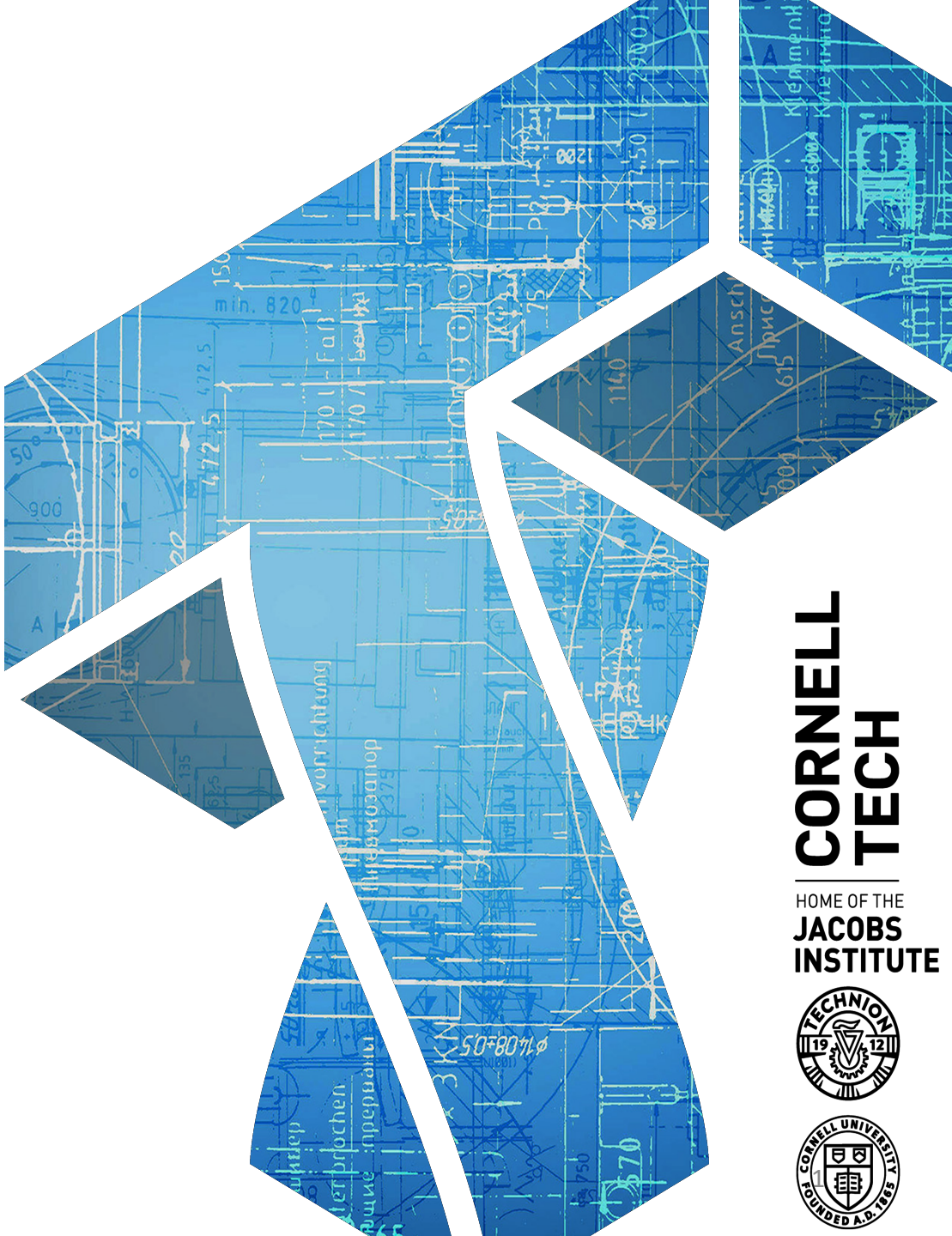
CS 6431: E-crime and botnet takeovers

Instructor: Tom Ristenpart

https://github.com/tomrist/cs6431-fall2021



CORNELL TECH
HOME OF THE JACOBS INSTITUTE

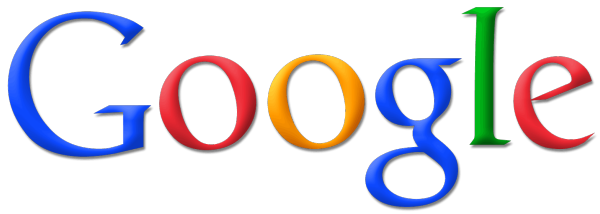
Summary from last few lectures

- Password literature rich and varied
  - Measurement studies, user behavior, NLP, cryptographic protocols, etc.
- One issue that came up:
  - What motivates credential theft?
  - For criminals, “underground” ecosystem of tools, communities, and attack tactics
- Study attackers
  - Methodologically challenging
  - One approach is to build new measurement instruments

Examples of economically-motivated abuse?

- Spam
  - unsolicited bulk emails
  - Illegal in USA since CAN-SPAM act of 2003
- Scams
  - Nigerian emails (advanced fee fraud / confidence trick)
  - Business email compromise (BEC)
- Phishing
  - trick users into downloading malware, submitting password to attacker, CC info to attacker, etc.
  - Spear phishing: targeted on individuals (used in high-profile intrusions)

What abuse vectors affect these companies?



