Chapter 1: Intro to "*Intro to CyberSecurity*"

Xianghang Mi

中国科学技术大学
University of Science and Technology of China

Outline

- Course FAQs (frequently asked questions)
- Motivations of CyberSecurity, i.e., why we need to learn cybersecurity?
- Overview of The Course Content

Outline

- Course FAQs
- Motivations of CyberSecurity
- Overview of This Course

Time and Location

- 学时：40理论 + 40实验
- Lectures
  - 9:45 AM – 12:10 PM, Tuesday
    - 09:45~10:30，10:35~11:20，11:25~12:10
  - 高新区 GT-B210
- Q&A
  - Office hours: 1:30pm ~ 3:00pm, Tuesday
    - Virtual access: https://meeting.tencent.com/dm/7MbjHGyc4ihV
    - Location: 高新校区第一学科楼B410
  - Online Discussion： Blackboard-> 讨论版

Course Staff: Instructor

- 糜相行 (Xianghang Mi) https://xianghang.me
- 计算机学院，特任教授
  - Assistant Professor, CSE at SUNY Buffalo
  - Research Scientist at Meta (Formerly Facebook)
- Research Interests: Network Security, IoT security, Cybercrime, AI for Security
- Email: xmi@ustc.edu.cn
- Reach out to me if you are interested in security research

Course Staff: Teaching Assistants

- 李骋昊： ericlichenghao@mail.ustc.edu.cn
- 吕凯盛： touko@mail.ustc.edu.cn

Important URLs

- Internal course website: https://www.bb.ustc.edu.cn/
- External course website: https://sec2024.xianghang.me/

Course Objectives

- Learn the fundamentals of cybersecurity
  - Problems, methodologies, and tools
  - Research frontiers
  - Challenges and future directions
- Practice various security tools and do hacking
- Have fun!

What You Will Not Learn From This Course

- A lot!!!
- Why?
  - There are many things I don’t know
  - Cybersecurity evolves across time with new attacks and defense

Course Material

- Recommended Materials
  - CyBok1.1, https://www.cybok.org/knowledgebase1_1/
  - Computer Security by David Wagner: https://textbook.cs161.org/
  - Computer Security: Principles and Practice, by William Stallings and Lawrie Brown, 4th version in 2018, 3th version in 2014
  - Security Engineering: A Guide to Building Dependable Distributed Systems, 3th version, 2020
  - A Graduate Course in Applied Cryptography, D. Boneh and V. Shoup http://toc.cryptobook.us/#toc

Course Workload

- **Start early, and try to enjoy it!!**
  - 5 security Labs/Homeworks
  - 1 programming assignment (an attack experiment)
  - ~~1 research project~~
  - 1 final exam
- You can work as a team of up to **3 members** for **all tasks except for the final**
  - You may **work as different teams** for separate assignments
  - Clearly clarify the individual contributions when working as a team

Grading (Tentative)

- Final: 30%
- Security Labs/HWs: 50%
- 1 Programming Assignments: 20%
- ~~1 Research Assignment 20%~~
- Bonus Points:
  - Positive/negative course participation ±5%

Grading (Tentative)

- Done individually or (highly recommended) in teams of up to 3 students
  - One submission per team
  - All team members get the same grade, unless the contribution is minor
- Late Policy: 0.9^X if X <=7 else 0
  - late submissions of X days: scored out of 100 * 0.9^X points instead of 100
  - For example, X=1, the scoring scale will be 90 instead of 100
- Handwritten work and legibility
  - Written assignments should be neat and legible
  - The TAs **reserve the right to assign a zero grade** to illegible handwritten homeworks/reports

Lectures

- Understand material in class (**ask questions!**)
  - Read relevant chapters, and the recommended reading materials
- Will also post extra readings for better understanding and additional concepts

The Final Exam

- No make up exam will be given without a valid excuse
- **No lame excuses please!!!**
  - I have to go home/start an internship early, can I take the exam earlier?
  - I had a fight with my girlfriend/boyfriend
  - I need a B+ to graduate this semester
  - I have a job interview on/before the exam/project due day
  - ….

The Final Exam

- *Time: 10:45AM - 12:10PM, Tuesday, June 11 (tentative)*
- Policies
  - This exam is open-book, but seeking help from any other individuals is NOT ALLOWED.
  - *You must not give or receive any unauthorized help on this exam. All work must be done on your own*
- Handwritten answering should be human readable.

Academic Integrity

- **Your honor is most valuable, don't take risk losing it!**
- **No tolerance on cheating/plagiarism!!!**
  - All academic integrity violation cases can be reported to the department, school, and university, and recorded
  - **0** on the particular assignment/exam on 1st offense
  - Fail the course on 2nd offense
- Team members are equally responsible!
- Students who share the work with others are as responsible for academic dishonesty as those receiving the material
- Again, no lame excuses!
  - I did not know/I was not sure/I forgot

Academic Integrity

Your honor is most valuable, don’t take risk losing it!

Academic Integrity

No tolerance on cheating/plagiarism

More on Academic Integrity

- Group study/discussions are encouraged but submission must be your own work
- Homeworks/Labs
  - No collaboration among individuals/teams
  - Use of reference material is allowed as long as you explicitly state the references
- Programming Assignments:
  - Discussion of ideas is welcome but **no sharing of code!**
  - Use of code found online is allowed as long as you explicitly state the reference

Class Participation

- Very important!
- Attend classes, participate in discussions, express your opinion, ask questions
- Feel free to share ideas, questions, articles on cybersecurity, etc. with whole class

How to make it interesting? How to do well?

- Ask questions, the more the better
  - "It is not the answer that enlightens, but the question"
- Give suggestions
  - I will consider them seriously
- Do the assigned readings and surf the web to read related things
- Start early on homeworks/labs/programming assignments

Where Do I Ask Questions About

- Lectures, homework?
  - The forum on Blackboard
  - Office hours
  - Email to the teaching team
- Programming assignments, labs?
  - Office hours
  - Email to the teaching team

The Use of AI Agents

- What's your opinion?
  - Should students be allowed to use AI agents for assignments?
  - How about the final?
  - What are the pros and cons?

The Use of AI Agents

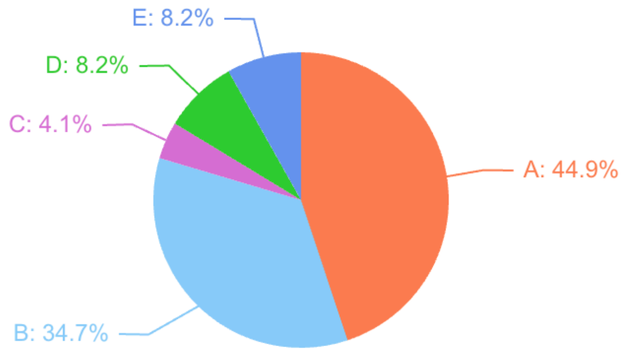
- Any AI agent is allowed for all tasks including the final exam
- Any negative impact?
  - It can generate factious errors or hallucinations
  - What can you really benefit from this course?

Historical Grading

中国科学技术大学课程成绩分析表

开课学期：	2023年春季学期	课程名称：	信息安全导论	开课学院：	011计算机科学与技术系
授课教师：	糜相行	应试人数：	49	平均成绩：	76.59
低分率：	8.2%（<40分）	及格率：	91.8%（≥60分）	优秀率：	44.9%（≥85分）

等级	分数段	人数	百分比	饼状图
A	85~100	22	44.9%	
B	75~84.9	17	34.7%	
C	65-74.9	2	4.1%	
D	60-64.9	4	8.2%	
E	0-59.9	4	8.2%	



备注：

Survivor Bias!

Outline

- Course FAQs
- Motivations of CyberSecurity
- Overview of This Course

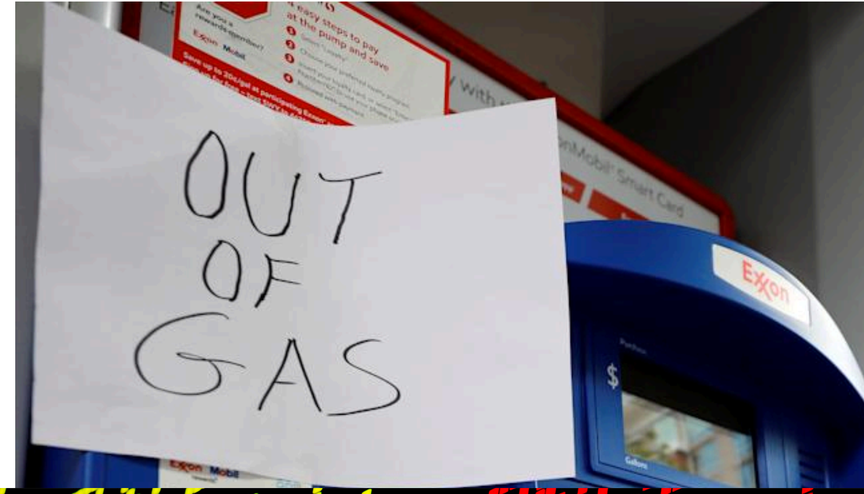
Weak VPN Password

VPN

Colonial Pipeline Company

Customer Billing

Pipeline Operation



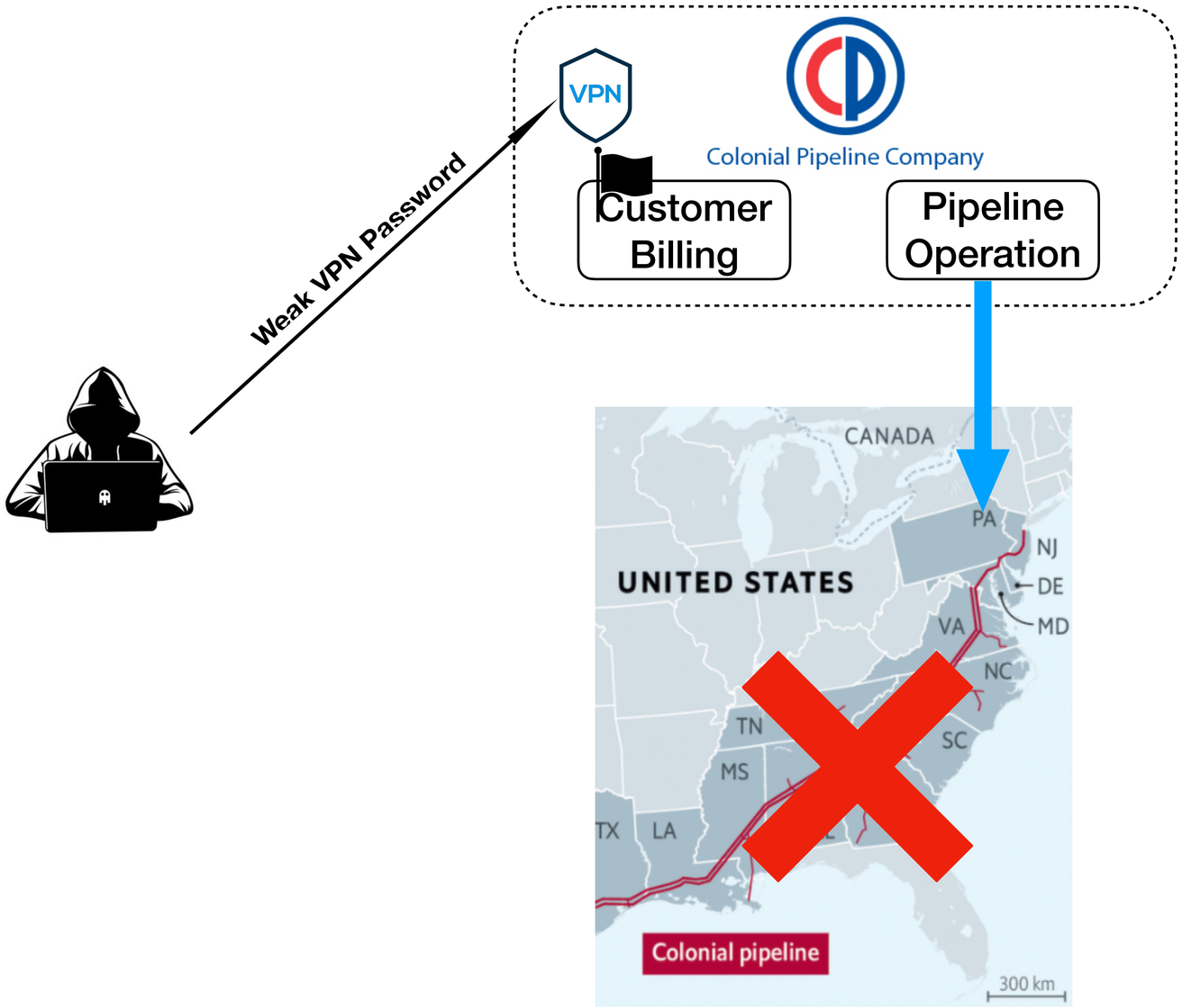
The FBI confirms that the Darkside ransomware is responsible for the compromise of the Colonial Pipeline networks. We continue to work with the company and our government partners on the investigation.

FBI FEDERAL BUREAU OF INVESTIGATION

STATEMENT

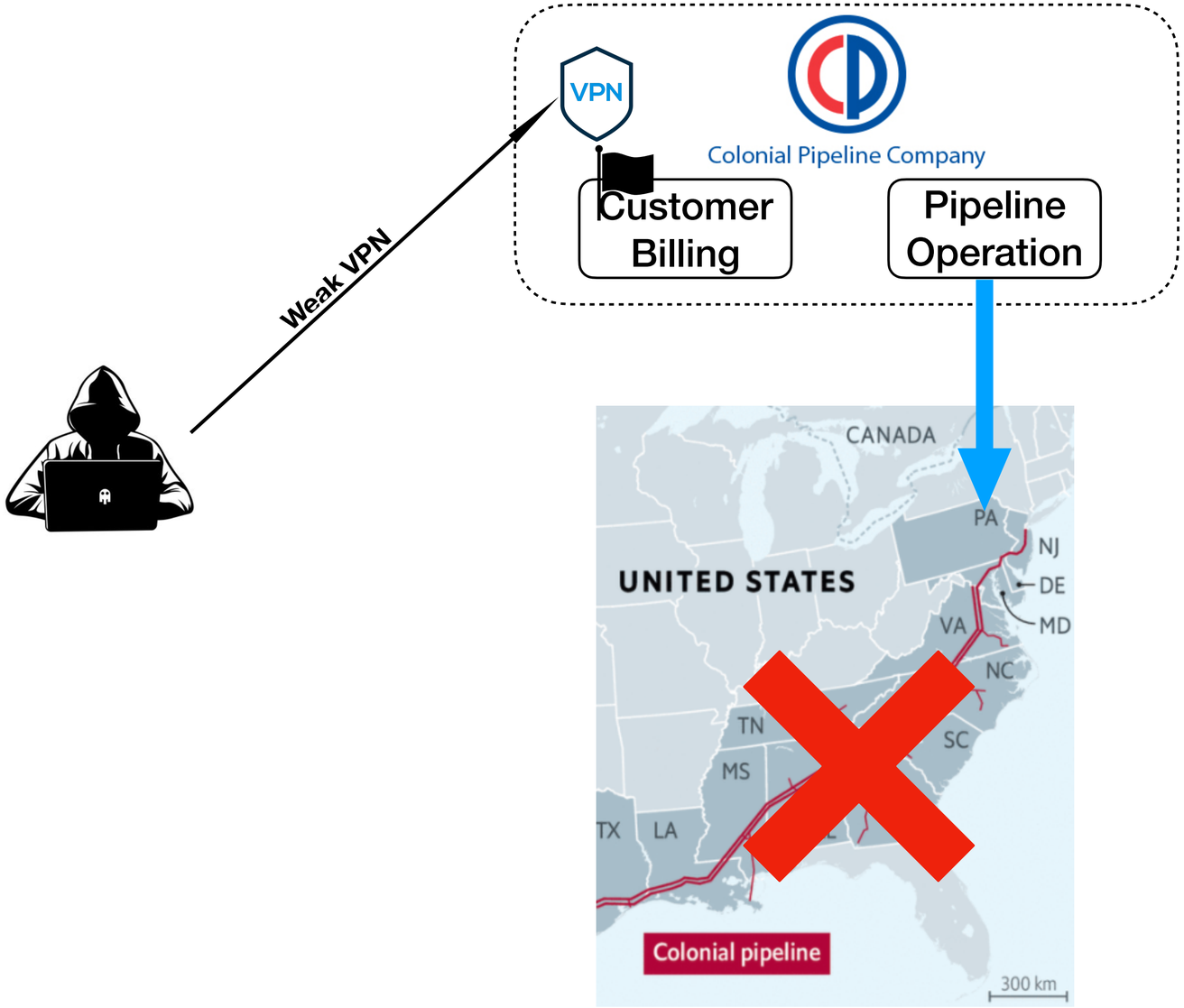
DarkSide Ransomware Creates Emegency State in the U.S.

Colonial Pipeline Hack: timeline



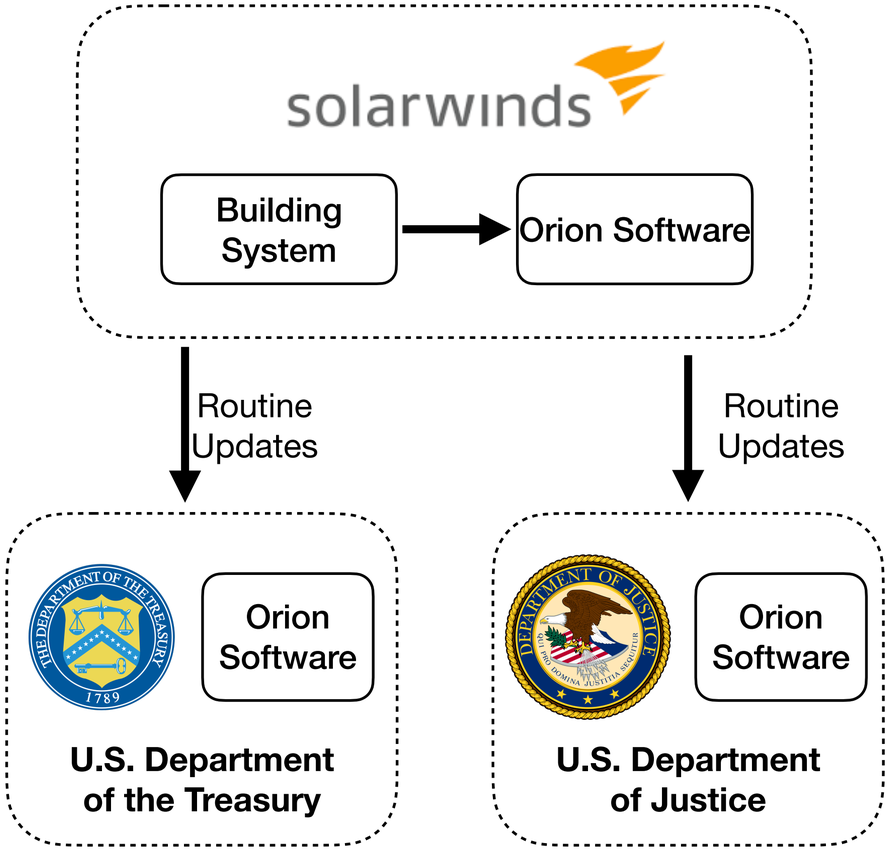
- **April 29, 2021:** the attacker got access to the networks of Colonial Pipeline Co, through a VPN account
- **5 a.m, May 7, 2021:** an employee in Colonial's control room saw a ransom note demanding cryptocurrency appear on a computer
- **6:10 a.m, May 7 2021:** the pipeline has been shut down, which was the first time Colonial had shut down the entirety of its gasoline pipeline system in its 57-year history
- Colonial paid the hackers, a $4.4 million ransom shortly within several hours after the hack, and The hackers also stole nearly 100 gigabytes of data from Colonial Pipeline
- **5 p.m., May 12 2021:** the restart of pipeline operations began, ending a six-day shutdown.

