Spamalytics

Steve Johnson

Introduction

- What percentage of people click on spam?
- How profitable is spam?
- Answer these questions for a better understanding of how to stop spam
- But how to answer them?

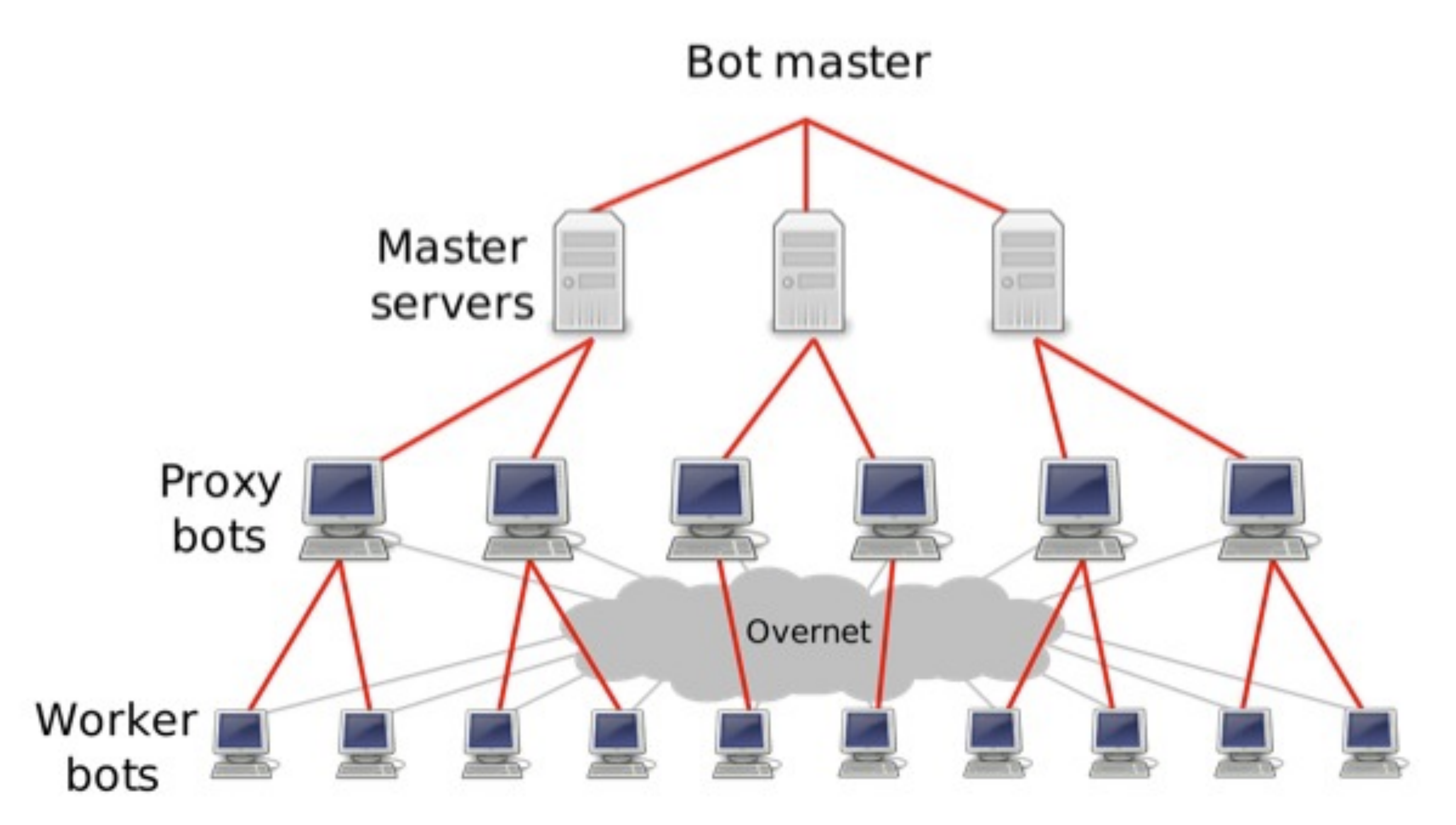
Overall Methodology

- Temporarily take control of part of the Storm botnet
- Send through spam, but change URLs to point to their own servers
- Analyze results using data from web sites, botnet workers

Economics of Spam

- Junk mail costs about $250-1000 per thousand to send with a conversion rate of 2.15%
- Ease of sending email begat spam on a huge scale, and a spam arms race
- Spam costs ??? per thousand with a conversion rate of ???
- Filling in ???s may help us win the arms race using economics