Removing bad actors from services is hard

- **Facebook**: spammer accounts, fake fraud accounts, …
- **Twitter**: illicit promotional accounts, …
- **Yelp:** fake reviews, fake restaurants, …
- **AirBNB**: scam rental or experience ads
- **Lyft**: colluding drivers + riders
- **WhatsApp**: scams, phishing, misinformation
- **Google**: spammers using Gmail, advertisers violating terms of service, SEO, Play store bad apps, …

In industry defense against abuse increasingly falls to trust and safety teams

Jon

Junk - Exchange August 24, 2019 at 3:16 PM

J

Official request

Reply-To: jocun01250@mail2banker.com

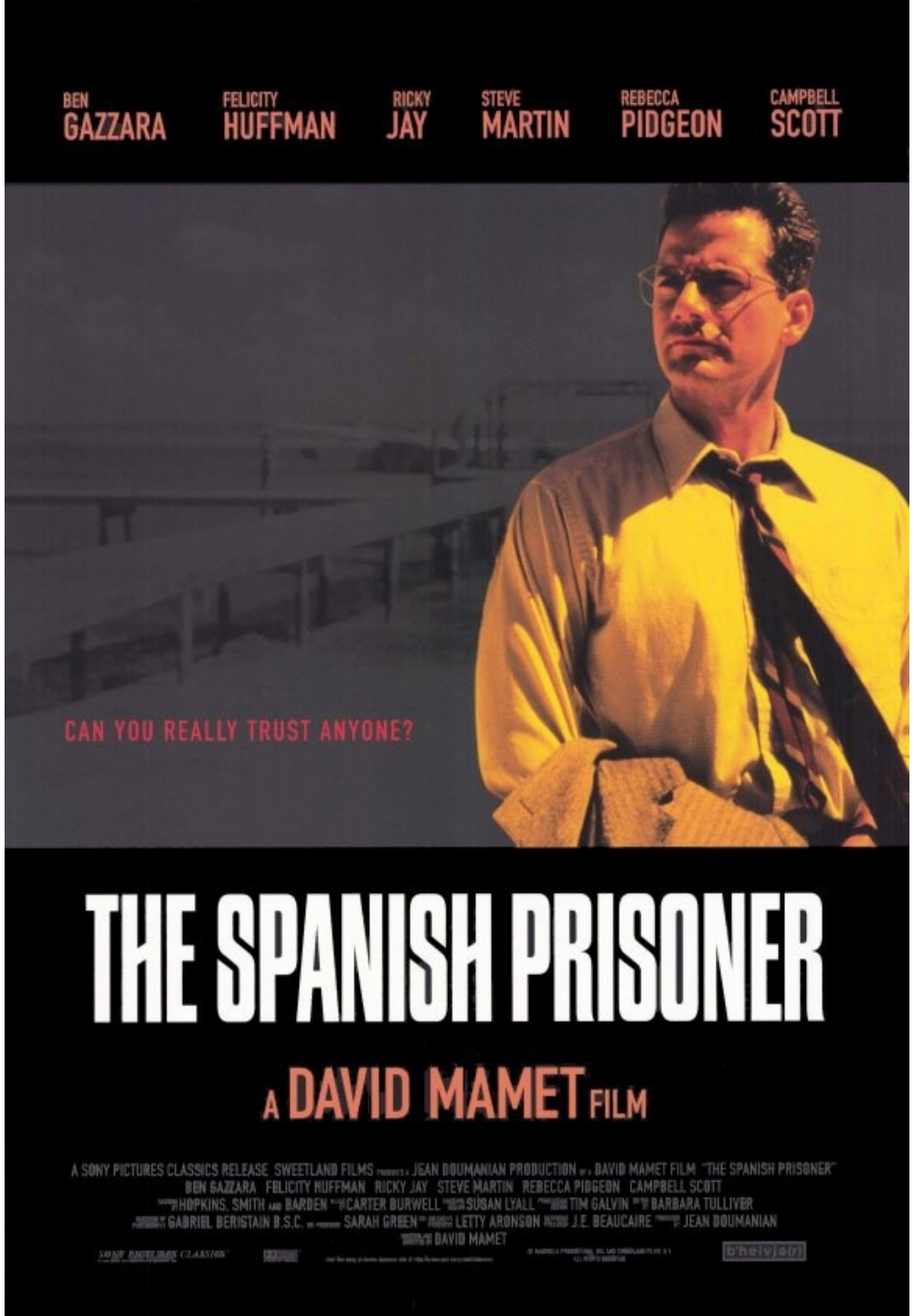
Hello,

This is an official request to you, we have a deceased client whose inheritance we wish to payout to avoid seizure. We are no longer paying for private searches to trace his relatives after today. We contacted you knowing you could be. Please treat urgent.