Colonial Pipeline Hack: summary



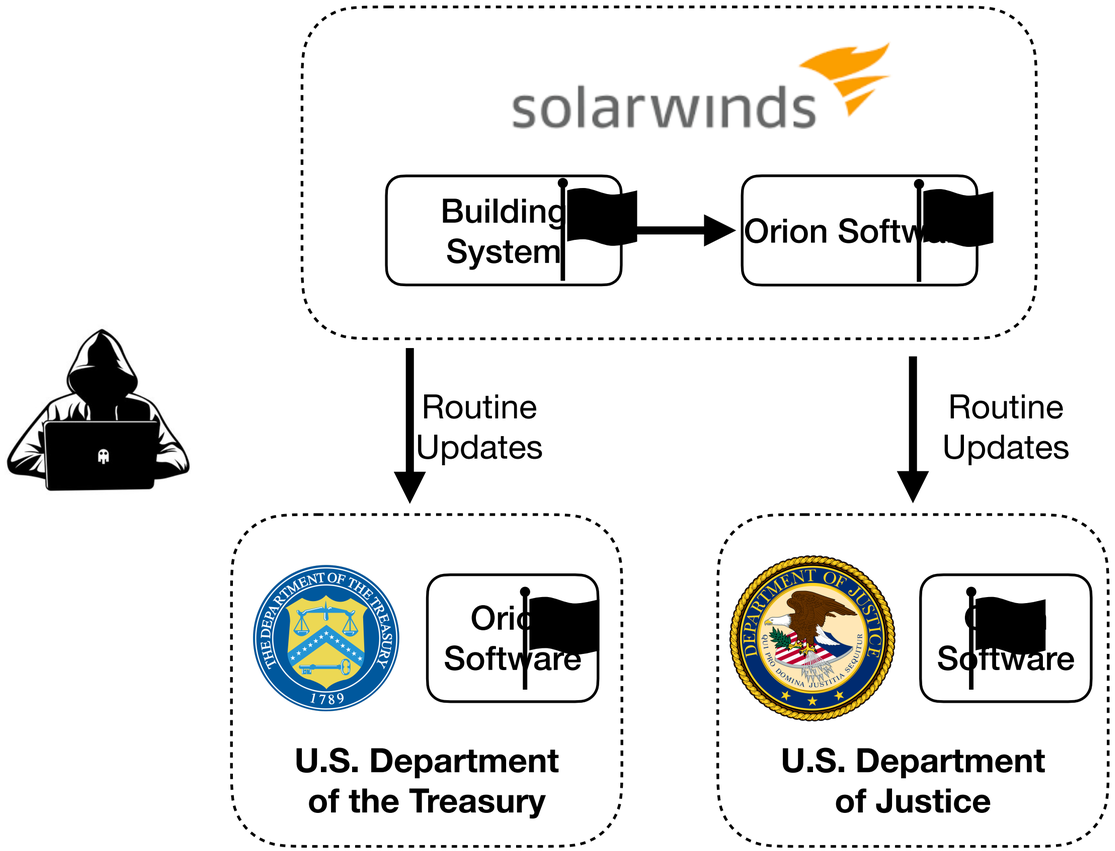
- A ransomware attack
- The largest cyberattack on an oil infrastructure target in the history of the United States
- Weak authentication (no multi-factor authentication) for the VPN access
- Fortunately, attackers compromised only the information technology systems, but not the operational technology systems which are responsible for controlling the flow of gasoline
- **We may see more attacks on industrial control systems (ICS)**

The SolarWinds Hack: background



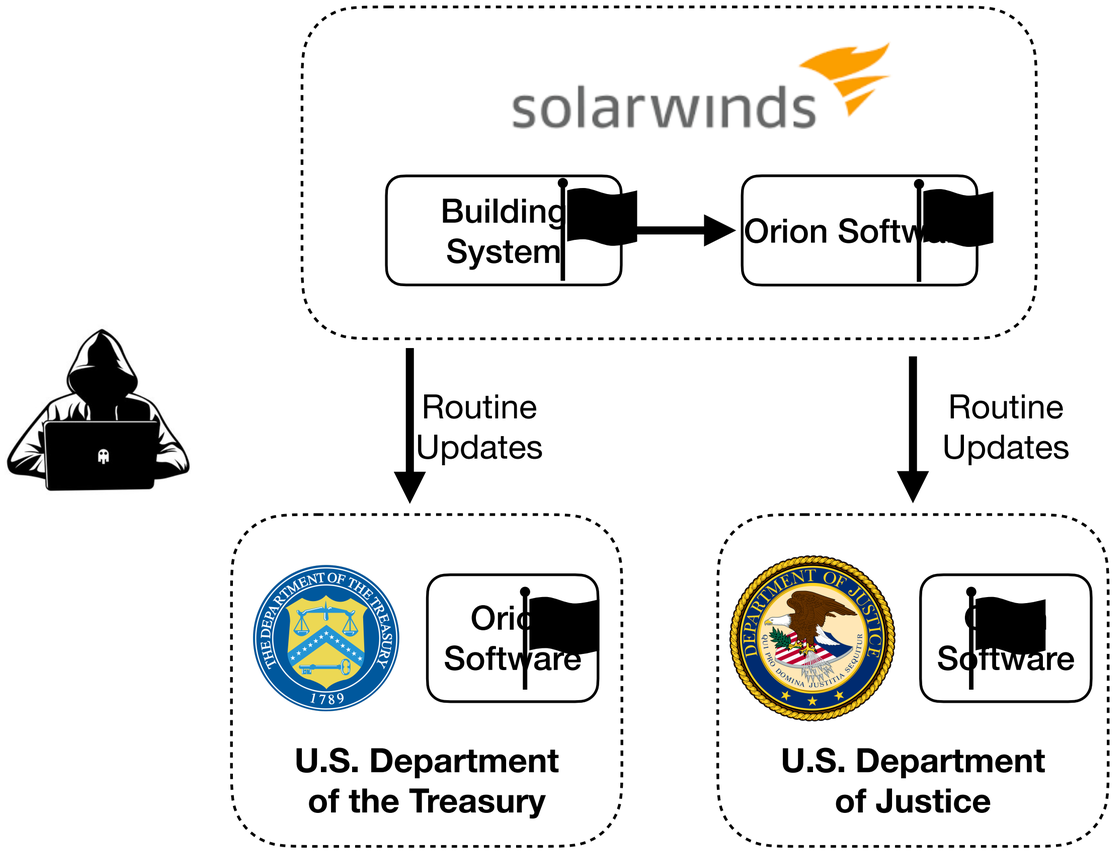
- SolarWinds, a Texas-based Software Company
- One of its major software products is Orion, a network management tool
- Orion is well adopted by more than 30K private and public organizations, e.g., U.S. Department of the Treasury
- Orion is routinely updated for new security patches, and performance enhancements, etc.

The SolarWinds Hack: timeline



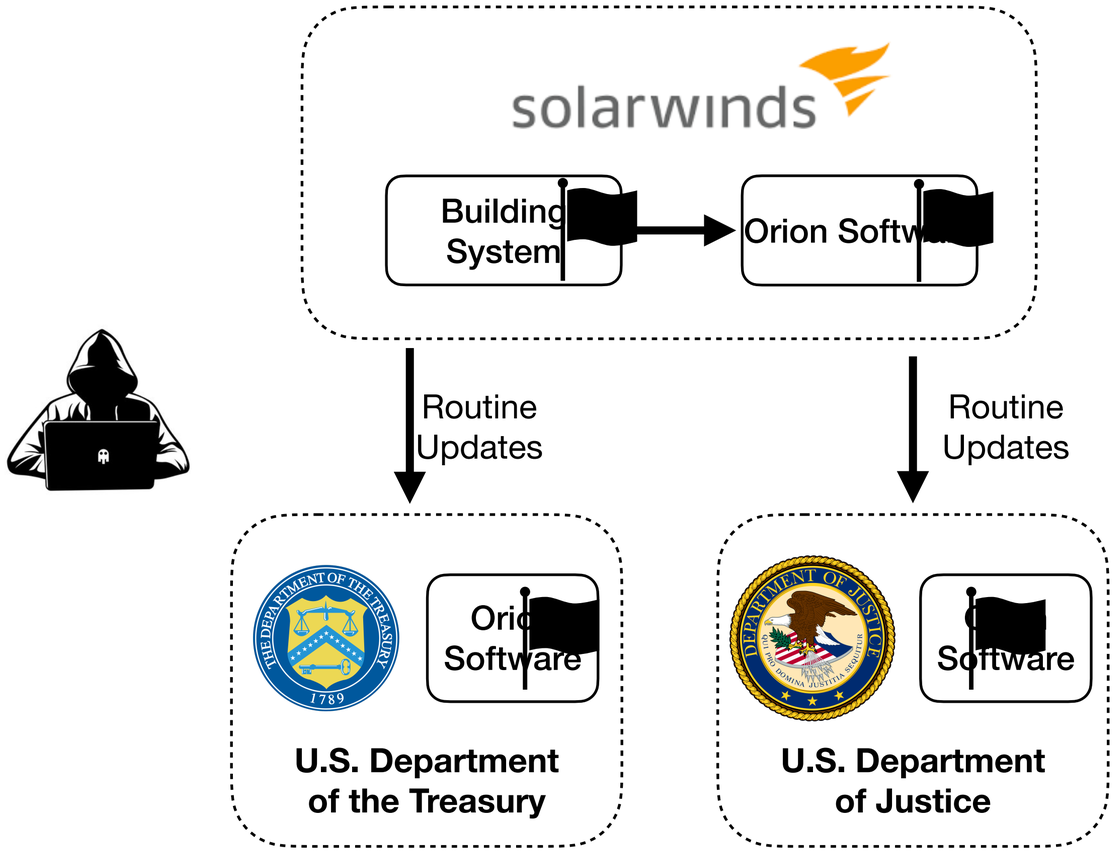
- **Sept 2019:** Attackers gained unauthorized access to SolarWinds's software building system
- **Oct 2019:** a proof-of-concept **tiny** code snippet was injected into the Orion Software, and got distributed to the clients.
- **Feb 20 2020:** the real malicious code, known as Sunburst was injected into Orion
- **Late March 2020**: the tainted Orion was distributed to clients
- **Late Dec 2020**: after being hacked for nine months, the attack was uncovered by FireEye

The SolarWinds Hack: attack techniques



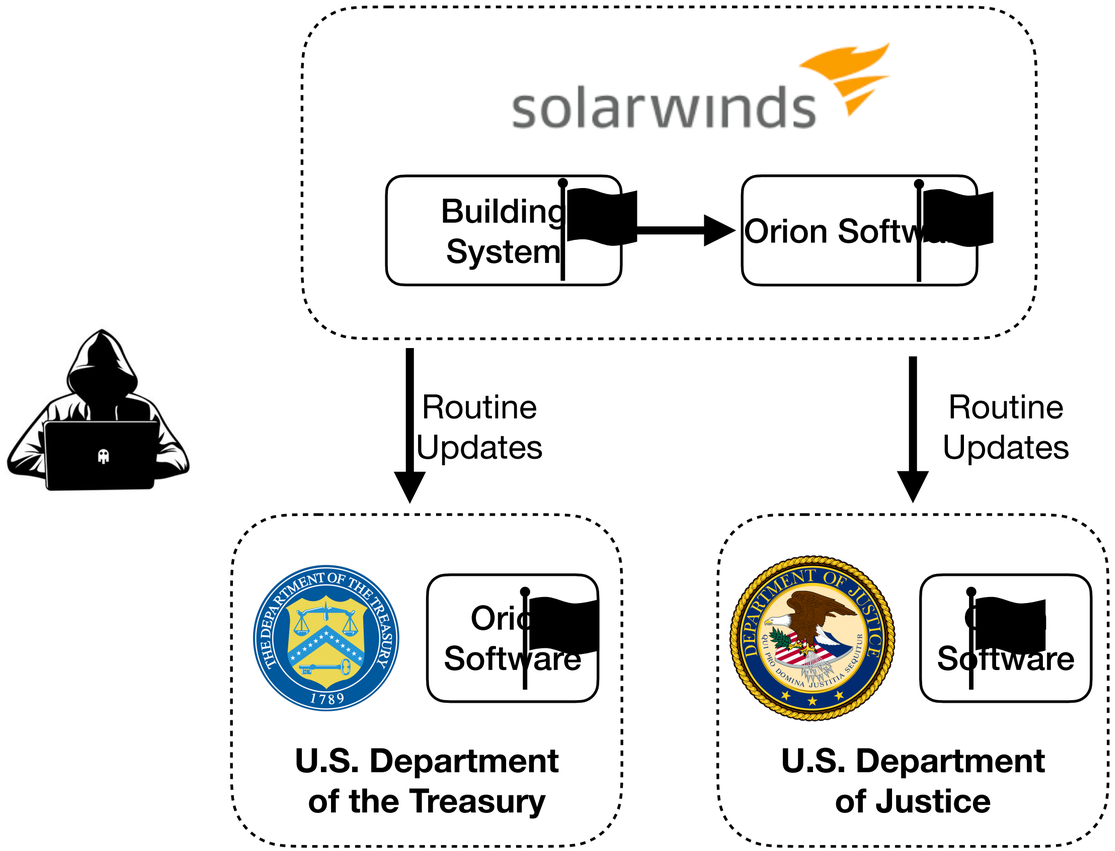
- Orion is a perfect target considering its critical role, and its customers listing on SolarWinds's website
- Mimicked the Orion software communication protocols
- Cleaned the crime scene so thoroughly investigators can't prove definitively who was behind it.

The SolarWinds Hack: results



- The hackers had gained access to the data and emails of at least nine US federal agencies, including the Department of the Treasury and the Department of Justice, and about 100 private companies.
- The hackers also punched a hole into the Cybersecurity and Infrastructure Security Agency (CISA), the office at the Department of Homeland Security
- The attack was there for nine months, it is unclear whether they were just reading emails

The SolarWinds Hack: summary



- This is a typical software supply chain attack
- Only the attackers know the whole picture, of which, many details are either undisclosed or not clear to the public
- There were early warning signs, but ignored or missed for various reasons

Security & Privacy Problems

– What else?

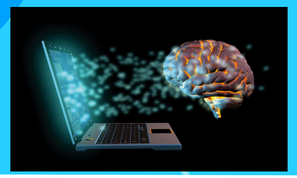
Security & Privacy Problems

- Misinformation or Disinformation
  - e.g., fraudulent messages, fake news, or hate speech
- User Tracking or Illicit Surveillance

- What else?

CCS'22: Hunting and Profiling SMS Spam from Twitter

CyberThreat Mining
&
Intelligent Security



Residential Proxies

100% anonymous proxies from all around the globe with zero IP blocking. Harness residential IPs to effortlessly capture accurate data from any source on the web.

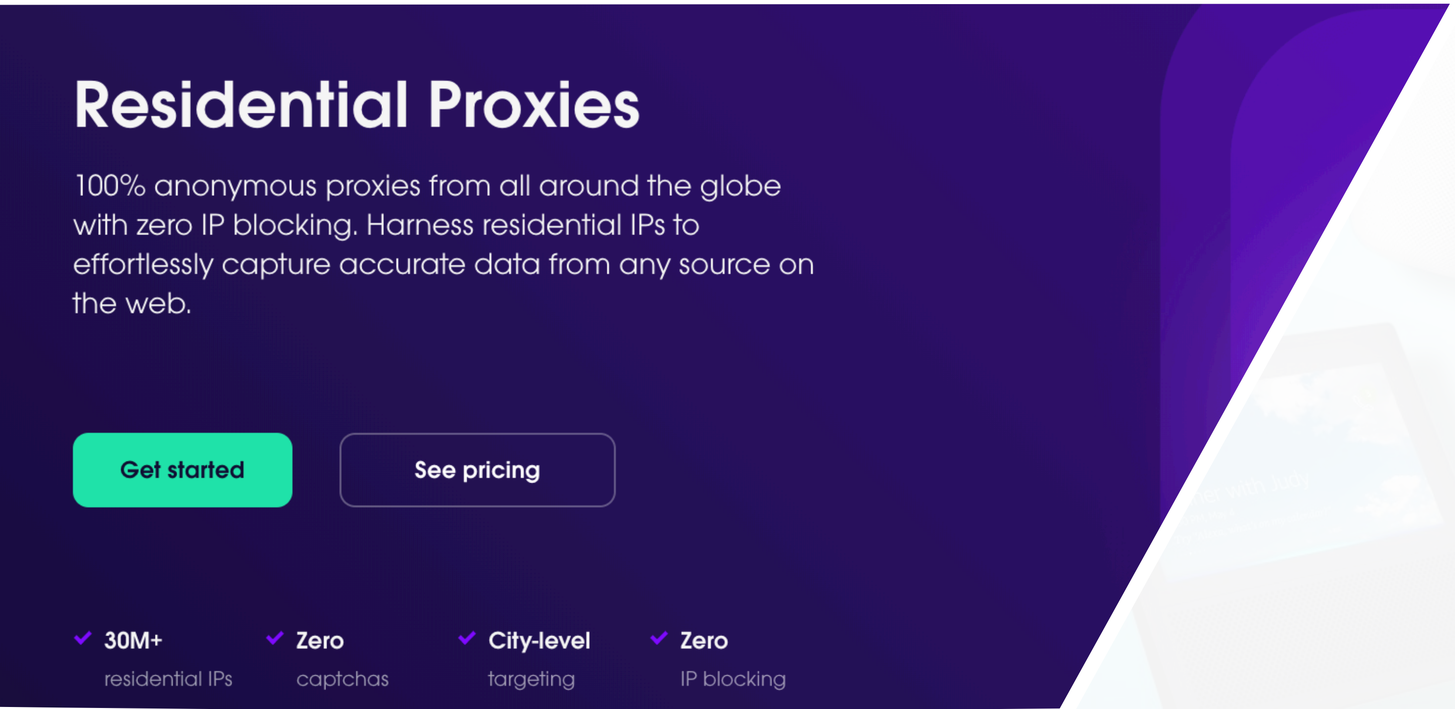
SP'19: Residential IP Proxy as a Service
NDSS'21 Your Phone is My Proxy
CCS'22 Residential Proxies in China

- ✓ **30M+** residential IPs
- ✓ **Zero** captchas
- ✓ **City-level** targeting
- ✓ **Zero** IP blocking

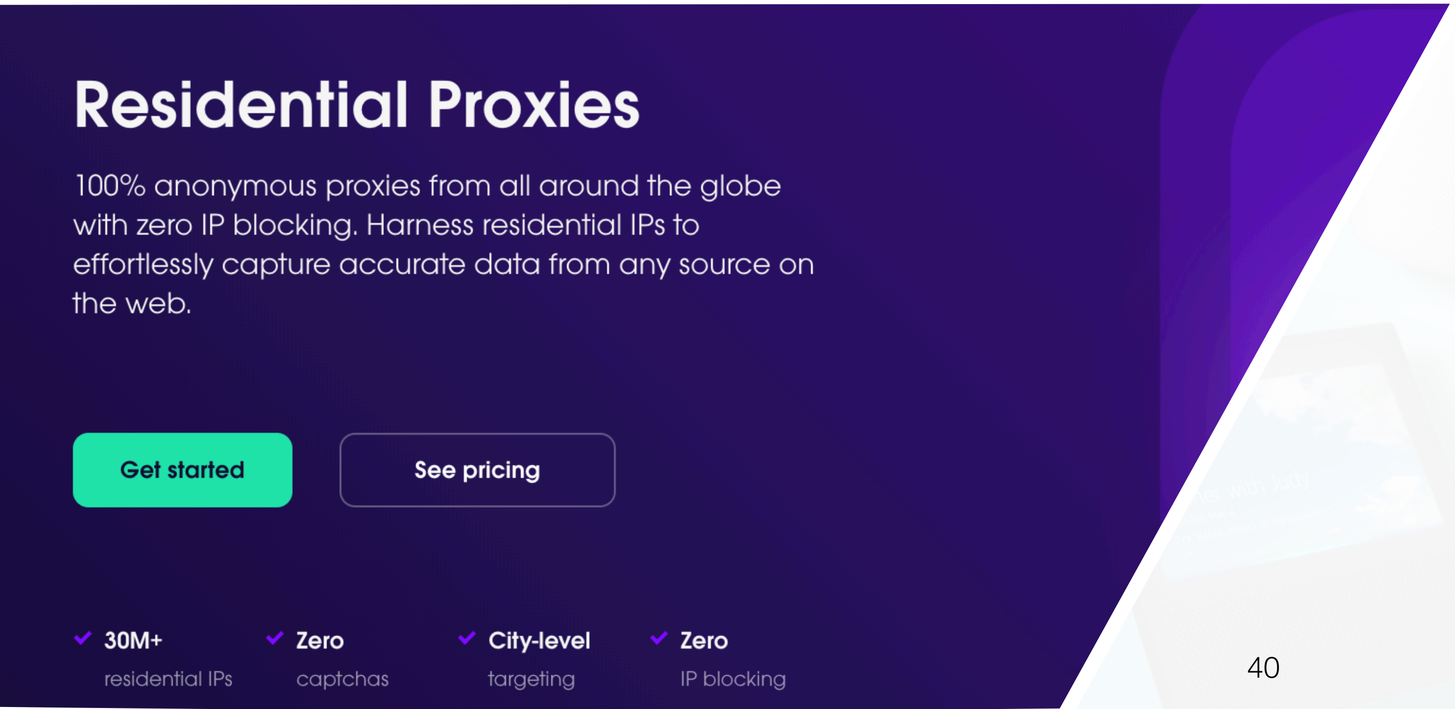
SP'19: Dangerous Skills on Google Home and Amazon Alexa

A continual security study on **residential proxies**:

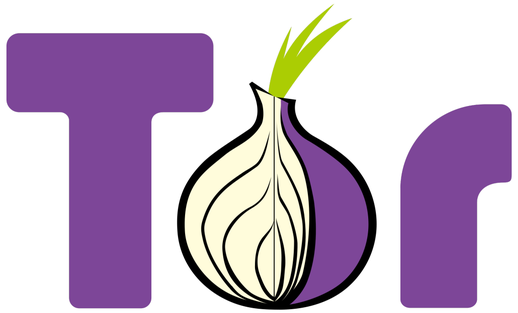
- Resident Evil: Understanding Residential IP Proxy as a Dark Service, **SP'19**
- Your Phone is My Proxy: Detecting and Understanding Mobile Proxy Networks, **NDSS'21**
- An Extensive Study of Residential Proxies in China, **CCS'22**
- Understanding and Classifying Network Traffic of Residential Proxies, **Under Submission**



Resident Evil: Understanding Residential IP Proxy as a Dark Service, **SP'19**



Web Proxies



HTTP/HTTPS /SOCKS

✗ **Exit nodes are constrained**

✗ **Exit nodes are distinguishable**

✗ **Exit nodes may be heavily abused**

Service blocking or degradation

Residential IP Proxy as a Service

bright data
(formerly named Luminati)

Proxy Web Data Resources Pricing Log in Sign up

RESIDENTIAL PROXY NETWORK

The best-performing Rotating Residential Proxies

Avoid restrictions and blocks with the fastest residential network in the industry

- Target any country, city, carrier & ASN
- 99.99% uptime - extremely stable
- 72+ million ethically-sourced IPs

