Measuring and Analyzing Search-Redirection Attacks in the Illicit Online Prescription Drug Trade

Nektarios Leontiadis (CMU/EPP/CyLab) **leontiadis@cmu.edu**

Joint work with Nicolas Christin (CMU) and Tyler Moore (Harvard)

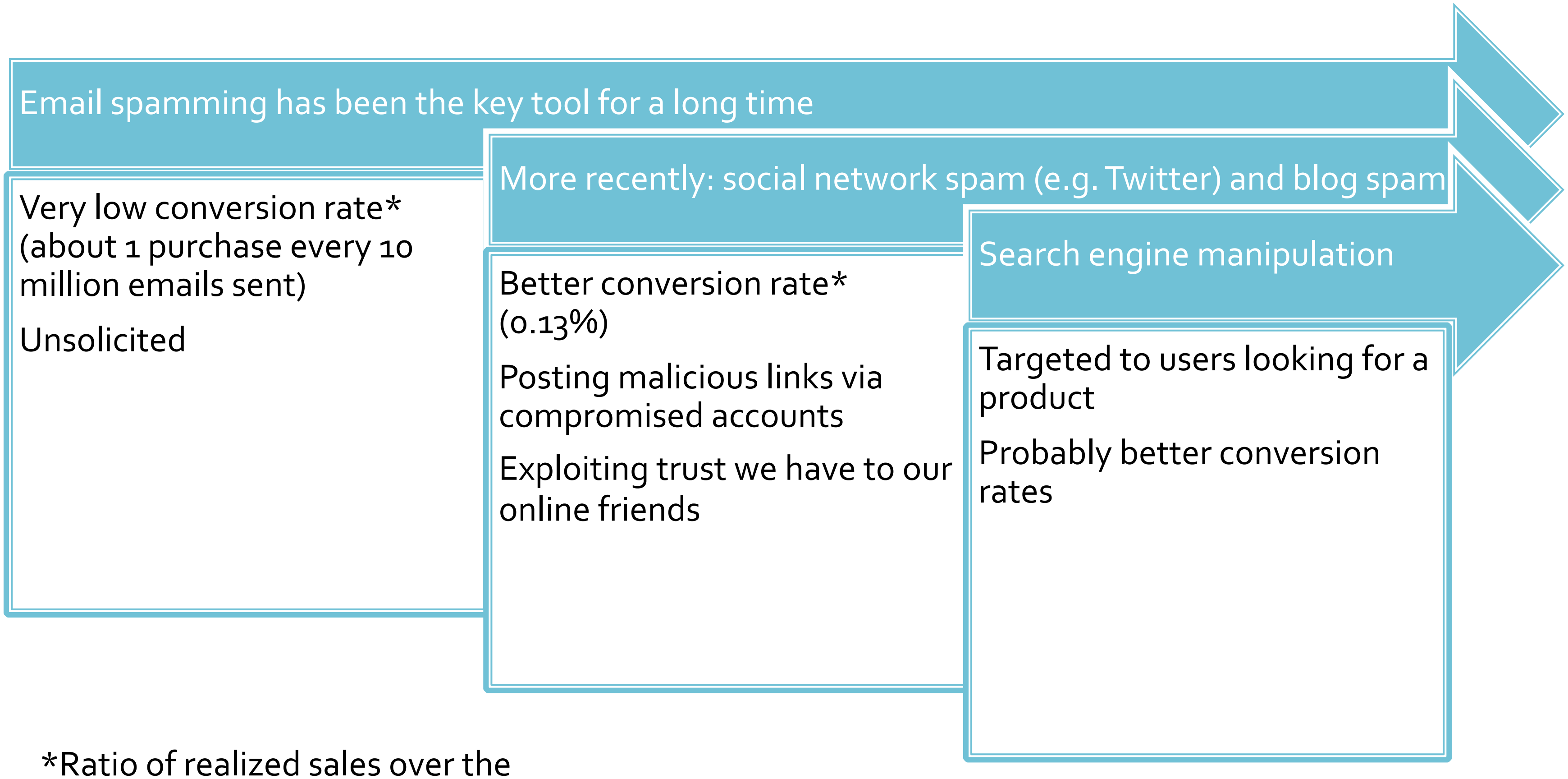
Motivation

- Online crime
  - Emergence of complex supply chains
  - Understanding economics is key to combat it

- Why focus on drugs?
  - What about counterfeit software, fake watches…?
  - Most dangerous form of online crime
    - Wrong dosage can kill, cf. Ryan Haight

- Method of exposure
  - Revealing interesting insights about the mechanics of the illicit trade

Illicit online advertising



Email spamming has been the key tool for a long time

Very low conversion rate* (about 1 purchase every 10 million emails sent)

Unsolicited

More recently: social network spam (e.g. Twitter) and blog spam

Better conversion rate* (0.13%)