Regards,
Jon

Spanish Prisoner confidence trick

- 19th century
- In contact with rich guy in Spanish prison
- Just need a little money to bribe guards, he'll reward you greatly
- ***Advance-fee scam***



https://www.ic3.gov/Media/PDF/AnnualReport/2020_IC3Report.pdf

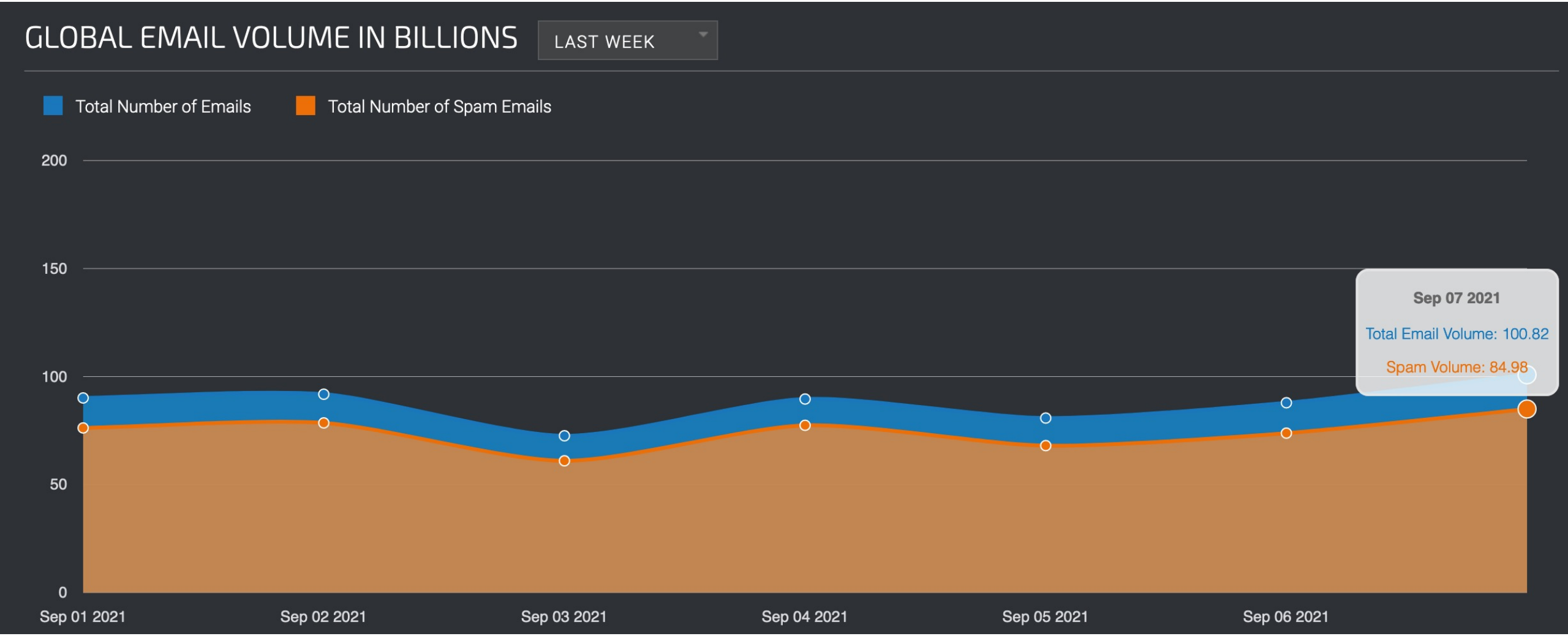
By Victim Loss

Crime Type	Loss	Crime Type	Loss
BEC/EAC	$1,866,642,107	Overpayment	$51,039,922
Confidence Fraud/Romance	$600,249,821	Ransomware	**$29,157,405
Investment	$336,469,000	Health Care Related	$29,042,515
Non-Payment/Non-Delivery	$265,011,249	Civil Matter	$24,915,958
Identity Theft	$219,484,699	Misrepresentation	$19,707,242
Spoofing	$216,513,728	Malware/Scareware/Virus	$6,904,054
Real Estate/Rental	$213,196,082	Harassment/Threats Violence	$6,547,449
Personal Data Breach	$194,473,055	IPR/Copyright/Counterfeit	$5,910,617
Tech Support	$146,477,709	Charity	$4,428,766
Credit Card Fraud	$129,820,792	Gambling	$3,961,508
Corporate Data Breach	$128,916,648	Re-shipping	$3,095,265
Government Impersonation	$109,938,030	Crimes Against Children	$660,044
Other	$101,523,082	Denial of Service/TDos	$512,127
Advanced Fee	$83,215,405	Hacktivist	$50
Extortion	$70,935,939	Terrorism	$0
Employment	$62,314,015		
Lottery/Sweepstakes/Inheritance	$61,111,319		
Phishing/Vishing/Smishing/Pharming	$54,241,075		

Descriptors*

Social Media	$155,323,073
Virtual Currency	$246,212,432

*These descriptors relate to the medium or tool used to facilitate the crime and are used by the IC3 for tracking purposes only. They are available only after another crime type has been selected. Please see Appendix B for more information regarding IC3 data.



https://www.talosintelligence.com/reputation_center/email_rep#tab=1

Botnets

- Botnets:
  - Command and Control (C&C)
  - Zombie hosts (bots)
- C&C type:
  - centralized, peer-to-peer
- Infection vector:
  - spam, random/targeted scanning
- Usage:
  - What they do: spam, DDoS, SEO, traffic generation, …

