Online Crime and Security Economics

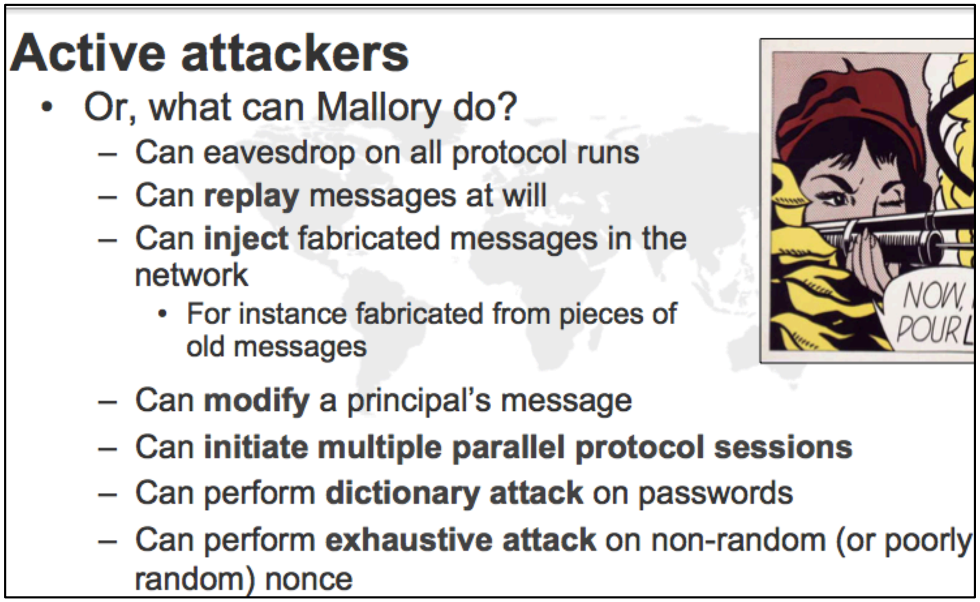
Nicolas Christin
Carnegie Mellon University
nicolasc@cmu.edu

International Conference on Data Economy

“Traditional” view of computer security

(from my intro. security class)

- Attackers are
  - Bound by computational and mathematical limitations…
  - … but by little else
  - High expertise assumed



Active attackers

- Or, what can Mallory do?
  - Can eavesdrop on all protocol runs
  - Can **replay** messages at will
  - Can **inject** fabricated messages in the network
    - For instance fabricated from pieces of old messages
  - Can **modify** a principal’s message
  - Can **initiate multiple parallel protocol sessions**
  - Can perform **dictionary attack** on passwords
  - Can perform **exhaustive attack** on non-random (or poorly random) nonce

- Sound security engineering shouldn’t rest on assumptions about possible attacker’s weaknesses
- Likewise, defenders are assumed to be security-conscious

Security in practice

vs

?

Security in practice

vs

(Most) attackers in practice

- Most security attacks carried out by entities that are
  - Financially interested
  - Economically rational
  - Not necessarily overly sophisticated

  - Heavily reliant on commoditization
    - Purchase "services" from others

(Exception: the (still fairly rare) nation-state actors that are not outsourcing to criminals)

- Defenders (end-users) are also subject to biases, lack of interest, …
  → story for a different talk

Research agenda

- **Understanding incentives of attackers and targets are critical to improving online security**
  - Useful to find where to target attackers
  - Useful to find how to deploy defenses

- How to discover and model these incentives?

- **Security analytics:** Assortment of different techniques
  - Game theory
  - Machine learning
  - **Network measurements**
  - Behavioral economics

Question

- How can we model attacker behavior?
- Attackers usually not keen on being interviewed
- Modeling based on utility assumptions needs to be grounded in empirical evidence

- ...however...
- Online attackers leave **lots** of data for us to analyze

Relevant papers (case studies)

- Online sale of prescription drugs
  1. Leontiadis, Moore and Christin. Measuring and analyzing search-redirection attacks in the illicit online prescription trade. *USENIX Security 2011*
  2. Leontiadis, Moore and Christin. A nearly four-year longitudinal study of search-engine poisoning. *ACM CCS 2014*
  3. Leontiadis, Moore and Christin. Pick Your Poison: Pricing and Inventories at Unlicensed Online Pharmacies. *ACM EC 2013*