The Storm Botnet



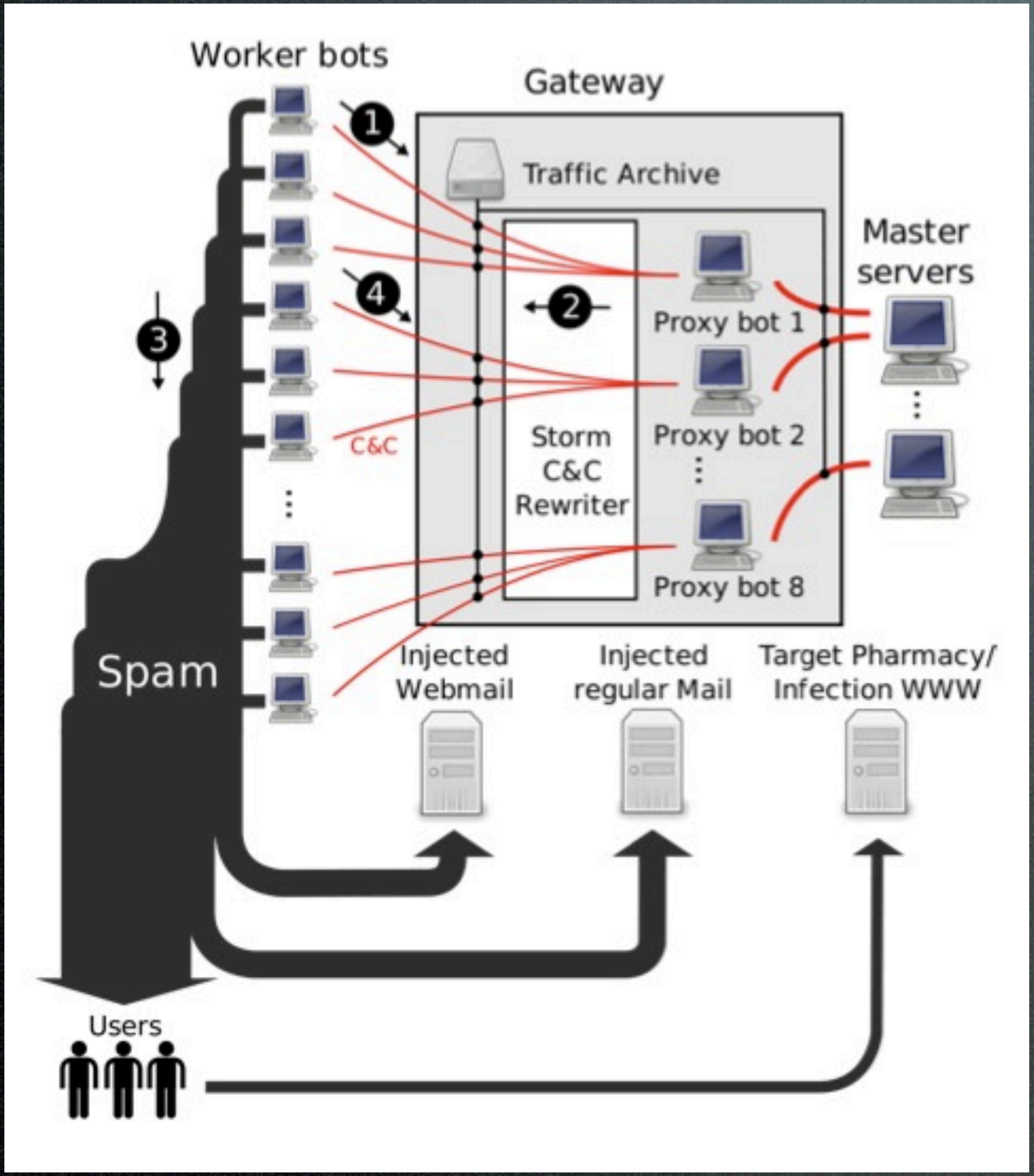
Storm: Connecting

- Populate “bootstrap list” from parent, from random IDs, and from found peers
- Connect to peers
- Publicize self to peers

Storm: Storing/Finding

- DHT interface
- Time-based “rendezvous code” to find each other. One for yesterday, today, and tomorrow.
- Combine date with random integer 0-31 for 32 total keys per day
- Used to rendezvous with C&C nodes, which publish their IP+port for others to find and connect to

Storm: Spamming



(2)

Emails:
stephen.r.johnson@case.edu,
barbara.snyder@case.edu,
misha@case.edu

Subject:
{adj} {synonym_for_viagra} for you

Body:
Two {pills} of {synonym_for_viagra}
10.99{!!!} {url}

(4)
stephen.r.johnson: success
barbara.snyder: success
misha: failure

Invading Storm

- Allow virtual machines to be infected and elevated to proxy status
- Route bot traffic through a gateway which rewrites URLs and blocks DDOS requests
- Now the workers are spamming with the researchers' URLs which they can analyze hits to

Measuring Delivery

- Ability to pass filters measured by setting up test email accounts and inserting the addresses into jobs
- Remove them from results to hide them from real Storm controllers
- Some extra email received there due to dictionary bots, “leakage” in Storm

Measuring Conversion

- URLs in dictionary rewritten to be researcher-controlled URLs with unique IDs appended
- Focus on two types of campaigns: self-propagation and pharmaceuticals
- Pharmaceutical campaigns point to affiliate web sites
- Self-propagation campaigns use executables disguised as greeting cards, April Fools jokes

Measuring Conversion

- To mimic pharmaceutical sites, entire sites cloned except for 404 instead of payment page
- To mimic self-propagation, replace Storm executable with program to send a single HTTP POST to researchers' servers and then quit (to confirm execution of program)