Posting malicious links via compromised accounts

Exploiting trust we have to our online friends

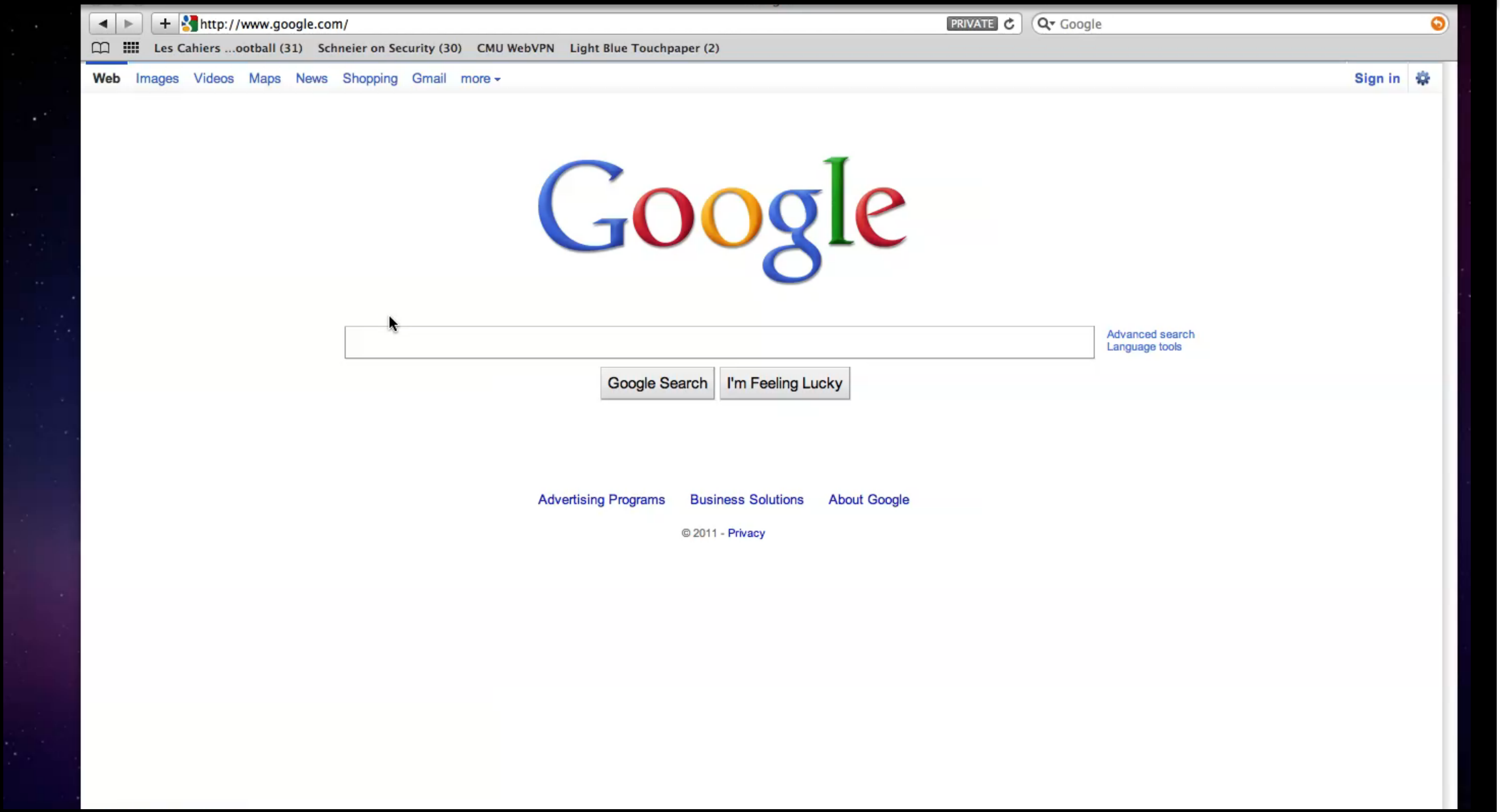
Search engine manipulation

Targeted to users looking for a product

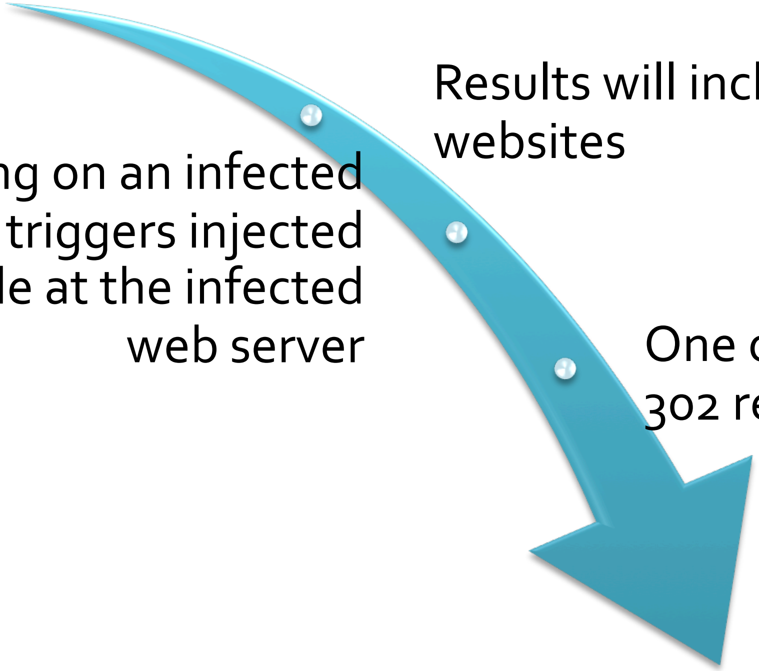
Probably better conversion rates

*Ratio of realized sales over the number of emails/clicks

Search-redirection attack



Attack modus operandi



Bob runs a query on Google (e.g. no prescription cialis)