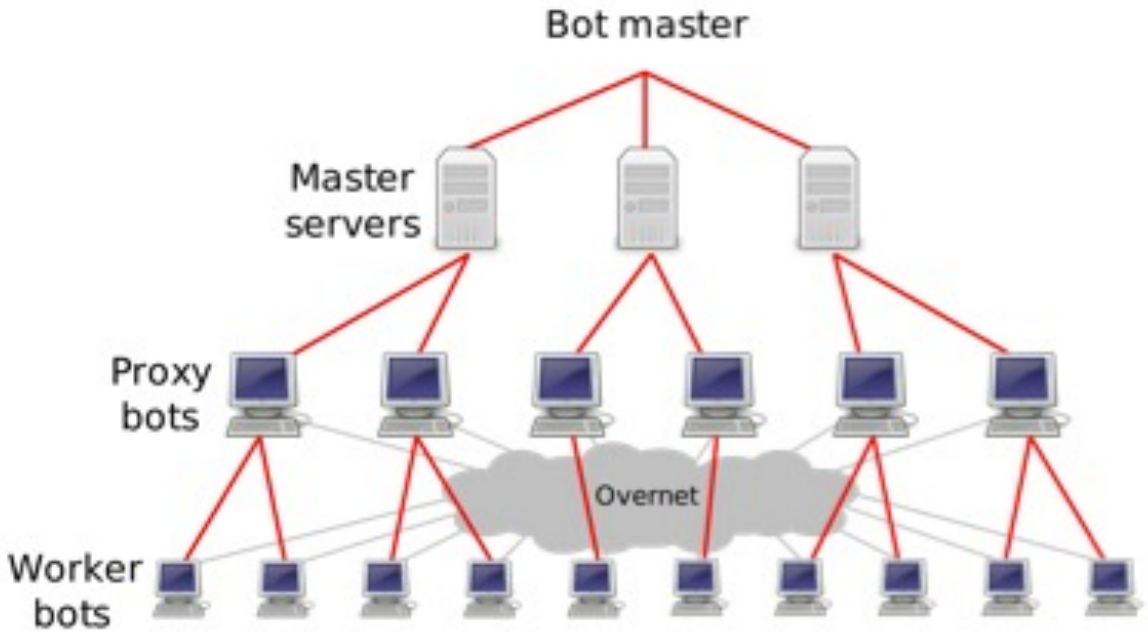


Figure 1: The Storm botnet hierarchy.

Storm botnet (2007-08)

- September 2007
  - Media: 1 – 50 million bots
  - More likely: 10,000s to 100,000s
- Early spam campaigns used titles such as "230 dead as storm batters Europe."
- Propagated via spam linking to malware
- Thought to be controlled by Russian Business Network

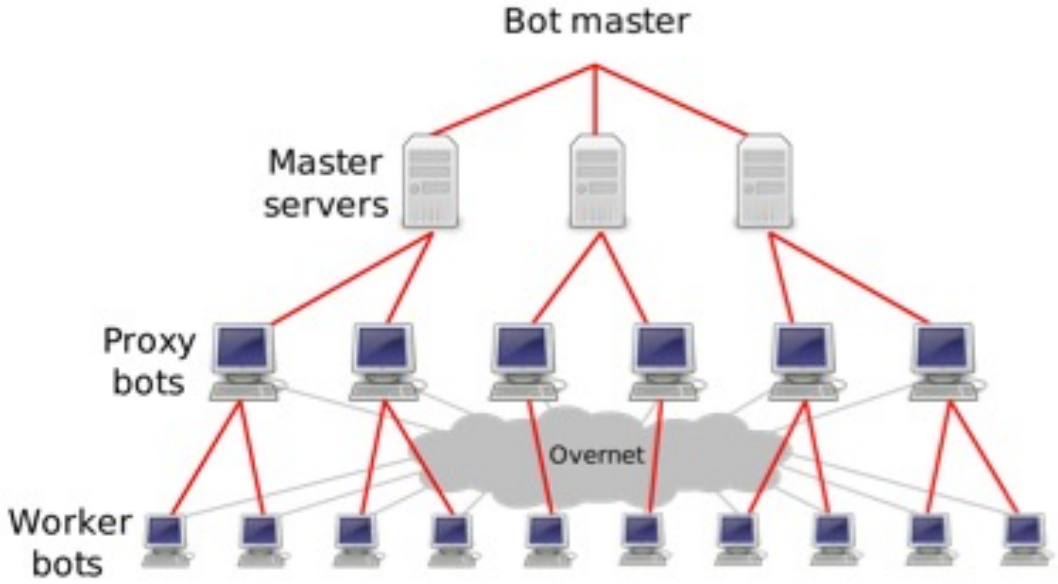


Figure 1: The Storm botnet hierarchy.