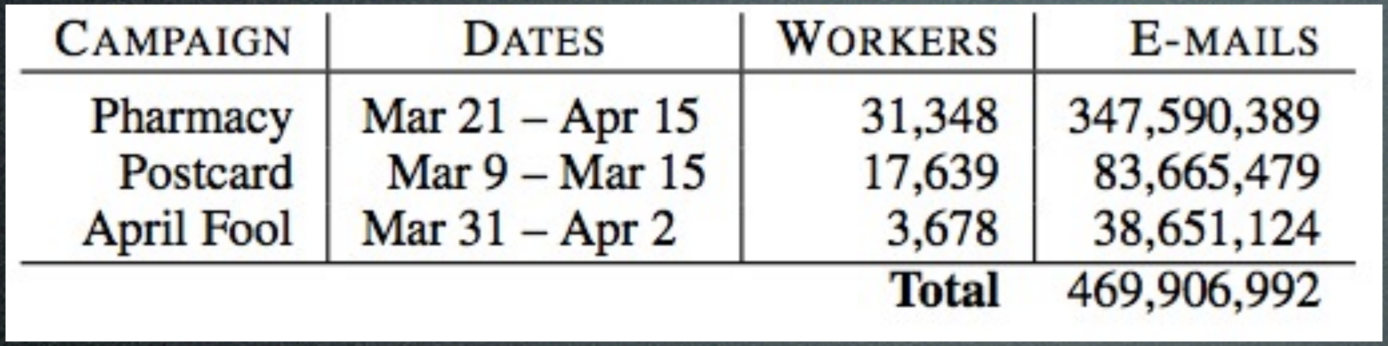
Behavior of Crawlers

- Access URL with no unique identifier
- Access robots.txt
- Disable Javascript and images
- IPs that access with multiple User-Agents
- Downloads executable 10+ times
- Add honeypot IPs to dictionaries that are not sent in spam

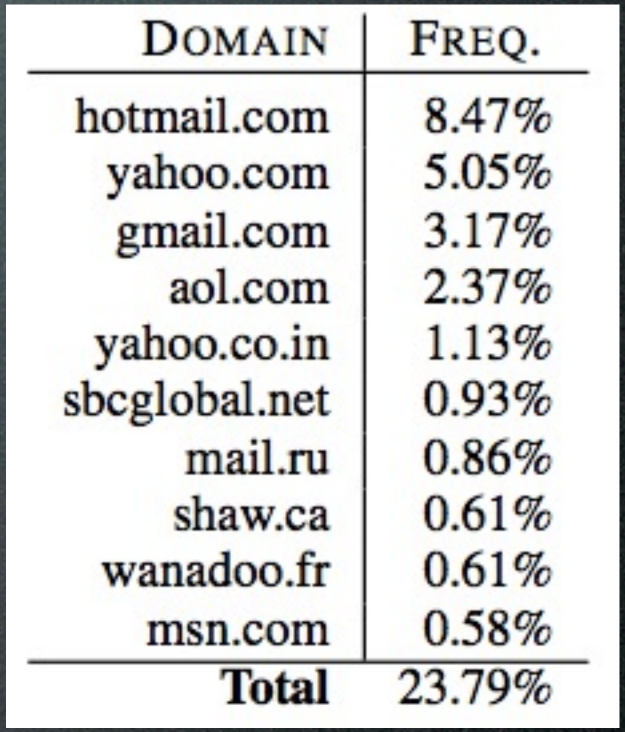
Ethics

- Strictly reduces harm
- Neuters spam messages
- Proxies do not pass through harmful jobs
- Proxies themselves do not participate in spam campaigns