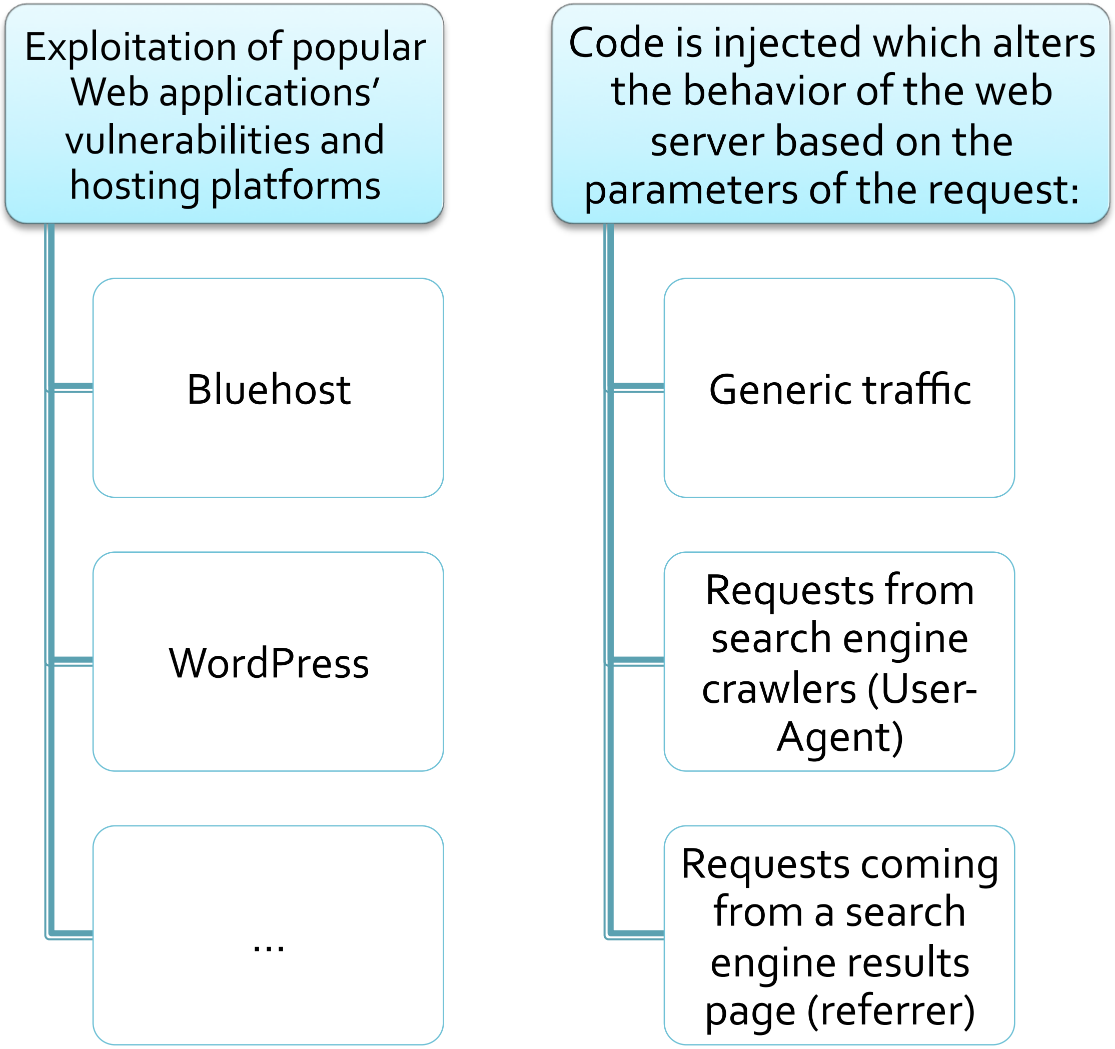
Results will include infected websites

Clicking on an infected result triggers injected code at the infected web server

One or more HTTP 302 redirections occur

Bob lands on an online pharmacy store

Compromise details



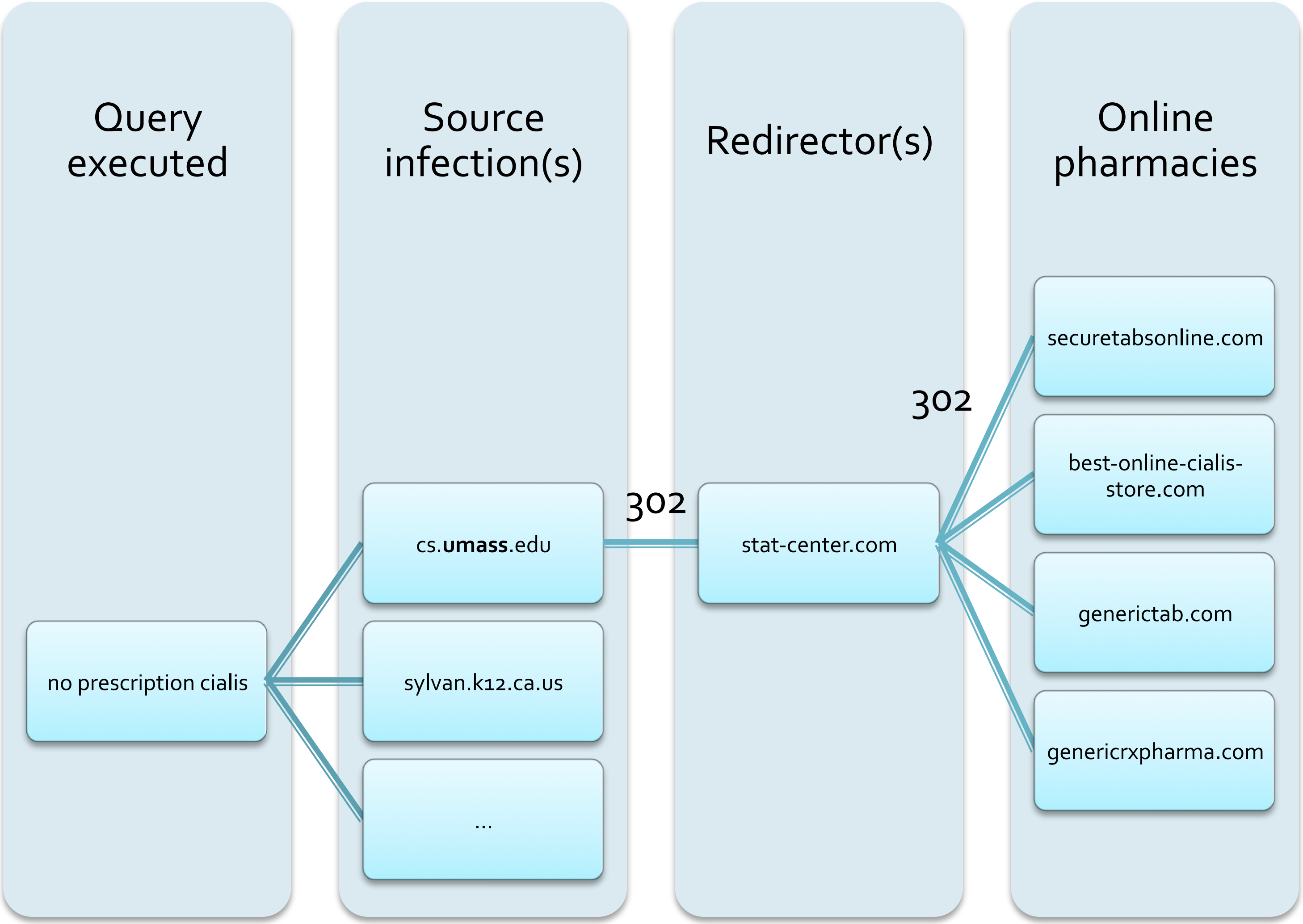
Exploitation of popular Web applications' vulnerabilities and hosting platforms

- Bluehost
- WordPress
- …

Code is injected which alters the behavior of the web server based on the parameters of the request:

- Generic traffic
- Requests from search engine crawlers (User-Agent)
- Requests coming from a search engine results page (referrer)

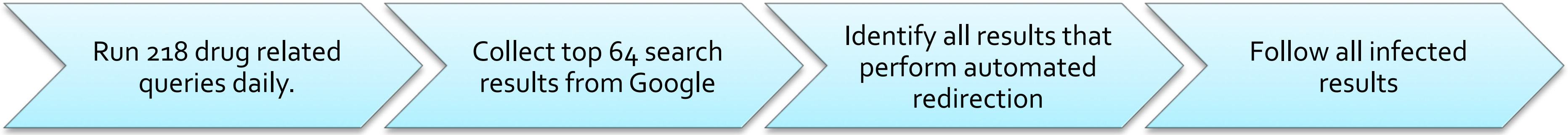
A redirection chain example



Outline of the rest of the talk

1. Experimental methodology
2. Effect of search-redirection attacks on search results
3. Delving into the RX network
4. Sketching conversion rates

Data collection process



Run 218 drug related queries daily.

- Daily collection from 4/12/2010.
- Using data until 10/21/2010.
- Complemented by a second 10-week dataset
- Collection process is still running.

Collect top 64 search results from Google

- The limit is defined by Google Search API
- Storing all results for later processing
- Will also examine position information

Identify all results that perform automated redirection

- A search result defines the website that a user will be redirected to when clicking on the link
- If the browser is redirected instead to a different website (domain), the result is infected.

Follow all infected results

- Follow each result identified as infected from previous step
- Follow all redirections that might occur
- Record all the redirection information

Some of the 218 queries used

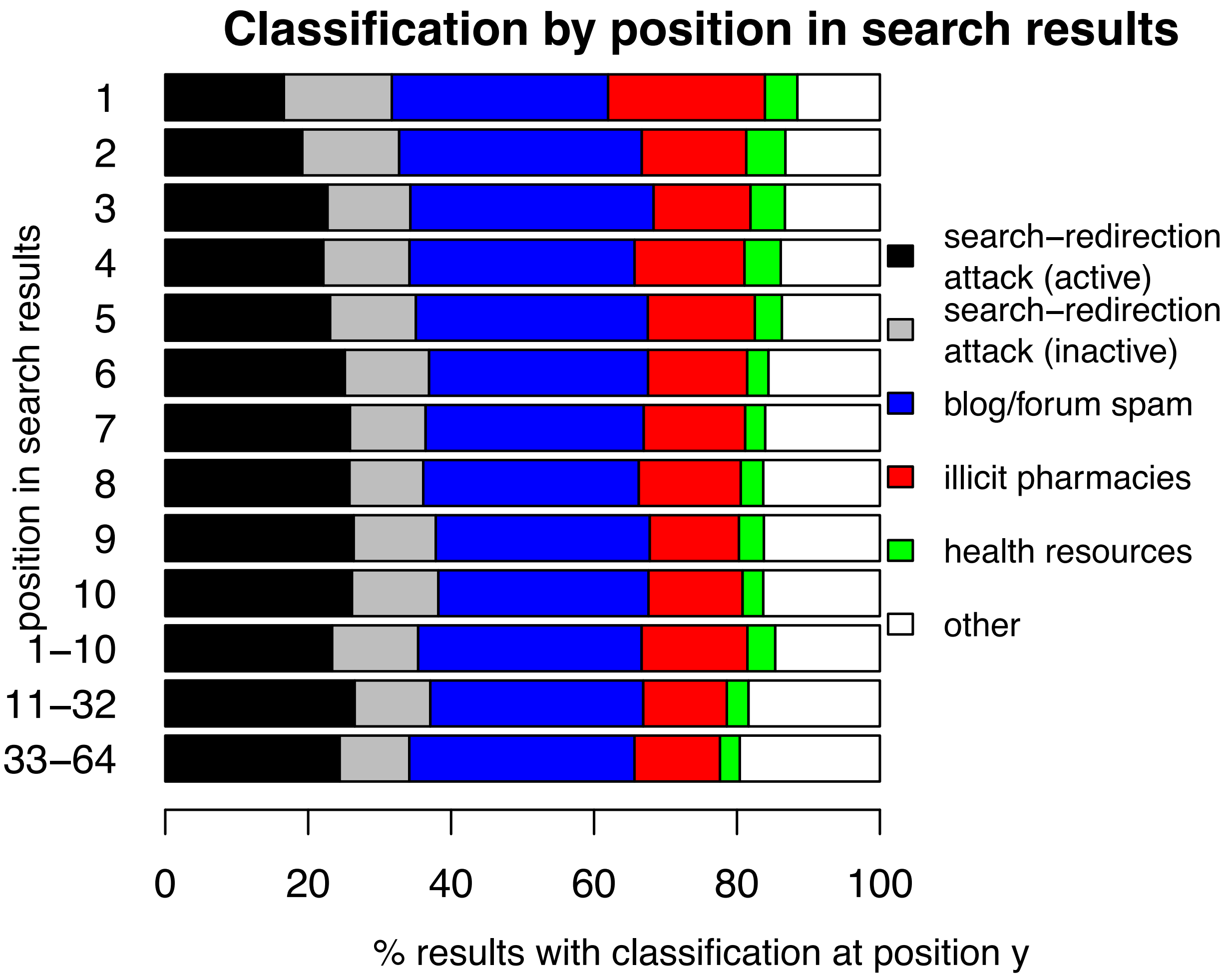
vicodin no prescription
cheap valium non prescription
buy ativan online injecting pills
buy xanax valium online florida
order vicodin si levitra online
buy xanax valium online florida
color of adipex pills safest place to buy online

vicodin without prescription
generic cialis free sample
cheap tadalafil
20 mg ambien overdose
prozac side effects
ambien buy online
alprazolam online without prescription buy cheap

