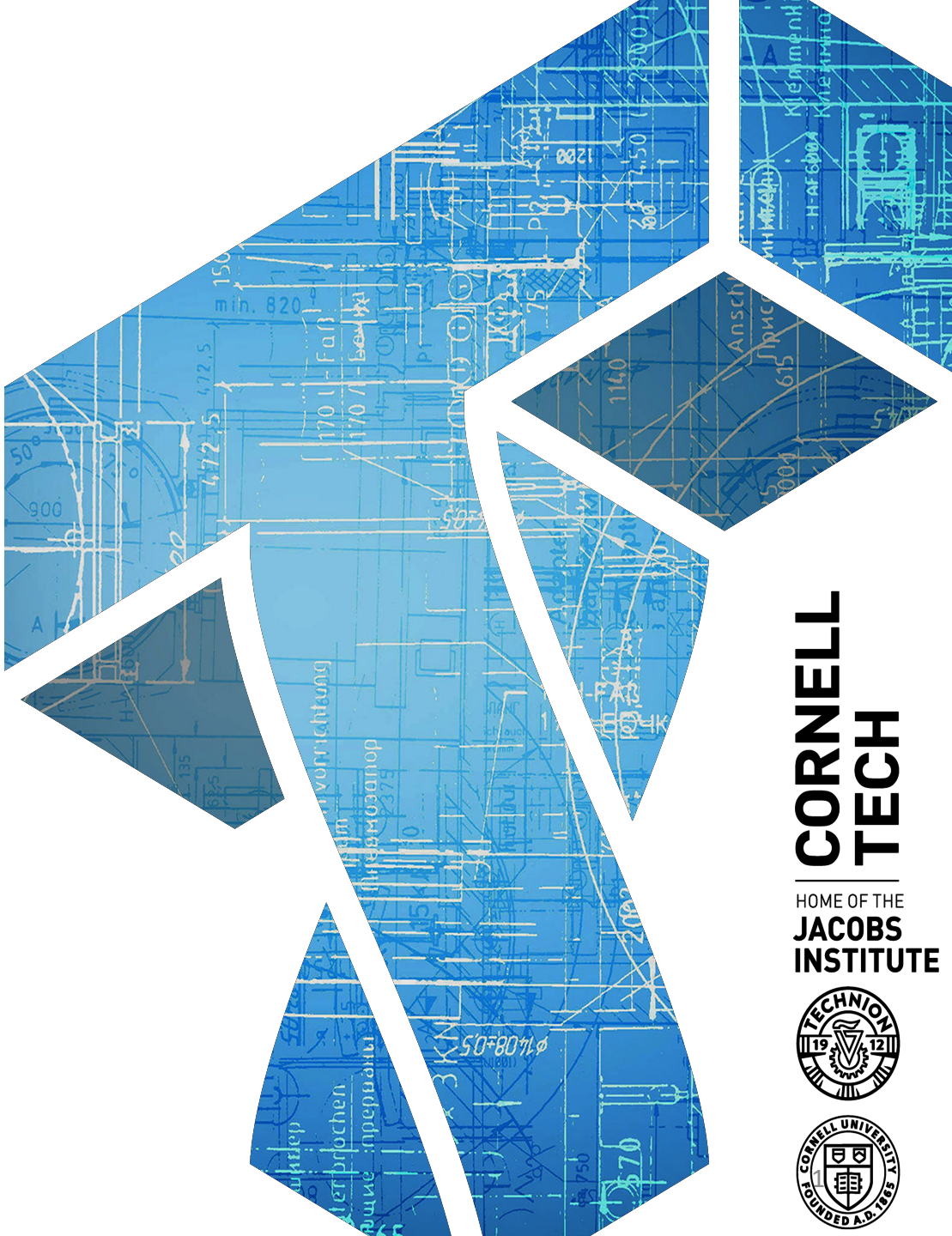
CS 5435: Non-profit-motivated abuse and harassment

Instructor: Tom Ristenpart

https://github.com/tomrist/cs5435-fall2024



CORNELL TECH

HOME OF THE JACOBS INSTITUTE

Trust & Safety

Focuses on how computing technology is abused to cause harm to people

Adversary

Computing systems

Defenders (designers, engineers, lawyers, etc.)

Targets / victims (Other users)

T&S goals can intersect with security goals (CIA) of underlying computing systems

Trust & Safety

Fraud

Stalking

Online harassment

Misinformation campaigns

Computer security

Denial of service attacks

Data breaches

Malware

Account compromise

Moderation

De-platforming

Abuse-aware design

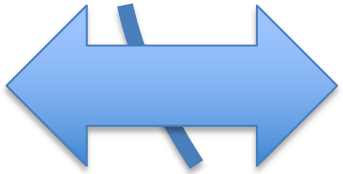
Clinical computer security

Access controls

Passwords

Cryptography

2FA and biometrics



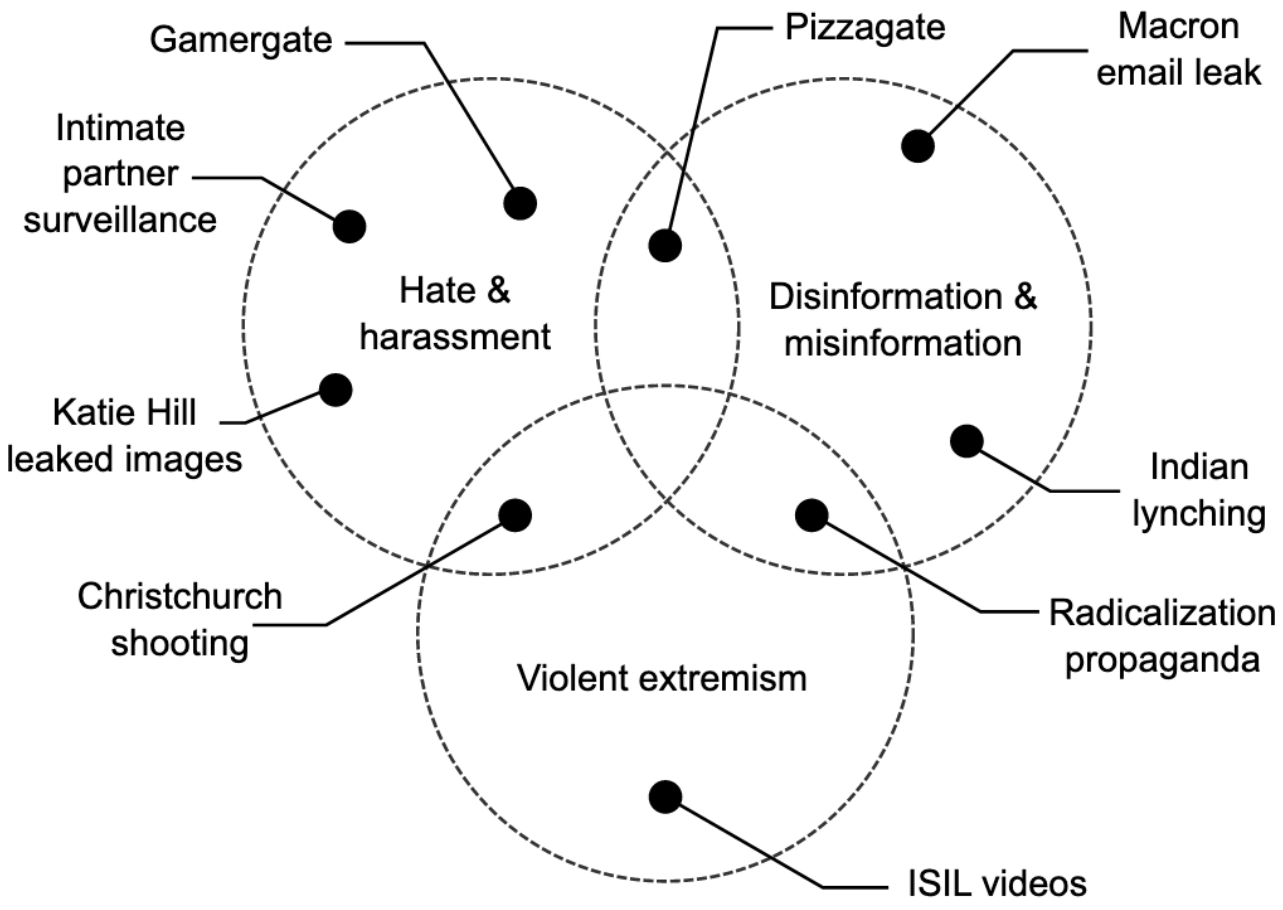
Trust & Safety topics

- Commercially motivated fraud (last lecture)
- Misinformation / disinformation
- Violent extremism (terrorism)
- Hate & harassment
  - "Hate and harassment occurs when an aggressor (either an individual or group) specifically ***targets another person or group*** to ***inflict emotional harm***, including coercive control or instilling a fear of sexual or physical violence."