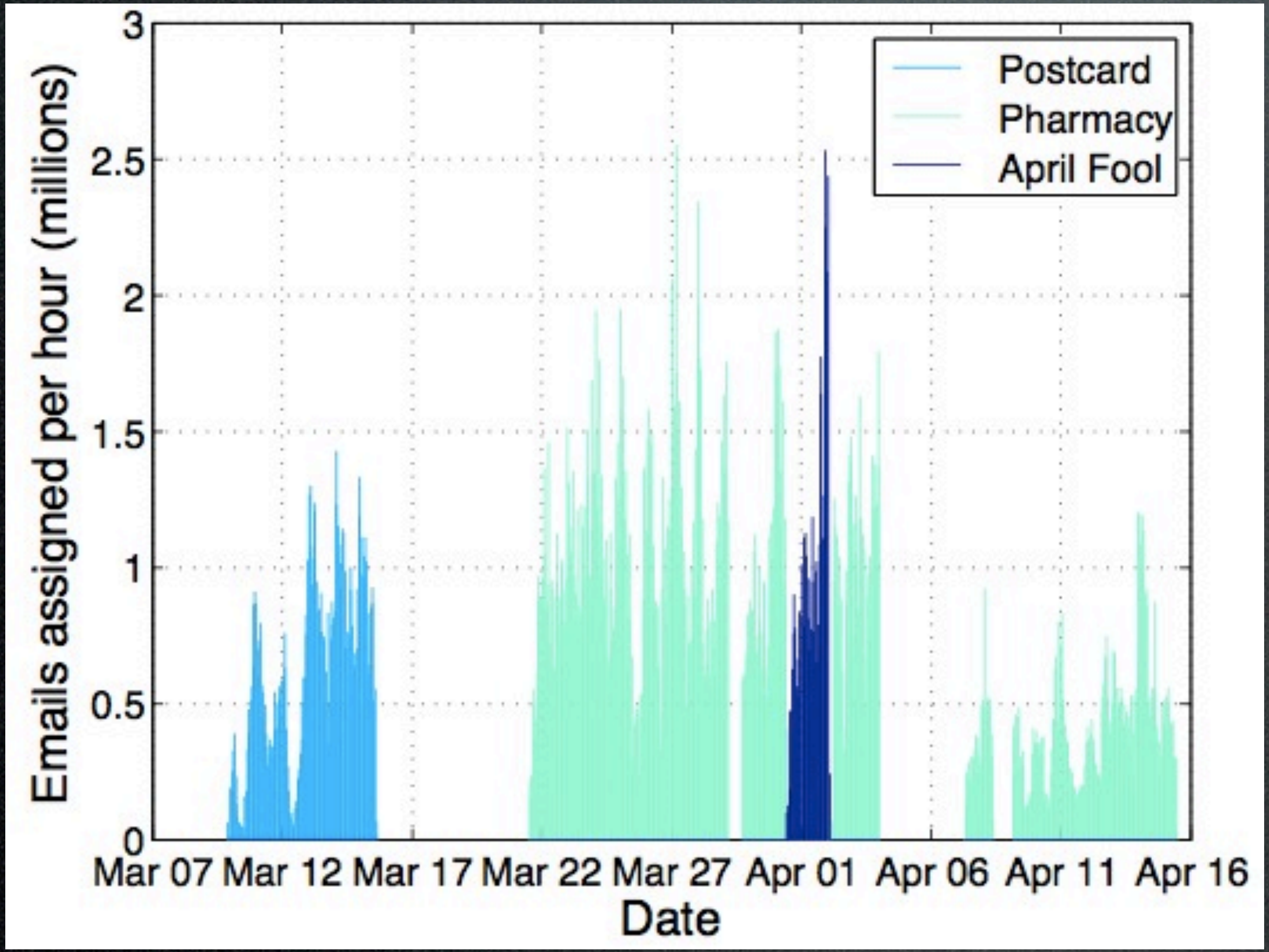
Experimental Results



CAMPAIGN	DATES	WORKERS	E-MAILS
Pharmacy	Mar 21 – Apr 15	31,348	347,590,389
Postcard	Mar 9 – Mar 15	17,639	83,665,479
April Fool	Mar 31 – Apr 2	3,678	38,651,124
		Total	469,906,992



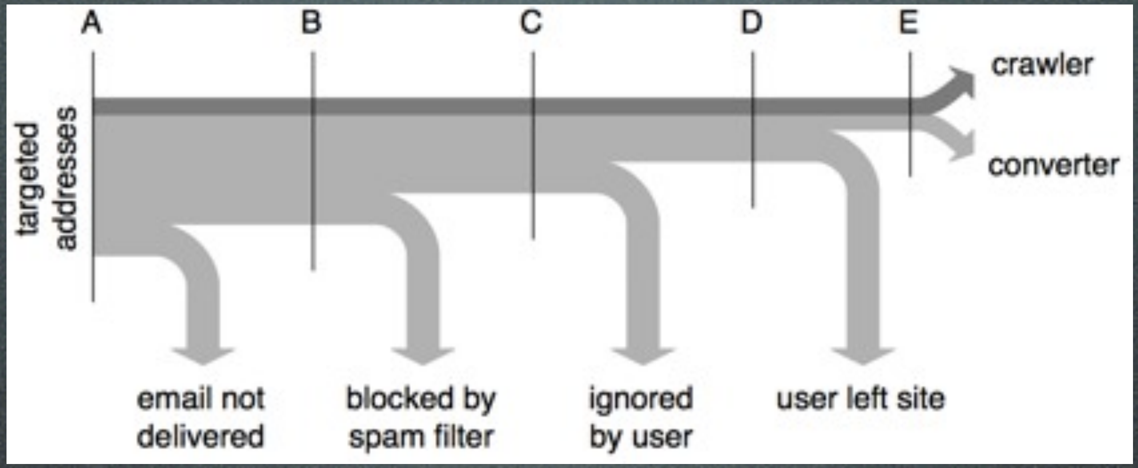
DOMAIN	FREQ.
hotmail.com	8.47%
yahoo.com	5.05%
gmail.com	3.17%
aol.com	2.37%
yahoo.co.in	1.13%
sbcglobal.net	0.93%
mail.ru	0.86%
shaw.ca	0.61%
wanadoo.fr	0.61%
msn.com	0.58%
Total	23.79%