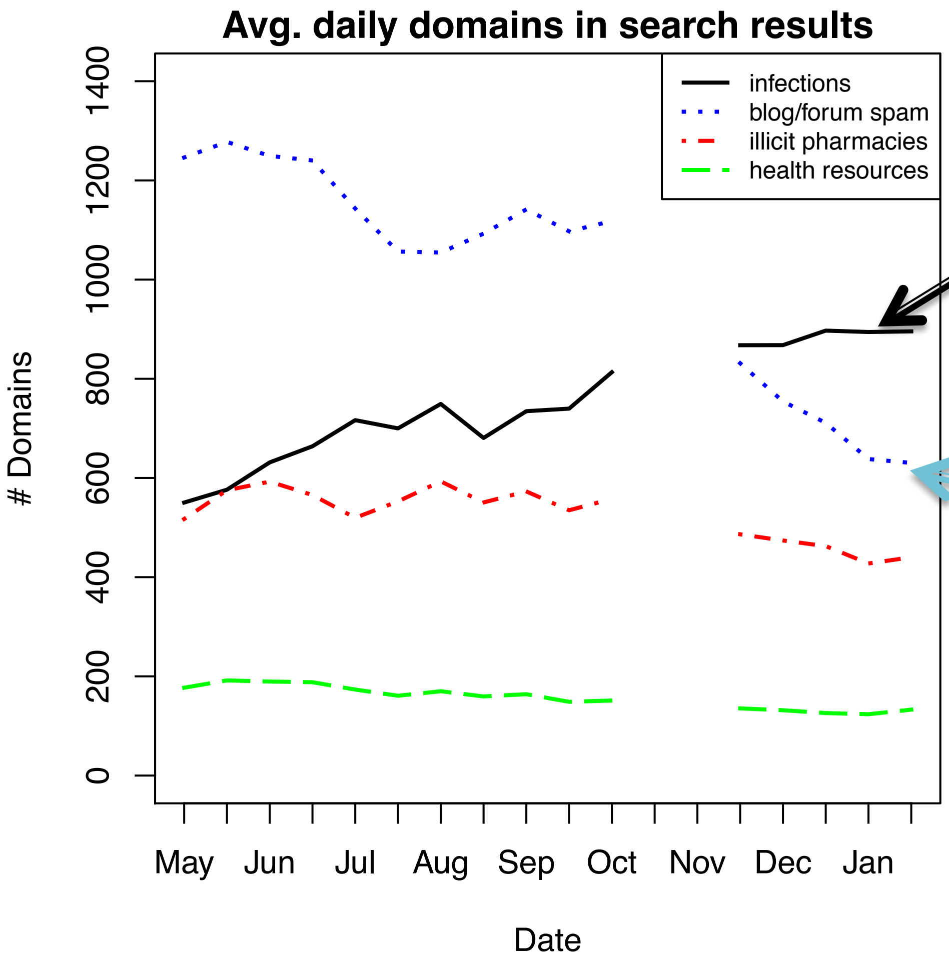
Search results classification

	URI (number)	URI (%)	Domains (#)	Domains (%)
Source infections	73909	53.8	4652	20.2
Active	*44503*	*32.4*	*2907*	*12.6*
Inactive	*29406*	*21.4*	*1745*	*7.6*
Health resources	1817	1.3	422	1.8
Pharmacies	4348	3.2	2138	9.3
Legitimate	*12*	*0.01*	*9*	*0.04*
Illicit	*4336*	*3.2*	*2129*	*9.2*
Blog/forum spam	41335	30.1	8064	34.9
Uncategorized	15945	11.6	7766	33.7
Total	**137354**	**100**	**23042**	**100**

