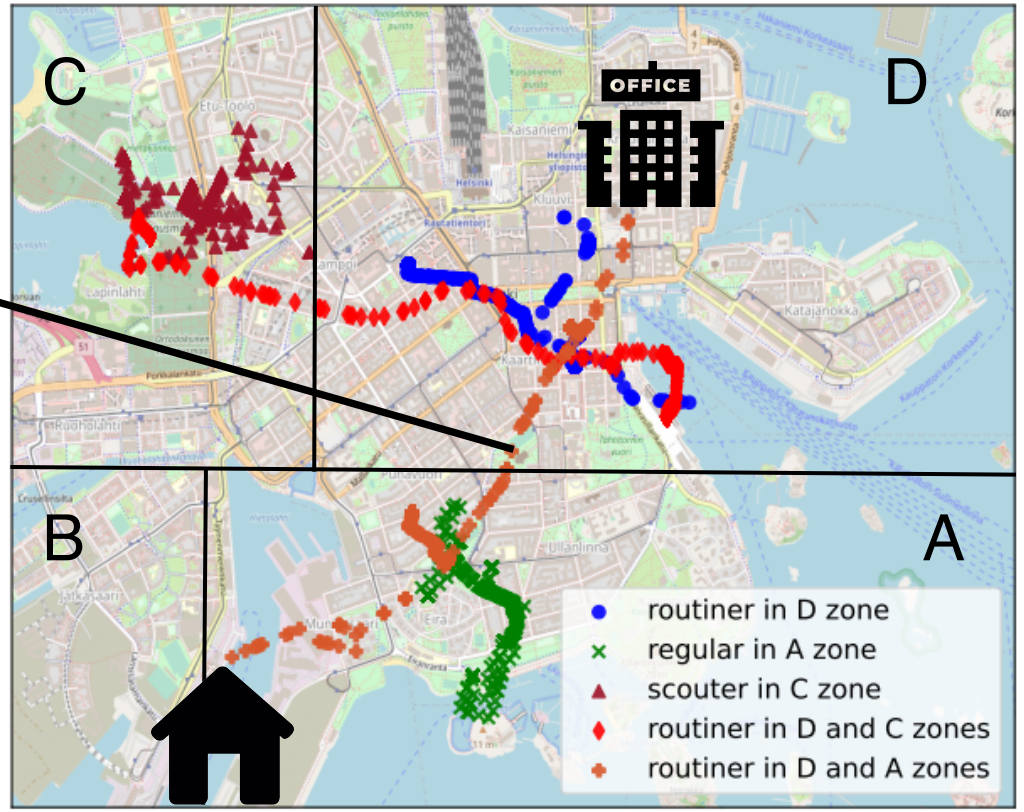
Implemented in C++, 19K code lines

Enhanced Working Day Mobility Model [1]

- Baseline : Working Day Mobility Model*
- WDM => En-WDM
  - Home & work clusters + Popularity
  - Exploration Profiling : *routiners/ regulars/scouters*
  - Displacement based profiling
  - Realistic parameterization : bus schedule, probability to have a car, ...



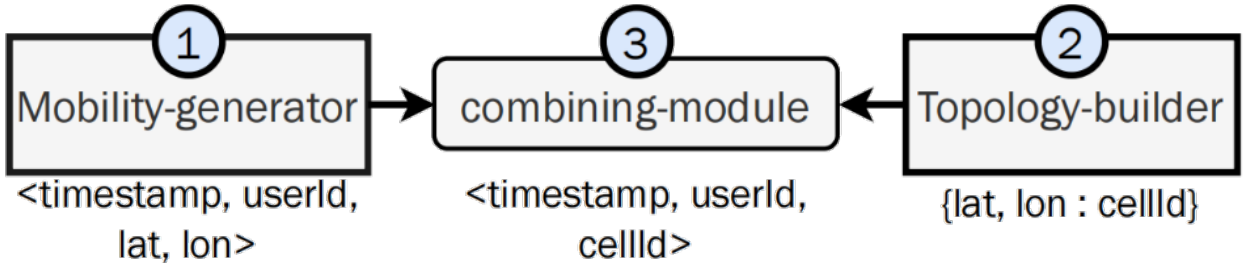
Displacements : bus/walk/car



[1] F. Ekman, A. Keränen, J. Karvo, and J. Ott. 2008. Working day movement model. In 1st ACM SIGMOBILE workshop on Mobility models (MobilityModels '08).