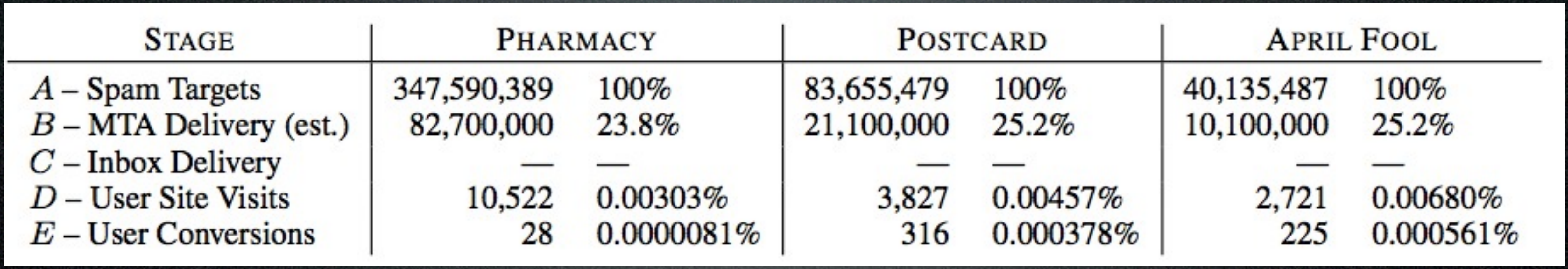


Workers and Spam

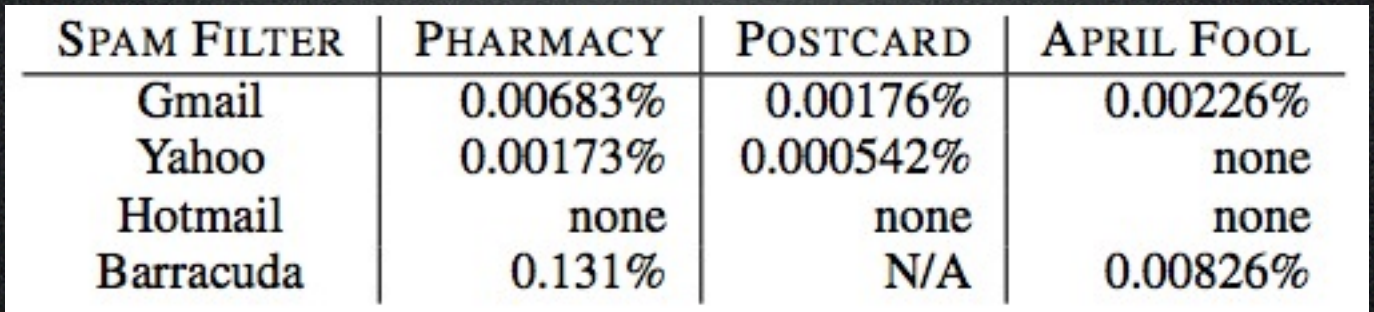
- 78% of workers connected to researchers' proxies once, 92% at most twice, 99% at most 5 times
- 81% connected to only a single proxy, 12% to two, 3% to four, 4% to 5+
- Self-propagation campaign dictionaries ~92% unique addresses
- Pharma dicts ~60% unique