Sign up for a free trial >

Residential IP Proxy as a Service

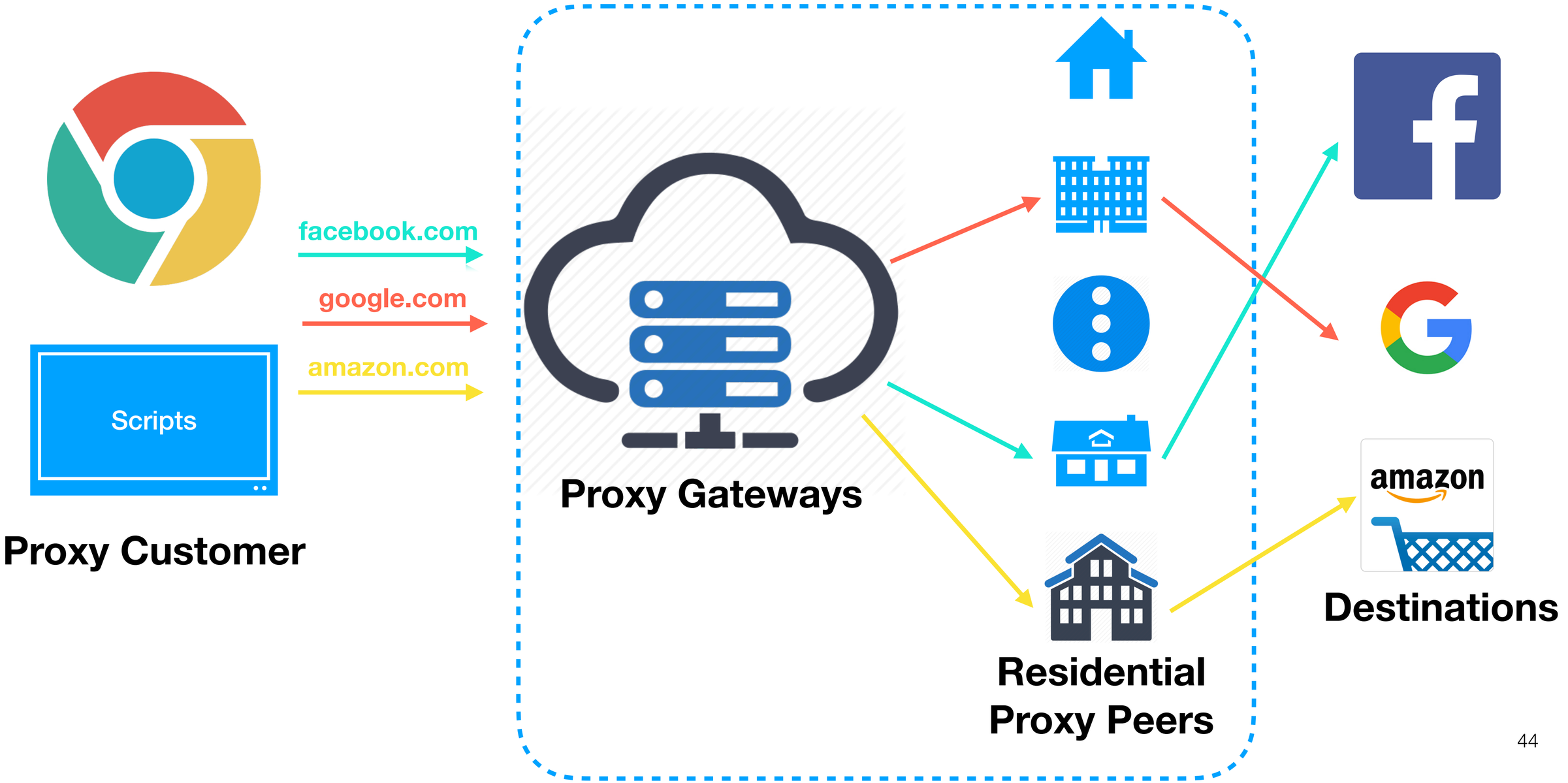
Millions of Residential IPs

Clean IPs, Never Get Blocked

Globally Distributed

No Traffic Limits

Abstract the Ecosystem



Reason about Security Problems

Assumptions — Millions of residential and clean proxy peers

Recruitment — How can millions of proxy peers get recruited?

- Legitimate channels to recruit volunteers?
- Monetization of malware or compromised devices?

Usage — What are those services used for?

- serving as infrastructure for online abuse activities?

Qualify and Quantify Security Problems

Verify Assumptions → Recruitment → Usage

Scale

Residential or Not

Evasiveness

Each request is identified by a unique subdomain

Each request/response has payload encrypted and signed

Provider	Price	Payment	Infiltration Period
Proxies Online	$25/GB	Paypal	07/06/2017 - 11/24/2017
Geosurf	$300/month	Paypal	09/17/2017 - 10/22/2017
ProxyRack	$40/month	Bitcoin	09/18/2017 - 11/24/2017
Luminati	$500/month	Paypal	09/25/2017 - 11/01/2017
IAPS Security	$500/month	Bitcoin	09/23/2017 - 11/01/2017

Qualify and Quantify Security Problems

Verify Assumptions → Recruitment → Usage

Scale

Residential or Not

Evasiveness

60+ millions of successful probes

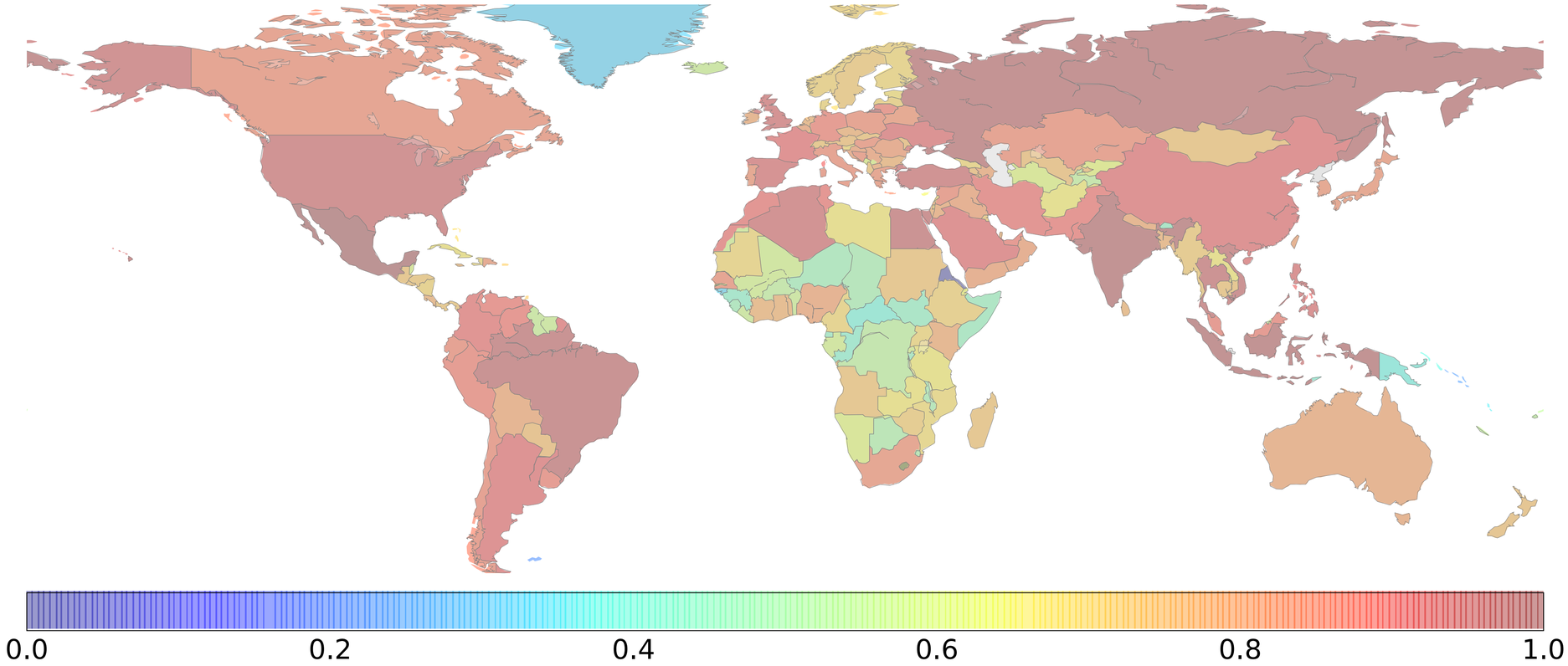
6.2 millions of unique IPv4 addresses

238 countries/regions, 52K+ ISPs.

Abstract the Ecosystem

Verify Assumptions → Recruitment → Usage

- Scale
- Residential or Not
- Evasiveness



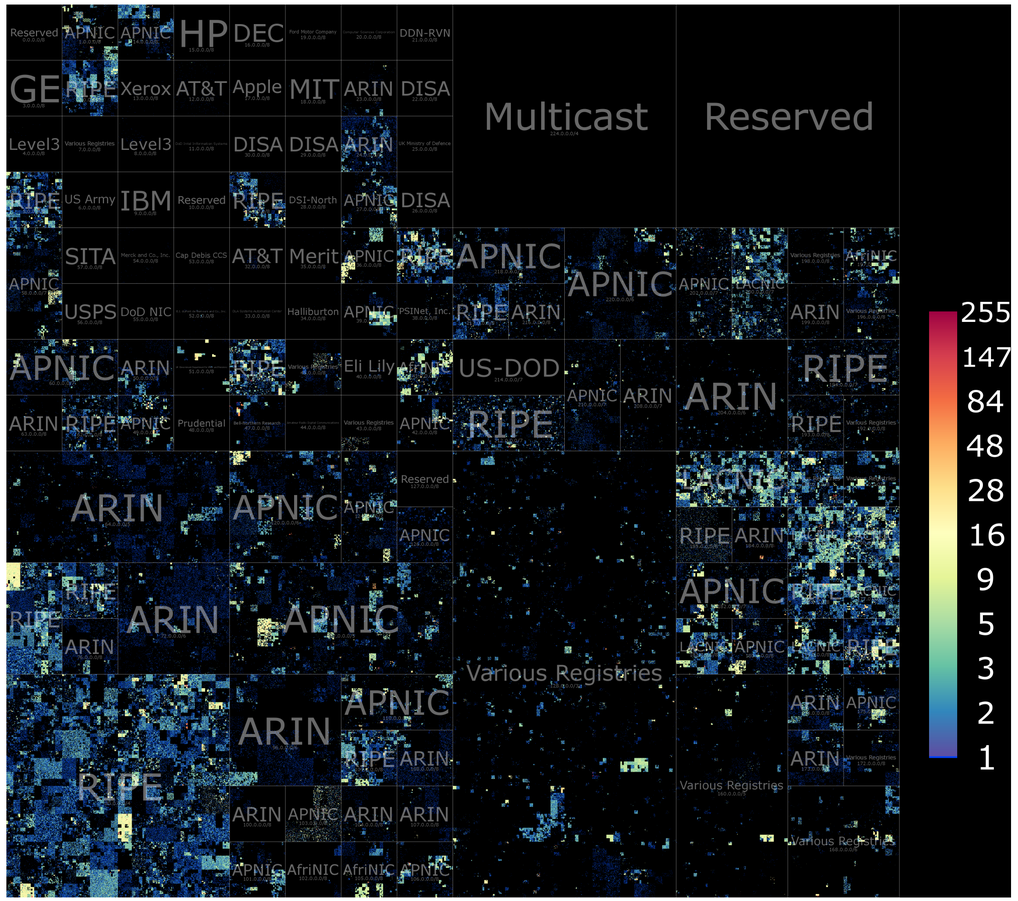
Abstract the Ecosystem

Verify Assumptions → Recruitment → Usage

Scale

Residential or Not

Evasiveness



4096 * 4096 bitmap

Each /24 IPv4 prefix is mapped to a pixel, using Hilbert curve of order 12

Different pixel colors denote # of proxy IPs for a given /24 prefix

Qualify and Quantify Security Problems

Verify Assumptions → Recruitment → Usage

Scale

Residential or Not

Evasiveness

10K residential &
10K non-residential IPs

ML Classifier Training/Tuning with 35 robust features

Random Forest Classifier

Recall: 97.12%

Precision: 95.61%

Qualify and Quantify Security Problems

Verify Assumptions → Recruitment → Usage

Scale

Residential or Not

Evasiveness

5.9M (95.22%) of 6.2M predicted as residential IPs

Qualify and Quantify Security Problems

Verify Assumptions → Recruitment → Usage

Scale

Residential or Not

Evasiveness

Publicly available proxy datasets

- Tor relays
- Free web proxies
- IP2Proxy LITE

→ **Only 0.06% of 6.2M IPs**

Qualify and Quantify Security Problems

Verify Assumptions → Recruitment → Usage

Scale

Residential or Not

Evasiveness

Publicly available IP threats

- Botnet bots
- Spamhaus EDROP
- Open Threat Exchanges

→ **Only 2.20% of 6.2M IPs**

Qualify and Quantify Security Problems

Verify Assumptions → Recruitment → Usage

Legitimate Recruitment

Proxy Devices

Proxy Programs

Only Luminati was found to recruit users through Hola programs

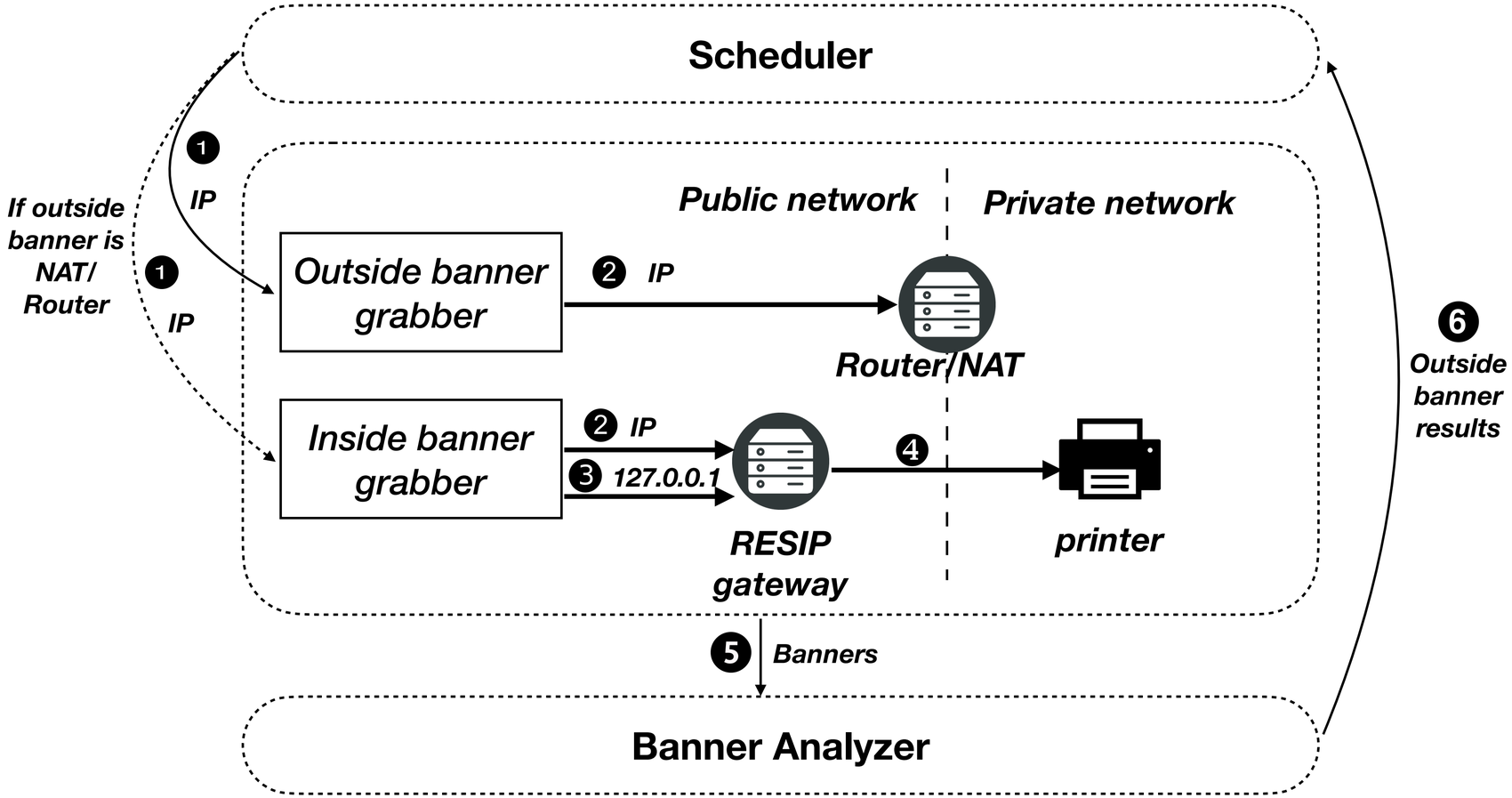
Qualify and Quantify Security Problems

Verify Assumptions | Recruitment | Usage

Legitimate Recruitment

Proxy Devices

Proxy Programs



Qualify and Quantify Security Problems

Verify Assumptions → Recruitment → Usage

Legitimate Recruitment

Proxy Devices

Proxy Programs

All providers got suspicious IoT devices identified for their proxy IPs, including Luminati

Device Type	Num	(%)
router	114,768	48.42
firewall	25,088	10.58
WAP	24,470	10.32
gateway	22,003	9.28
broadband	17,358	7.32
webcam	13,024	5.49
security-misc	10,608	4.48
DVR	4,249	1.79
media device	2,589	1.09
storage-misc	1,988	0.84

Qualify and Quantify Security Problems

Verify Assumptions → **Recruitment** → Usage

Legitimate Recruitment

Proxy Devices

Proxy Programs

Traffic logs of Infiltration probes

[uuid].[timestamp].
[providerId].[gwId]._____

Accurate Correlation

Traffic logs of potentially unwanted programs (PUP)

Program Hash, Device Id,
URL, IP, Timestamp, etc

Qualify and Quantify Security Problems

Verify Assumptions → **Recruitment** → Usage

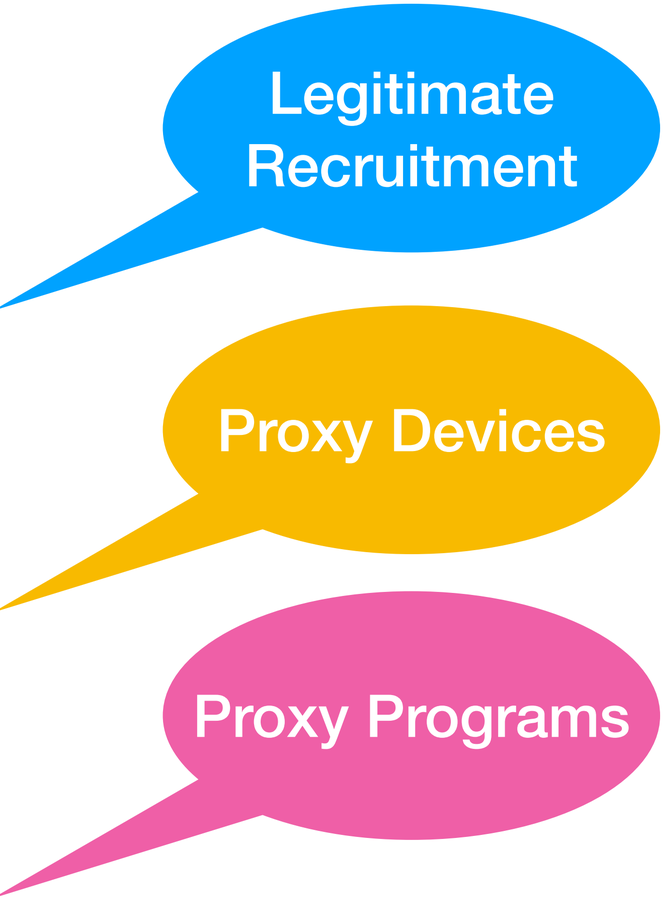
Legitimate Recruitment

Proxy Devices

Proxy Programs

- 67 PUP (potentially unwanted programs) samples identified on the Windows platform
- Proxy programs are found for all 5 providers
- 50 of these PUPs were flagged by anti-virus engines as malicious

Qualify and Quantify Security Problems

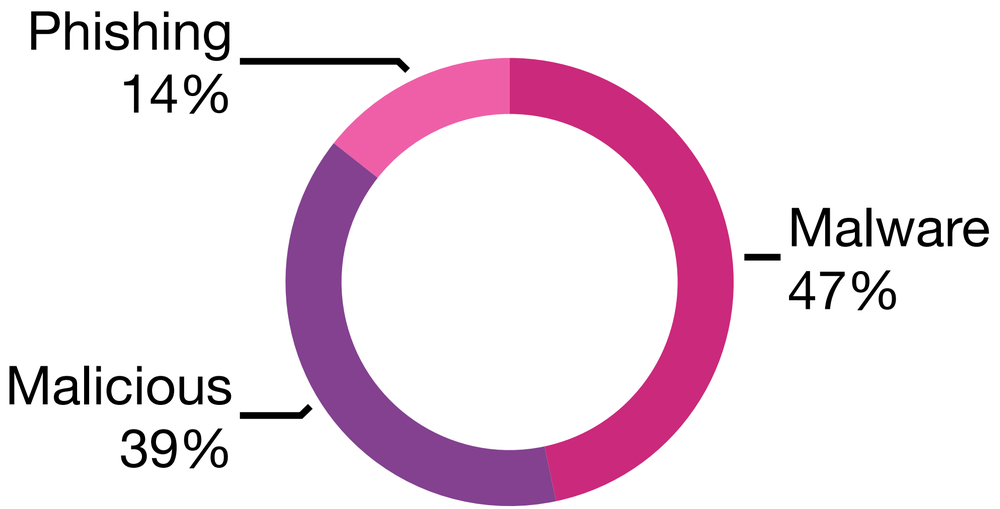


The recruitment process is very suspicious, with a mix of legitimate and malicious channels