Enhanced Working Day Mobility Model [1]

- Baseline : Working Day Mobility Model*
- WDM => En-WDM
  - Home & work clusters + Popularity
  - Exploration Profiling : *routiners/ regulars/scouters*
  - Displacement based profiling
  - Realistic parameterization : bus schedule, probability to have a car, ...



[1] F. Ekman, A. Keränen, J. Karvo, and J. Ott. 2008. Working day movement model. In 1st ACM SIGMOBILE workshop on Mobility models (MobilityModels '08).

Hackathon: Directives & Goals

FraudZen Hackathon

Publication process in academia

1. Researcher find an interesting and open challenge and works on it
2. They write a paper reporting the research results
3. Paper is submitted to conference
4. Paper is reviewed by other researchers (peer-reviewed)
5. In case paper gets accepted…
6. It is presented at the conference

With **topical** research challenges related to your field

FraudZen Hackathon

4 Research poles

Mobile Network architecture and Knowledge Extraction

Privacy protection through Anonymization

Offensive Security

Defensive Security

FraudZen Hackathon

Your tasks

1. Pick a pole from those presented on which you want to work
2. … propose a project aligned with the pole's topics
3. … review the literature of your project and refine ideas
4. … work on the given project and get preliminary results
5. … write a paper on the preliminary results (4 pages)
6. … submit the paper to me
7. … **To a real conference** (17 January 2025)

FraudZen Hackathon

Your tasks - Timeline

18.11.2024	Pole Assignment
25.11.2024	Project idea submission and discussion
10.01.2025 23h59	Submit paper
17.01.2025	[For 5 best papers] Submission to Algotel&Cores
25.02.2025	Closing meeting & Final presentation
14.03.2025	Algotel&Cores Notification

FraudZen Hackathon

Mobile Network Architecture and Knowledge Extraction

Exploring network architecture and data dynamics to enhance mobility and traffic simulations.

- **Object of Study:** Mobile network modeling and simulator; no fraud simulation required
- **Projects guiding lines**
  - Integrate 4G/5G/6G features into FraudZen for improved mobile network simulation
  - Extract insights from user data on social, mobility, and traffic patterns
  - Propose and validate new human mobility models or contexts within En-WDM or FraudZen
  - Develop and integrate new traffic models into the simulator (e.g., FraudZen)
- **Starting point(s):**

- Anne Josiane Kouam, Aline Carneiro Viana, and Alain Tchana. 2023. Zen: LSTM-based generation of individual spatiotemporal cellular traffic with interactions. arXiv preprint arXiv:2301.02059. https://arxiv.org/abs/2301.02059
- 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). Association for Computing Machinery, New York, NY, USA, 33–40. https://doi.org/10.1145/1374688.1374695
- Eduardo Mucceli, Aline Carneiro Viana, Carlos Sarraute, Jorge Brea, and José Ignacio Alvarez-Hamelin. 2016. On the Regularity of Human Mobility. Pervasive and Mobile Computing (Dec. 2016). https://inria.hal.science/hal-01367825
- Eduardo Mucelli Rezende Oliveira, Aline Carneiro Viana, Kolar Purushothama Naveen, and Carlos Sarraute. 2015. Measurement-driven mobile data traffic modeling in a large metropolitan area. In PerCom 2015. https://inria.hal.science/hal-01089434

FraudZen Hackathon

Privacy Protection through Anonymization

Applying techniques to protect user identities in network data.

- **Object of Study:** Regular mobile datasets; no fraud simulation required.
- **Projects guiding lines**
    - Implement privacy techniques to obscure user identities while addressing dataset challenges.
    - Investigate privacy attacks on pseudonymized datasets for potential leaks.
    - Propose strategies to balance privacy and data usability.
- **Starting points:**
    - F. Jin et al., "A Survey and Experimental Study on Privacy-Preserving Trajectory Data Publishing" in IEEE Transactions on Knowledge & Data Engineering, vol. 35, no. 06, pp. 5577-5596, June 2023, doi: 10.1109/TKDE.2022.3174204.

FraudZen Hackathon

Offensive Security

Simulating fraud tactics to understand and combat SIMBox fraud.