Conversion Rates



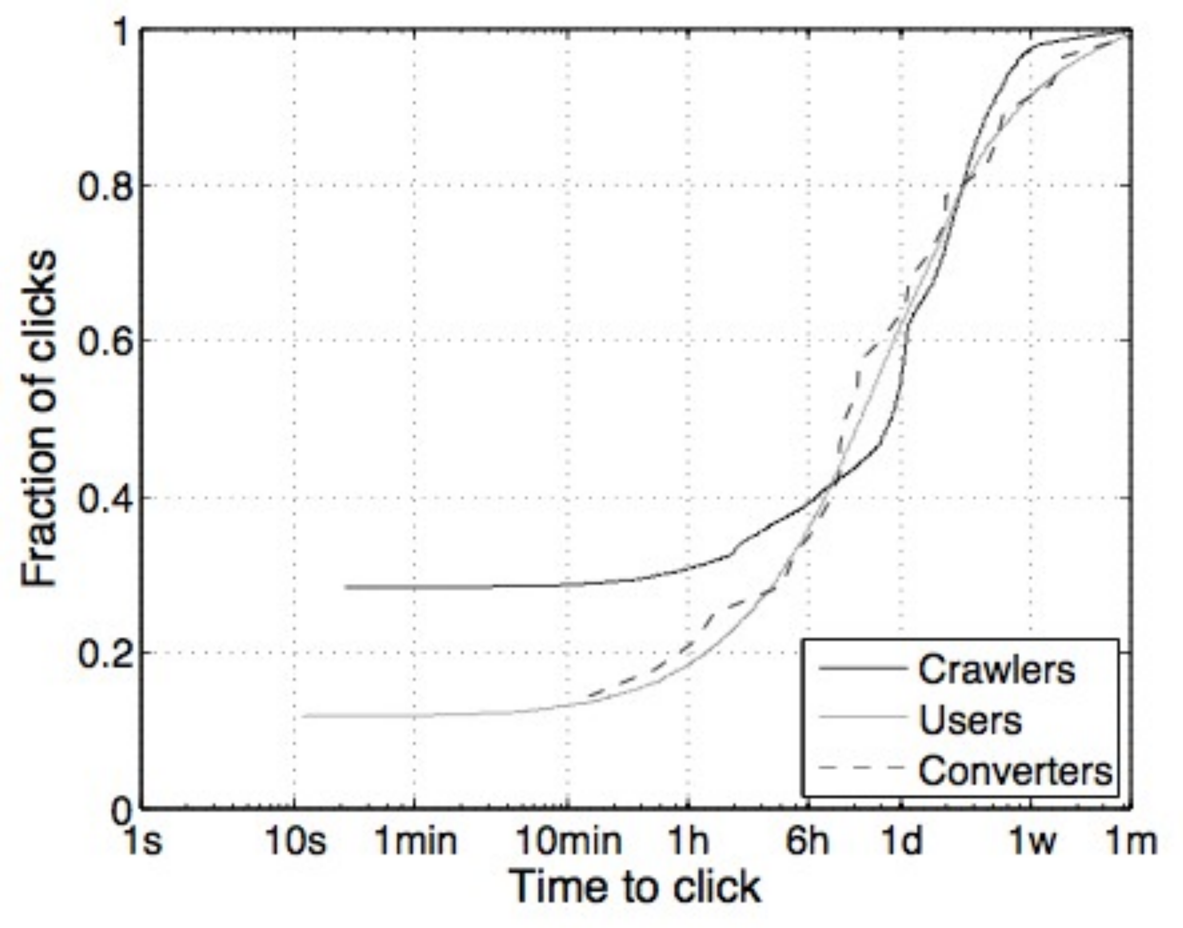


Stage	Pharmacy		Postcard		April Fool	
A – Spam Targets	347,590,389	100%	83,655,479	100%	40,135,487	100%
B – MTA Delivery (est.)	82,700,000	23.8%	21,100,000	25.2%	10,100,000	25.2%
C – Inbox Delivery	—	—	—	—	—	—
D – User Site Visits	10,522	0.00303%	3,827	0.00457%	2,721	0.00680%
E – User Conversions	28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



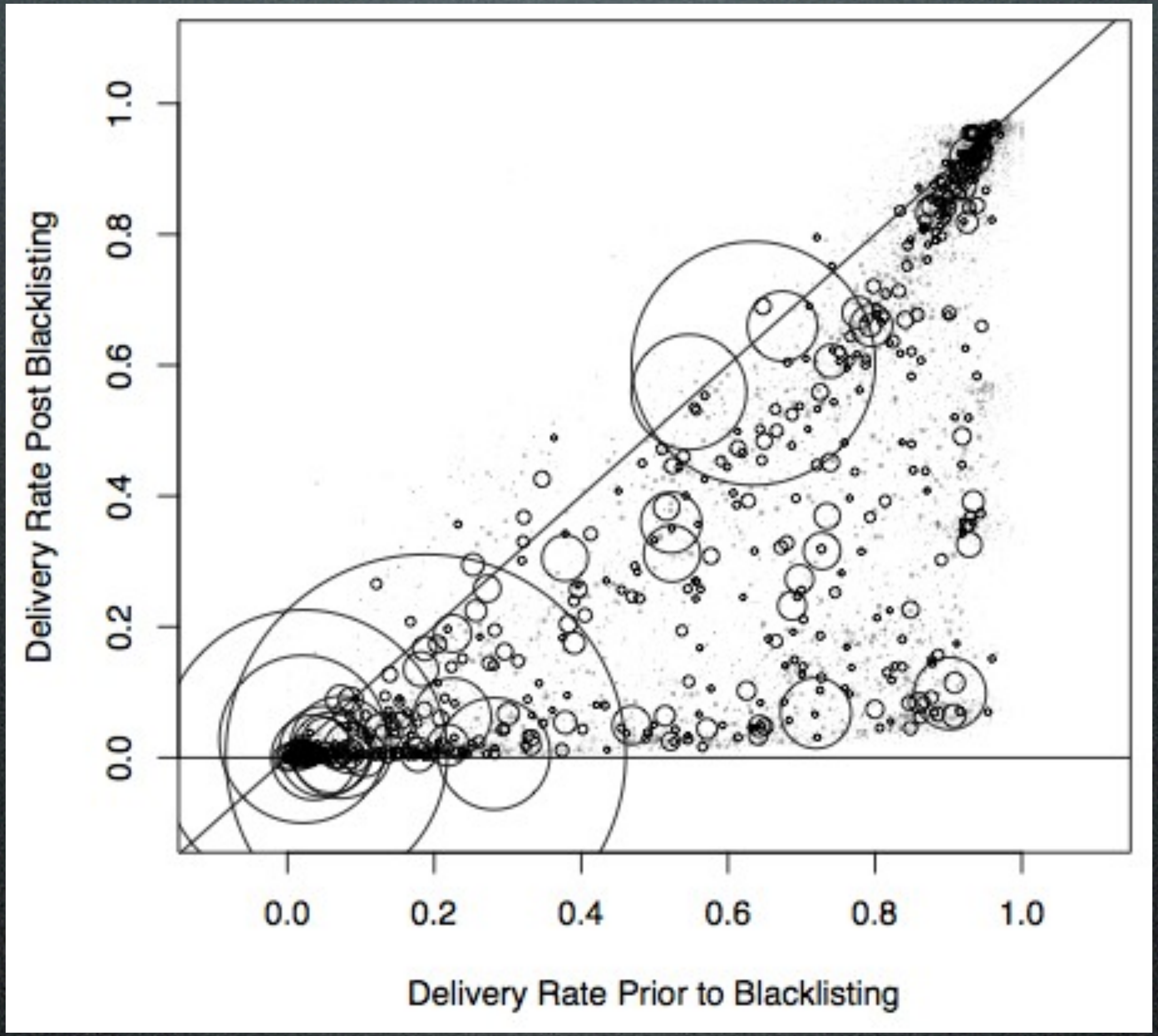
Spam Filter	Pharmacy	Postcard	April Fool
Gmail	0.00683%	0.00176%	0.00226%
Yahoo	0.00173%	0.000542%	none
Hotmail	none	none	none
Barracuda	0.131%	N/A	0.00826%