Qualify and Quantify Security Problems

Verify Assumptions | Recruitment | Usage

- For the 67 proxy programs, 5M traffic logs were sampled to study usage
- 9.36% of the destinations were reported to be malicious by VirusTotal

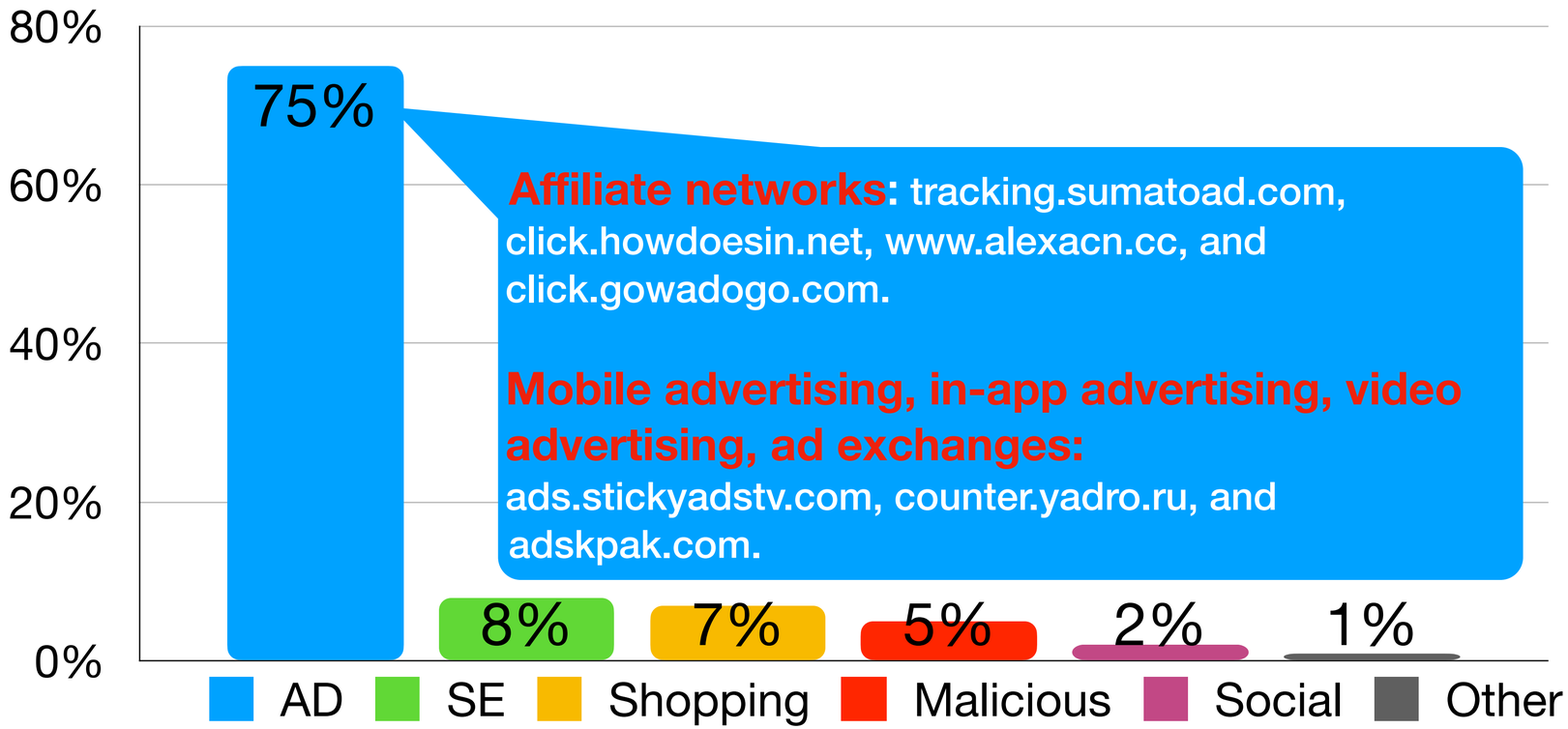


ntkrnlpa.cn,
gwf-bd.com,
fadergolf.com,
www.2345jiasu.com,
www.pf11.com,

Qualify and Quantify Security Problems

Verify Assumptions → Recruitment → **Usage**

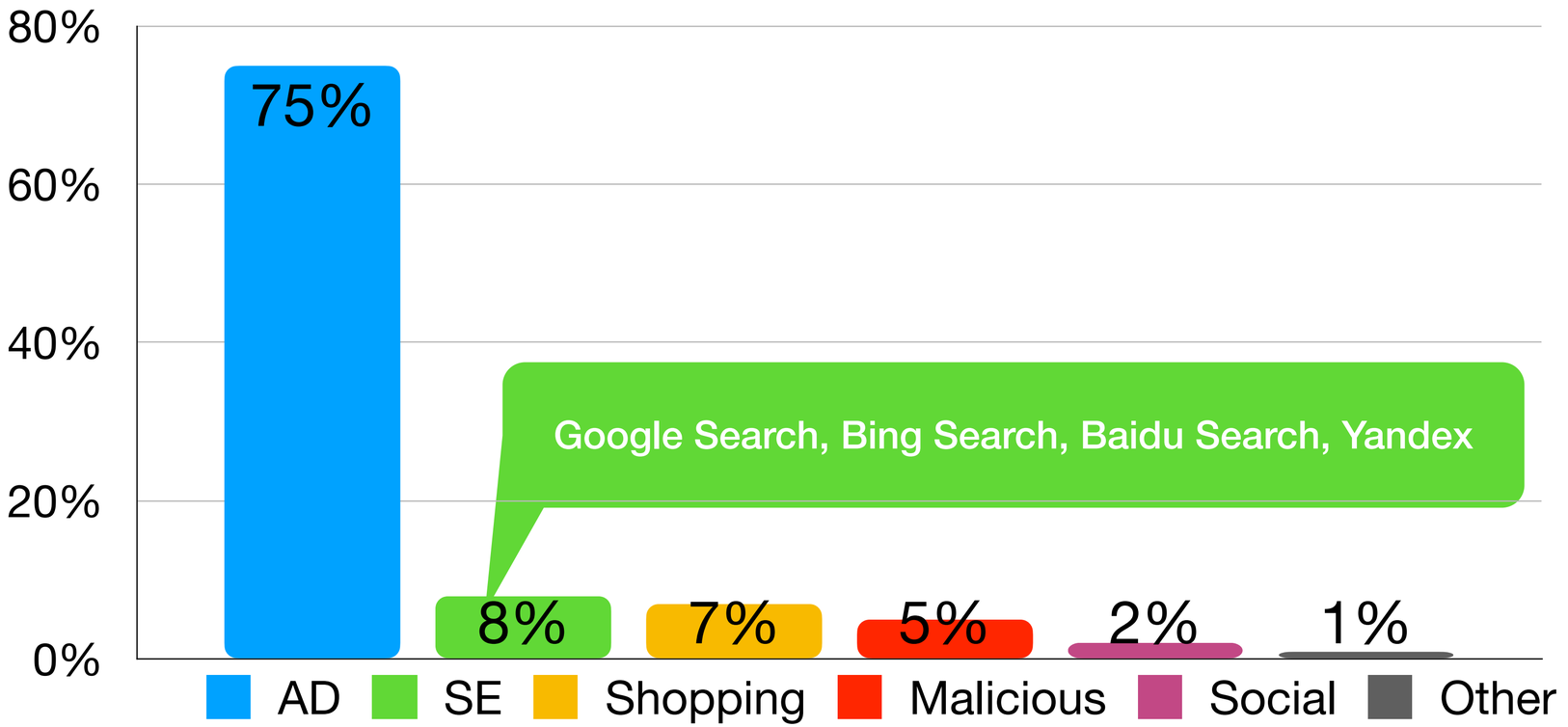
Top 1000 traffic destinations were manually studied.



Qualify and Quantify Security Problems

Verify Assumptions → Recruitment → **Usage**

Top 1000 traffic destinations were manually studied.



Qualify and Quantify Security Problems

Verify Assumptions → Recruitment → Usage

We indeed observed a large volume of abuse traffic including ad fraud, Black Hat SEO, etc.

Misc. Findings

Connection between proxy providers

Risk to the local network

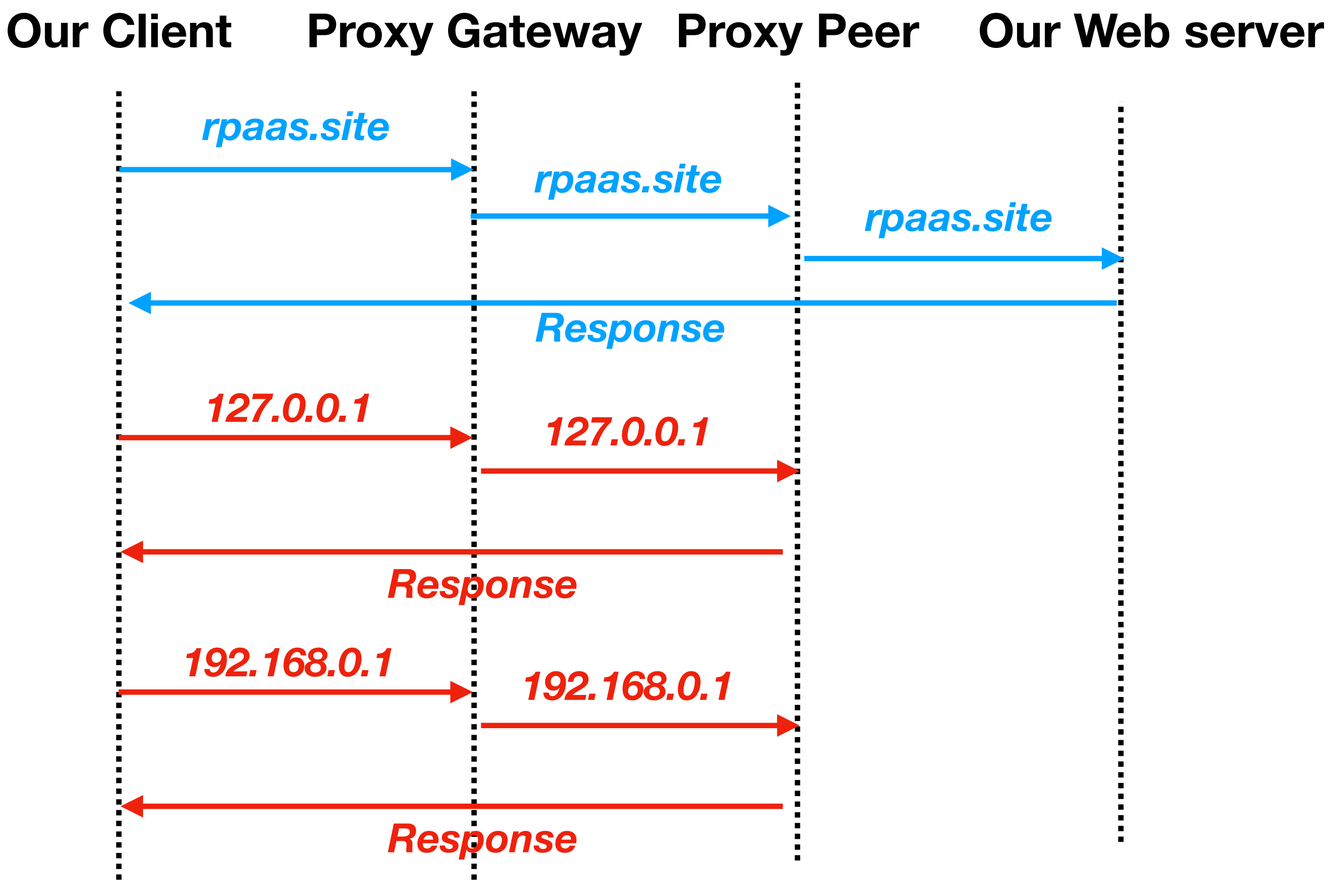
- **Proxies Online and Geosurf are the same proxy provider**
- **IAPS Security is some kind of reseller for Luminati**

Misc. Findings

Connection between proxy providers

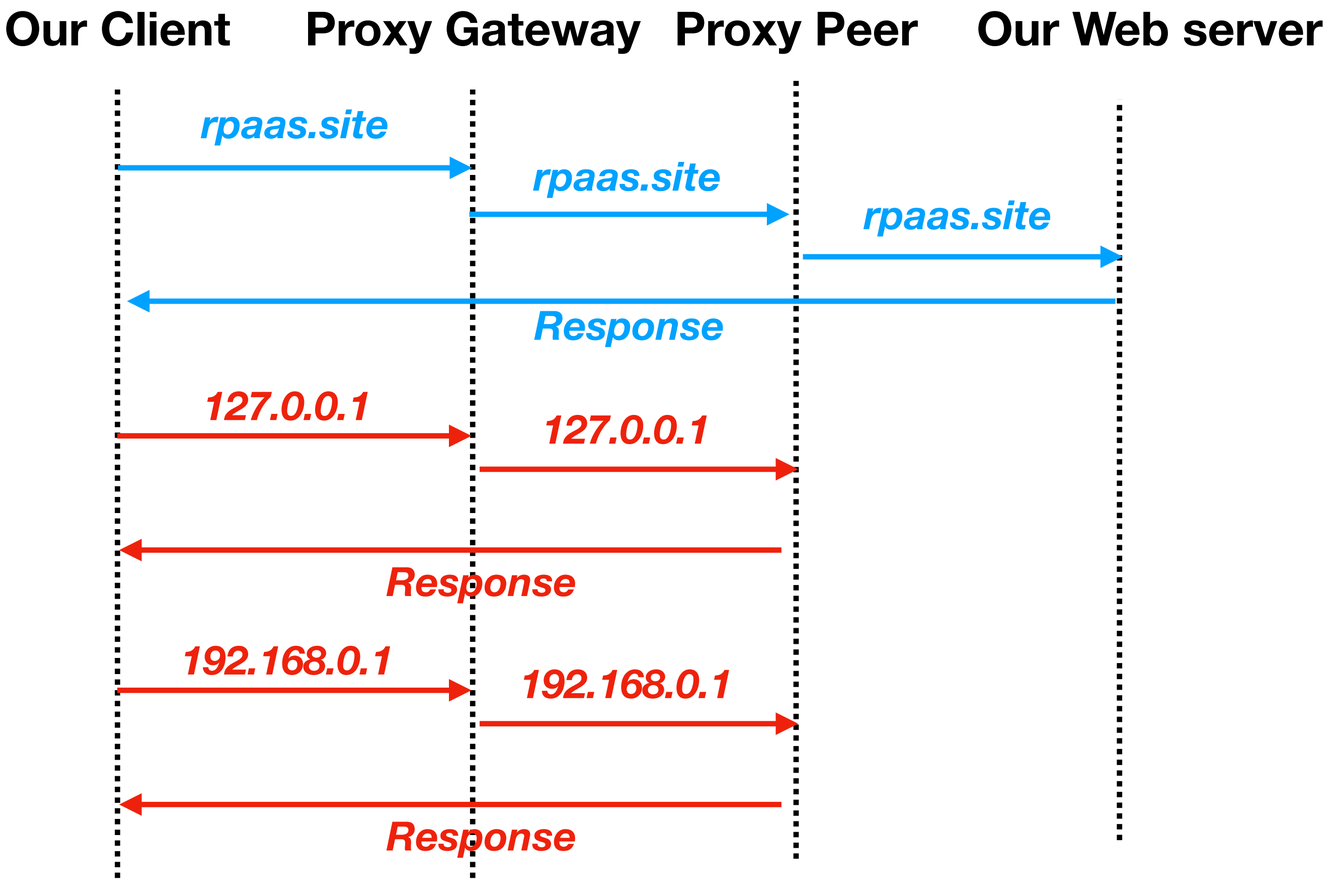
Risk to the local network

3 out of 5 providers allow local traffic



Risk to the local network

3 out of 5 providers allow traffic towards local hosts



Looking Back

Millions of residential IPs with high evasiveness

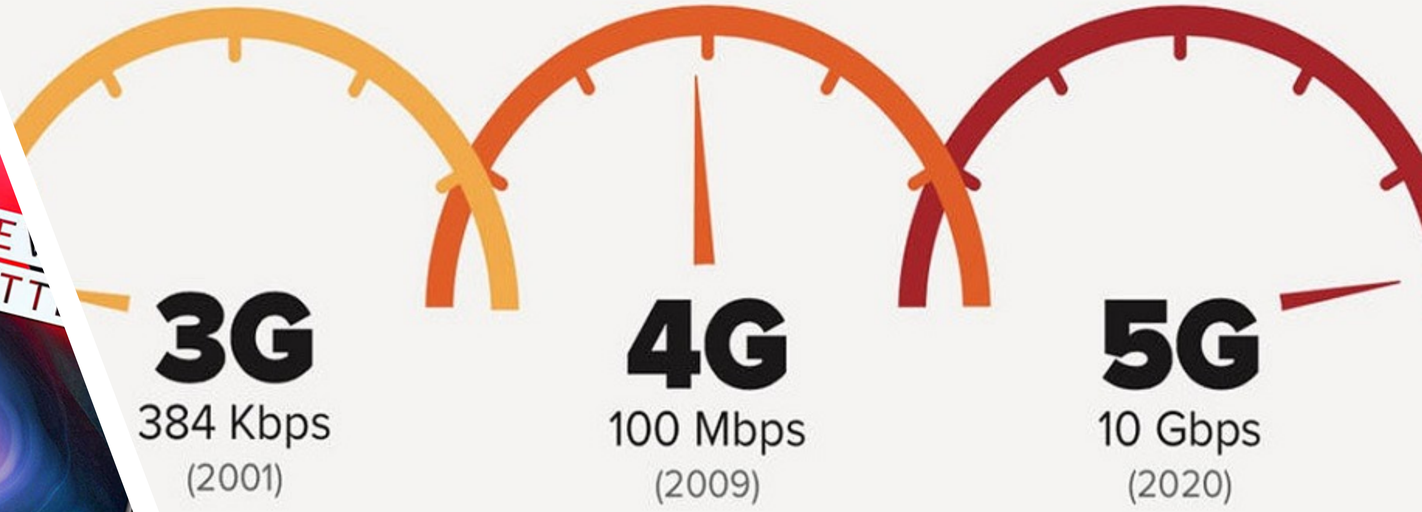
A prosperous ecosystem with higher prices and more service providers

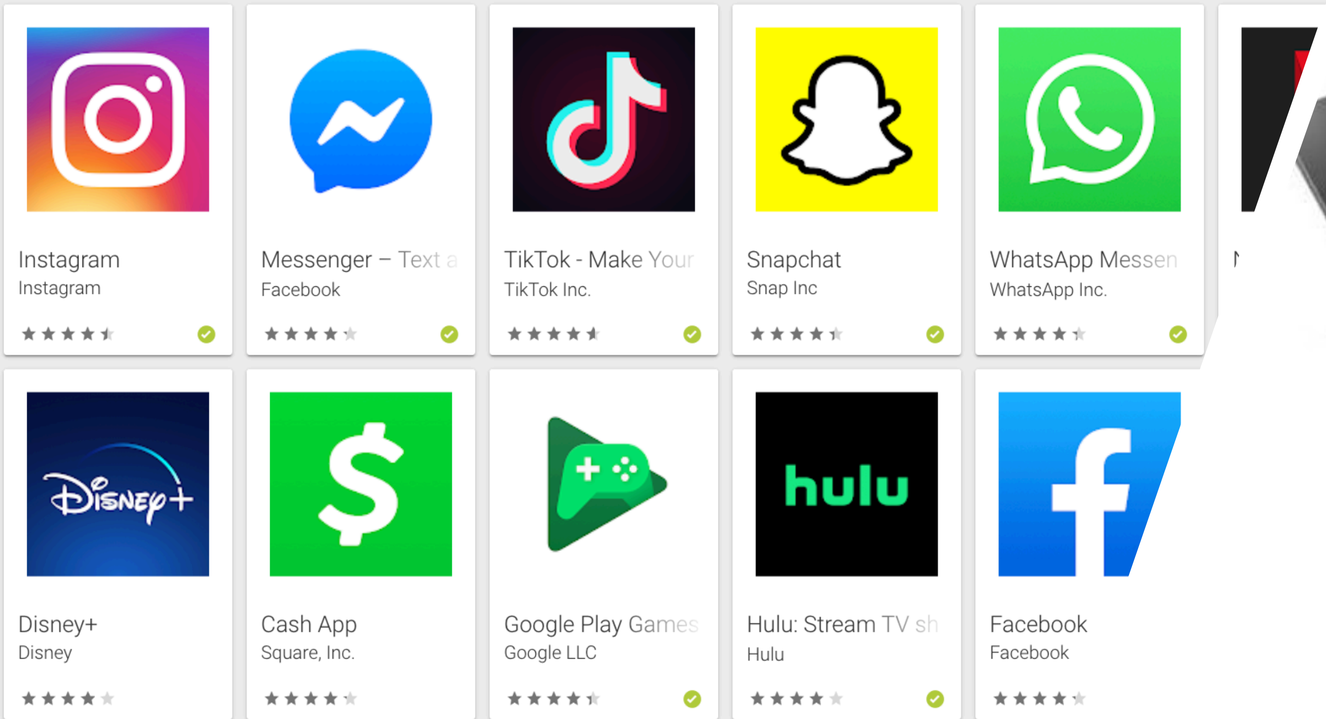
Potential threats to local network environments