- **Object of Study**: SIMBox fraud strategies simulation
- **Projects guiding lines**
  - Implement SIMBox fraud models that simulate realistic user behaviors and strategies
  - Evaluate the effectiveness and cost-efficiency of these models for fraudsters
  - Analyze implications for network security, including detection feasibility and integration challenges
- **Starting point(s):**
  - Anne Josiane Kouam, Aline Carneiro Viana, Alain Tchana. 2024. Battle of Wits: To What Extent Can Fraudsters Disguise Their Tracks in International bypass Fraud?, ASIACCS 2024 - 19th ACM Asia Conference on Computer and Communications Security, Jul 2024, Singapore, Singapore. https://doi.org/10.1145/3634737.3657023

FraudZen Hackathon

Defensive Security

Developing advanced AI-based techniques for detecting SIMBox fraud.

- **Object of the Study**: SIMBox fraud detection solutions.
- **Projects guiding lines**
  - Develop adaptive detection methods using AI to address evolving fraud behaviors
  - Integrate meta-learners targeting specific fraudulent behaviors into a comprehensive detection system
  - Establish metrics to evaluate the effectiveness of detection methods against emerging fraud strategies
- **Starting point(s):**
  - Anne Josiane Kouam, Aline Carneiro Viana, Alain Tchana. 2024. Battle of Wits: To What Extent Can Fraudsters Disguise Their Tracks in International bypass Fraud?, ASIACCS 2024 - 19th ACM Asia Conference on Computer and Communications Security, Jul 2024, Singapore, Singapore. https://doi.org/10.1145/3634737.3657023

Guidelines

Research

- *Is my Project Idea relevant?*
- **Comprehension**
  - Understand your main papers
  - .. but also explore the overall research topic!
- **Look for additional resources: Where to start**
  - https://scholar.google.de/
  - https://dl.acm.org/
  - https://ieeexplore.ieee.org/
  - https://dblp.uni-trier.de/

Guidelines

Writing

- *Your Focus is on implementation !!*
- Clear, logic structure
- Consistent notation
- Examples and figures (at least one)
- References to all figures and tables in text
- Spelling and grammar
- Citations and consistent bibliography!
- Use LATEX
- A template will be provided - Use English or French

Rule of thumb: Your paper should be the starting point for fellow students who are non-familiar with the topic

Guidelines

Writing

- **Introduction**
  - What is the problem? Why is it relevant ?
  - What is the current state of the art ?
- **Main part**
  - Structured overview of existing literature
  - How are the papers connected / different to each other ?
- **Discussion / Conclusion**
  - Discuss limitations
  - Online future work

The paper can be in English/French and should comprise 4 pages plus max. 2 pages of references and appendix

Practical organization

1- Pole Assignment

2- Project submission

1. **Deadline** Reminder: 25.11.2024 (**One week only**)
2. Form groups of 3/4 people in your pole
3. Propose an idea and send me an email (kouam.djuigne@tu-berlin.de) to have a feedback
4. **Upload** PDF file on the website
   - Idea title
   - Short description of what you want to do and how to plan to do it (2 paragraphs)
   - Group members

2- Paper writing and submission

1. **Deadline** Reminder: 10.01.2025
2. The template will be put on the website and notified (use Overleaf for instance)
3. **Upload** PDF file on the website
   - Mention the group members as authors in order of contribution
   - The star (*) indicates equivalent contribution

Thanks for your attention

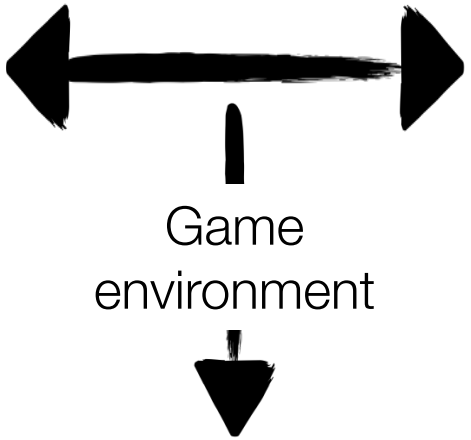
2- POW: Leveraging Mobile Human Behavior

Game-theoretic approach

Attacker

Goal:
Be indistinguishable in CDR records

SIMBox fraud model:
Selection of algorithms for automatic fraud behavior implementation



Investigator

Goal: Build a vector space maximizing distance btw fraudulent and legitimate users

- **Features set,** i.e.**,** user communication behavior
- **Observation period**, e.g., day, week
- **Binary classifier,** e.g., RF, SVM