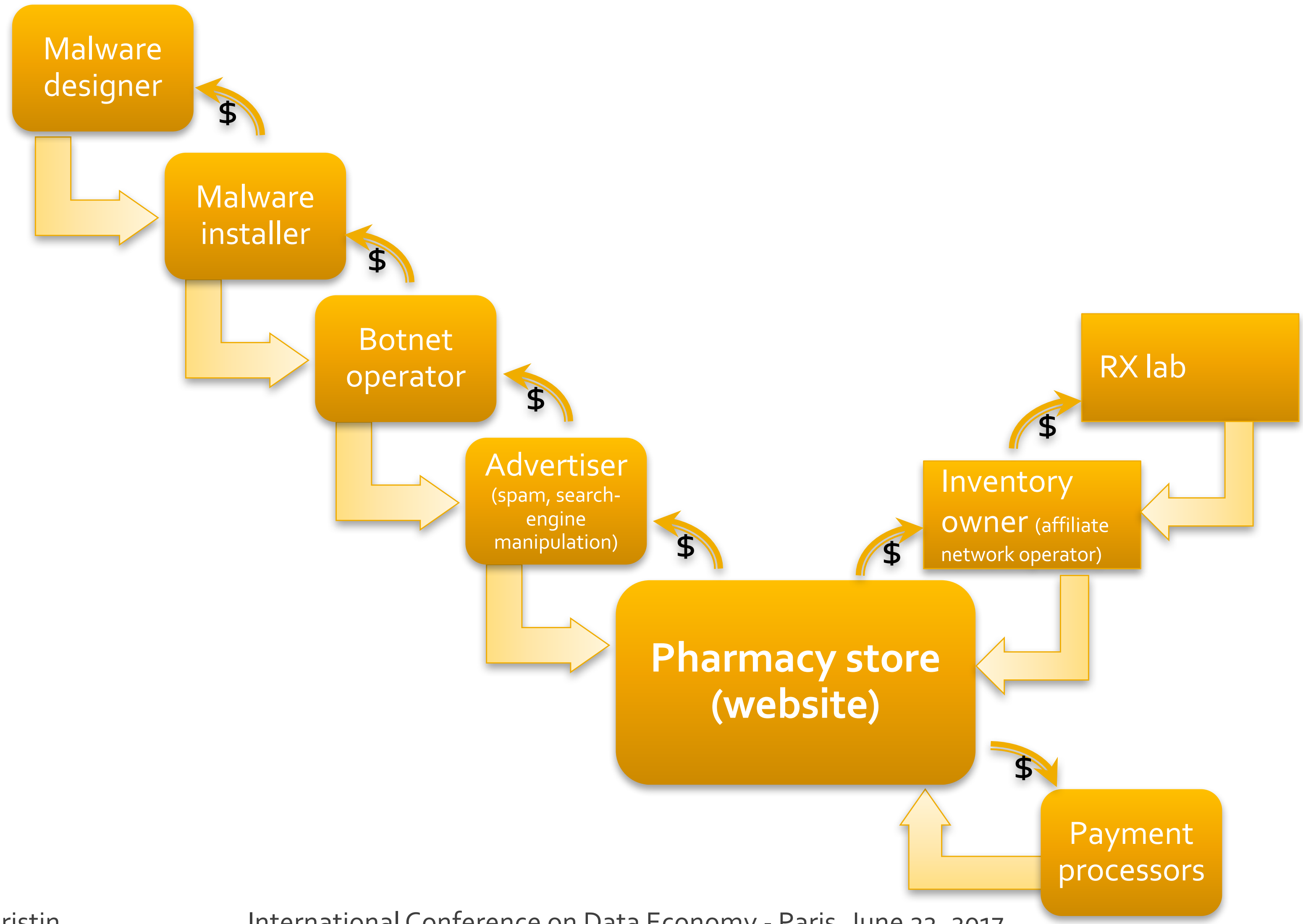
- Online anonymous markets
  1. Christin. Traveling the Silk Road: A measurement study of a large anonymous online marketplace. *WWW'13*
  2. Soska and Christin. Measuring the longitudinal evolution of the online anonymous marketplace ecosystem. *USENIX Security 2015*

Case study: Online sale of drugs

- One of the best known illicit online trades
  - Who hasn’t received email spam for prescription drugs?