Problematic recruitment: a mix of legitimate and suspicious channels

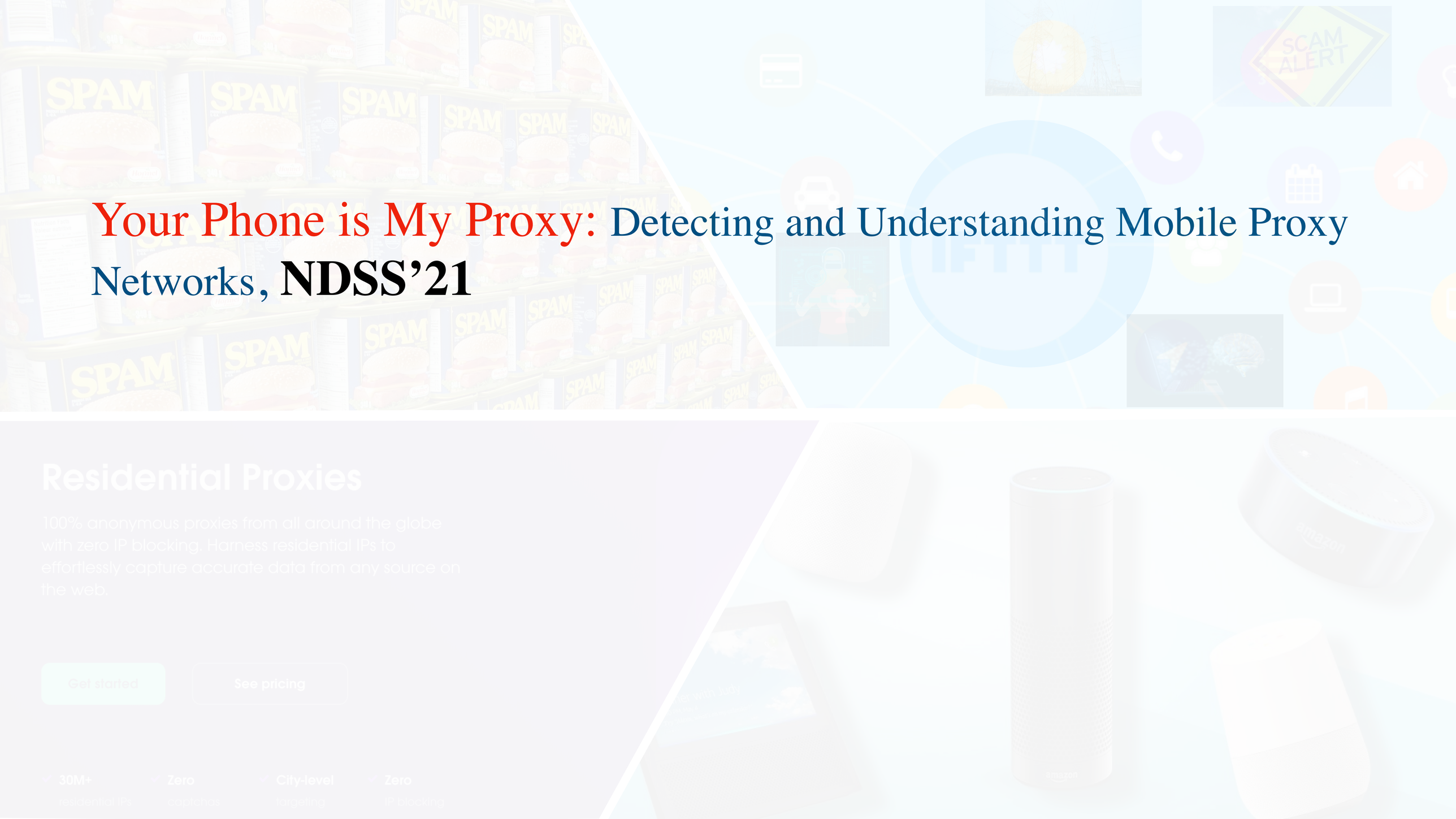
Powerful infrastructure for online abuse activities

Promising and stealthy monetization channels for compromised devices



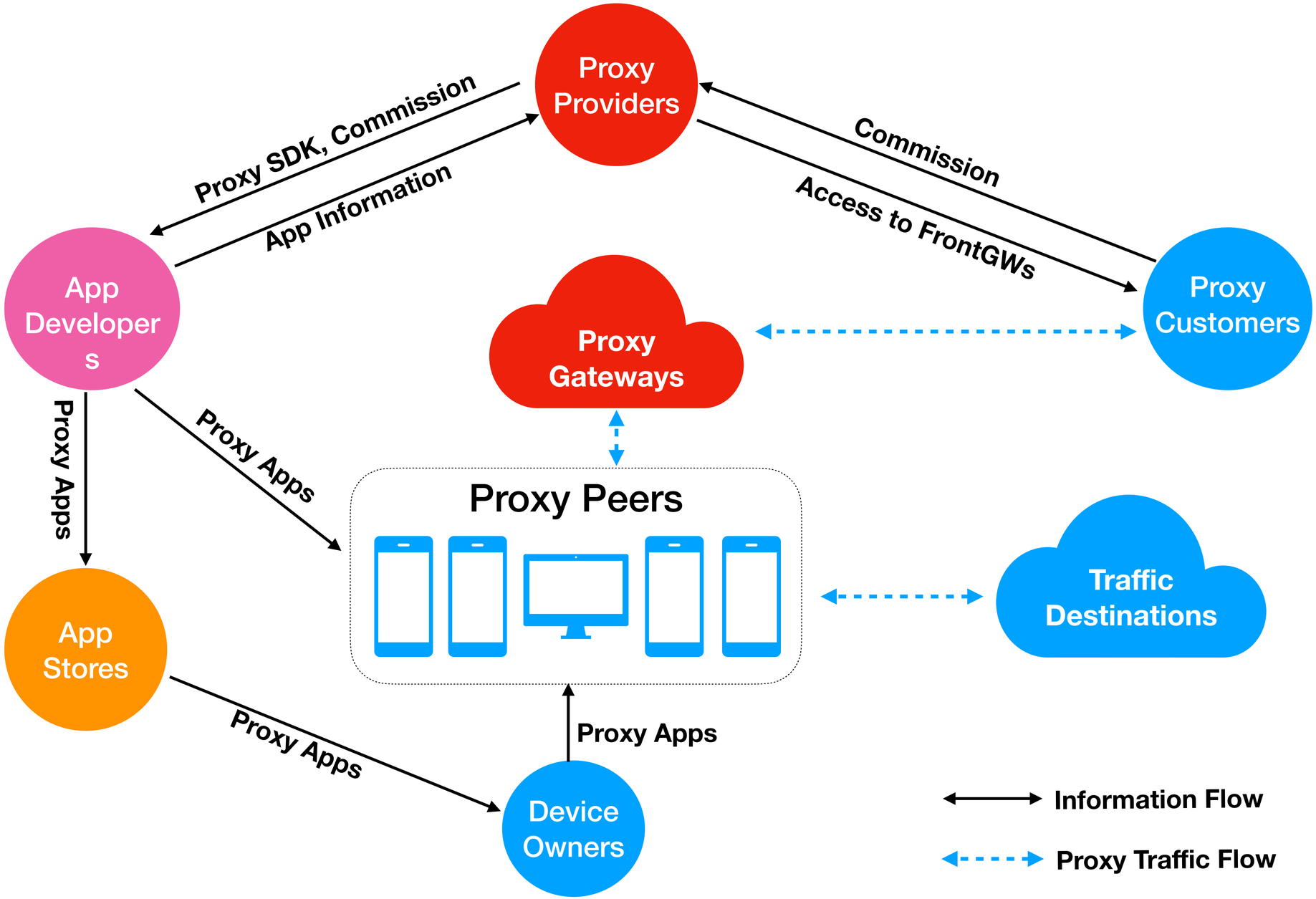


Pictures from Google Images



Your Phone is My Proxy: Detecting and Understanding Mobile Proxy Networks, **NDSS'21**

Abstract the Ecosystem



Reason about Security Problems

Assumptions → **Proxy SDK is an attractive monetization option**

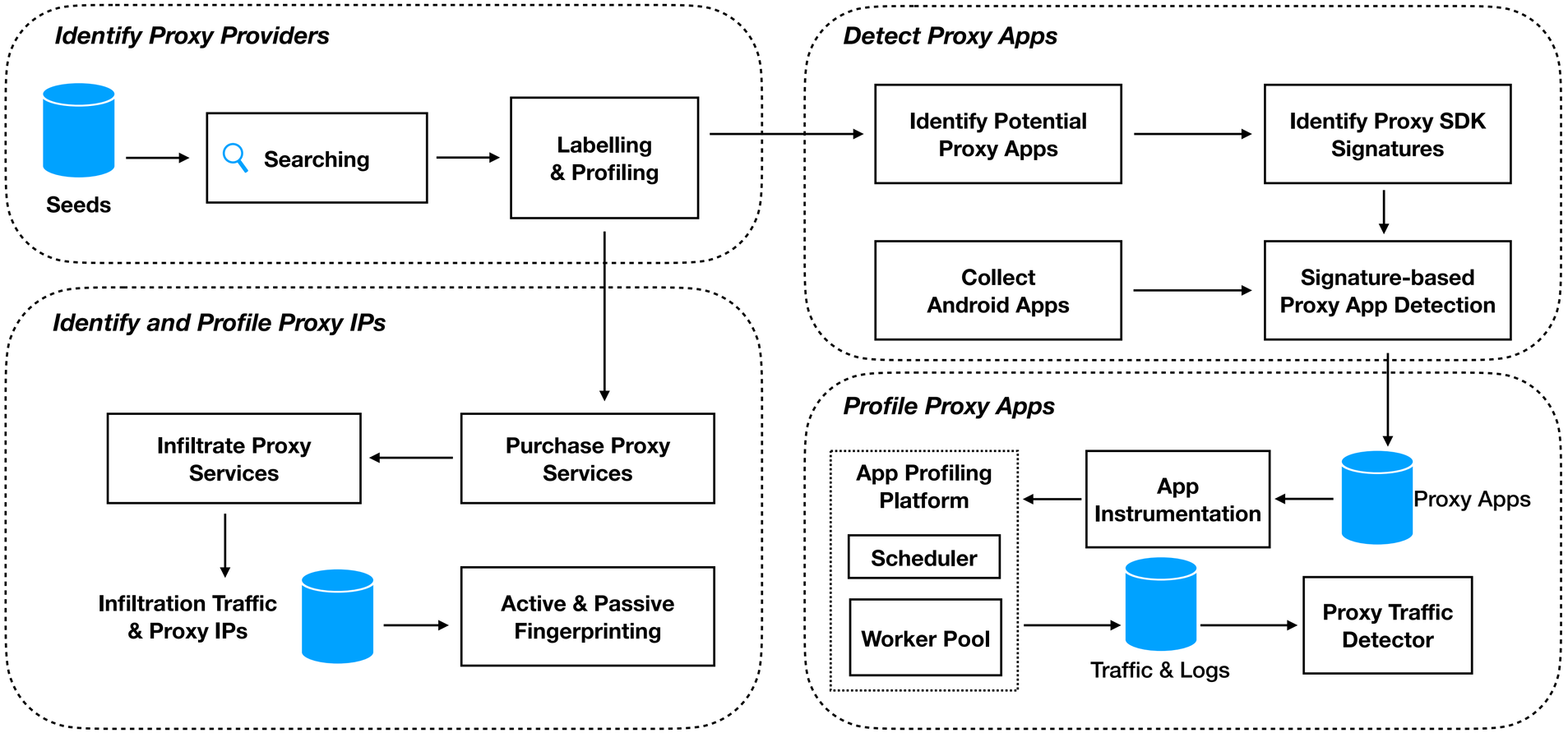
3rd-party Apps (Proxy Apps) → **Security risks to and from proxy apps**

- **Malicious apps?**
- **Damage the reputation of benign apps?**

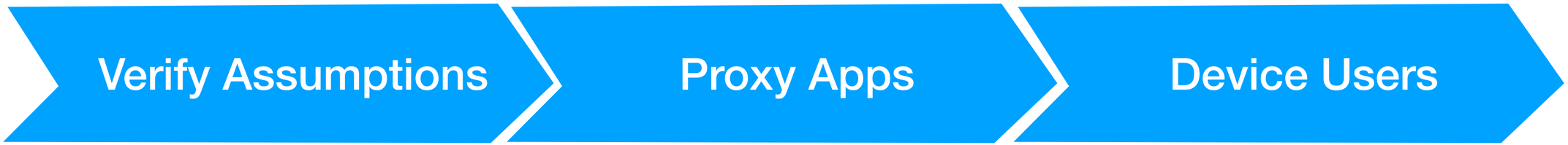
Device Users (Device Owners) → **What are the security and privacy risks to devices and their users?**

- **Users' awareness and willingness of relaying traffic using proxy SDKs?**
- **Any abuse of users' device resources especially cellular data?**

Methodology



Qualify and Quantify Security Problems



Qualify and Quantify Security Problems

Verify Assumptions → Proxy Apps → Device Users

Proxy Provider	Supported Platforms	Pricing Policy
Luminati	Android, Windows, Mac OS	$50K/1M MAU (Monthly Active Users) By Nov 2019
MonkeySocks	Android	
Oxylabs	Android	
IPninja	Android, Windows, Chrome Extension	

Qualify and Quantify Security Problems

Verify Assumptions → Proxy Apps → Device Users

Proxy SDK is an attractive and profitable app monetization option

Qualify and Quantify Security Problems

Verify Assumptions → **Proxy Apps** → Device Users

Scale

Malicious Apps

Damage to App Reputation

- A signature-based proxy app detector
- Scanned 2M Android APKs
- 1, 701 Android proxy APKs of 963 proxy Apps
- All ever in Google Play, but only 171 left by Aug 2019.

Qualify and Quantify Security Problems

Verify Assumptions → Proxy Apps → Device Users

Scale

Malicious Apps

Damage to App Reputation

300M+ installations in sum

Qualify and Quantify Security Problems

Verify Assumptions → **Proxy Apps** → Device Users

Scale

Malicious Apps

Damage to App Reputation

- Most were removed from Google Play
- For each proxy app, extract all its historical versions
- Analyzing their VirusTotal analysis reports

Before integrating proxy SDKs, 39.50% got flagged as malicious, among which, many are phishing apps

Qualify and Quantify Security Problems

Verify Assumptions → **Proxy Apps** → Device Users

Scale

Malicious Apps

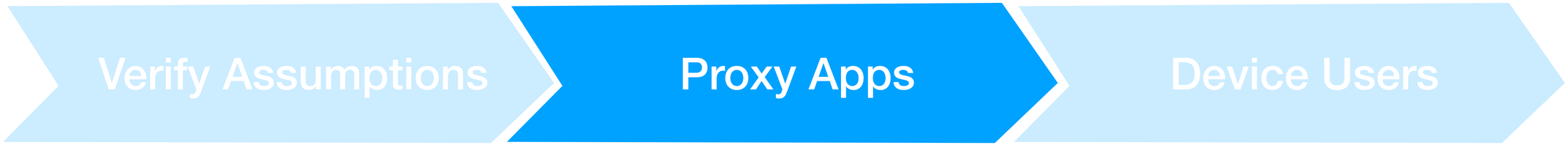
Damage to App Reputation

After integrating proxy SDKs, 86.32% of the proxy APKs got flagged as malicious.

Google Play has adapted their developer's policies since June 2019, and consider this as device and network abuse

"Apps that facilitate proxy services to third parties may only do so in apps where that is the primary, user-facing core purpose of the app."

Qualify and Quantify Security Problems



Security risks are bi-directional, incurred to and from proxy apps

Qualify and Quantify Security Problems

Verify Assumptions → Proxy Apps → Device Users

User Consent

Abuse of Device Resource

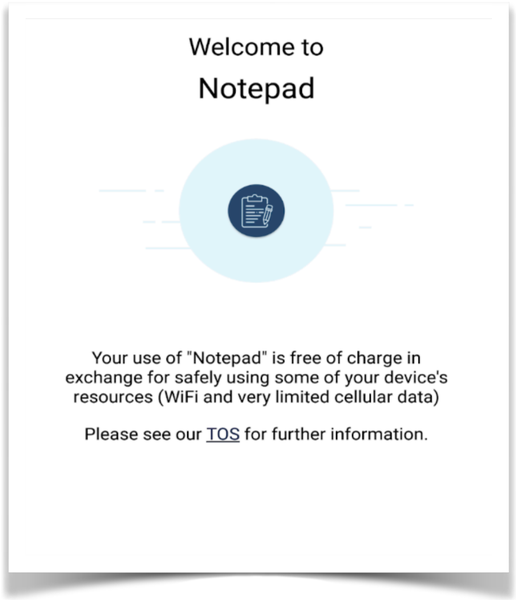
Qualify and Quantify Security Problems

Verify Assumptions → Proxy Apps → Device Users

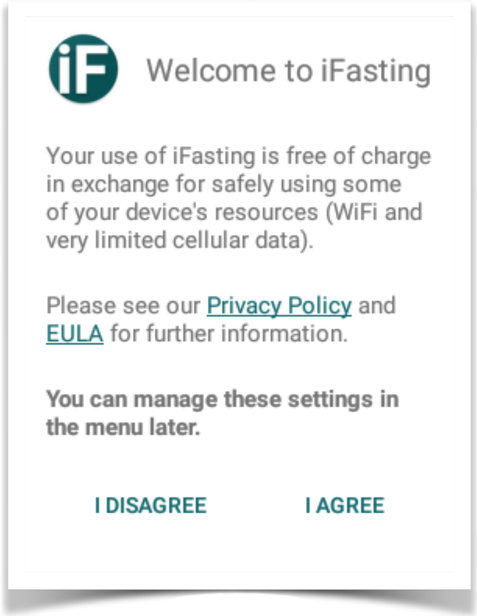
User Consent

Abuse of Device Resource

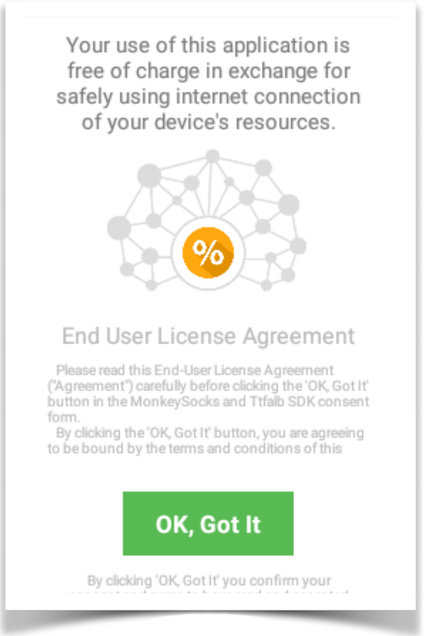
Your use of "App XX" is free of charge in exchange for safely using some of your device's resources (WiFI and very limited cellular data).



Luminati



Oxylabs



Monkeysocks

Qualify and Quantify Security Problems

Verify Assumptions → Proxy Apps → Device Users

User Consent

Abuse of Device Resource

User awareness of traffic relaying consent

User willingness of traffic relaying in exchanges of rewards

90% participants cannot interpret the consent correctly.

34% are “not at all” willing, only 8% allows relaying using cellular data.

Qualify and Quantify Security Problems

Verify Assumptions → Proxy Apps → Device Users

User Consent

Abuse of Device Resource

We identified a secret proxy SDK X

No user consent, aggressively relay traffic!

Colocated in 24 proxy apps of Luminati, and 3 proxy apps of Monkeysocks

Qualify and Quantify Security Problems

Verify Assumptions → Proxy Apps → Device Users

User Consent

Abuse of Device Resource

Adapt relaying behaviors based on system conditions

Consume both WIFI and cellular data

It is out of control for the users

Qualify and Quantify Security Problems

Verify Assumptions → Proxy Apps → Device Users

User Consent

Abuse of Device Resource

Either no user consent or it is very ambiguous

Resource usage is out of control for users

Residential Proxies

100% anonymous proxies from all around the globe with zero IP blocking. Harness residential IPs to effortlessly capture accurate data from any source on the web.

SP'19, Resident Evil: Understanding Residential IP Proxy as a Dark Service

- 30M+ residential IPs
- Zero captchas
- City-level targeting
- Zero IP blocking

Understanding and Classifying Network Traffic of Residential Proxies

Mobile Proxy Network

NDSS'21, Your Phone is My Proxy: Detecting and Understanding Mobile Proxy Networks

Start Now!

Pricing

CCS'22 An Extensive Study of Residential Proxies in China

Looking Back

- The ecosystem of RESIPs serves as a monetization channel for malware across platforms
- RESIPs provide a powerful infrastructure for cybercrime activities
- RESIPs incur non-negligible risks to individuals (their devices), corporate networks, and web services
- Our studies are well-acknowledged by Abuse Research at Google, Google Play Security, and Location Integrity at Facebook, etc.

The datasets and code are made publicly available at .

Why Can This happen?