Crawlers, Time to View



- 87% of page views were from crawlers
- 10% of viewing IPs were crawlers

Effects of Blacklisting



Extrapolation

- Authors make huge disclaimers about all analysis based on sample size
- 28 “sales” for 350,000,000 emails over 26 days
- Average sale price ~$100, so about $140/day
- Researchers controlled 1.5% of proxies, so real revenue probably about $7,000

Extrapolation

- Yearly revenue $3.5M, split 50/50 with affiliates is $1.75M
- "Retail" price of spam delivery $80/M, so $25,000 to send 350M emails which is **not** cost-effective
- Conclusion: Storm controllers are spammers themselves
- Therefore, spammers must be vertically integrated

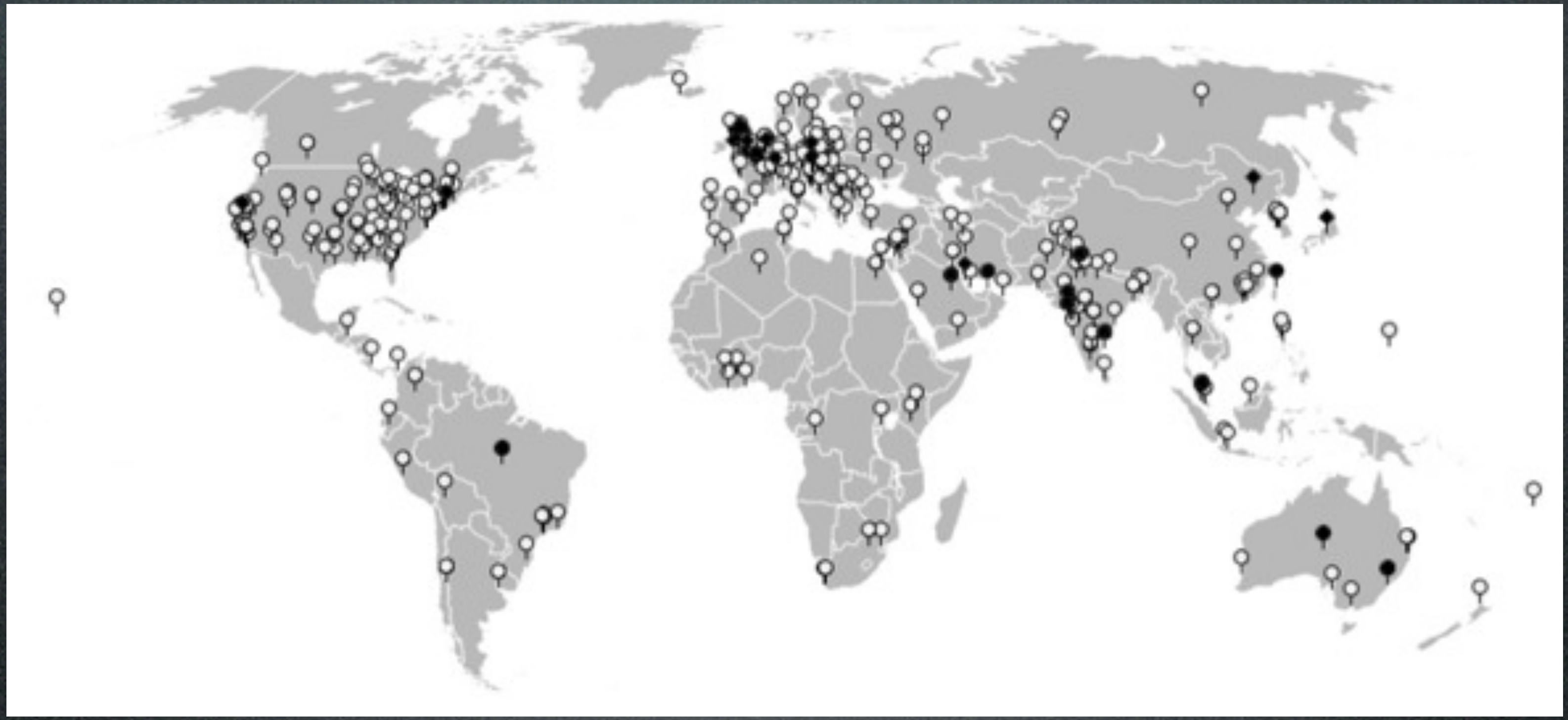
Issues and Questions

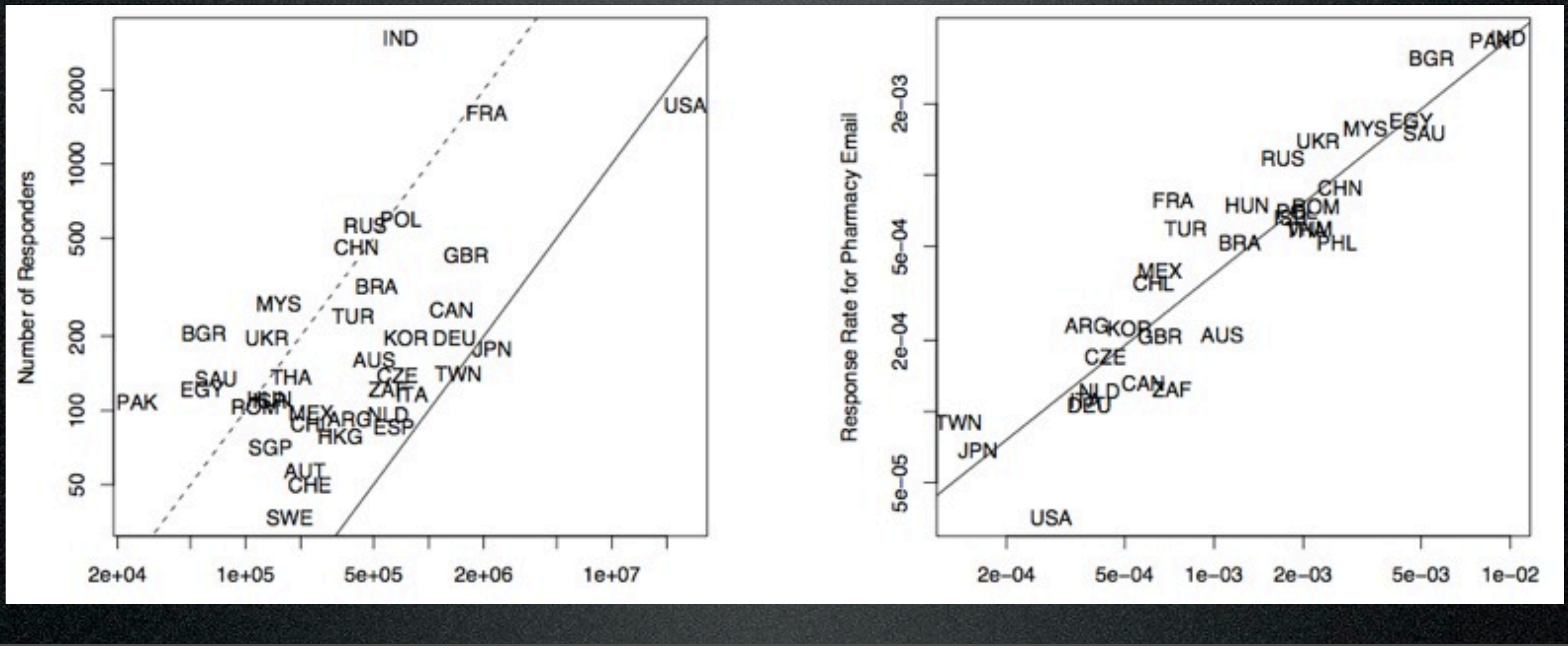
- Lots of extrapolation based on small sample size and anecdotes, even with disclaimers
- Ethics
- If they can detect other researchers, can the botnet controllers detect them?
- How much data needed for statistical significance?

More Questions

- Do you think the reasoning for their extrapolations is fair?
- How representative of spam is their sample?

Geography of Conversions




Number of Responders
2e+04
1e+05
5e+05
2e+06
1e+07
IND
FRA
USA
POL
RUS
CHN
GBR
BRA
MYS
TUR
CAN
BGR
UKR
KOR
DEU
JPN
AUS
SAU
THA
CZE
TWN
EGY
ZAF
ITA
PAK
HUN
ROM
MEX
CHL
ARG
NLD
ESP
HKG
SGP
AUT
CHE
SWE
Response Rate for Pharmacy Email
5e−05
2e−04
5e−04
2e−03
2e−04
5e−04
1e−03
2e−03
5e−03
1e−02
PAK
IND
BGR
MYS
EGY
SAU
UKR
RUS
CHN
FRA
HUN
ROM
TUR
BRA
PHL
MEX
CHL
ARG
KOR
GBR
AUS
CZE
CAN
ZAF
NLD
DEU
TWN
JPN
USA