An equal opportunity attack…



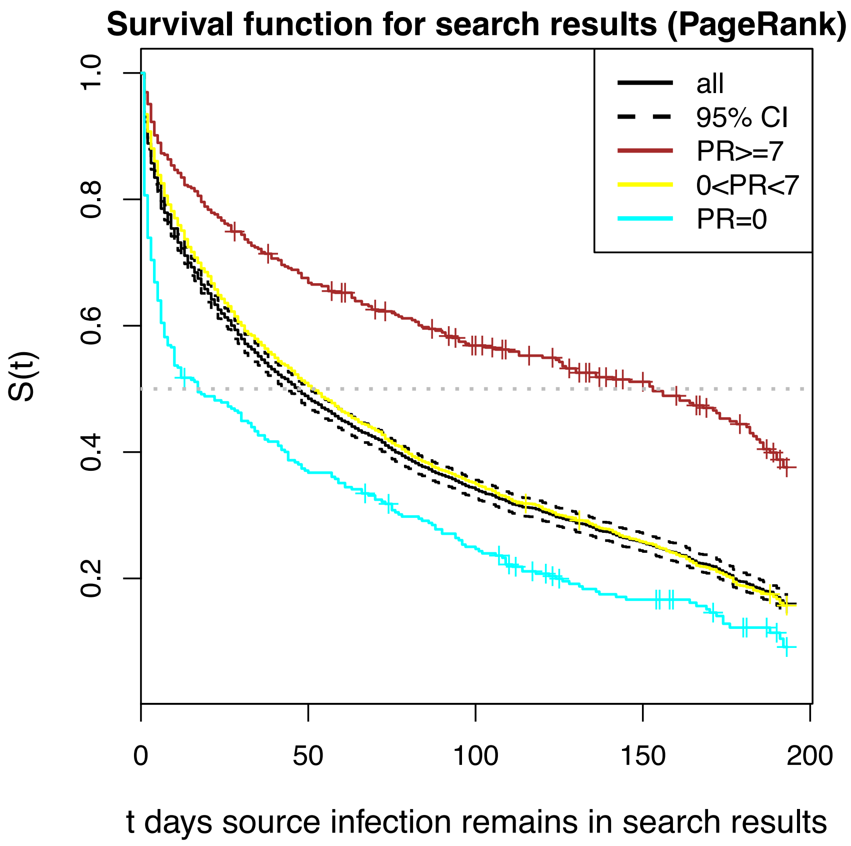
… with no signs of slowing down



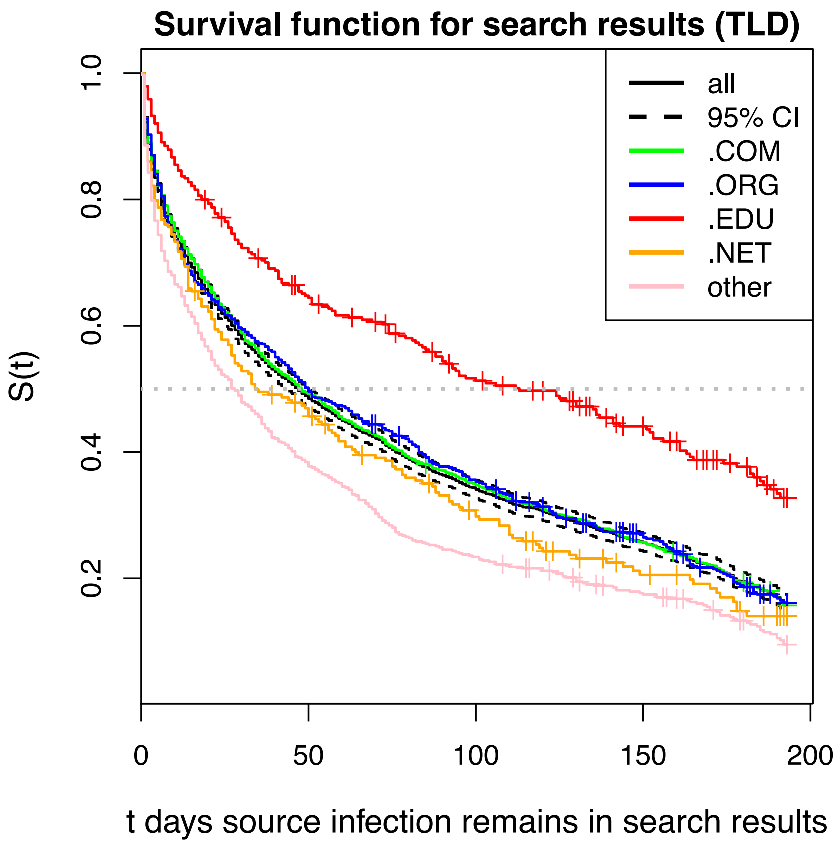
Search redirection attack is rising

Blog spam is declining

Infections last long time



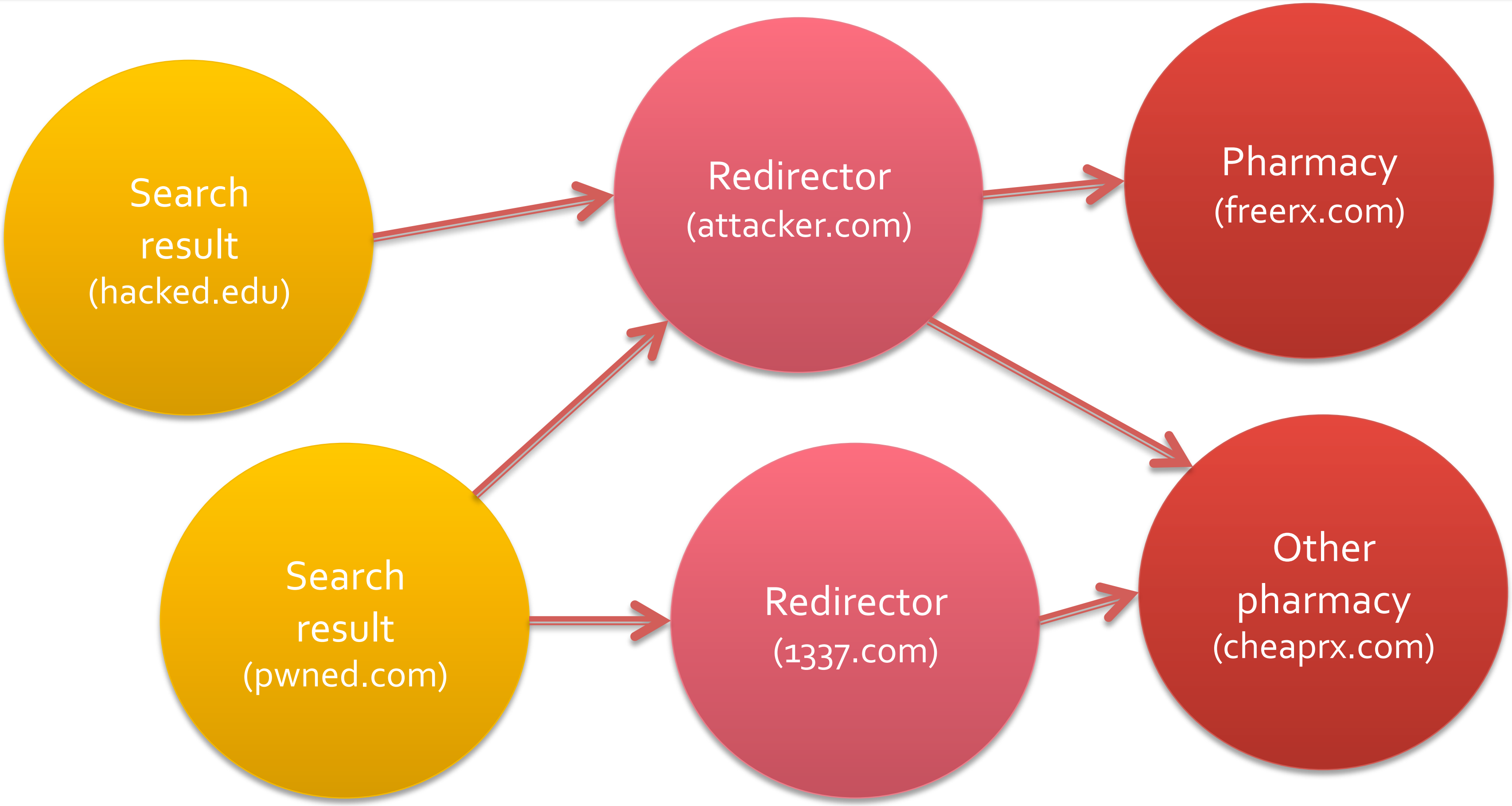
Survival function for search results (PageRank)



Survival function for search results (TLD)

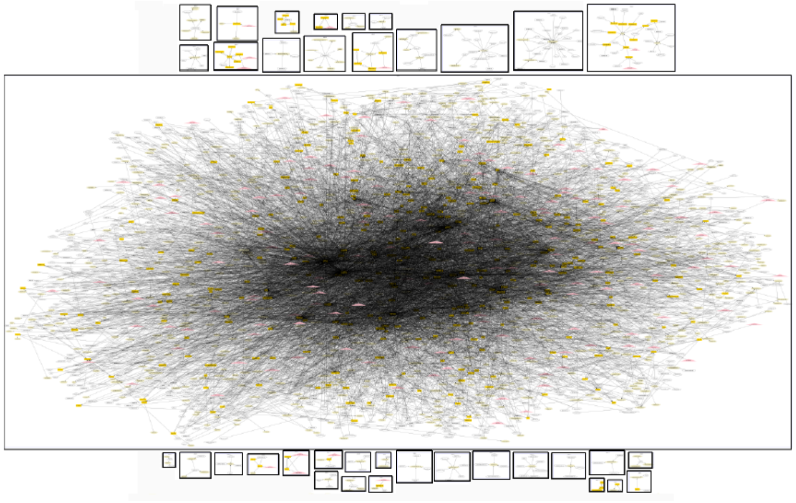
.edu sites particularly attractive, as well as high PageRank sites (often sites fall in both categories)

Uncovering relationships in search results


Search result (hacked.edu)
Redirector (attacker.com)
Pharmacy (freerx.com)
Search result (pwned.com)
Redirector (1337.com)
Other pharmacy (cheaprx.com)

Connected components in the graph evidence "some" level of business relationships between the nodes they connect