- Coarse-grained access control for network/bandwidth on edge devices
  - Any remote network addresses to visit
  - Any size of bandwidth to consume
- The lack of remote attestation for the fine-grained authorization of the device owners

A series of study on **emerging IoT platforms**:

- An Empirical Characterization of IFTTT: Ecosystem, Usage, and Performance, **IMC'17**
- Dangerous Skills: Understanding and Mitigating Security Risks of Voice-Controlled Third-Party Functions on Virtual Personal Assistant Systems, **SP'19**
- Command Hijacking on Voice-Controlled IoT in Amazon Alexa Platform, **Under Submission**

Dangerous Skills: Understanding and Mitigating Security Risks of Voice-Controlled Third-Party Functions on Virtual Personal Assistant Systems, **SP'19**

★**CSAW 2019 Best Paper Award**

Voice Assistant Devices

Alexa, play Today's Hits on Pandora

Alexa, turn on Living Room lights

Alexa, ask PayPal to send 10 dollars to Sam

Alexa, ask Medical Assistant to give me my diagnosis

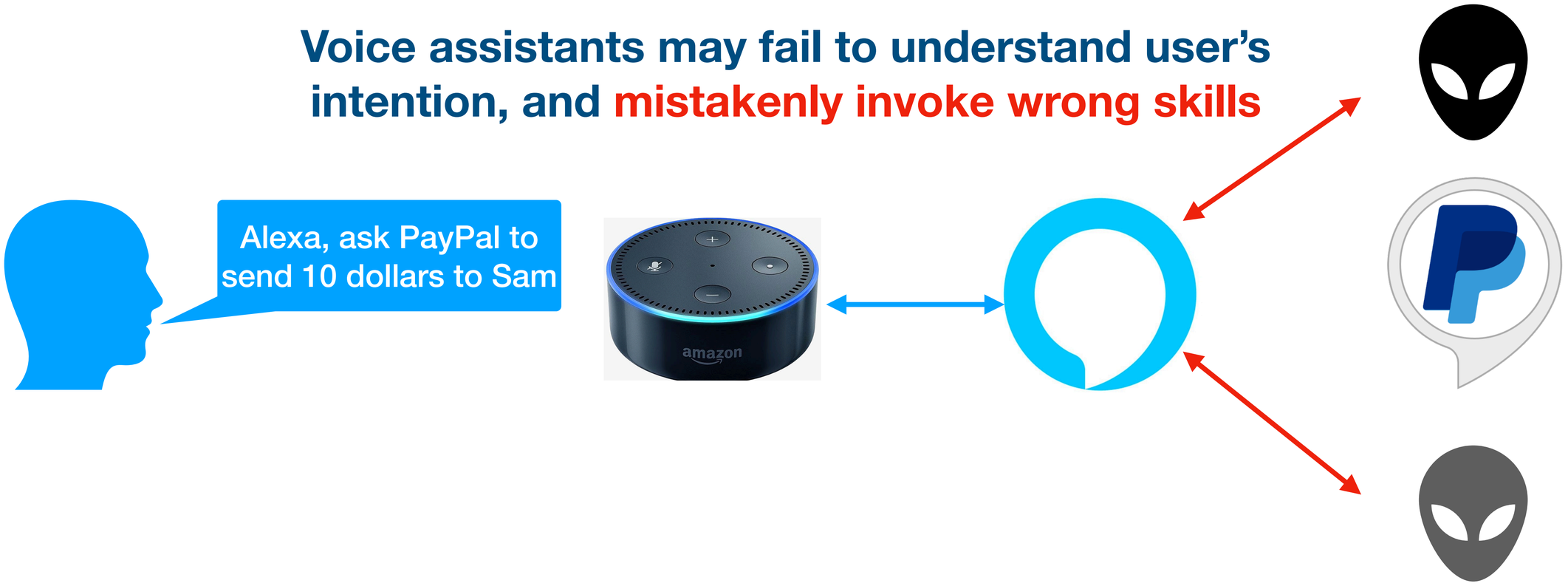
Abstract the Ecosystem

Voice assistants work like a relay, proxying and translating conversation between users and skills

User | **Smart Speaker** | **Voice Assistant Cloud** | **Third-party Skill Clouds**

Voice Squatting

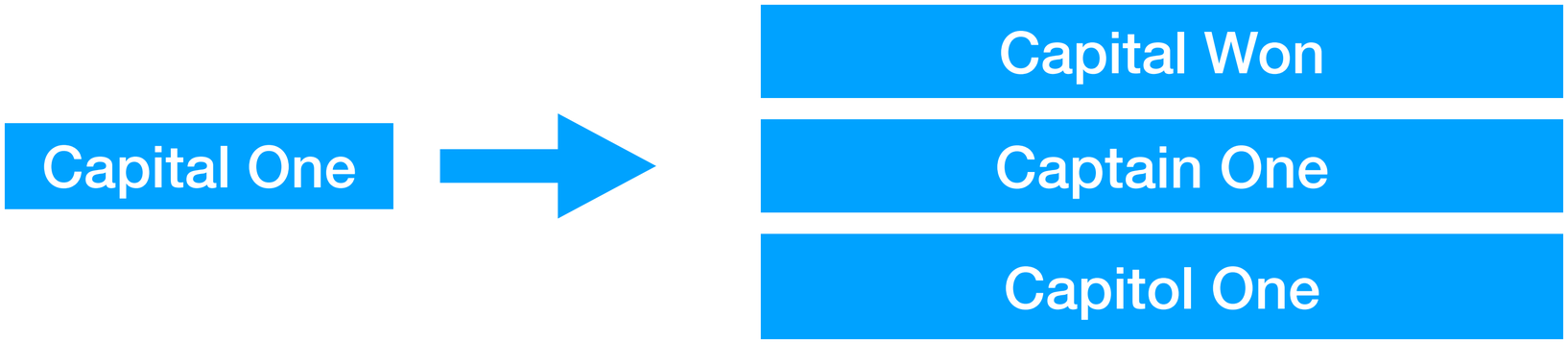
Voice assistants may fail to understand user's intention, and mistakenly invoke wrong skills



User | **Smart Speaker** | **Voice Assistant Cloud** | **Third-party Skill Clouds**

Qualify and Quantify Security Problems

Voice Squatting through similar pronunciation



Alexa			Google		
Amazon TTS	Google TTS	Human	Amazon TTS	Google TTS	Human
10/17	12/17	> 50%	4/7	2/4	> 50%

Qualify and Quantify Security Problems

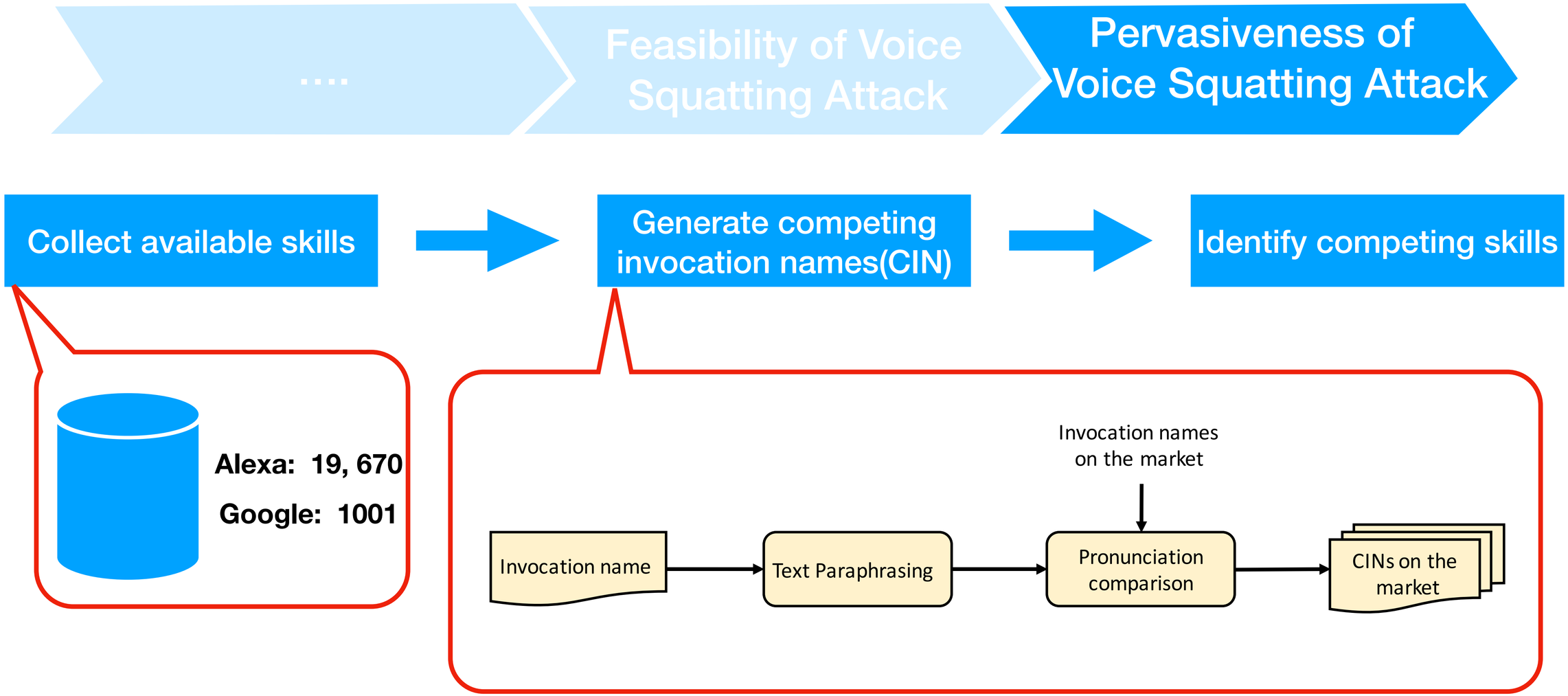
Voice Squatting through invocation name extending

Qualify and Quantify Security Problems

Voice Squatting through invocation name extending

	Alexa	Google
invocation name + “please”	10/10	0/10
“my” + invocation name	7/10	0/10
“the” + invocation name	10/10	0/10
invocation name + “app”	10/10	10/10
“mai” + invocation name	-	10/10
invocation name + “plese”	-	10/10

Qualify and Quantify Security Problems

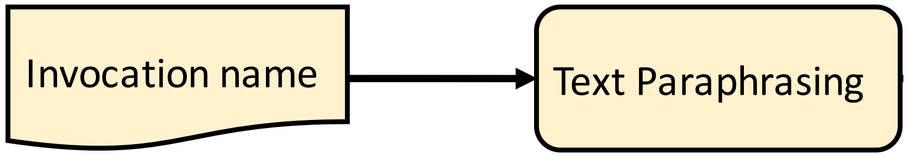


Qualify and Quantify Security Problems

Invocation name

Capital One

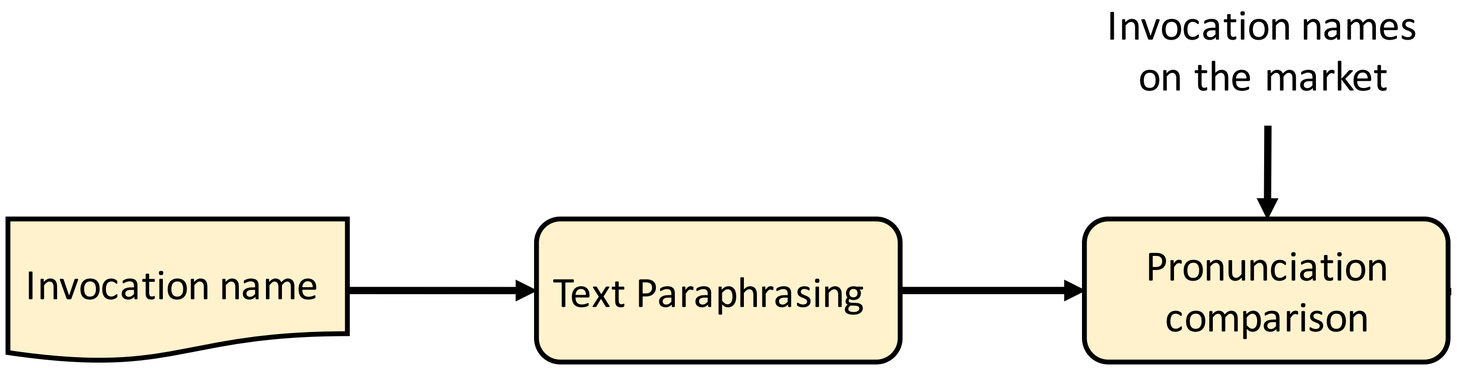
Qualify and Quantify Security Problems

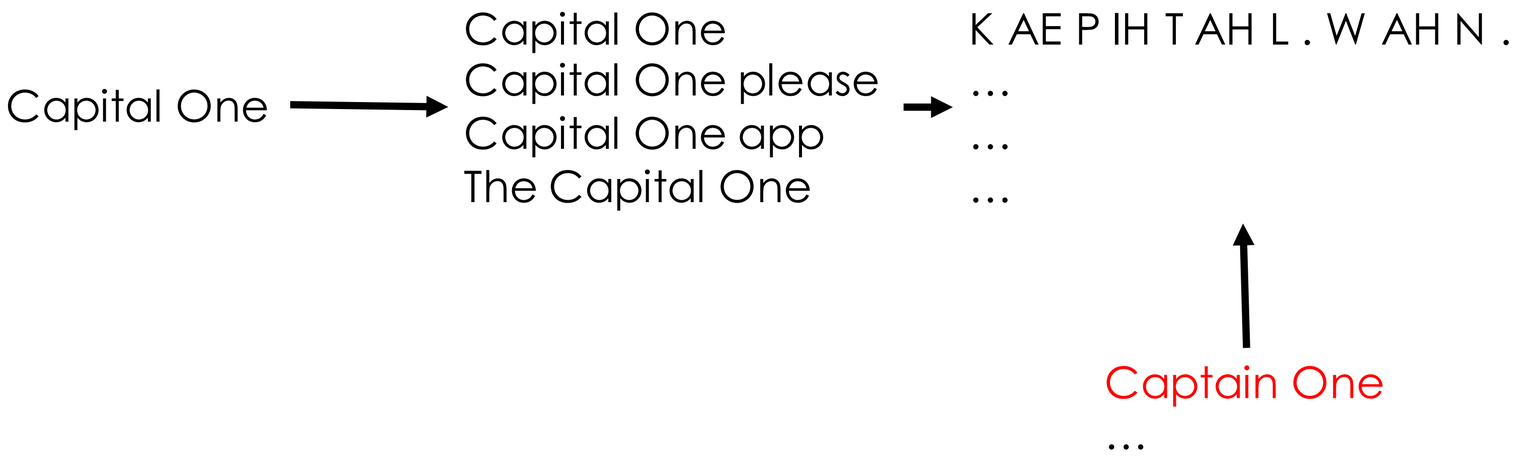


Capital One → Capital One, Capital One please, Capital One app, The Capital One

Qualify and Quantify Security Problems

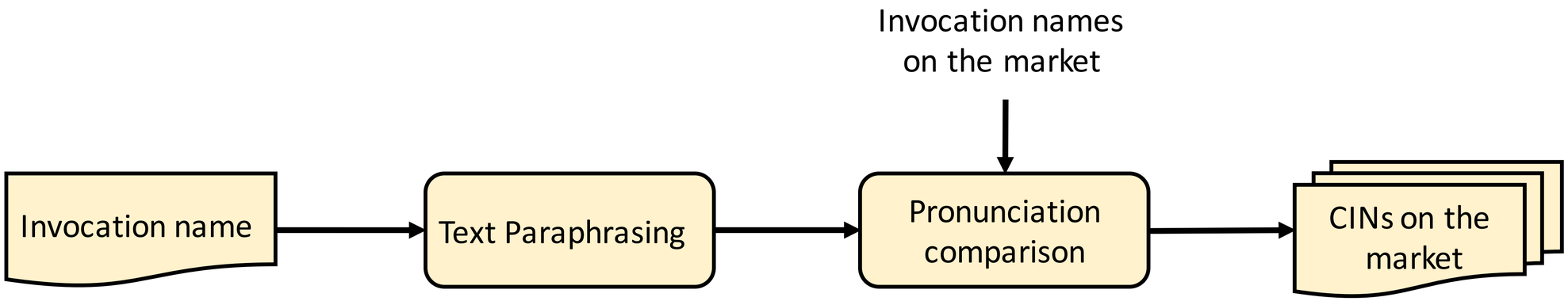
…. | Feasibility of Voice Squatting Attack | Pervasiveness of Voice Squatting Attack

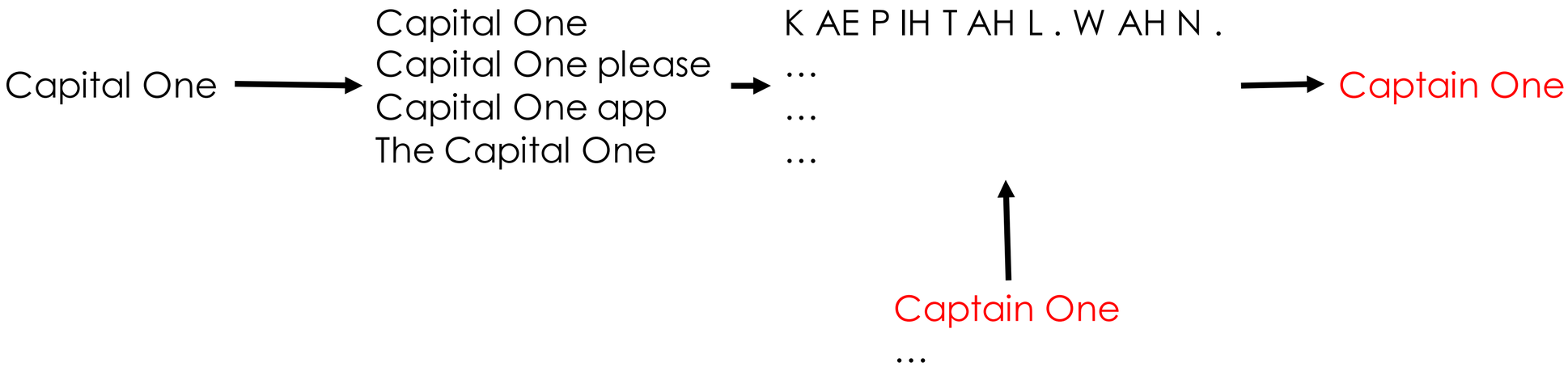




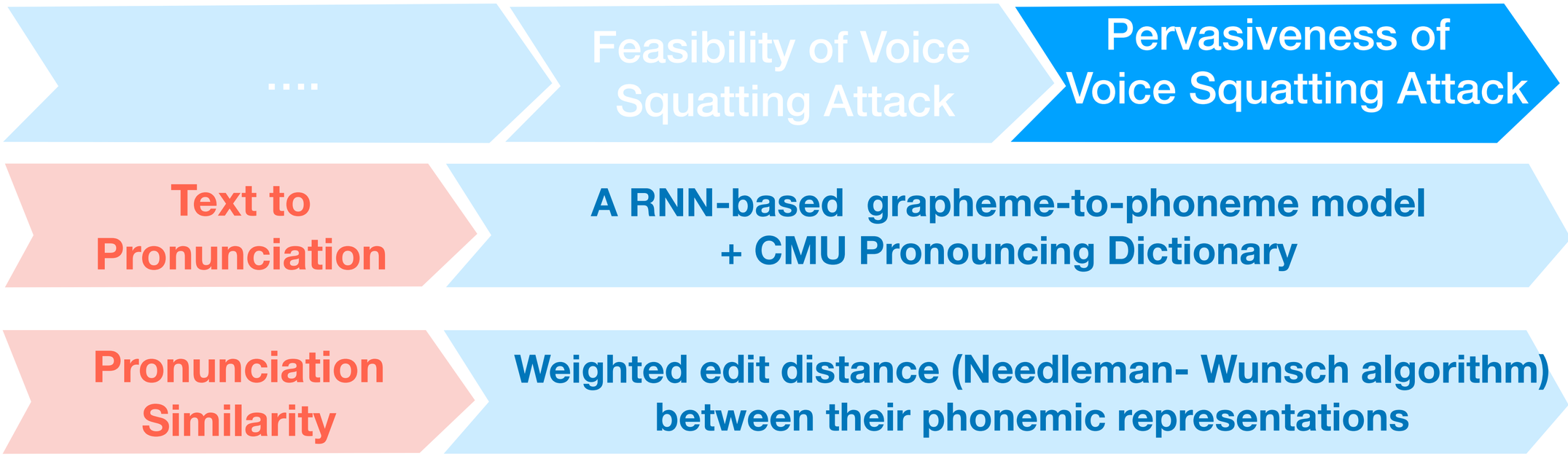
Qualify and Quantify Security Problems

…. | Feasibility of Voice Squatting Attack | Pervasiveness of Voice Squatting Attack





Qualify and Quantify Security Problems



A phoneme edit operation is less significant when it frequently appears between two alternative pronunciations of a given English word.