Features:

- Uses P2P (Overnet/Kademlia)
- Uses fast-flux DNS for hosting on named sites
- Binary has gone through many revisions
- Features of P2P network have evolved with time
- Hides on machine with rootkit technology

[Enright 2007]

See also: http://cseweb.ucsd.edu/~savage/papers/login08.pdf

Storm botnet (2007-08)

- Used a Distributed Hash Table (DHT) protocol Overnet/Kademilia
- Three types of bots:
  - Master servers (presumably managed by criminals)
  - Proxy bots (can send/receive incoming traffic)
  - Worker bots (can't receive incoming traffic)
- Worker bots find proxies (via DHT) and poll for spam tasks. Send spam
- Proxies route spam campaign templates to workers

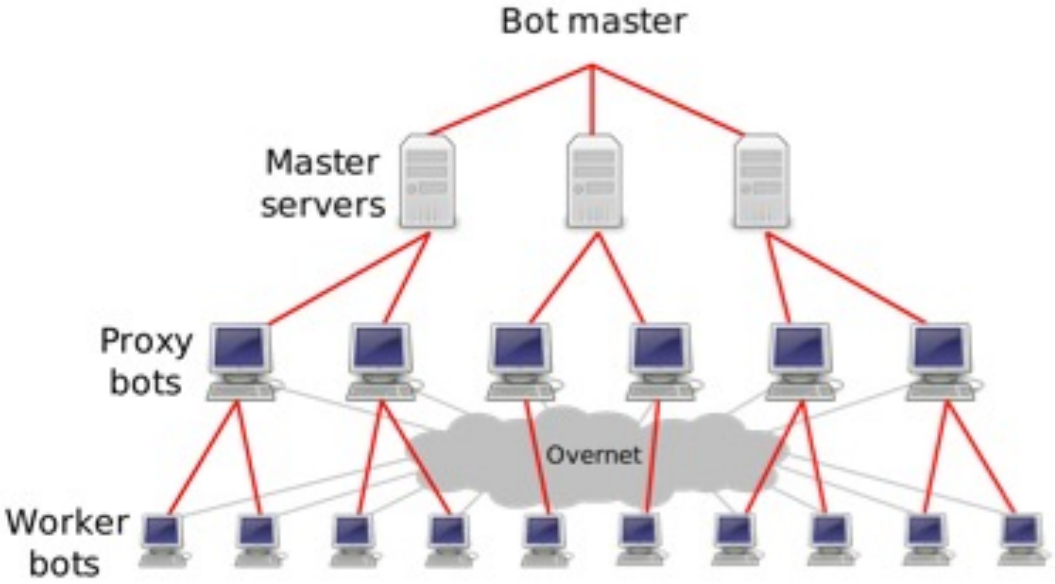
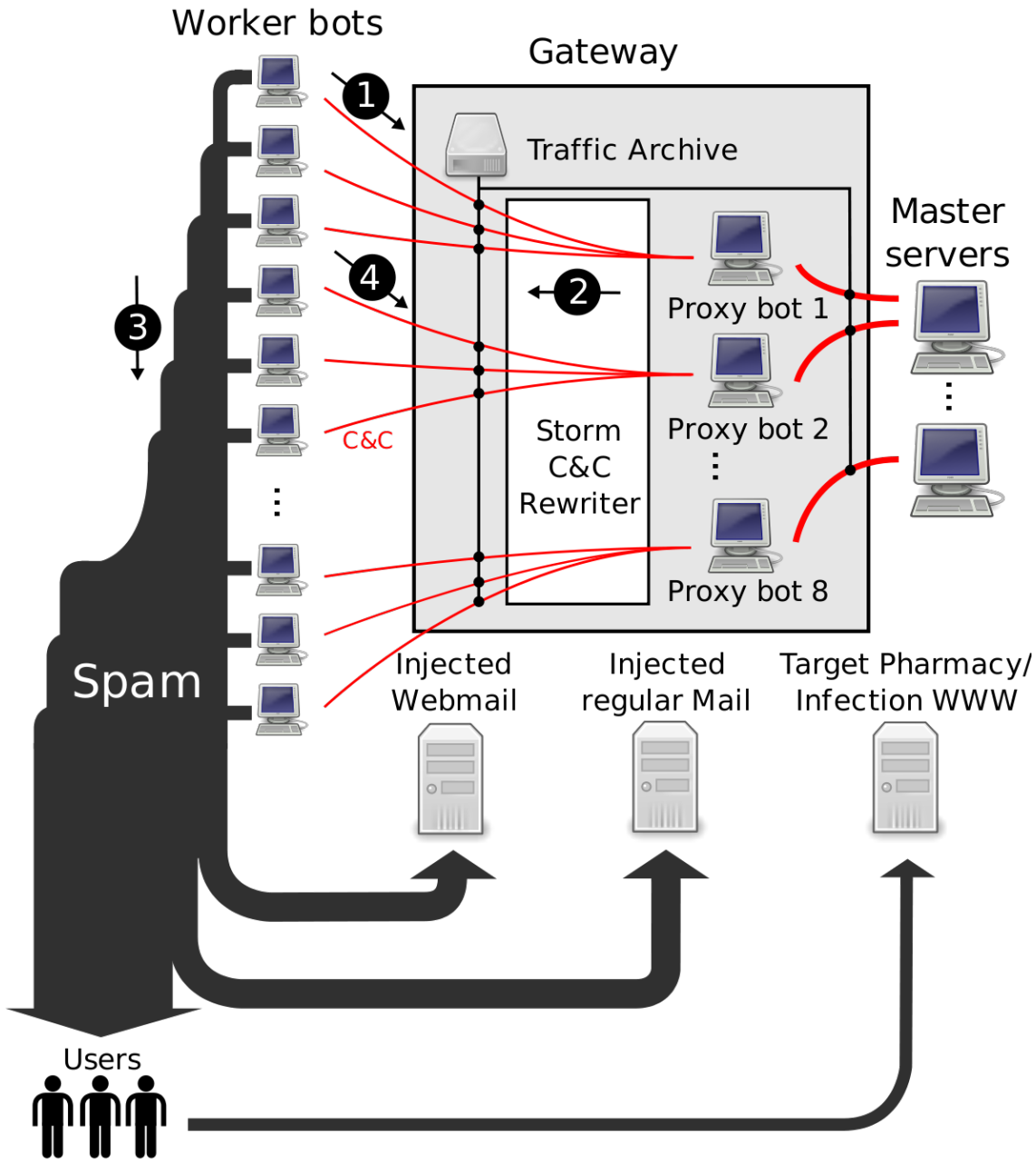