    Definition from [Citron 2014], quote from [Thomas et al. 2021]

Trust & Safety topics

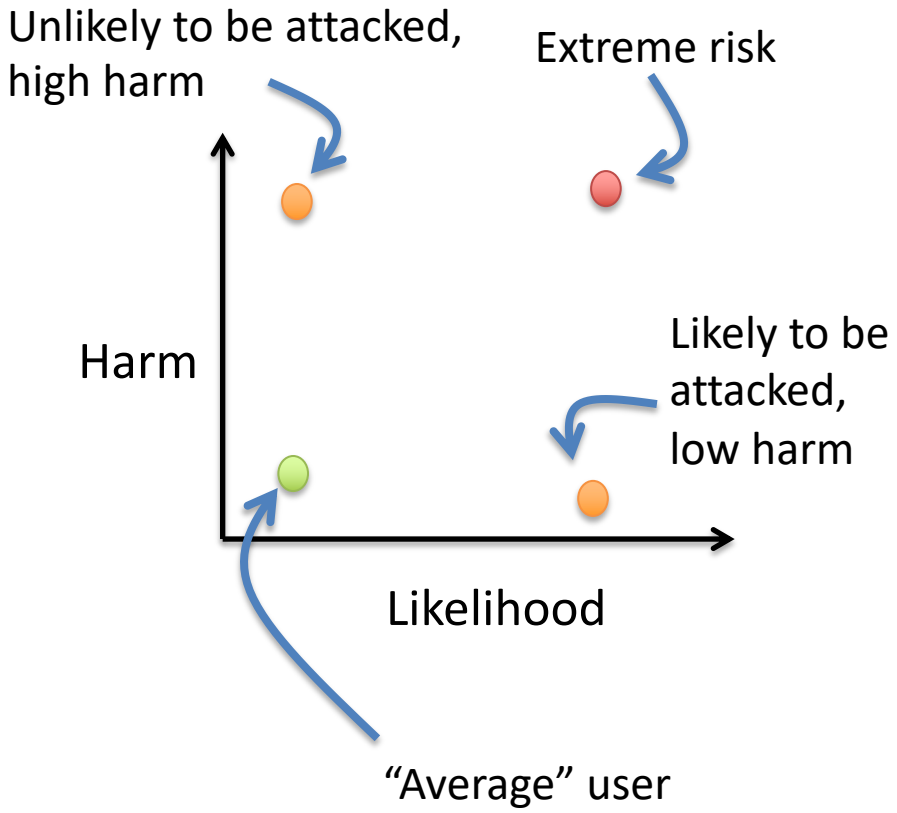
- Thomas et al. 2021 systematization paper on hate and harassment
- 50k person global survey
- 48% report experience digital H&H threats
  - sustained bullying, stalking, account takeover, …
- LGBQT+ people have higher rates



https://rist.tech.cornell.edu/papers/sok-abuse.pdf

At-risk people and groups

- An ***at-risk person*** is someone who is:
  - disproportionately harmed by an attack, and/or
  - disproportionately likely to suffer attack
- An ***at-risk population/group*** is any group for which a large fraction of its members are at-risk persons
- At-risk groups often related to identity, job role, socio-economic status, etc.



What are some other examples of possible at-risk groups?

Examples of other forms of abuse

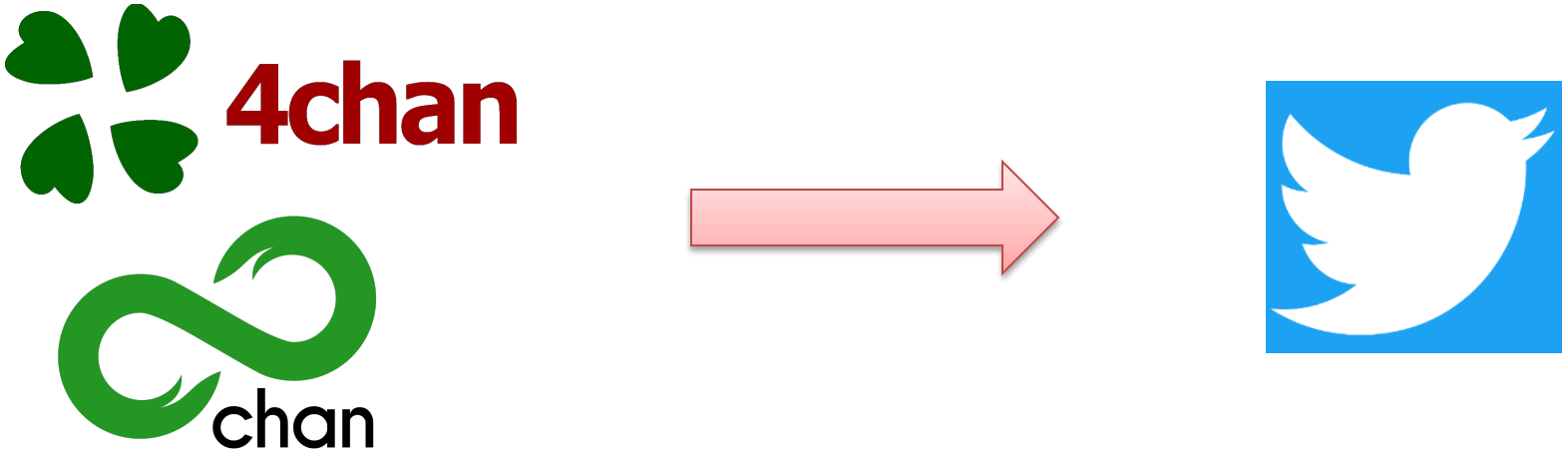
- Online harassment and bullying
  - Coordinated campaigns
  - Doxxing, raiding (other websites), …
- Misinformation campaigns
- Targeted attacks and RATs (remote access trojans)
- Tech abuse in intimate partner violence

Coordinated harassment campaigns

- Anonymous bulletin board
- Generated lots of memes, known to host some child porn (though technically against site policy)
- Associated with Anonymous hacker group
- Used to coordinate harassment, bullying, hacking
- Associated with alt-right groups
- Involved in Gamergate, banned Gamergate discussion

- Similar to 4chan (different site operator)
- No longer available due to CloudFlare and other providers dropping it

Coordinated harassment campaigns: Gamergate

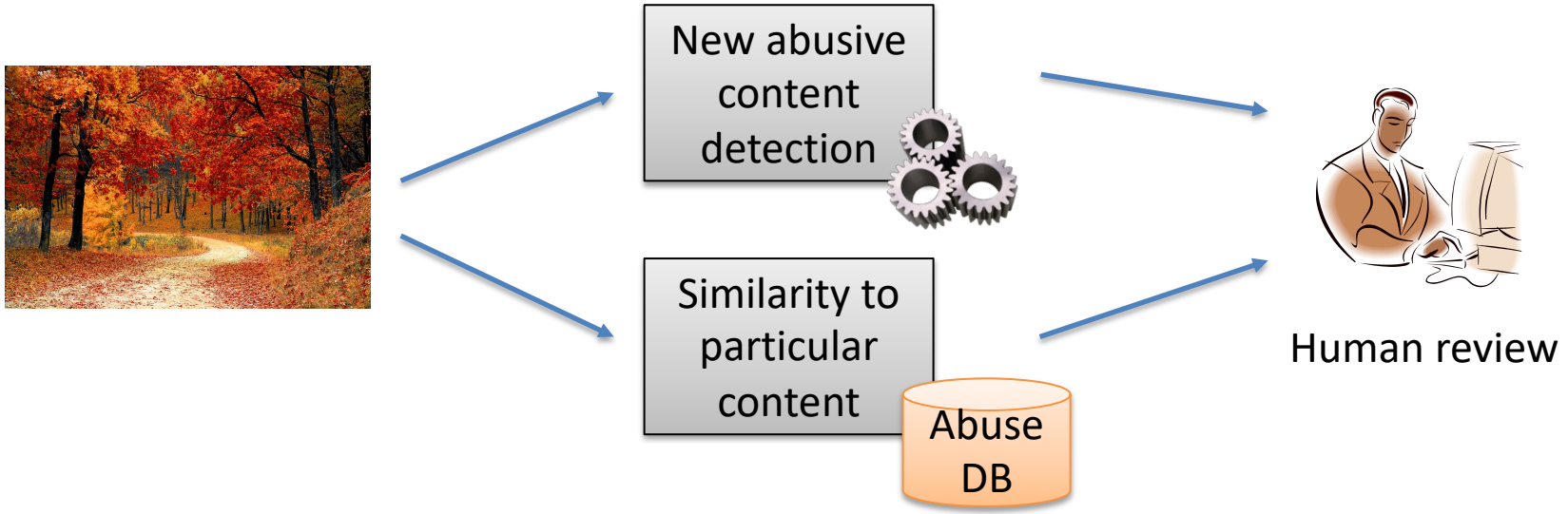


- Sustained harassment campaign against female gaming developers and others
  - Rape threats, murder threats, doxing (home address, other personal information disclosed)
  - Astroturfing, sock-puppet campaigns on Twitter
- Organized on 4chan & (after banning on 4chan) 8chan, most harassment played out on twitter (#gamergate)
- Why? Ex-boyfriend of initial victim posted false claims about victim
  - "Right-wing backlash against progressivism" (in gaming)