Weighted cost to replace phoneme α with phoneme β

$$WC(\alpha,\beta) = 1 - \frac{SF(\alpha,\beta) + SF(\beta,\alpha)}{F(\alpha) + F(\beta)}$$

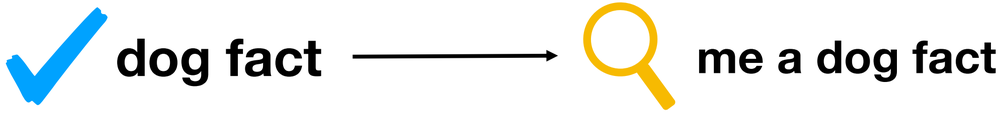
Frequency of replacing β with α in all pronunciation pairs

Frequency of β in all pronunciation pairs

Qualify and Quantify Security Problems

.... → Feasibility of Voice Squatting Attack → **Pervasiveness of Voice Squatting Attack**

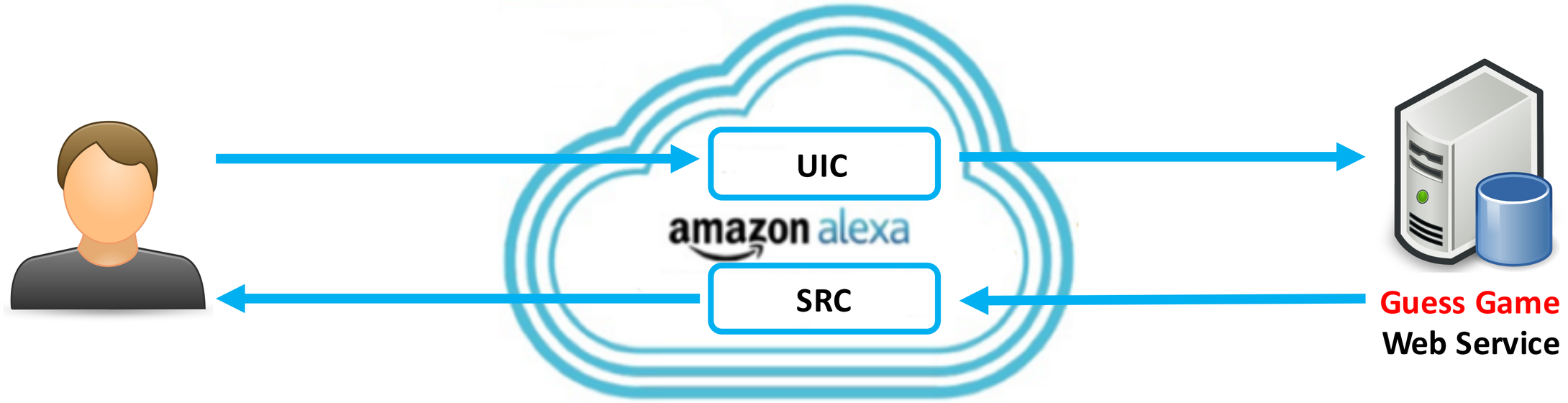
- **19% (3718) skills： similar pronunciation**
  - 66 skills were named as “cat facts”, and provided similar functions.
- **2.7% (531) skills: similar pronunciation， but different spelling**
- **1.8% (345) skills: longest prefix matching (name extending)**
- **Interesting cases**



Voice Masquerading Attack (VMA)

Skill switching is not well supported, allowing a skill to masquerade itself as other skills or even the system

Defend VMA through Runtime Auditing



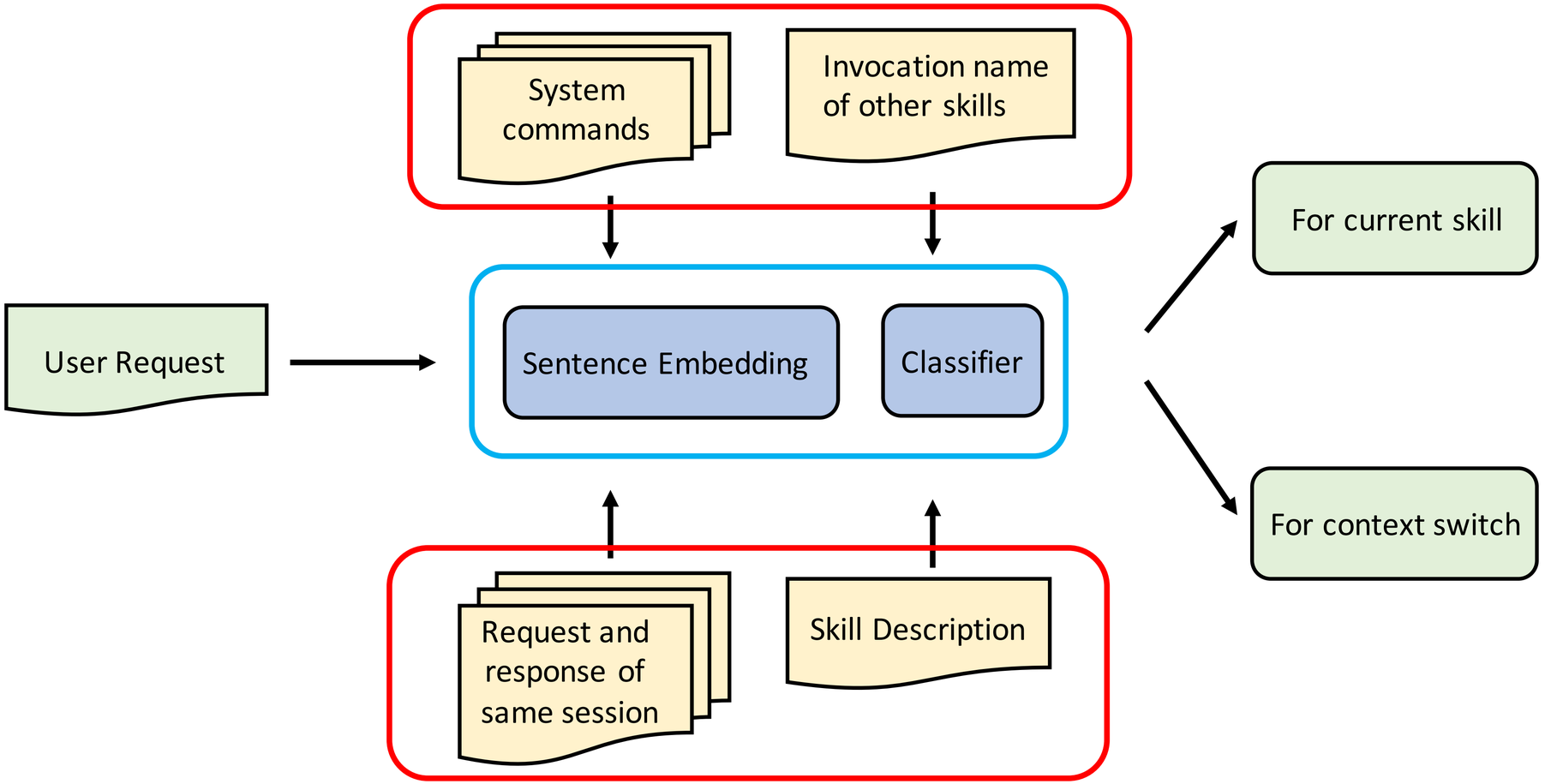
UIC: User Intention Classifier

Classify user's intention as context switching or not

SRC: Skill Response Checker

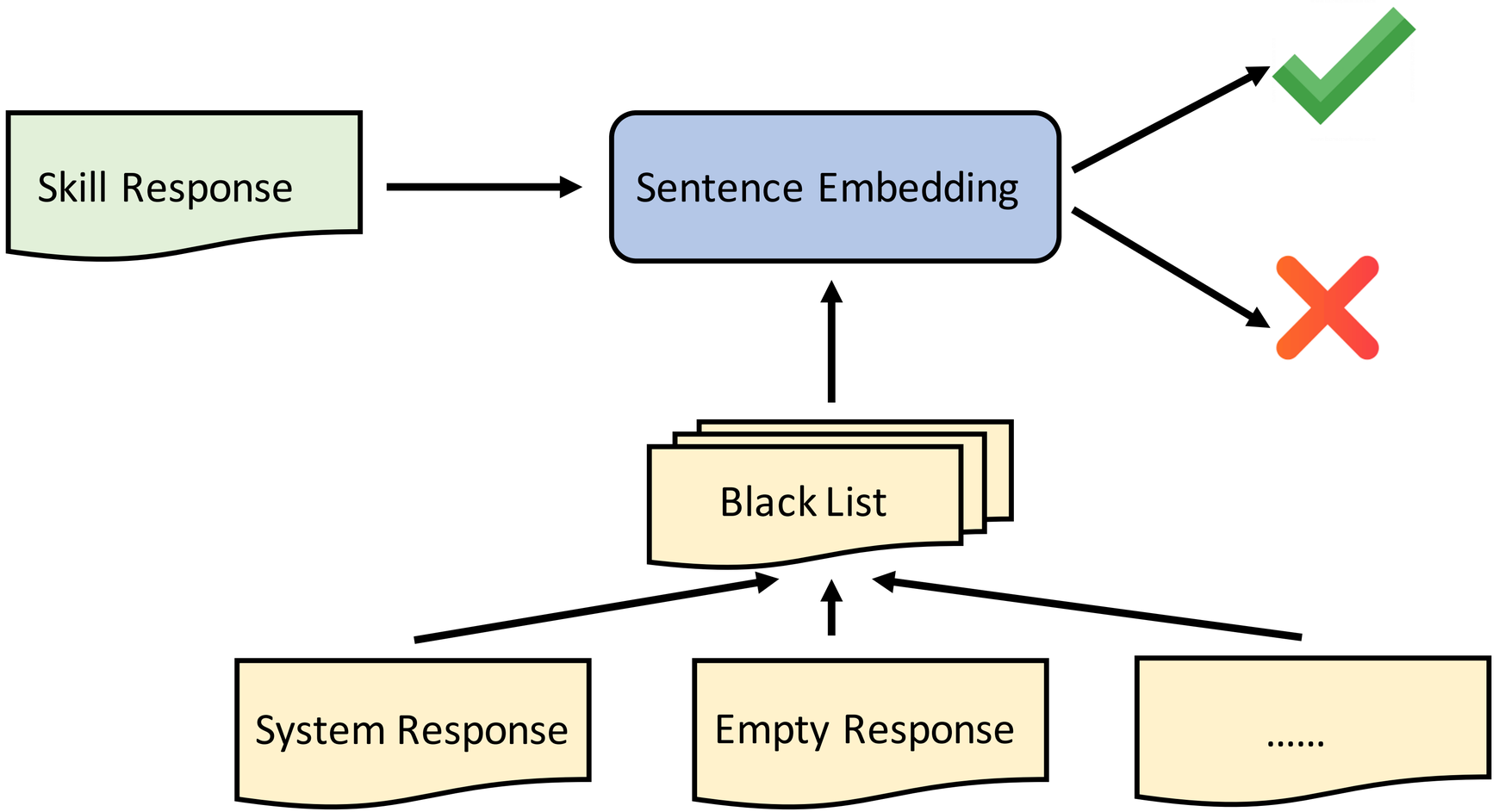
Identify suspicious skill response, such as fake skill recommendation

Defend VMA through Runtime Auditing



User Intention Classifier (UIC)

Defend VMA through Runtime Auditing



Skill Response Checker (SRC)

Looking Back

- Two attack scenarios: Voice Squatting & Voice Masquerading
- Both attacks were found to be feasible or even pervasive
- We explored a set of mitigation solution: CIN generator, User Intention Classifier, and Skill Response Checker.
- Both platform vendors have acknowledged our attacks

Attack Demos: https://sites.google.com/site/voicevpasec/

Why Can This happen?

- The low visibility of voice user interfaces (VUIs)
- A design tradeoff between usability and security
- AI systems are known to be vulnerable to adversarial examples

A series of study on cybercrime, e.g., spam/fraud, search poisoning, etc

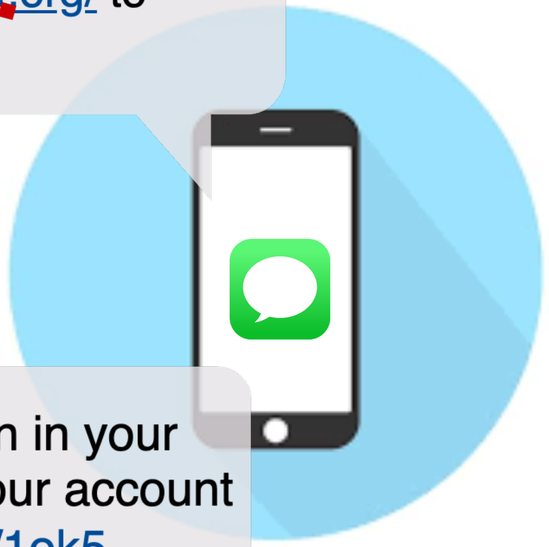
- Under the Shadow of Sunshine: Understanding and Detecting Bulletproof Hosting on Legitimate Service Provider Networks, **SP'17**
- Game of Missuggestions: Semantic Analysis of Search-Autocomplete Manipulations, **NDSS'18**
- Cracking Wall of Confinement: Understanding and Analyzing Malicious Domain Takedowns, **NDSS'19**
- Clues in Tweets: Twitter-Guided Discovery and Analysis of SMS Spam, **CCS'22**

Clues in Tweets: Twitter-Guided Discovery and Analysis of SMS Spam

, **CCS'22**

URGENT: UKGOV has issued a payment of 458 GBP to all residents as part of its promise to battle COVID 19. TAP here https://uk-covid-19.webredirect.org/ to apply

Scam!



You have (1) BitCoin in your account. Confirm your account here: http://cutt2l.ink/1ok5 Current market value 7393.67 GBP

Scam!

Complaints about spam texts were up 146% last year. Now, the FCC wants to take action

October 19, 2021 · 12:48 PM ET

BECKY SULLIVAN

On Monday, the agency's acting chairwoman, Jessica Rosenworcel, announced she will ask the commission to begin creating a new set of federal rules that would govern spam texts, like those in place now for robocalls.

Chip Somodevilla/Getty Images

Source: https://www.npr.org/2021/10/19/1047303425/complaints-about-spam-texts-fcc-robocalls

Publicly available SMS spam datasets are small-scaled, out-of-dated, or region-specific

SMS Spam Collection Data Set

Download: Data Folder, Data Set Description

Abstract: The SMS Spam Collection is a public set of SMS labeled messages that have been collected for mobile phone spam research.

Data Set Characteristics:	Multivariate, Text, Domain-Theory	**Number of Instances:**	5574	**Area:**	Computer
Attribute Characteristics:	Real	**Number of Attributes:**	N/A	**Date Donated**	2012-06-22
Associated Tasks:	Classification, Clustering	**Missing Values?**	N/A	**Number of Web Hits:**	447821

Introduction of FBS Spam SMS Dataset

This released dataset is used to support the CCS'20 paper *Lies in the Air: Characterizing Fake-base-station Spam Ecosystem in China*.

> **ACM Reference Format**: Yiming Zhang, Baojun Liu, Chaoyi Lu, Zhou Li, Haixin Duan, Shuang Hao, Mingxuan Liu, Ying Liu, Dong Wang and Qiang Li. 2020. Lies in the Air: Characterizing Fake-base-station Spam Ecosystem in China. In Proceedings of the 2020 ACM SIGSAC Conference on Computer and Communications Security (CCS '20), November 9–13, 2020, Virtual Event, USA. ACM, New York, NY, USA, 14 pages. https://doi.org/10.1145/3372297.3417257

How can we continuously collect fresh spam messages without incurring privacy risks?

[redacted]
@[redacted]

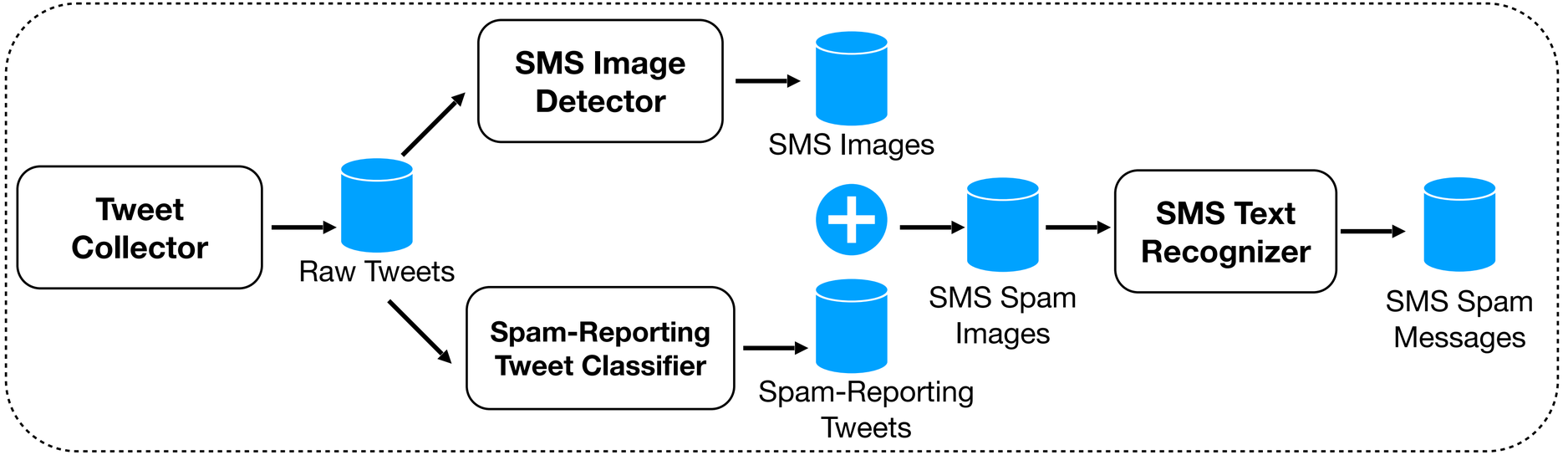
New rule for when a market becomes mainstream: when you start getting spam SMS messages about it

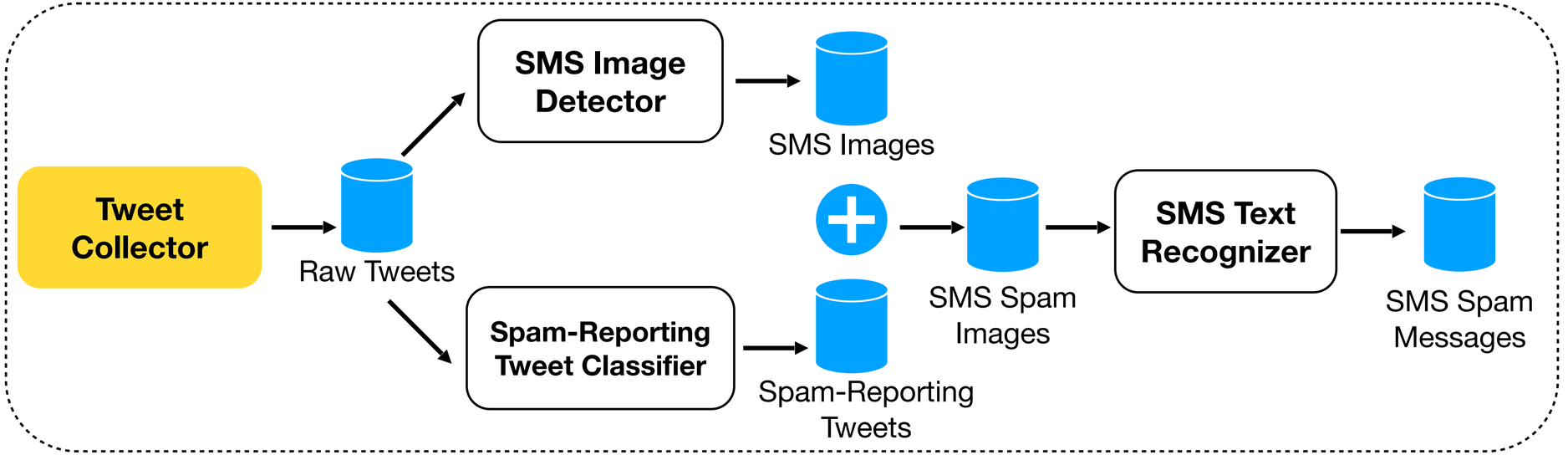
[redacted]
@[redacted]

Watch out the spammers are out! They have no morals! Please check before u click.. if in doubt Google it!