Game metrics

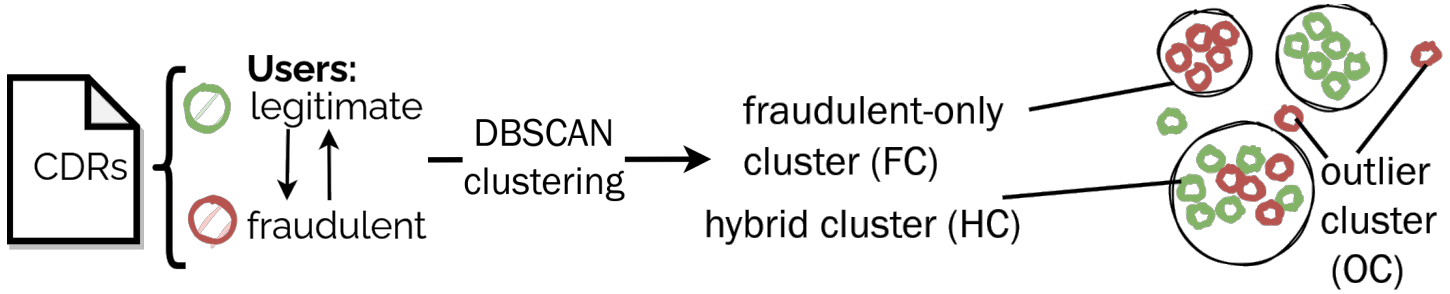
- **Classification capability:** Balanced accuracy, Recall, Precision
- **Adversary capability:** in-crowd-blending capability

2- POW: Leveraging Mobile Human Behavior

Game-theoretic approach

Game metrics

- **Classification capability:** Balanced accuracy, Recall, Precision
- **Adversary capability:** in-crowd-blending capability



$$\frac{|HC|}{|HC| + |FC| + |OC|}$$

2- POW: Leveraging Mobile Human Behavior

SIMBox fraud mitigation practical study

1- Human Communication Behavior Mimicking

2- Impacted **CDRs fields**, **Motives** = seeked impact, **Algorithm** = how, **Parameters**

Com. Behavior	CDR field	Adversary's strategy		
		Motive (M_i)	**Algorithms ($M_i.j$)**	**Parameters $(P_i)^X$**
Traffic	time (t_p)	M1- SIM activity limitation (time-related)	1- Working period per day	Day period
			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	/
		M2- Network activity generation (Text)	1- Activate	/
		M3- Network activity generation (Calls)	1- Activate	/

2- POW: Leveraging Mobile Human Behavior

SIMBox fraud mitigation practical study

Com. Behavior	CDR field	Motive (M_i)	Adversary's strategy: Algorithms ($M_i.j$)
Traffic	time (t_p)	M1- SIM activity limitation (time-related)	1- Working period per day
			2- Working period per week
		M2- Network activity generation (data/text/calls)	1- Fixed inter-event-time (IET)
			2- Random IET inside a fixed interval
			3- Multiple IET per day period
			4- Triggered by a metric threshold (#call, call duration, allocation time)
	Event Type (et_p)	M1- Network activity generation (Data)	1- Activate
		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)
		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
			2- Any except previous
			3- Any except previous zone ID
			4- Specified order
		M2- Short Base Station (BS) movements, i.e., choice of the next location	1- Random
			2- Default
			3- Specified order
			4- Manually fixed
		M3- SIM to module allocation, i.e., choice of when to move	1- Periodic
			2- Metric threshold (call duration, #calls)
			3- Specified duration
		M4- Short movements in the surroundings, i.e., choice of when to move	1- Fixed duration
			2- Threshold of metric (call duration, #calls)
			3- Specified duration
		M5- Practical gateway deployment	1- Most visited locations
		M6- Displacement mode	1- Automatic (handled by the SIMBox)
			1- Physical (with a car/motobike)
		M7- Mobility uniqueness	1- *SIMBox* architectural organization
Social	Contact (con_p)	M1- Incoming traffic routing	1- History
			2- Random
			3- In-turn, i.e., first available SIM card
			4- Sequence
			5- Balance
		M2- Network activity generation (contact ctrl.)	1- Activate
Device Property	Device ID (dev_p)	M1- IMEI modification generation rule	1- None, i.e., no IMEI modification
			2- One generation (SIM)
			3- Periodic generation (SIM/GSM mod.)
			4- Metric-based generation (GSM mod.)
		M2- IMEI value setting	1- Random
			2- Tac-based IMEI
			3- Prefix-based IMEI
			4- Registry-based IMEI