Mininformation campaigns

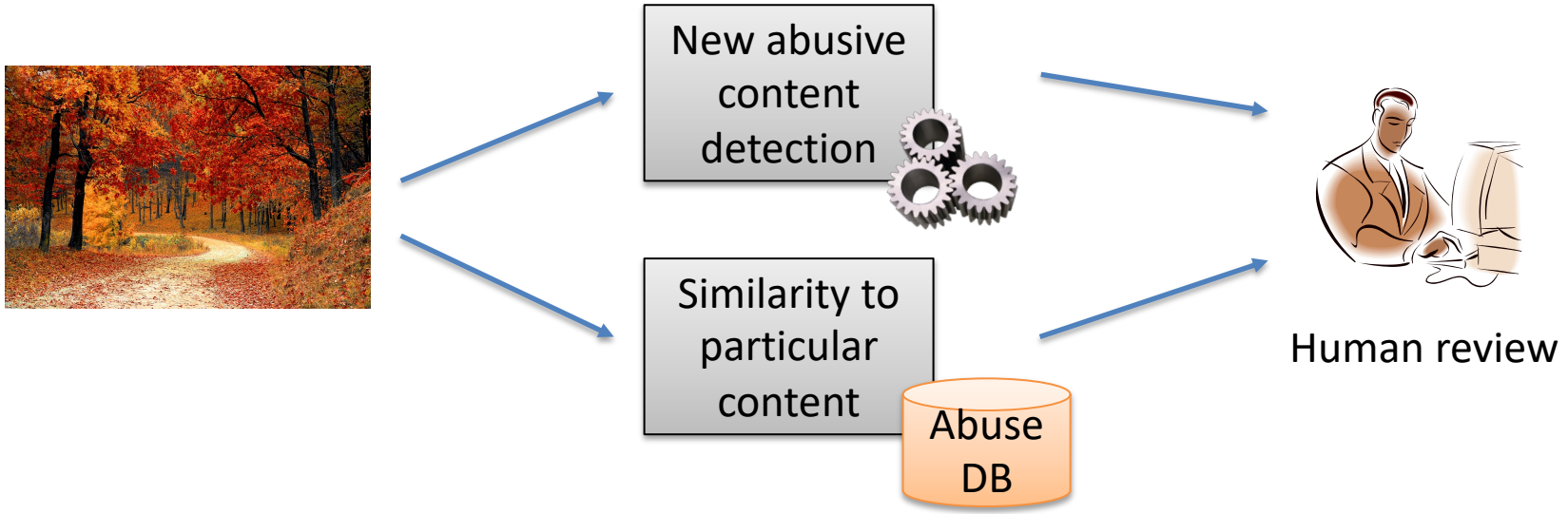
- Pizzagate conspiracy
- QAnon conspiracy
- 2016 election interference
  - Cambridge Analytica
  - Russian influence operations
- Research studies indicate falsehoods spreads faster than truth on social media

Identifying abuse content



- Use locality-sensitive hashing to check if similar to known bad content
  - H such that |H(image) – H(image')| < threshold if image, image' very similar
- Use machine learning to try to identify patterns indicative of abusive content
  - Is image a picture of a naked child?
- Refer out to human for review (Facebook has 10,000s of moderators)
- Content flagged by users also sent through reviewing pipelines
  - Flagging mechanisms also subject to abuse (illicit takedowns)

Identifying abuse content



- PhotoDNA is system for similarity matching of child sexual abuse media (CSAM) run by Microsoft
  - Used widely in industry
- Legal requirements around CSAM strict: must report detected content to National Center for Missing and Exploited Children (NCMEC)

Targeted attacks

- Dissidents, journalists, activists targeted by nation-states
  - Phishing attacks, botnet-style C&C servers to collect data
  - Remote Access Trojans (RATs)
- Small industry of companies providing “lawful access” tools


From: Melissa Chan <melissa.aljazeera@gmail.com>
To:
Sent: Tuesday, 8 May 2012, 8:52
Subject: Torture reports on Nabeel Rajab

Acting president Zainab Al Khawaja for Human Rights Bahrain reports of torture on Mr. Nabeel Rajab after his recent arrest.

Please check the attached detailed report along with torture images.

1 attachment: Rajab.rar 1.4 MB Save

Figure 1: E-mail containing FinSpy.

https://www.usenix.org/system/files/conference/usenixsecurity14/sec14-paper-blond.pdf
https://www.usenix.org/system/files/conference/usenixsecurity14/sec14-paper-marczak.pdf
https://www.usenix.org/system/files/conference/usenixsecurity14/sec14-paper-hardy.pdf

Commoditization of APT Tools: Pegasus

- Ahmed Mansoor (UAE) human rights activist
  - "On August 10 and 11, 2016, Mansoor received SMS text messages on his iPhone promising "new secrets" about detainees tortured in UAE jails if he clicked on an included link."
- Analysis indicated that link connects to exploit chain to remotely jailbreak iPhone
  - Three zero-days (WebKit browser vuln, ASLR bypass, kernel vuln)
  - Remotely installs implant (spyware)
  - Remote code execution (RCE)
- Multiple Pegasus versions, some have "no click" exploits

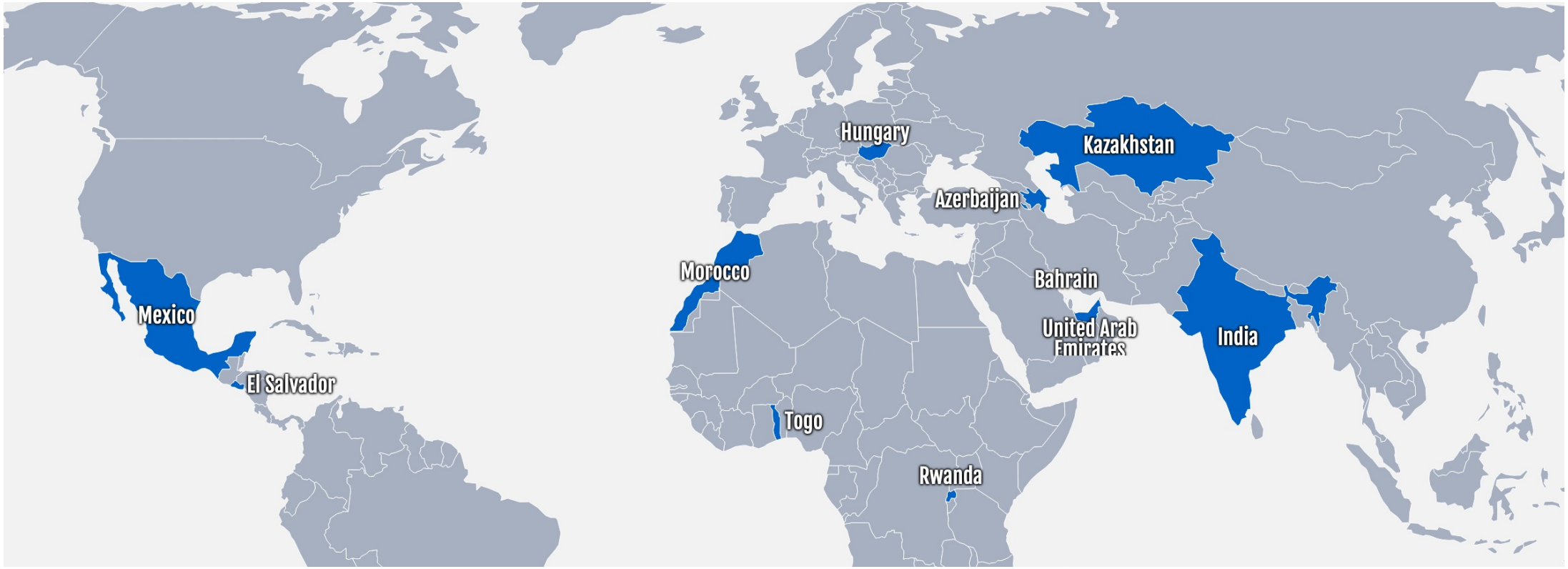
https://citizenlab.ca/2016/08/million-dollar-dissident-iphone-zero-day-nso-group-uae/

Commoditization of APT Tools: Pegasus

- Analysis indicated that link connects to exploit chain to remotely jailbreak iPhone
  - Three zero-days (WebKit browser vuln, ASLR bypass, kernel vuln)
  - Remotely installs implant (spyware)
  - Remote code execution (RCE)
- Multiple Pegasus versions, some have “no click” exploits
- Apple Lockdown mode:
  - https://support.apple.com/en-us/105120

https://citizenlab.ca/2016/08/million-dollar-dissident-iphone-zero-day-nso-group-uae/

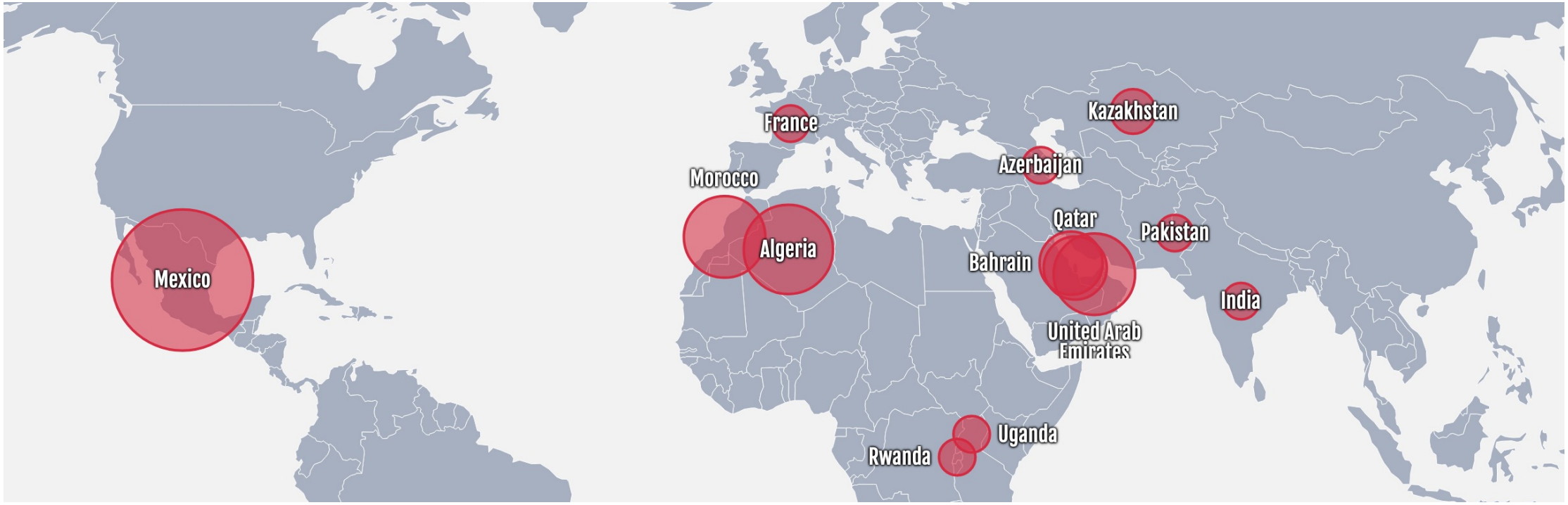
Commoditization of APT Tools: Pegasus



Pegasus Project identified 12 governments as NSO customers (blue) and ~50,000 potential targets of (red circles)