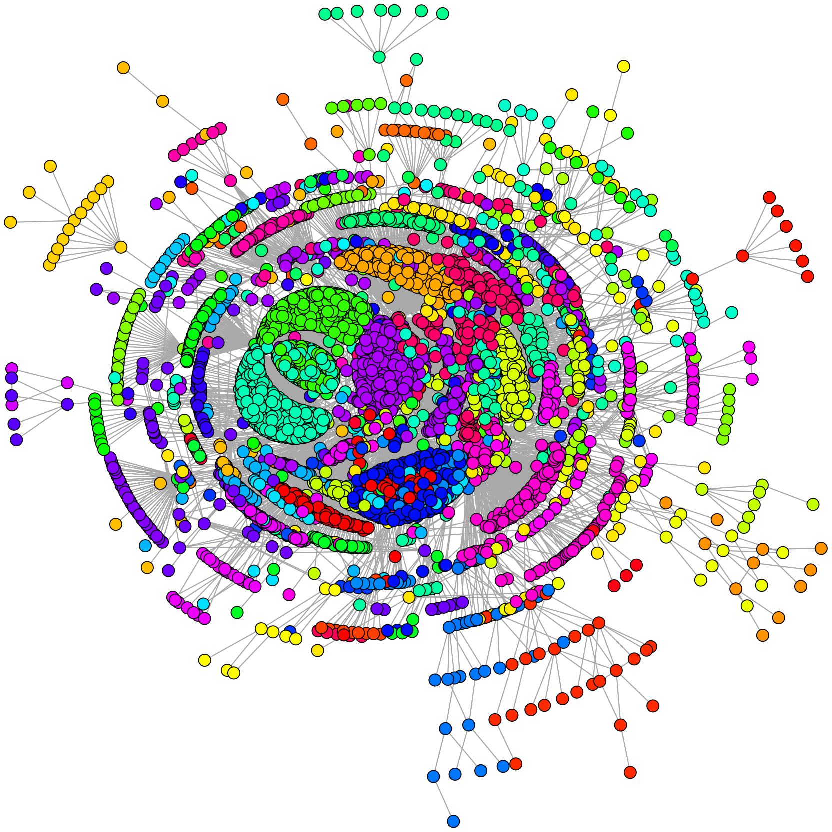
Connected components



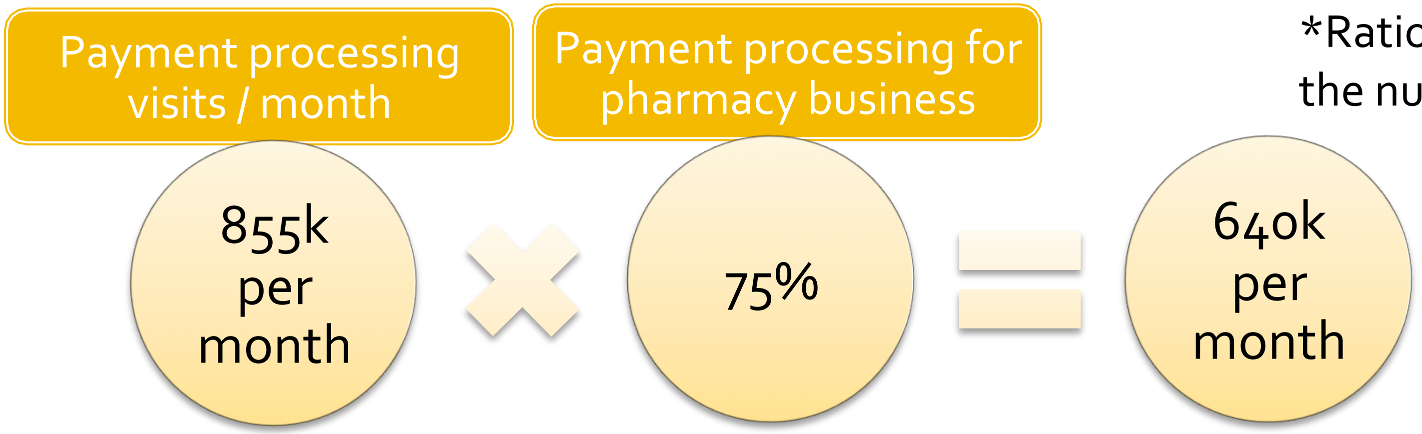
- 34 connected components
- One connected component contains
  - 96% of all infected domains
  - 90% of all redirection domains
  - 92% of all pharmacies
- Is one person responsible for all of this?!
  - Not necessarily, but evidence of partner relationships

Identifying the main players

- Run (spinglass) clustering algorithm in big connected component
- Evidence of separate organized groups/ campaigns more loosely connected to each other
- Interesting AS/registrar patterns.
  - 11 ASes host most redirect servers
  - Some are over-represented



Conversion Rate*



Payment processing visits / month

855k per month

×

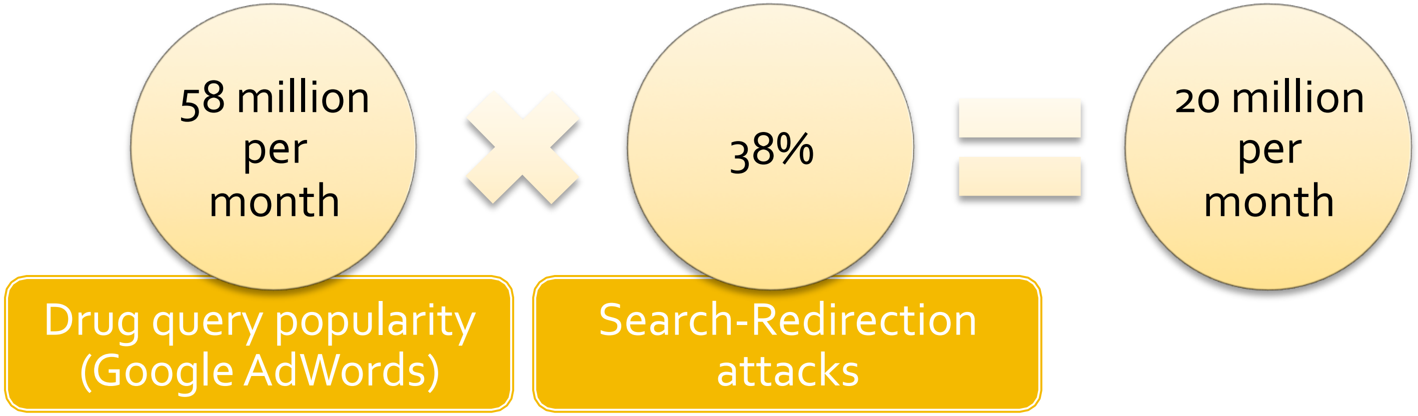
Payment processing for pharmacy business

75%

=

640k per month

$$Conversion\_rate \geq \frac{640{,}000}{20{,}000{,}000} K \xrightarrow{K=0.1} 0.32\%$$



58 million per month

Drug query popularity (Google AdWords)