345,600,000 possible

Fraud model $= \{$

$algT* = M1.1,$

$algET* = null,$

$algEM* = M1.1 \times M2.1,$

$algCID* = M1.1 \times M5.1 \times M6.1 \times M7.1,$

$algCON* = M1.2,$

$algDEV* = M1.3 \times M2.1\}$

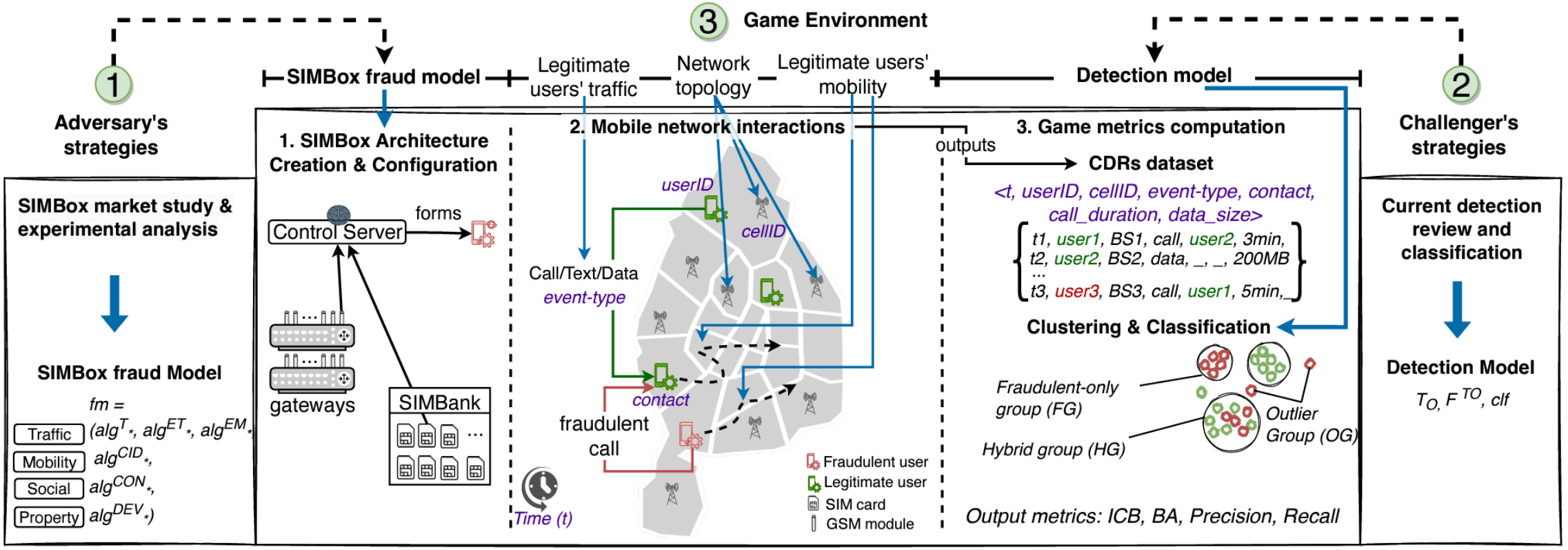
Adversary's strategy: Algorithms ($M_i.j$)
1- Working period per day
2- Working period per week
1- Fixed inter-event-time (IET)
2- Random IET inside a fixed interval
3- Multiple IET per day period
4- Triggered by a metric threshold (#call, call duration, allocation time)
1- Activate
1- Activate
1- Activate
1- Metric (call duration/ #calls) threshold per period (day/week/month)
1- Balance (to the SIM with fewest historical calls)
1- Manually fixed, i.e., no change
2- Any except previous
3- Any except previous zone ID
4- Specified order
1- Random
2- Default
3- Specified order
4- Manually fixed
1- Periodic
2- Metric threshold (call duration, #calls)
3- Specified duration
1- Fixed duration
2- Threshold of metric (call duration, #calls)
3- Specified duration
1- Most visited locations
1- Automatic (handled by the SIMBox)
1- Physical (with a car/motobike)
1- *SIMBox* architectural organization
1- History
2- Random
3- In-turn, i.e., first available SIM card
4- Sequence
5- Balance
1- Activate
1- None, i.e., no IMEI modification
2- One generation (SIM)
3- Periodic generation (SIM/GSM mod.)
4- Metric-based generation (GSM mod.)
1- Random
2- Tac-based IMEI
3- Prefix-based IMEI

infinite

Fraud model **instances**

2- POW: Leveraging Mobile Human Behavior

SIMBox fraud mitigation practical study



Cellular network simulator, i.e.,
Network infrastructure & topology

Implemented in C++, 19K code lines

2- POW: Leveraging Mobile Human Behavior

Experimental setup

- **5 selected *SIMBox* fraud models**
  - fd_naive
  - fd_traffic
  - fd_mobility
  - fd_social
  - fd_all
- **Other Fraudulent parameters**
  - %diverted incoming intl. calls: 3% =>one call/7min, 12% =>one call/2min
  - #fraudulent UEs: 50, 100, 150, 200