https://forbiddenstories.org/pegasus-project-impacts-map/

Commoditization of APT Tools: Pegasus



Pegasus Project identified 12 governments as NSO customers (blue) and ~50,000 potential targets of (red circles)

https://forbiddenstories.org/pegasus-project-impacts-map/

Targeted inter-personal attacks: IPV

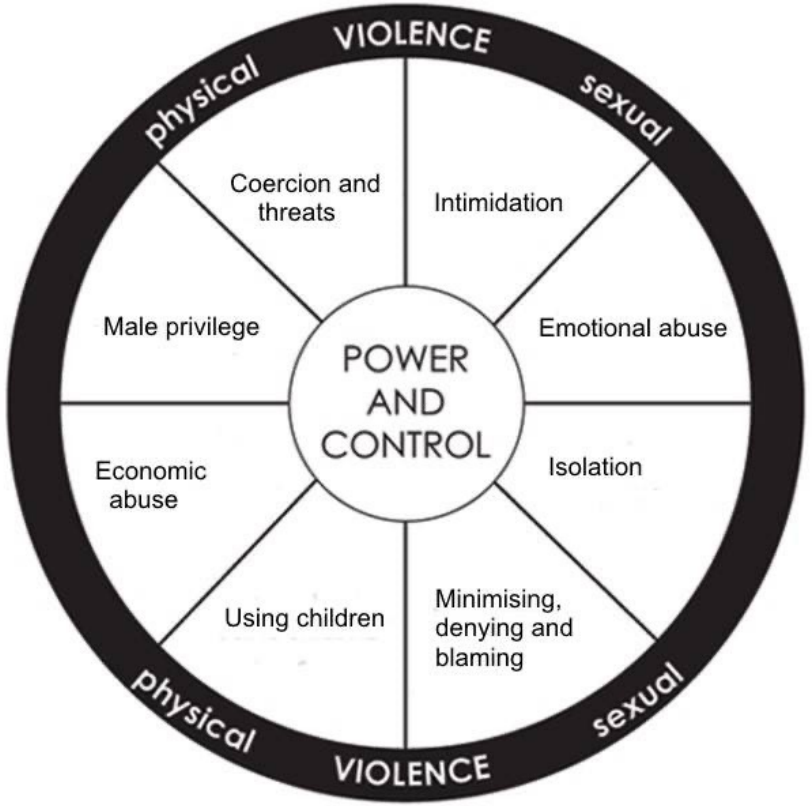
25% of women
11% of men

suffered rape, physical violence, and/or stalking **by an intimate partner**

[USA CDC's National Intimate Partner and Sexual Violence Survey 2010-2012]

National Domestic Hotline definition:

"pattern of behaviors used by one partner to maintain power and control over another partner in an intimate relationship"



Variant of the Duluth Power and Control Wheel

Targeted inter-personal attacks: IPV

25% of women
11% of men

suffered rape, physical violence, and/or stalking **by an intimate partner**

[National Intimate Partner and Sexual Violence Survey 2010-2012]

Intimate partner violence (IPV) abusers exploit technology:

- Harassing texts/messages
- GPS devices & spyware apps
- Victim accounts being “hacked”
- Physical device access
- …

Research at Cornell Tech

Mayor's Office to End Domestic and Gender-Based Violence

Qualitative studies of IPV ecosystem in NYC

Freed, Palmer, Minchala, Levy, R., Dell. *Digital Technologies and Intimate Partner Violence: A Qualitative Analysis with Multiple Stakeholders*, CSCW 2017

Freed, Palmer, Minchala, Levy, R., Dell. *"A Stalker's Paradise": How Intimate Partner Abusers Exploit Technology,* CHI 2018

IPV spyware & anti-spyware measurement studies

Chatterjee, Doerfler, Orgad, Havron, Freed, Levy, Dell, McCoy, R. *The Spyware Used in Intimate Partner Violence*, Oakland 2018

Clinical computer security models and tools

Havron, Freed, Chatterjee, McCoy, Dell, R. *Clinical Computer Security for Victims of Intimate Partner Violence*, USENIX Security 2019

Freed, Havron, Chatterjee, McCoy, R., Dell. *"Is my phone hacked?" Analyzing Clinical Computer Security Interventions with Survivors of Intimate Partner Violence*, CSCW 2019

New York City Family Justice Centers

Mayor's Office to End Domestic and Gender-Based Violence

ENDGBV runs ***Family Justice Centers***
One in each neighborhood of NYC

Range of services for clients: victim-survivors of domestic violence, sex trafficking, and elder abuse:

- Civil / legal services
- Counseling & safety planning
- New York Police Department (NYPD)
- District Attorney's offices
- Access to emergency shelter
- Non-profit organizations

2022 { **50,798** client visits to an FJC
119,204 IPV police reports

https://www.nyc.gov/assets/ocdv/downloads/pdf/2022-ENDGBV-Annual-Fact-Sheet_Final.pdf

Year-long qualitative study: methods

Clients (Survivors / Victims)	11 focus groups with 39 women (English & Spanish) ages 18-65 (average 42) from 15 different countries with range of education levels most no longer living with abusive partner
Professionals	Semi-structured interviews with 50 professionals female (45) and male (5) case managers, social workers, attorneys/paralegals, & police officers

Largest and most demographically diverse study to date

Client Story #1

"I was raised in a country in which women are trained to serve men. I'll say that he never hit me, basically, but he did the worst thing ever. Honestly, I would probably rather that he hit me instead of the things that he did to me.

Use shared ownership to install spyware

So he put a spyware in [our] computer. Obviously, I didn't know because he studied computer programming, so he was very savvy about it. …

Non-consensual intimate images

Account compromise

So he went into my computer, he got my Facebook password, email password, and he shared naked pictures of me. He sent them from his Facebook to my bosses. He took my phone and he sent them through private messages to several friends, but also through my email and my Facebook because he had the password.

Device compromise

Impersonate victim

The embarrassment that I went through, the public humiliation, it... beat me to the ground"

Four categories of common attacks

Ownership-based

- Abuser owns device/account
- Shared account/device
- Buying children device

- Prevent use / destroy device
- Digitally control access
- Track location, monitor usage

Account/device compromise

- Physical access to unlocked device
- Force password / pin revelation
- Remotely “hack” via security questions / passwords

- Install spyware / “dual-use” app
- Track location, monitor victim
- Steal or delete info
- Lock victim out of account
- Impersonate victim

Harmful messages or posts

- Call/text/message victim (from spoofed account)
- Post harmful content (e.g., threaten violence)
- Harass victim’s friends/family
- Proxy harassment

Exposure of private information

- Blackmail by threat of exposure
- “Doxxing” victim
- Non-consensual intimate images
- Fake profiles/advertisements of sexual services

Attacks not technically sophisticated, but succeed because threat models don't match threats

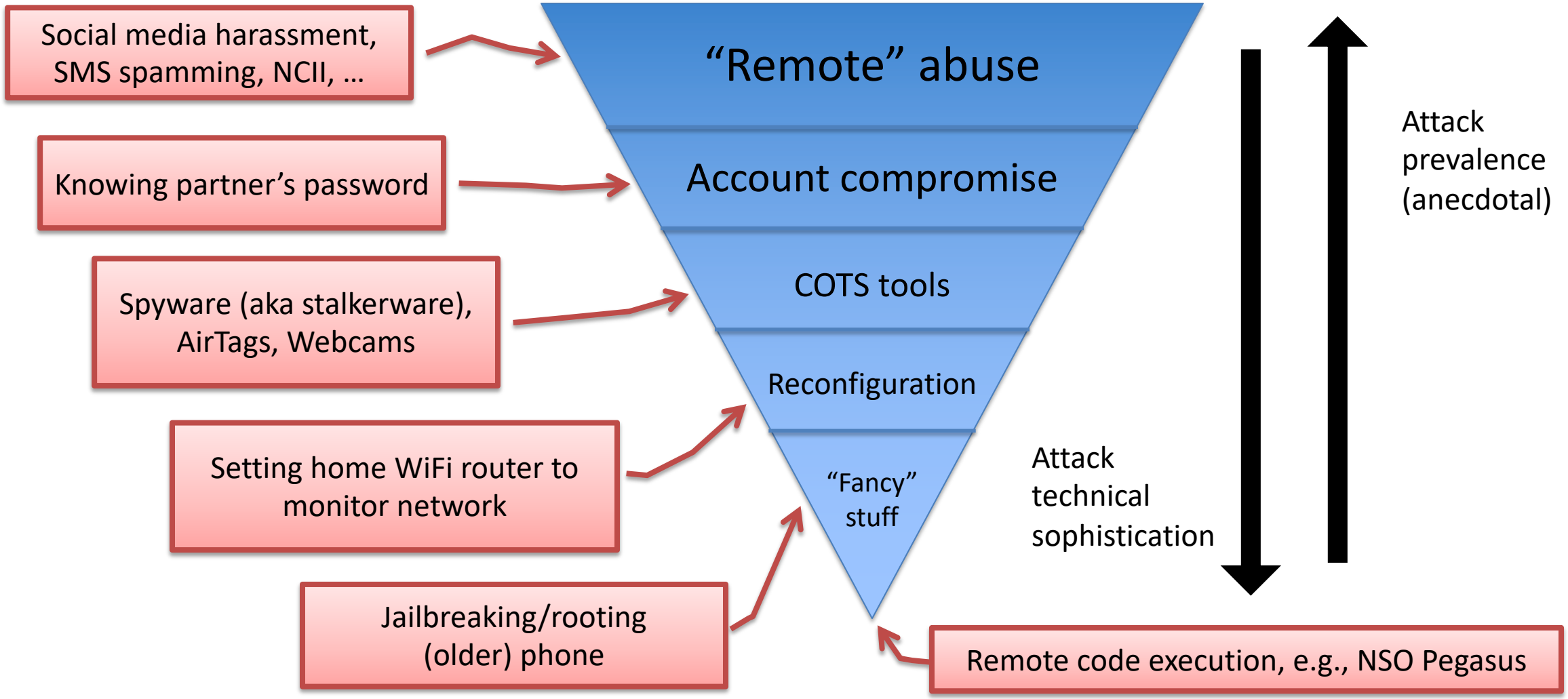
Threat models specify: (1) Who are Attacker and Victim
(2) Attacker's capabilities & goals