×

38%

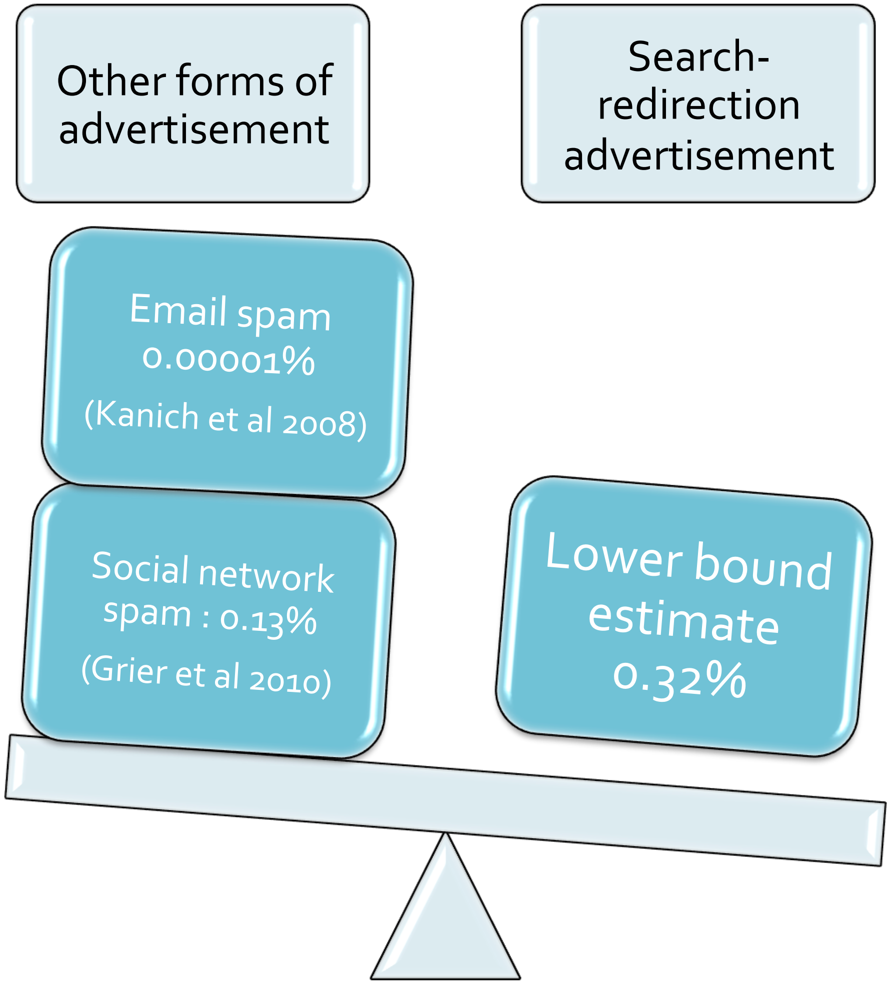
Search-Redirection attacks

=

20 million per month

*Ratio of realized sales over the number of visitors

Comparing conversion rates



Other forms of advertisement

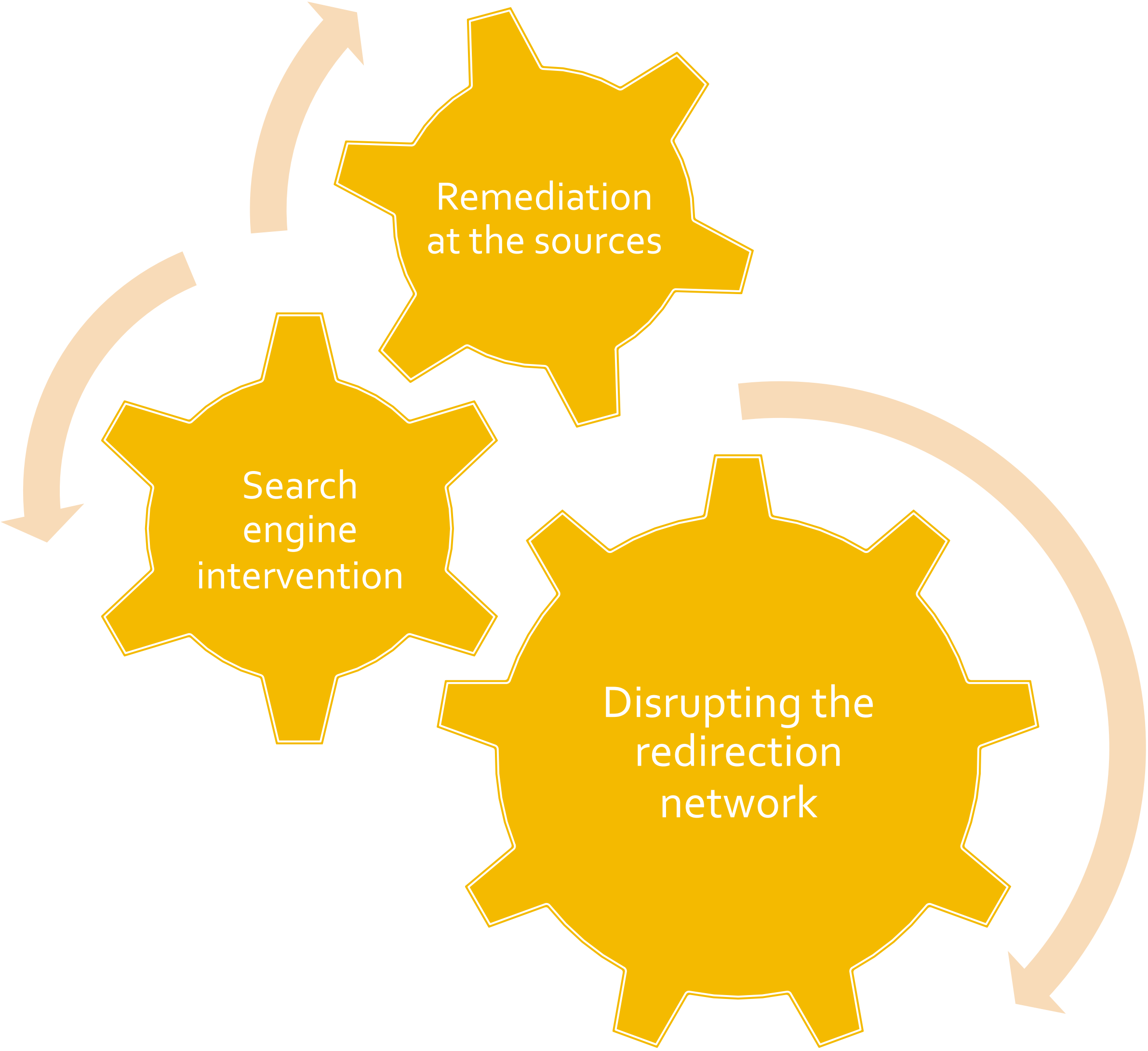
Search-redirection advertisement

Email spam
0.00001%
(Kanich et al 2008)

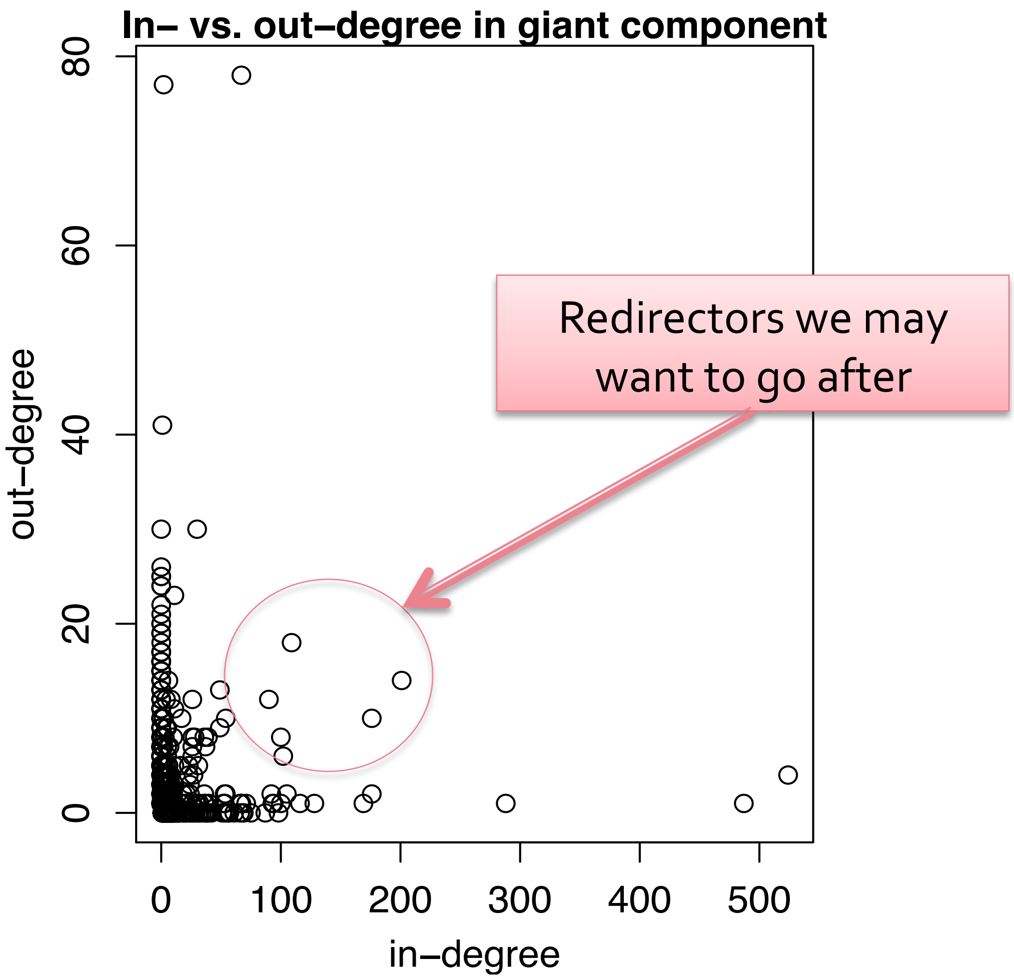
Social network spam : 0.13%
(Grier et al 2010)