Figure 1: The Storm botnet hierarchy.

Botnet takeover studies

- ***What are spam delivery, response, & conversion rates?***
  - Delivery rate: fraction of spam that ends up in user inbox
  - Response rate: fraction of spam for which user clicks link
  - Conversate rate: fraction of spam leading to purchase (or infection)

- Spamalytics [Kanich et al., 2008]
  - Storm botnet
  - Rewrote spam to redirect to researcher-controlled websites



Kanich et al., Spamalytics: An Empirical Analysis of Spam Marketing Conversion, 2008

Conversion rate measurements

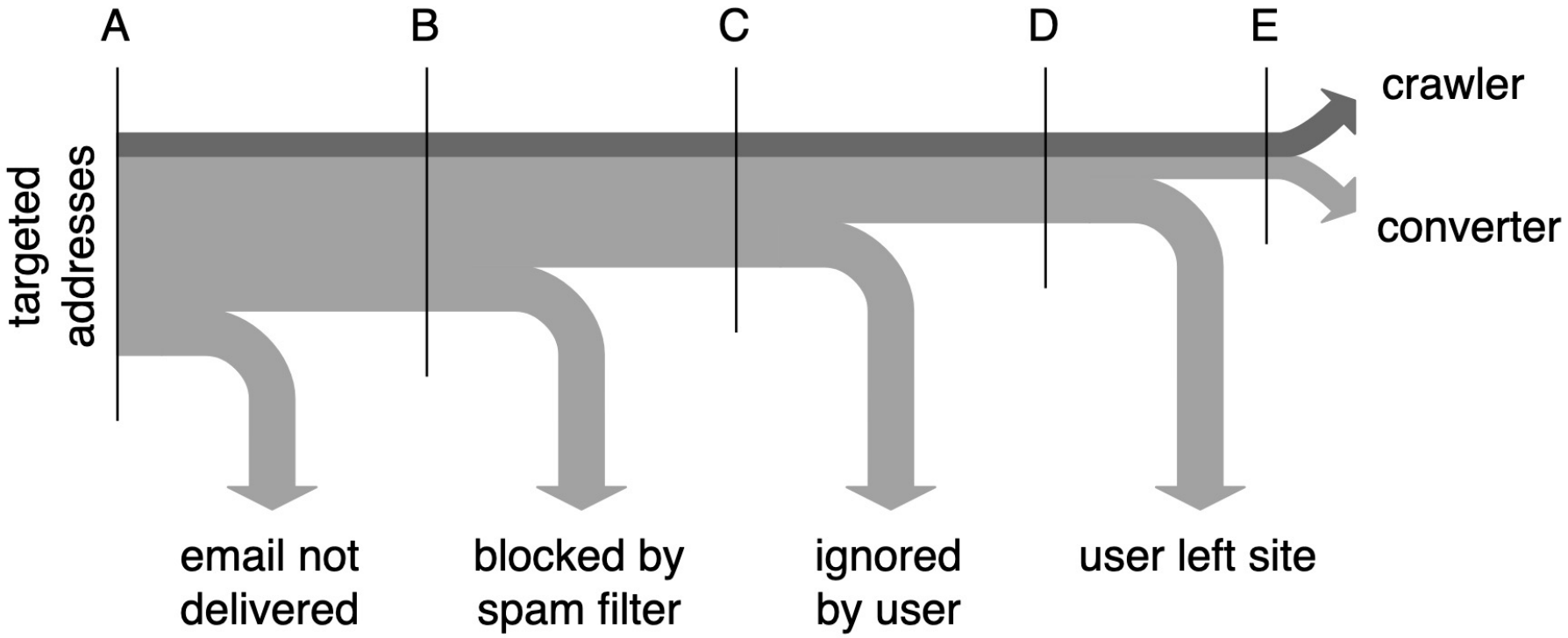


Figure 6: The spam conversion pipeline.