Account login – IPV threats

Threat	Description	Defense?
Abuser knows password (different location)	Abuser knows password and logs in from different location	Two-factor authentication; device authentication
Abuser knows password (same location)	Abuser knows password and logs in from same household or even device	Change password; use two-factor authentication
Abuser compels password disclosure	Abuser forces password disclosure (emotional or physical coercion)	Decoy passwords?
Abuser uses device to gain access	Session cookies allow abuser to login from device without auth checks	Delete session cookies

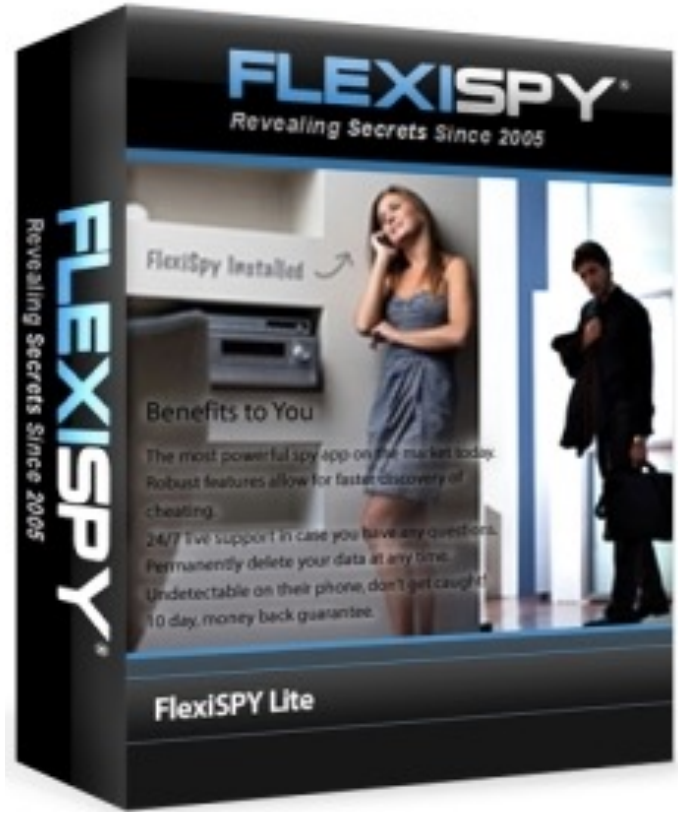
"[The abuser] stole her computer and was able to access all this information . . . her school applications, her bank accounts, all sorts of things, and gain access and control of these things. That . . . had a totally traumatizing effect."
- Case manager

The tech abuse inverted pyramid



The spyware (aka stalkerware) problem

"An abusive partner kicked in our front door and wound up in the lobby of our building by tracking her phone . . . ***it was some secondary application that the abuser had put on it and knew exactly where she was****. He literally kicked our front door open. We called the police … it was scary."* – Case manager



https://www.spymasterpro.com/

A typical off-store spyware app

Spyzie

Real-time + historical

All-Inclusive Mobile Phone Spy

- ✓ Spy on all iOS and Android devices
- ✓ Track SMS, Call logs, App chats, GPS etc.
- ✓ No Rooting or Jailbreaking required
- ✓ Invisible mode, monitor the activity remotely

Covert

Remotely record audio, video, and alter settings

$39.99 ~~$49.99~~ / Month

Buy Now

Required: Physical access to the target android phone.

Source: https://www.spyzie.com/ad/phone-spy.html

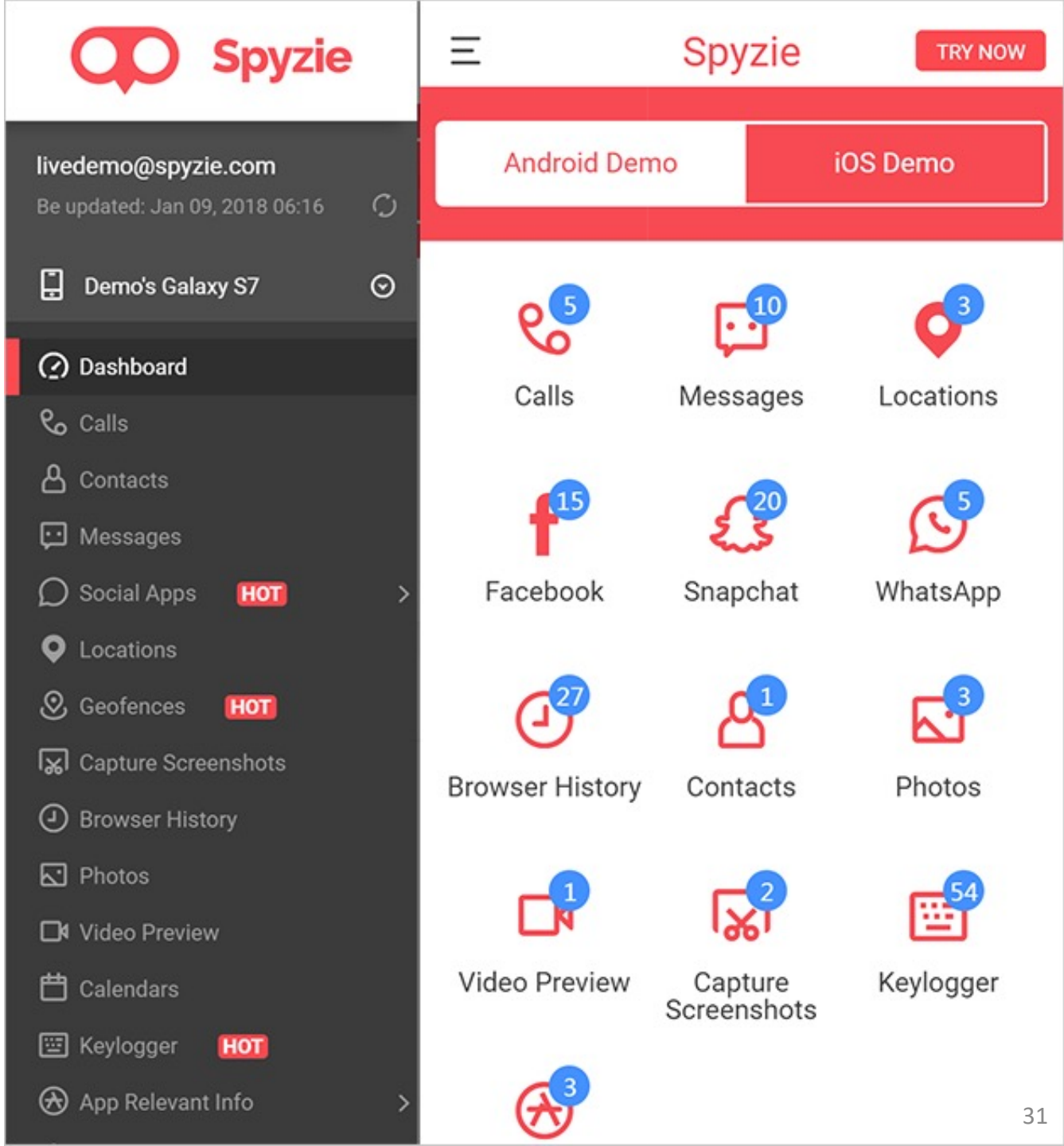
Smooth (ab)user experience with off-store spyware