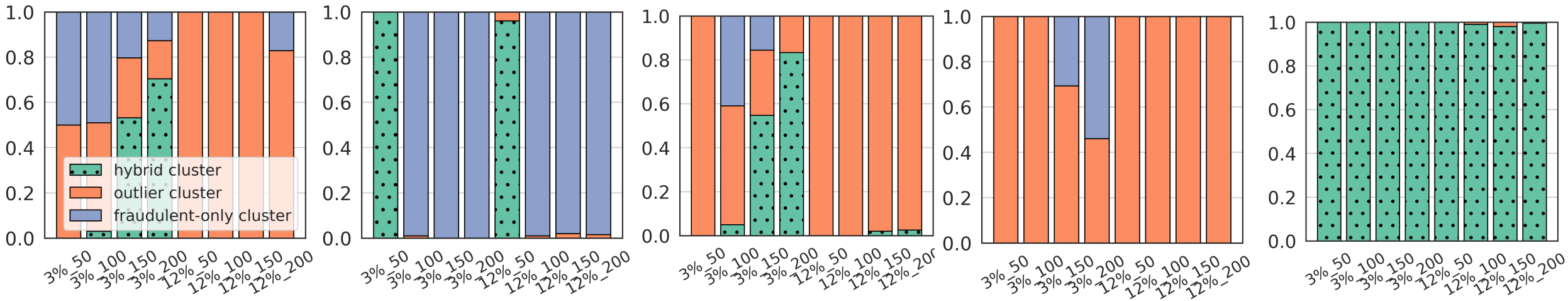
- **Legitimate Behavior:**
  - real-world, non-public, and fully anonymized CDRs from a major telecom operator.
  - One month duration
  - 21K users
- **Literature detection implementation**
  - observation period, e.g., *day, week*
  - set of features, i.e., traffic_based, traffic+mobility, traffic+social, all
  - binary classifier, e.g., ANN, SVM, RF, GBDT

Total: 1280 scenarios tested

2- POW: Leveraging Mobile Human Behavior

Insights for detection

In-crowd-blending capability



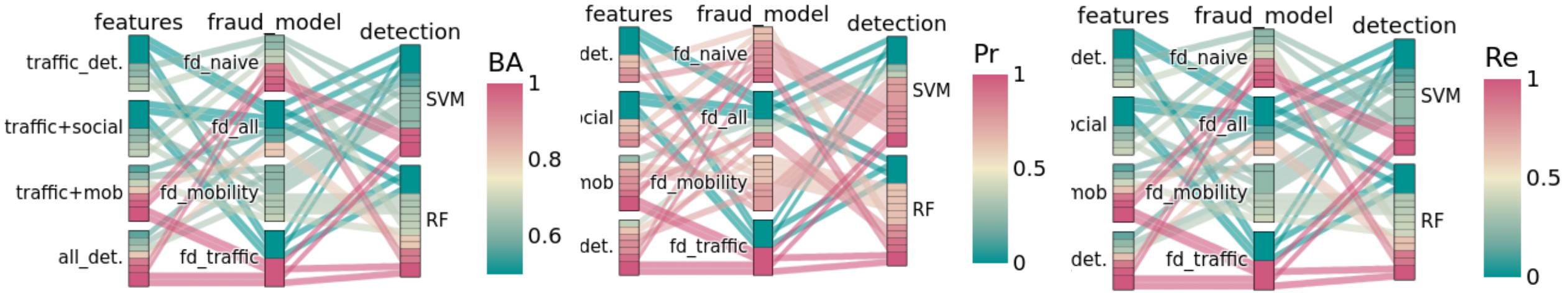
fd_{naive} $fd_{traffic}$ $fd_{mobility}$ fd_{social} fd_{all}

Insight: fd_{all} ~ almost 100% efficiency w.r.t. investigator strategies

2- POW: Leveraging Mobile Human Behavior

Insights for detection

Feature set, Fraud model, Binary classifier



Balanced Accuracy | Precision | Recall

Insights:

Features set: Mobility behavior is the best detection facet

Classifier: simple decision rules combination is better than global behavior