STAGE	PHARMACY		POSTCARD		APRIL FOOL	
A – Spam Targets	347,590,389	100%	83,655,479	100%	40,135,487	100%
B – MTA Delivery (est.)	82,700,000	23.8%	21,100,000	25.2%	10,100,000	25.2%
C – Inbox Delivery	—	—	—	—	—	—
D – User Site Visits	10,522	0.00303%	3,827	0.00457%	2,721	0.00680%
E – User Conversions	28	0.0000081%	316	0.000378%	225	0.000561%

The victims

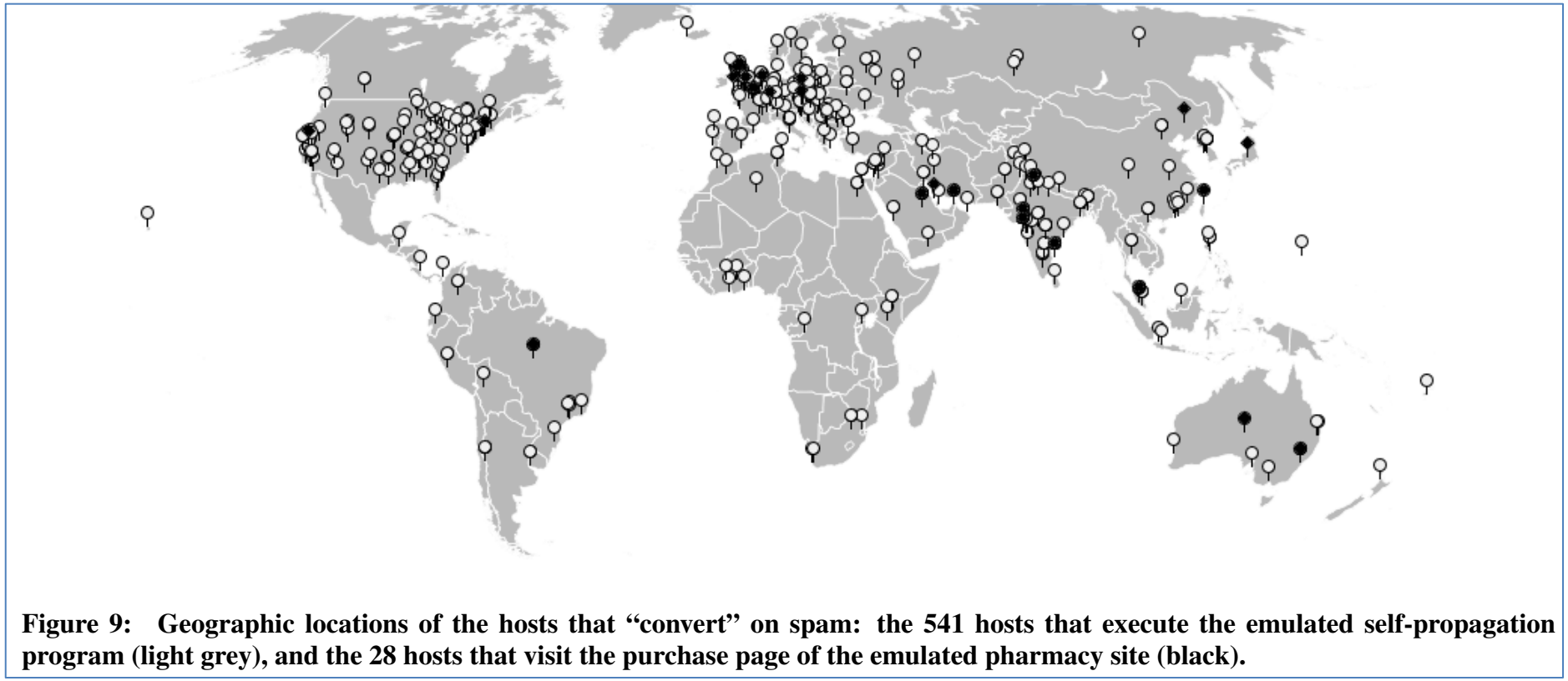


Figure 9: Geographic locations of the hosts that "convert" on spam: the 541 hosts that execute the emulated self-propagation program (light grey), and the 28 hosts that visit the purchase page of the emulated pharmacy site (black).

Kanich et al., Spamalytics: An Empirical Analysis of Spam Marketing Conversion, 2008