How-to guides

Installing is easy
Whole process takes ~5 min

On victim's device:
No app icon, no notification

On abuser end:
Fancy web UI or apps



Initial investigation into IPV spyware

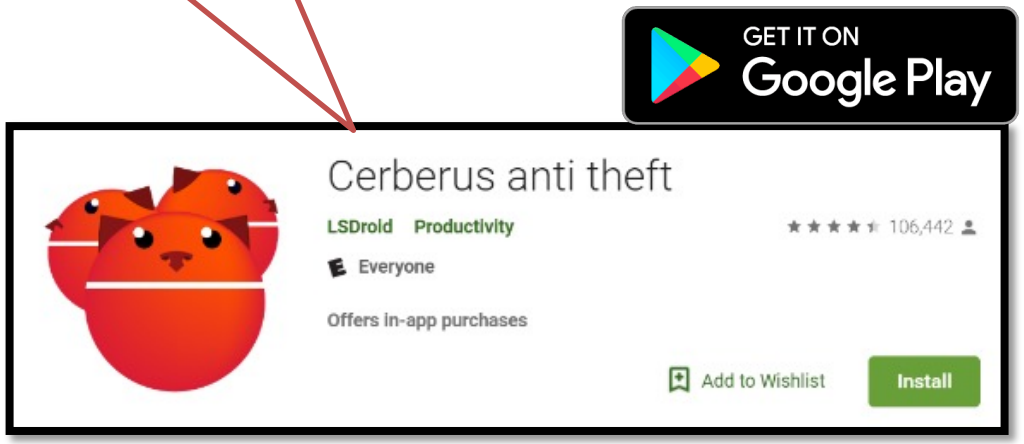
[Chatterjee et al. 2018] crawled Google & Apple app stores:

- Machine learning + hand-labeling to classify as IPV-relevant
- Thousands of IPV-relevant apps: mostly ***dual-use apps***

*"I'm looking for an app I can install on my wife's phone that is **hidden** so that I can see **where she is or has been** via cell towers or gps."*

*"[Install] **Cerberus** from the market. Once installed and configured, can be **set to be invisible in the app drawer**.*

*You can also **record audio** and **take pictures remotely** with it!"*



Source: forum.xda-developers.com/showthread.php?t=1266874

Investigation into IPV spyware

[Chatterjee et al. 2018] crawled Google & Apple app stores:

- Machine learning + hand-labeling to classify as IPV-relevant
- Thousands of IPV-relevant apps: mostly ***dual-use apps***

Reported findings to Google who took down some apps & content and strengthened policies

Symantec

KASPERSKY lab ANTIVIRUS

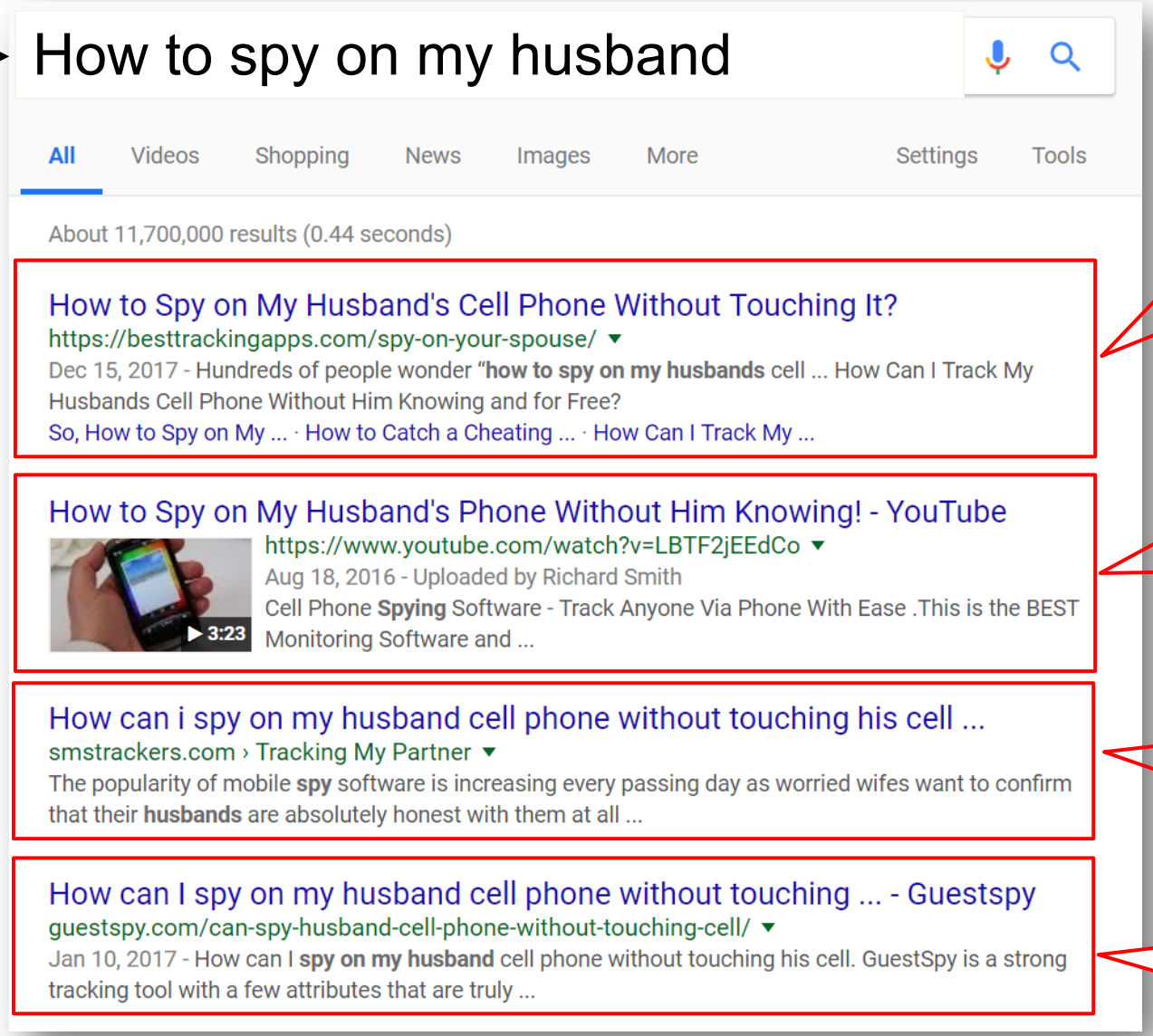
Helped antivirus companies to change products to warn about IPV spyware

avast

COALITION AGAINST STALKERWARE

How do abusers find spyware apps?

Need large number of query terms that an abuser might use



Blogposts comparing spy apps

Video how-to guides

Ad funnels for spy apps

Spy app websites

Search engines give query suggestions

how to spy on my

how to spy on my **husband**
how to spy on my **boyfriends phone**
how to spy on my **boyfriends iphone**
how to spy on my **girlfriends iphone**
how to spy on my **husband iphone**
how to spy on my **boyfriends facebook messages**
how to spy on my **husband phone**
how to spy on my **son's iphone**
how to spy on my **bf phone**
how to spy on my **wifi network**

Searches related to how to spy on my husband

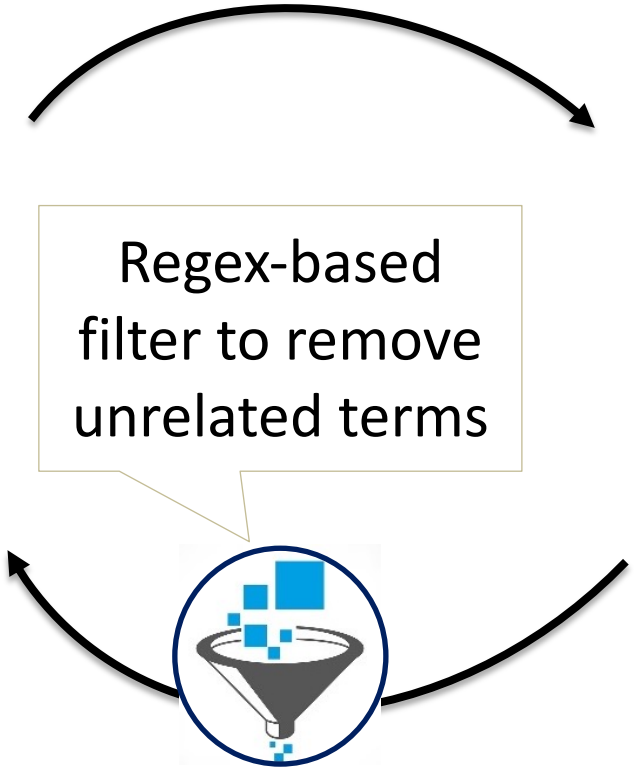
track my **husbands phone for free**
track my **husbands phone without him knowing**
how **can i find out who** my husband **is texting for free**
cheaters spy **app free**
how **can i track** my **husbands cell phone without him knowing and for free**
how to **hack** my **husbands whatsapp**
how **can i track** my **husbands phone location**
spy on **cell phone without touching target phone**

“Snowball” searching

how to spy on my girlfriend,
track my girlfriend android,
how to spy on my husband
…

Query set

Regex-based filter to remove unrelated terms



how to spy on my

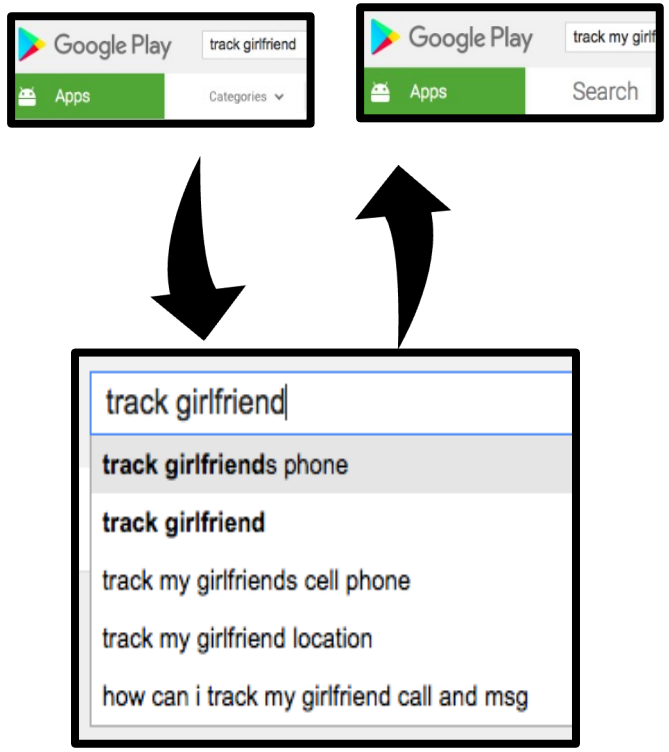
how to spy on my **husband**
how to spy on my **boyfriends phone**
how to spy on my **boyfriends iphone**
how to spy on my **girlfriends iphone**
how to spy on my **husband iphone**
how to spy on my **boyfriends facebook messages**