#COVID19 #scam #spam #SMS #gov

1:34 PM · Mar 22, 2020 · Twitter for Android

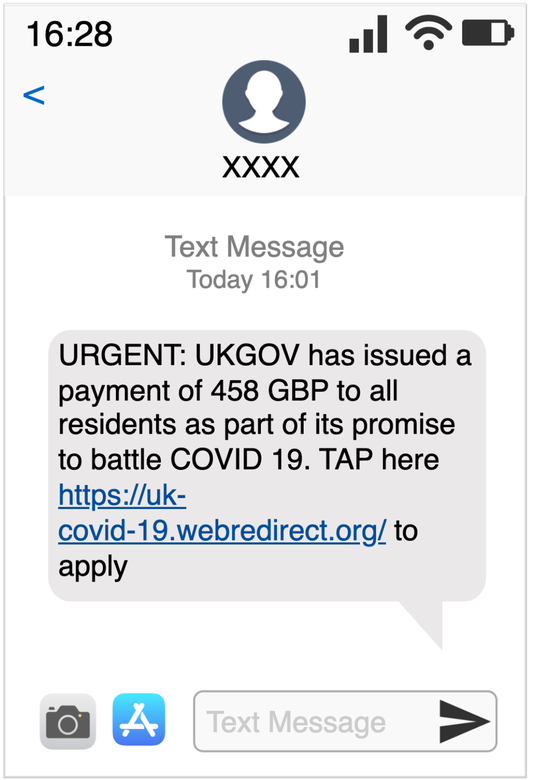

SMS Image Detector
SMS Images
Tweet Collector
Raw Tweets
SMS Spam Images
SMS Text Recognizer
SMS Spam Messages
Spam-Reporting Tweet Classifier
Spam-Reporting Tweets

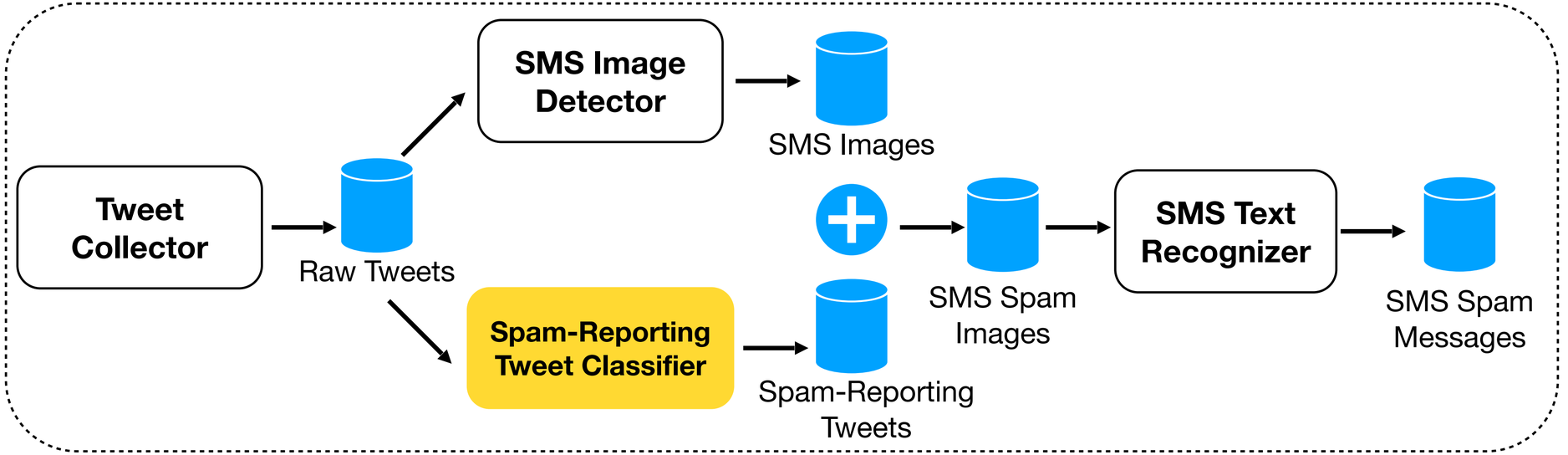


Twitter API

Watch out the spammers are out! They have no morals! Please check before u click.. if in doubt Google it!

#COVID19 #scam #spam #SMS #gov

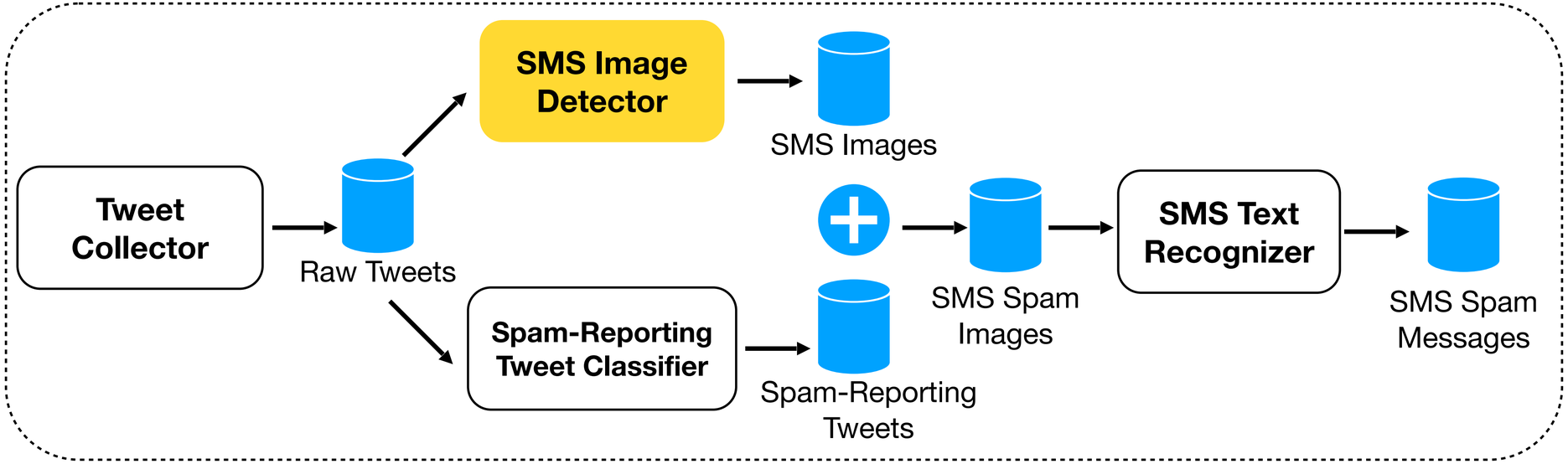


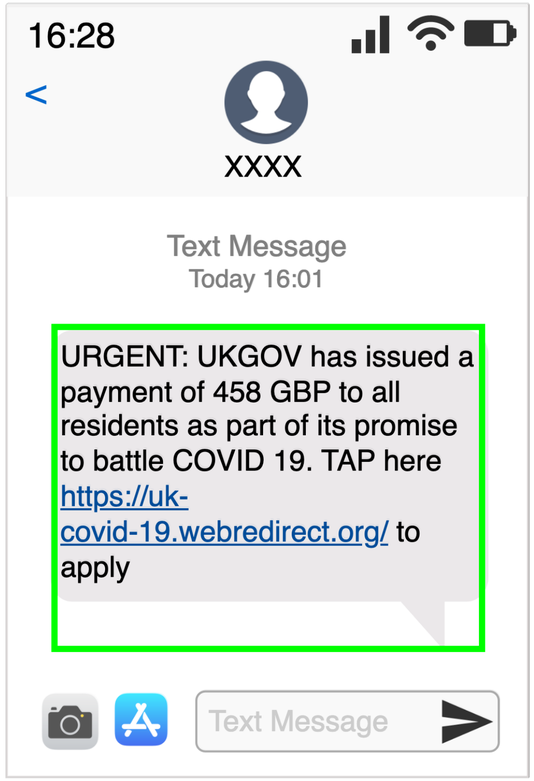


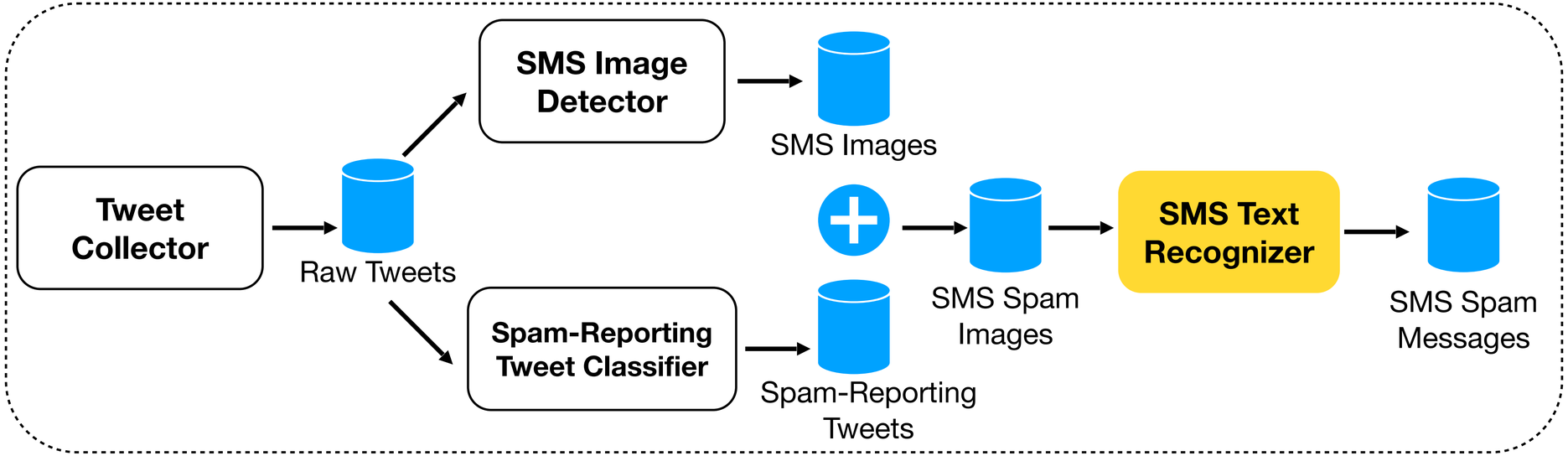
Watch out the spammers are out! They have no morals! Please check before u click.. if in doubt Google it!

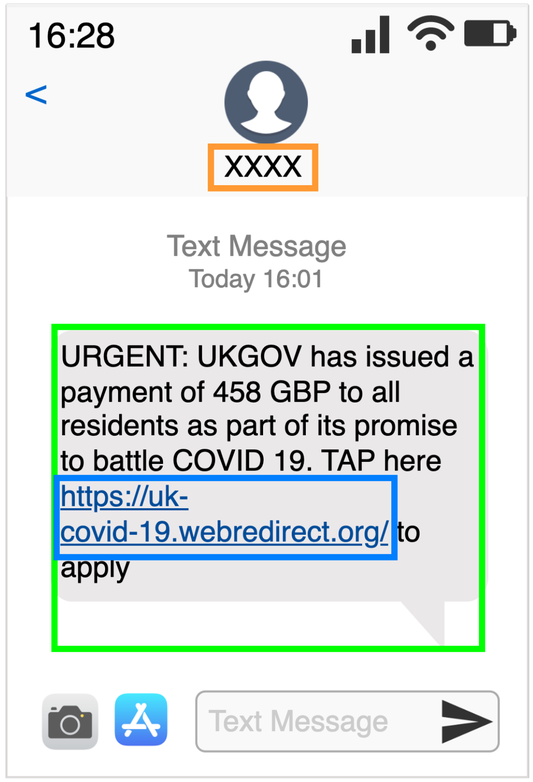
#COVID19 #scam #spam #SMS #gov

Spam-reporting ?


SMS Image Detector
SMS Images
Tweet Collector
Raw Tweets
SMS Text Recognizer
SMS Spam Images
SMS Spam Messages
Spam-Reporting Tweet Classifier
Spam-Reporting Tweets


16:28
<
XXXX
Text Message
Today 16:01
URGENT: UKGOV has issued a payment of 458 GBP to all residents as part of its promise to battle COVID 19. TAP here https://uk-covid-19.webredirect.org/ to apply
Text Message

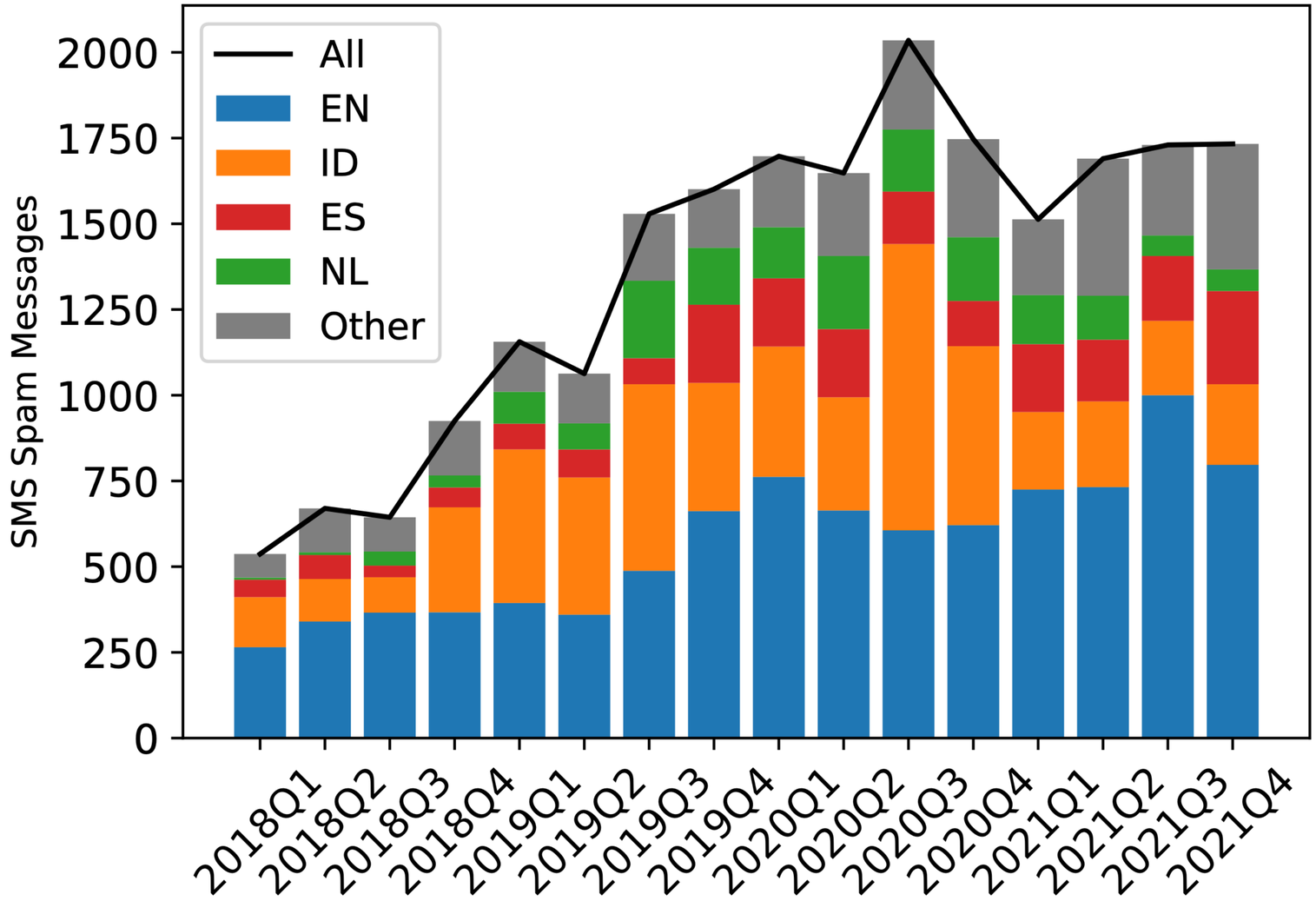

SMS Image Detector
SMS Images
Tweet Collector
Raw Tweets
SMS Text Recognizer
SMS Spam Images
SMS Spam Messages
Spam-Reporting Tweet Classifier
Spam-Reporting Tweets


16:28
<
XXXX
Text Message
Today 16:01
URGENT: UKGOV has issued a payment of 458 GBP to all residents as part of its promise to battle COVID 19. TAP here https://uk-covid-19.webredirect.org/ to apply
Text Message

Public SMS Spam Datasets.

	Volume	Period	Langs	Fraud Ratio	Spam Distribution	Source	Extendable
Ours	22K	2018Q1~2021 Q4	70	62%	General	Twitter	Yes
SMS Spam Collection	747	Before 2010	English	32%	General	A UK forum	No
FBS	14K	Three months	Chinse	38%	Fake Base Stations	A security company	No

Insights for Up-to-Date SMS Spam



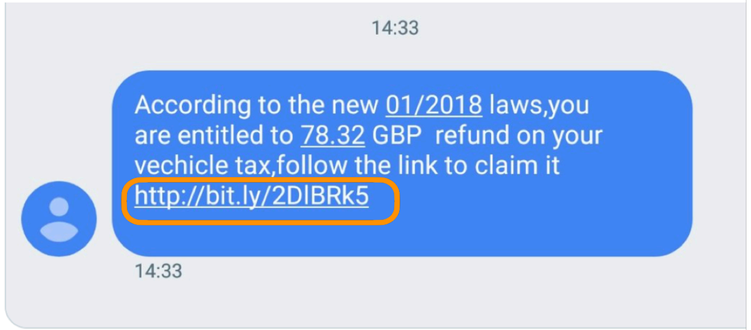
Finding

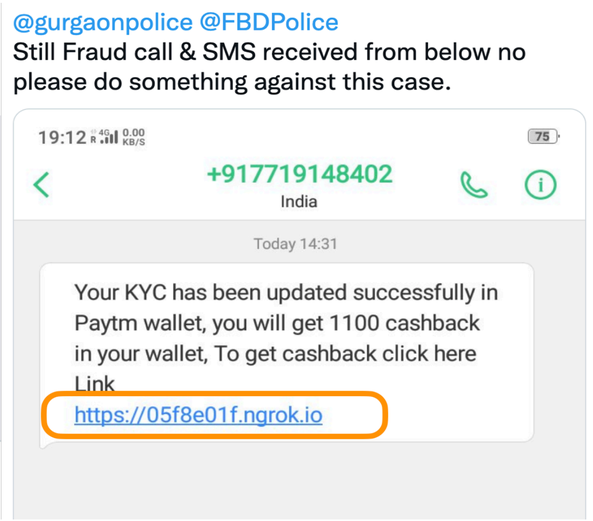
Reported Spam SMS on Twitter increases by a factor of 2X in Q4 2021, compared to Q1 2018.

Spam Category	Subcategory	Spam Ratio	# Labelled Messages
Fraud	Account alert	24.60%	233
	Finace	10.14%	96
	Prize	8.24%	78
	Delivery	6.12%	58
	Credit/Debit	5.49%	52
	Tax refund	2.53%	24
	COVID-19	1.69%	16
	Other	3.48%	33
Ads	Promotion	19.75%	187
	Loan/Gamble	9.40%	89
	Politics	2.53%	24
	Other	6.02%	57

Finding

Account alert (Fraud) is the most popular SMS spam category as reported on Twitter.





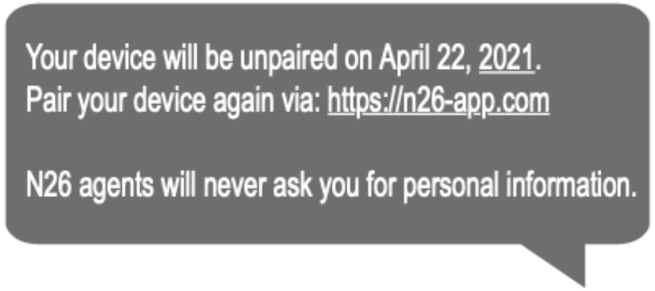
bitly

ngrok

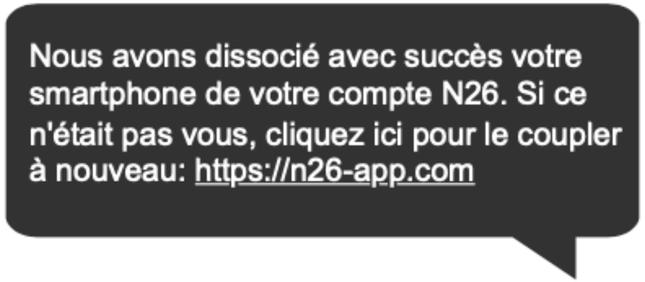
DuckDNS

Finding

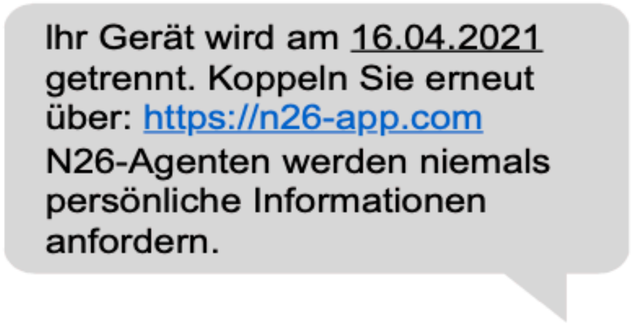
SMS spammers are increasingly adopting **novel evasion techniques**, e.g., dynamic DNS and port forwarding services.



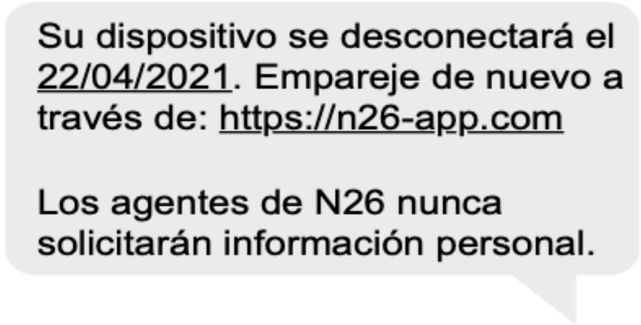
(a) Spam in English



(b) Spam in French



(c) Spam in German



(d) Spam in Spanish

Finding

A SMS spam campaign can spread similar messages in multiple languages.

Evaluate Spam Countermeasures

Anti-Spam Services

Perspective

95%

OOPSpam

91%

PLINO

79%

High detection rate but also high false alarm rate

Bulk SMS Services

6%

12%

17%

Low block rate

Text Messaging Apps

Android Messages

0%

iMessage

0%

AntiNuisance

0%

RoboKiller

48%

VeroSMS

69%

SMS Shield

73%

smsBlocker

73%

No alarms

Medium detection rate and acceptable false alarm rate

Looking Back

- The ever-updating SMS spam dataset is released on
  along with the source code
- We are working towards hunting up-to-date threat data in different categories, from different sources (e.g., cybercrime promotion on social networks)
- We are exploring privacy-preserving threat sharing mechanisms, e.g., federated learning and trusted execution environments.

Why Do Spam/Fraud Keep Happening?

- The weak (zero) authentication of content providers in communication systems, e.g., E-mail, SMS, telephone, search engines
- The insufficient international collaboration to fight against cross-border spam/fraud
- The rapid evolution of operators as driven by economic forces, to evade blocking, reach a wider victim base, and to bring in higher revenue

Outline

- Course FAQs
- Motivations of CyberSecurity
- Overview of This Course
  - Cryptography
  - Access Control
  - Network/Web Security
  - Software/System Security
  - Privacy
  - AI & Security

Q&A