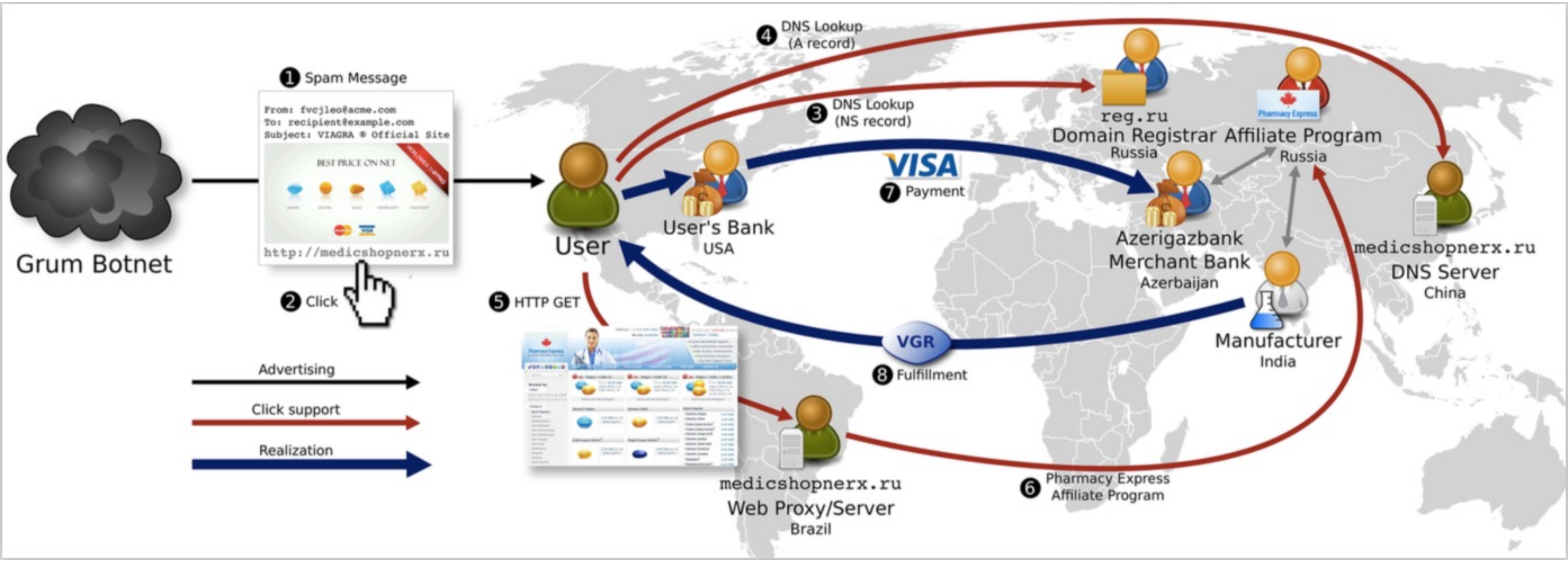
Take-aways

- Exceedingly low conversion rate
  - But, high volume: estimate $7,000 USD revenue per day
  - "*Put another way, the profit margin for spam (at least for this one pharmacy campaign) may be meager enough that spammers must be sensitive to the details of how their campaigns are run and are economically susceptible to new defenses*"
- What defenses are authors alluding to?

(Part of) the research arc by this group

- Measurement studies (scanning of bots to count infections)
- Infiltration studies
  - Better infection counts
  - Conversion rate estimates
- Intervention design studies
  - What parts of spam operations most amenable to disruption?
  - [Levchenko et al. 2011] Click Trajectories paper: 95% of spam-related purchases routed through handful of banks
- Intervention measurements
  - [McCoy et al. 2012]: partner with brand holders to request merchant banks shut down spammers' accounts. Measured efficacy: very effective!

Example spam-advertised goods backend



From Levchenko et al., “Click Trajectories: End-to-End Analysis of the Spam Value Chain”, IEEE Symposium on Security and Privacy, 2011

Ethics of botnet infiltration studies

- What principles does [Kanich et al. 2008] use to justify study?
  - "strictly reduce harm": users less at risk of harms due to study than if study had not been performed
  - "neutral actions": users never worse off due to study mechanisms
- What do you think? Let's discuss

Stone-gross et al. infiltration paper

- Infiltration techniques for P2P botnets like Storm not applicable to botnets with centralized C&C
- Suggest botnet hijack: take over C&C of a botnet
  - Botnets often use changing DNS names (fast flux) to allow look-up of C&C server IP address
  - Take over DNS entries for future names, redirect to researcher IPs. Called “sinkholing”
- Measurement:
  - 10 day hijack of Torpig botnet
  - Gathered data sent by ~180,000 infected hosts: credentials, credit card & bank accounts,

Stone-gross et al. infiltration paper

Data Type	Data Items (#)
Mailbox account	54,090
Email	1,258,862
Form data	11,966,532
HTTP account	411,039
FTP account	12,307
POP account	415,206
SMTP account	100,472
Windows password	1,235,122

Table 1: Data items sent to our C&C server by Torpig bots.

- 28% of victims reuse passwords

Legal & ethical justification for paper

- Cite Burstein 2008: “Conducting Cybersecurity Research Legally and Ethically”
- Principles:
  - The sinkholed botnet should be operated so that any harm and/or damage to victims and targets of attacks would be minimized
  - The sinkholed botnet should collect enough information to enable notification and remediation of affected parties.
- Did design realize these principles?

Some questions I had reading paper

- Paid hosting providers "*well-known to be unresponsive to abuse complaints*"
  - This sounds like grey-market or criminal "bullet proof" hosting
- Data handling underspecified
  - Why download the data? Could instead calculate & record summary statistics. This would seem to reduce harm much more
  - How was research access secured and what procedures in place?

Menlo report

- The Menlo Report: Ethical Principles Guiding Information and Communication Technology Research. 2012
- Adapts the Belmont Report's principles to computing technology research
  - Material differences in new domain
- Now a go-to reference for trying to understand and argue for experiment design

Studying e-crime

- Commercially/economically-motivated attacks widespread.
- New measurement techniques needed to understand economics, attacker behaviors, and disruption opportunities
- Ethically fraught landscape for research