Searches related to how to spy on my husbar

track my **husbands phone for free**
track my **husbands phone without him knowing**
how **can i find out who** my husband **is texting for free**
cheaters spy **app free**
how **can i track** my **husbands cell phone without him kno**

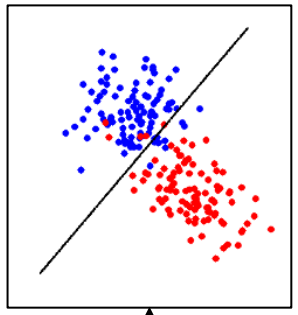
Finding IPV-relevant apps in official app stores

Snowball searching over
3 months on Play Store



15,000+ apps

Prune using
ML classifier



~4,500 apps

Filter using
human verifier

3,500+ apps

Looks at app title, description, permissions, and genre as reported in Play Store

Smaller study (2 weeks) on iTunes App Store: **451**

Taxonomy of on-store apps

Personal tracking

- Find my phone or Anti-theft
- Automatic call recorder
- Automatic data/SMS syncing
- Phone control
- Personal safety

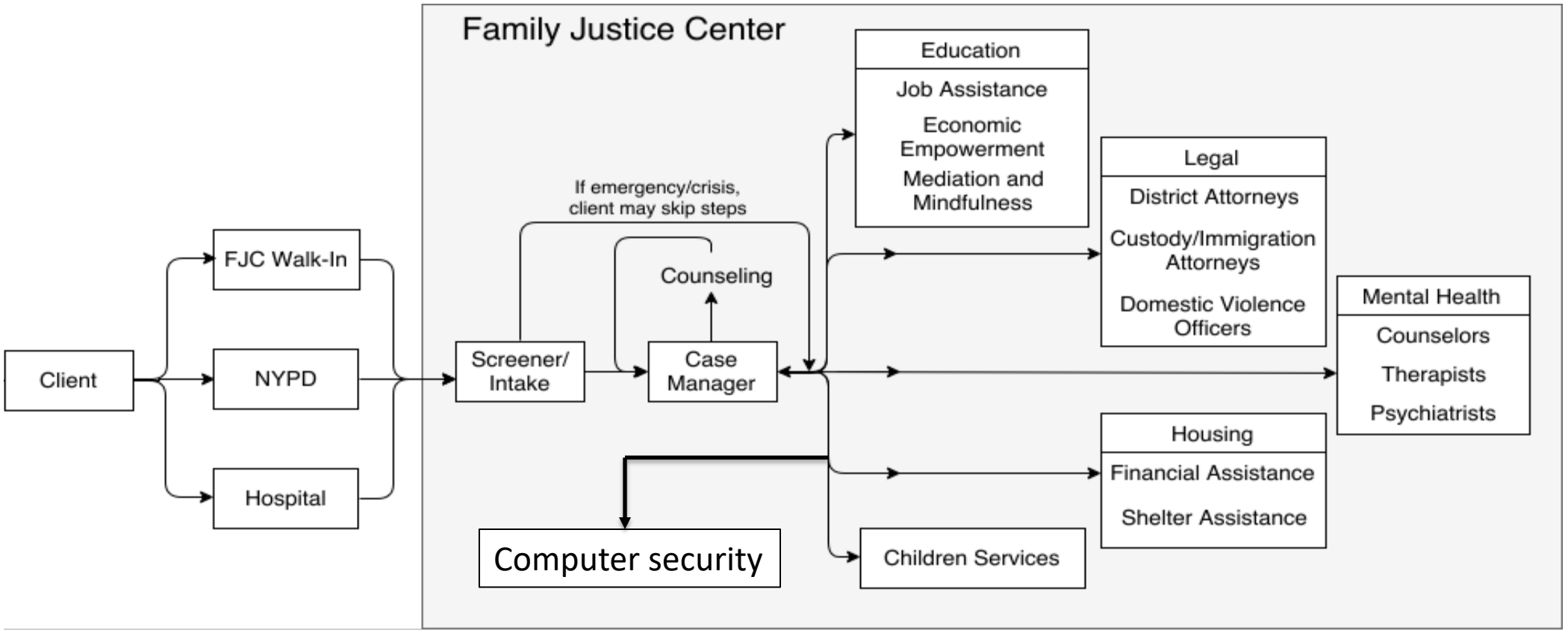
Mutual tracking

- Find-my-family / friend
- Couple tracking

Subordinate tracking

- Employee tracking
- Parental control

How are victims obtaining help?



2017: No best practices for evaluating tech risks

Victims & professionals overwhelmingly report having insufficient tech knowledge

Victims, professionals feel they lack needed technology expertise

Victims overwhelmingly report having insufficient technology understanding to deal with tech abuse

"[The victim is] absolutely not savvy on technology."
– Case manager

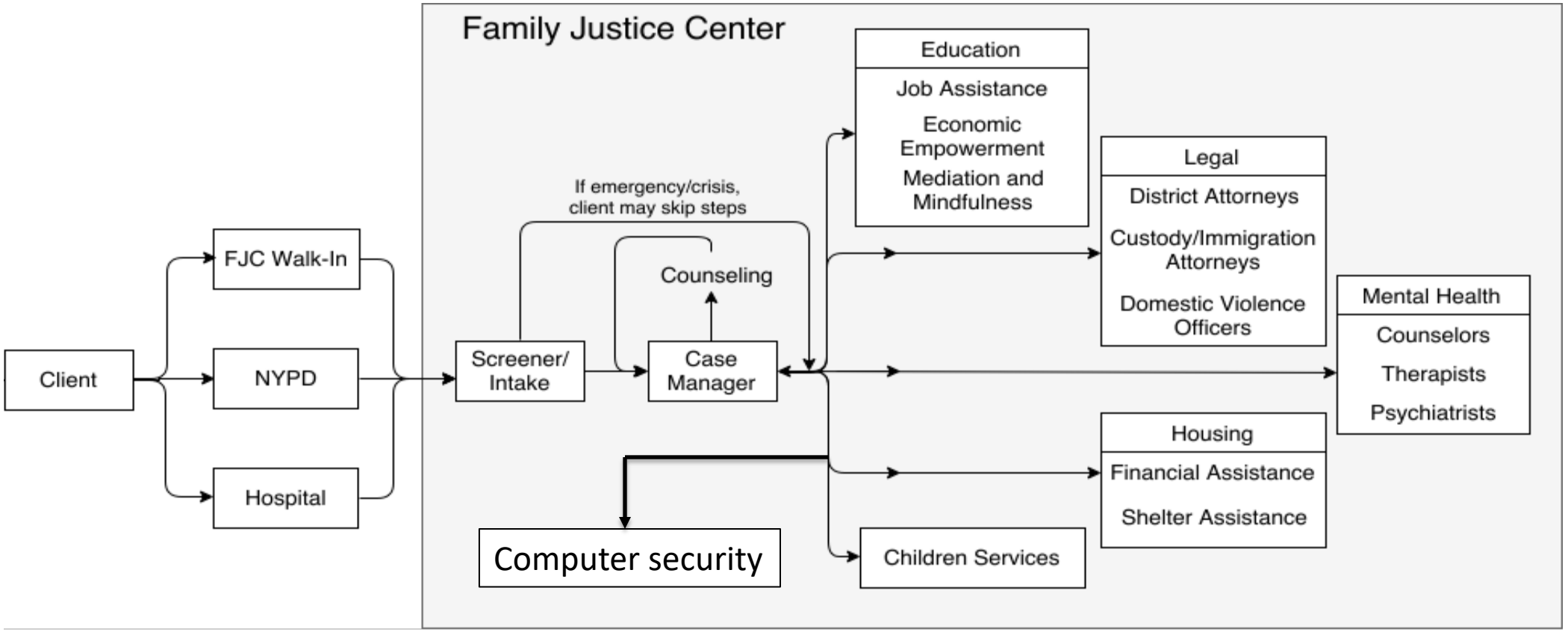
Abusers usually considered to be "*more tech-savvy*" than victims

Professionals overwhelmingly report having insufficient technology understanding to help with tech abuse

"I end up Googling it. And then I'll deal with [the client]. But I think… I don't know how to do it so we'll just Google it together." – Case manager

"Google-as-they-go"

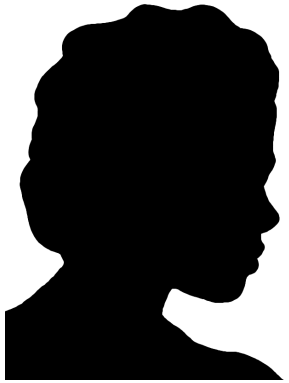
Our idea: clinical computer security



Clinical computer security

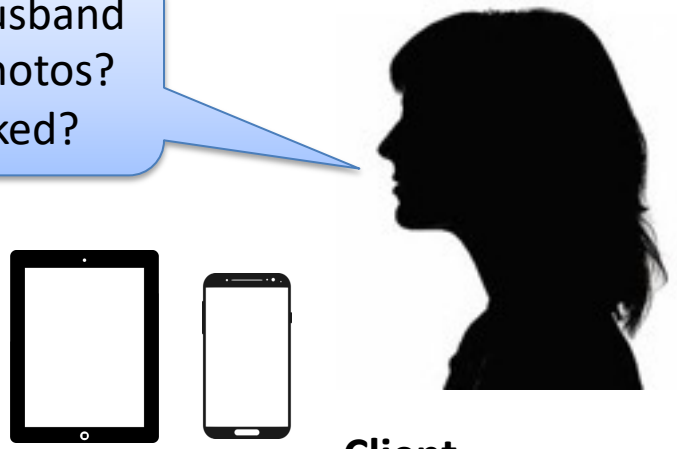
[Havron et al. Security 2019]
[Freed et al. CSCW 2019]

- Developed procedures with help of 14 focus groups with IPV support professionals
- Custom spyware detection tool (ISDi)



IPV tech consultant

How does my ex-husband keep getting my photos? Is my phone hacked?