Lower bound estimate
0.32%

Possible technical/policy remedies

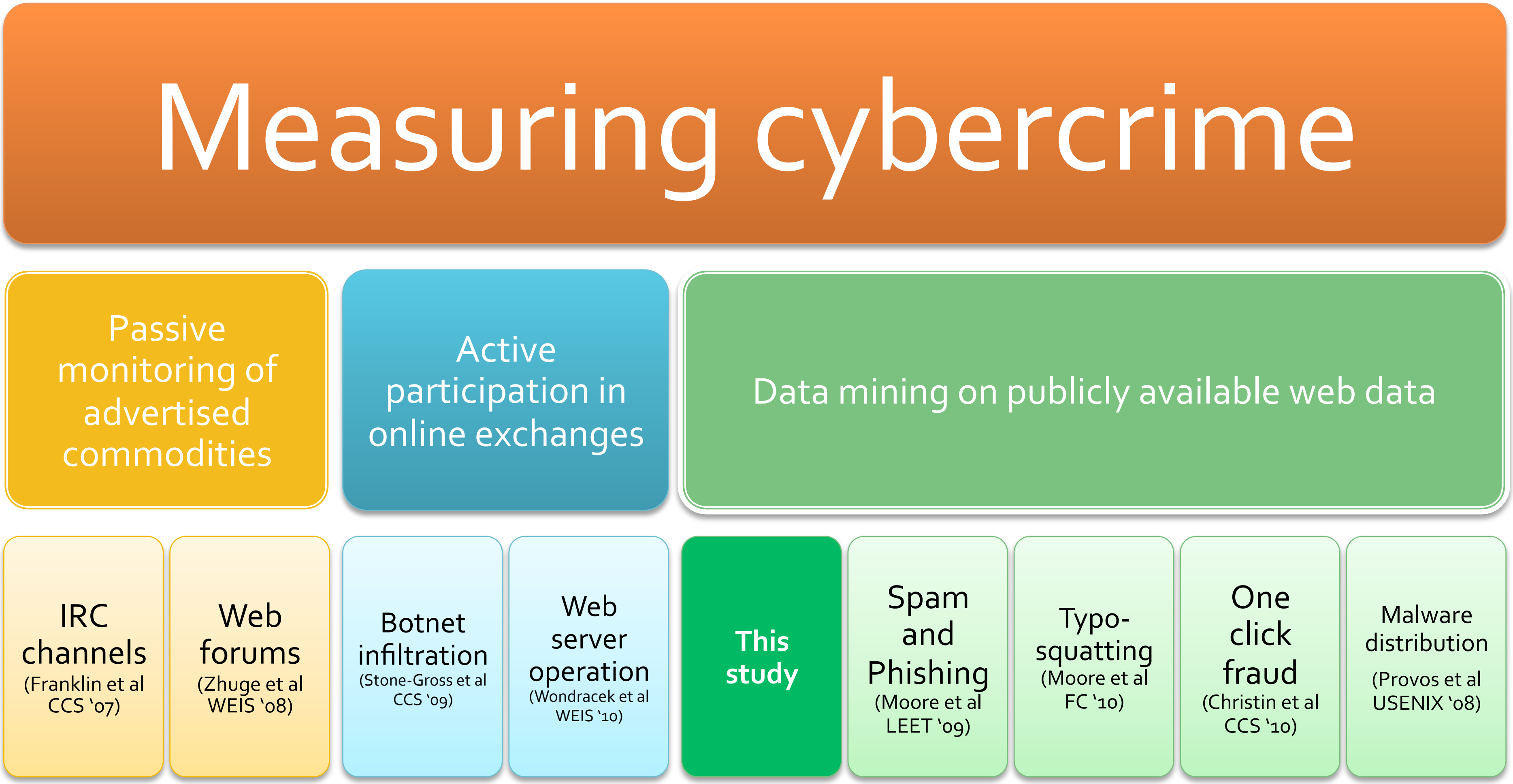

Remediation at the sources
Search engine intervention
Disrupting the redirection network

Breaking the redirection chains



Redirectors we may want to go after

Related work



Measuring cybercrime

Passive monitoring of advertised commodities

- IRC channels (Franklin et al CCS '07)
- Web forums (Zhuge et al WEIS '08)

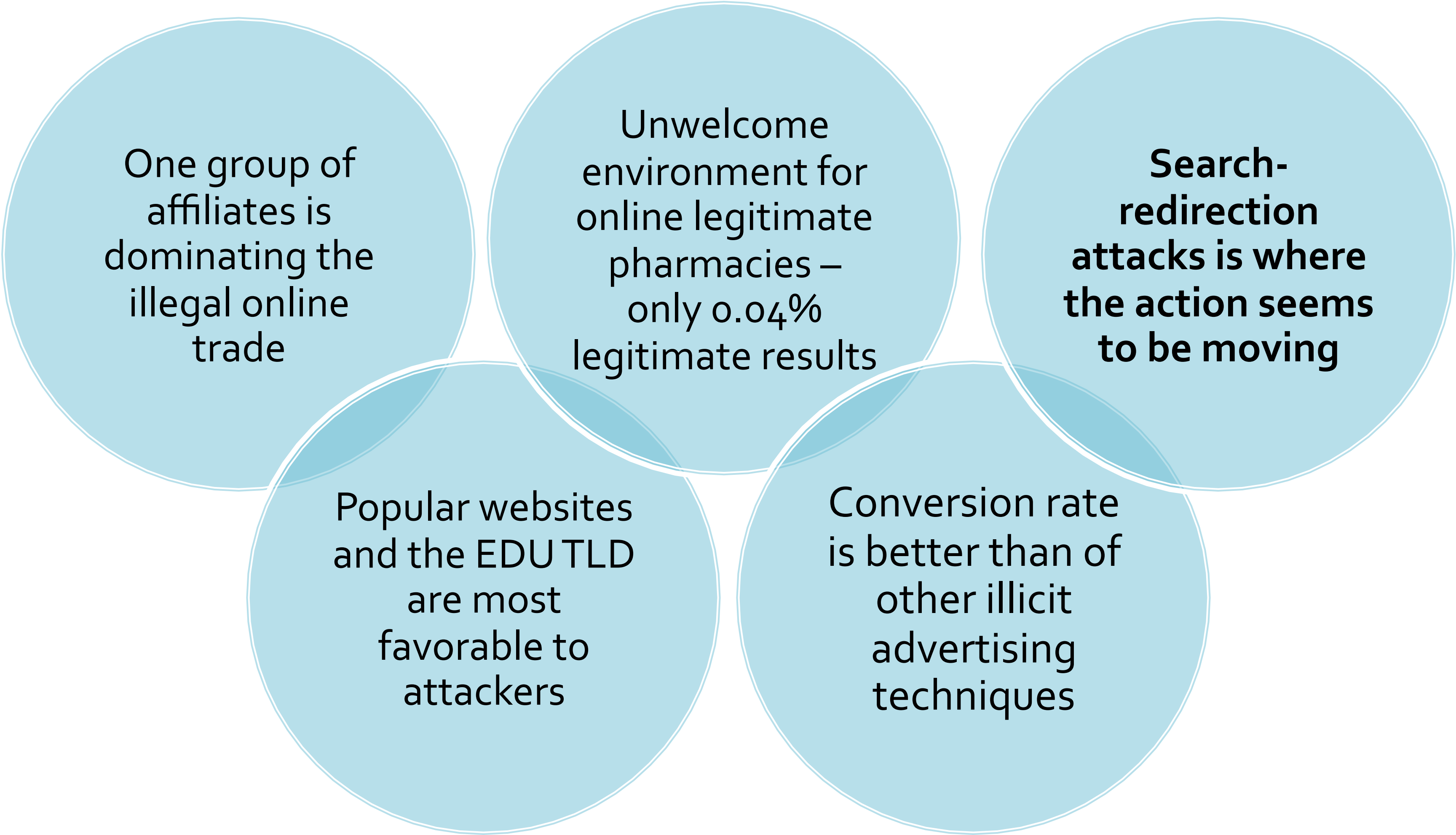
Active participation in online exchanges

- Botnet infiltration (Stone-Gross et al CCS '09)
- Web server operation (Wondracek et al WEIS '10)

Data mining on publicly available web data

- **This study**
- Spam and Phishing (Moore et al LEET '09)
- Typo-squatting (Moore et al FC '10)
- One click fraud (Christin et al CCS '10)
- Malware distribution (Provos et al USENIX '08)

Conclusions



One group of affiliates is dominating the illegal online trade

Unwelcome environment for online legitimate pharmacies – only 0.04% legitimate results

Search-redirection attacks is where the action seems to be moving

Popular websites and the EDU TLD are most favorable to attackers

Conversion rate is better than of other illicit advertising techniques

Questions?

Thank you!

Nektarios Leontiadis leontiadis@cmu.edu
Carnegie Mellon University/EPP/CyLab