Client (IPV victim/survivor)

Prototyping clinical services in New York City at FJCs

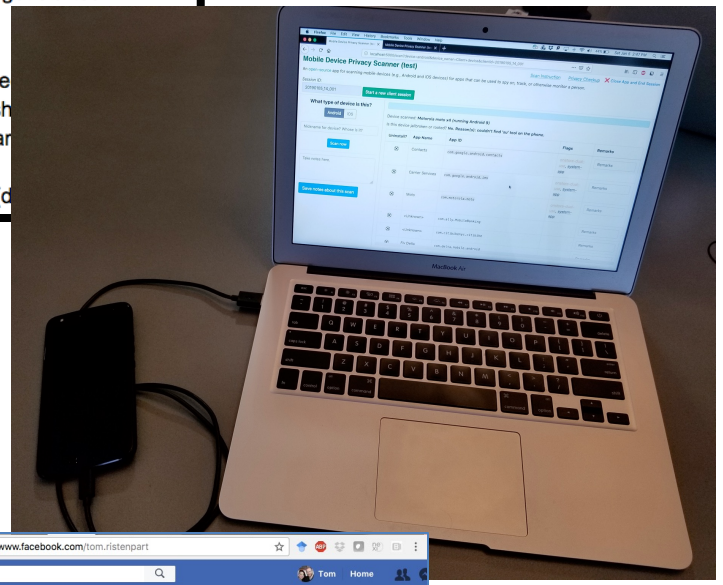
Case manager, lawyer, etc.

referral

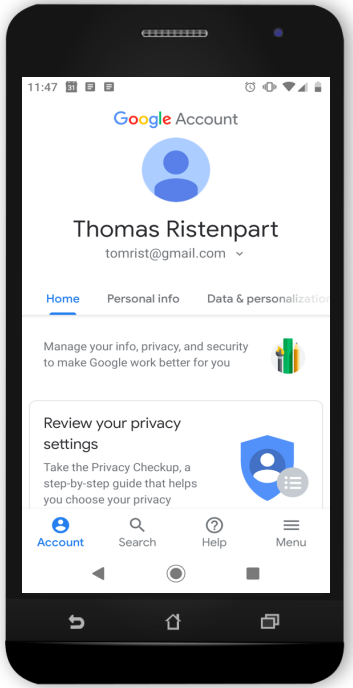
Technology Assessment Questionnaire (TAQ10)

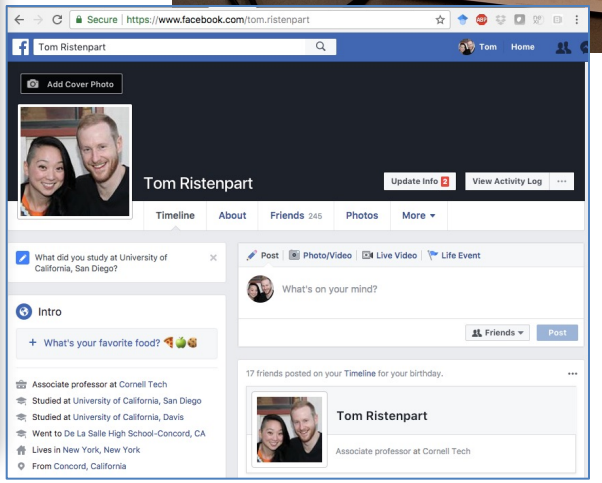
I am going to ask questions that might be helpful if you think your abuser
your phone, or have access to information on your phone or other techno
knowledge. This is technology abuse. Today we are going to be asking y
this topic. Do you have any questions for me? Let's get started.

1. Do you worry that your device(s) is being use
   show up unexpectedly or know things they sh
2. What devices do you use in your home or car
   desktop, laptop, etc.)
3. Do you currently (or have in the past) share(d



- Referral from FJC professional
  1. ***Understand***
  2. ***Investigate***
  3. ***Advise***
- Safety-planning hand-offs





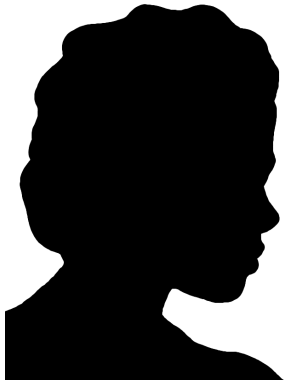
hand-off

Case manager, lawyer, etc.

Clinical computer security

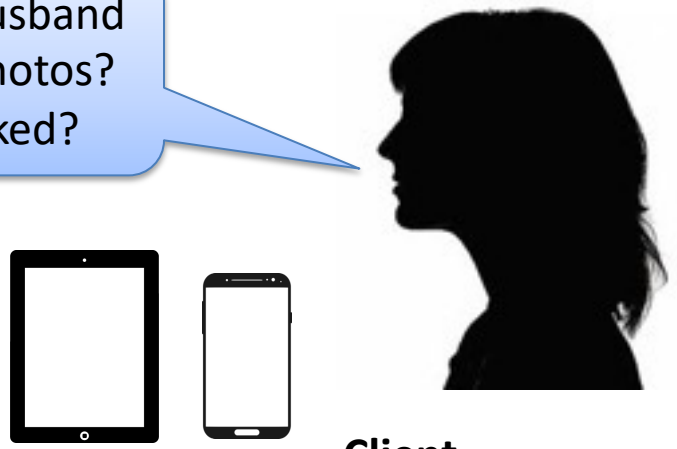
[Havron et al. Security 2019]
[Freed et al. CSCW 2019]

- Developed procedures with help of 14 focus groups with IPV support professionals
- Custom spyware detection tool (ISDi)



IPV tech consultant

How does my ex-husband keep getting my photos? Is my phone hacked?



Client (IPV victim/survivor)

Initial field study started in 2018 (44 clients)

- 105 devices, scanned 75 for spyware
- (potential) spyware apps or browser extensions on devices of 3 clients
- 1 rooted iPhone with spyware
- many account compromises
- reassured many clients about lack of tech problems or by improving their security

Clinical computer security

[Havron et al. Security 2019]
[Freed et al. CSCW 2019]

- Field study segued into us becoming a ***client service provider*** in NYC
- CETA is free clinic we have operated since 2018 study
  - Help NYC survivors navigate tech abuse
    - ~30 volunteers, handled >500 referrals
    - Referrals from FJCs, Anti-Violence Project
  - Trained 677+ people on tech abuse
  - Advocate of New York City 2019 Award
  - ***Instrumented:*** enables ongoing research on clinic service delivery and tech abuse

https://ceta.tech.cornell.edu

Dealing with abuse huge part of tech

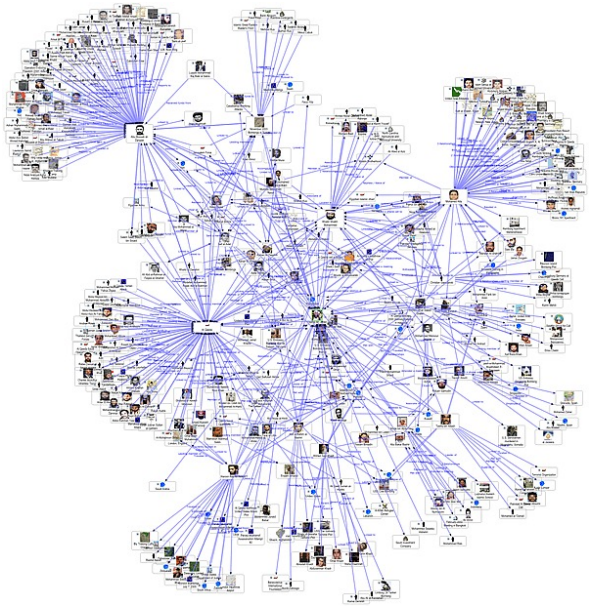
- **Abuse vs. computer security**
  - Abuse: using system's functionality
  - Computer security: exploiting "unintended" functionality
  - Not a clear dividing line, and fascinating interplay
- **Real harms to people**
  - Commercial scamming, fiscal harm
  - Debilitating personal attacks
- **Need socio-technical approaches**
  - Ex: IPV safety review legal restitution, policy, clinical computer security, …

Cambridge Analytica data scandal

- Paid Facebook users to use app and answer survey
- App collected data on each user's Facebook friends
  - 270k people downloaded app
  - 87 million users' data exposed
  - Data used to build psychographic profiles on users
- Profiles used in Trump 2016 presidential campaign to target political ads

Give me data on user and ***all of user's friends***

Facebook OpenGraph API



Cambridge Analytica data scandal

Cambridge Analytica

Discussion question:
Was this a data breach?

Give me data on user and ***all of user's friends***

Facebook OpenGraph API