- Potentially most dangerous form of online crime
  - Wrong dosage can kill: cf. Ryan Haight
- Complex supply chain that can tell us a lot about the online criminal ecosystem

Supply chain: high-level overview



Malware designer

$

Malware installer

$

Botnet operator

$

Advertiser (spam, search-engine manipulation)

$

Pharmacy store (website)

$

Inventory owner (affiliate network operator)

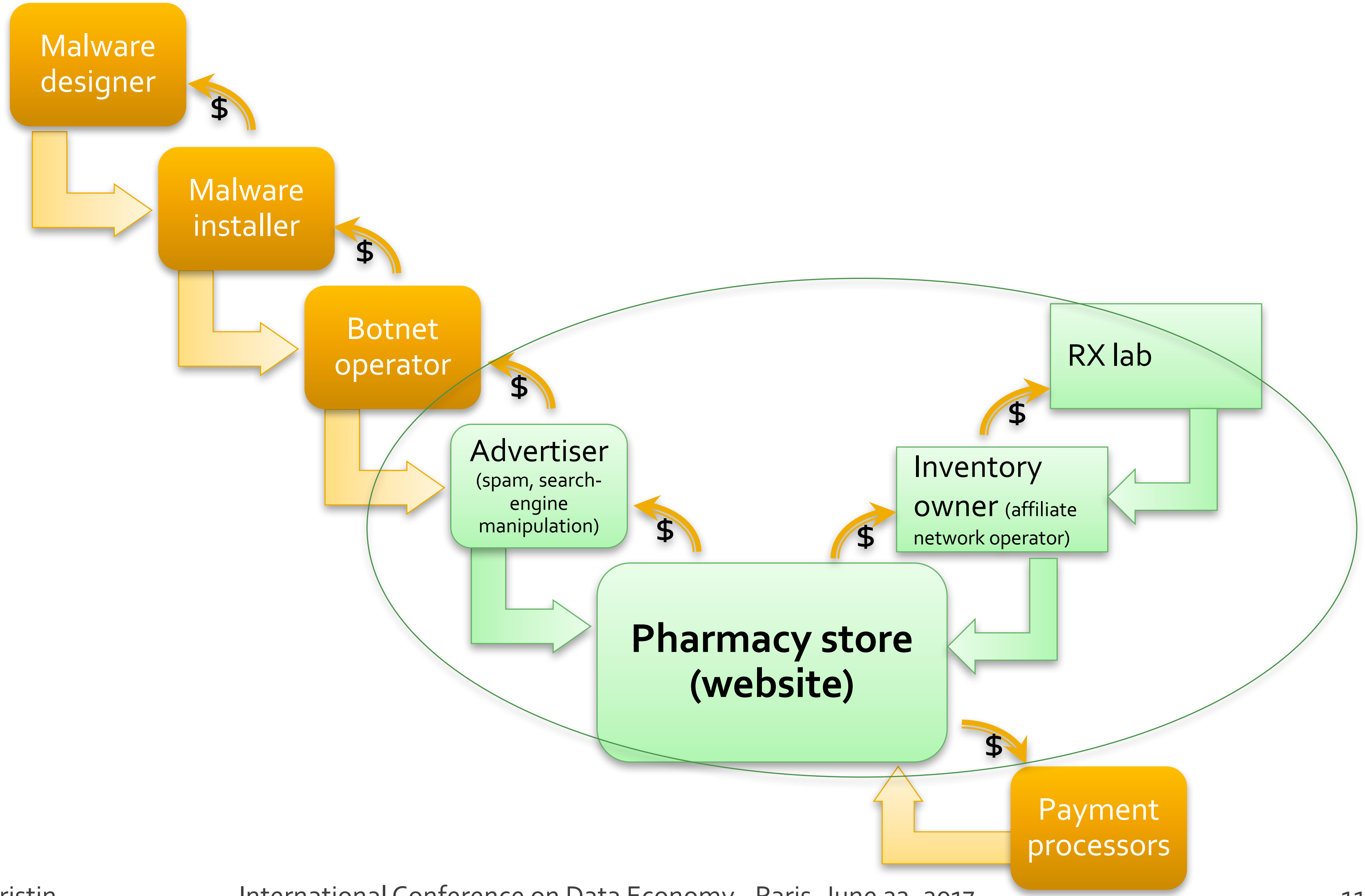
$

RX lab

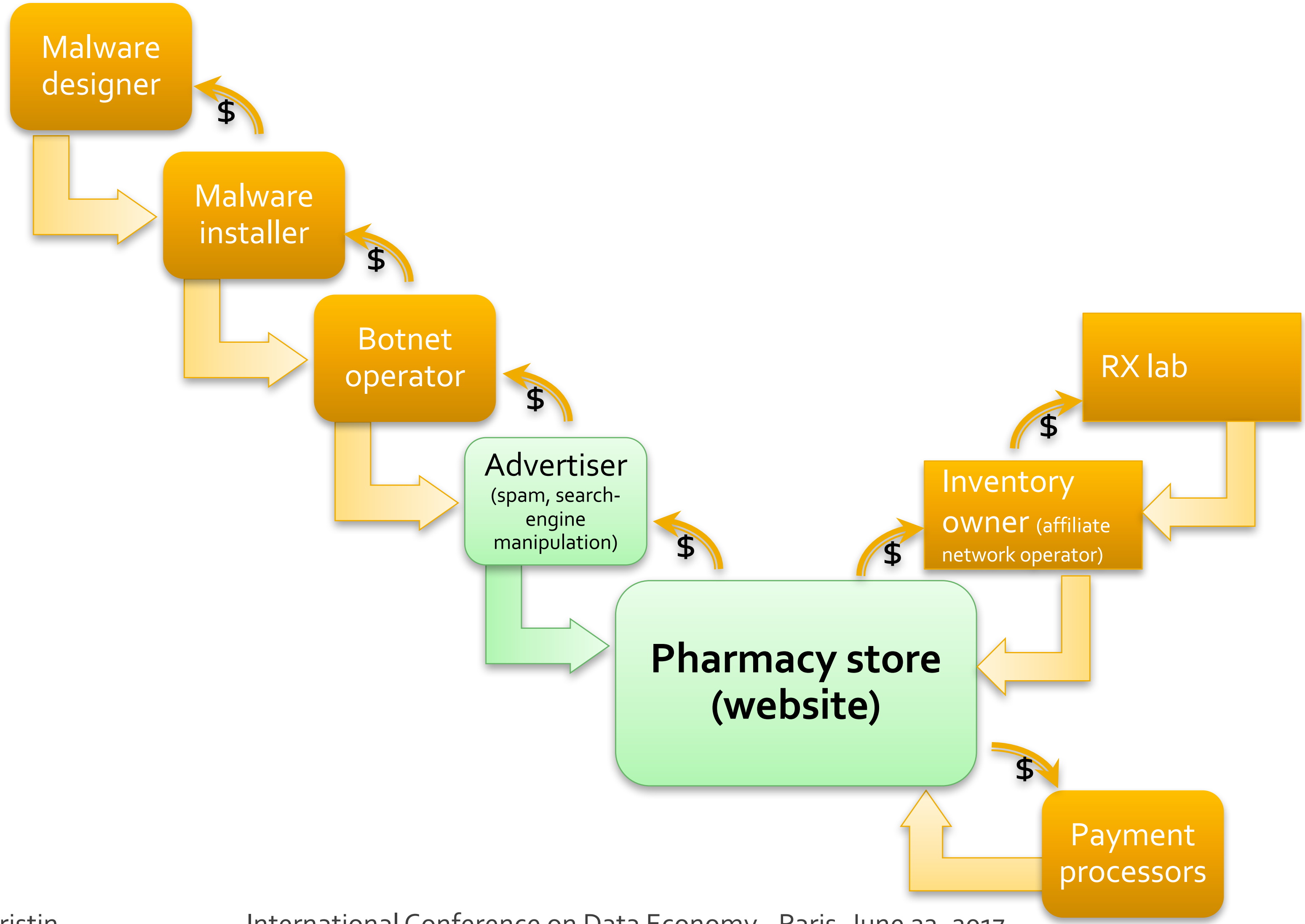
$

Payment processors

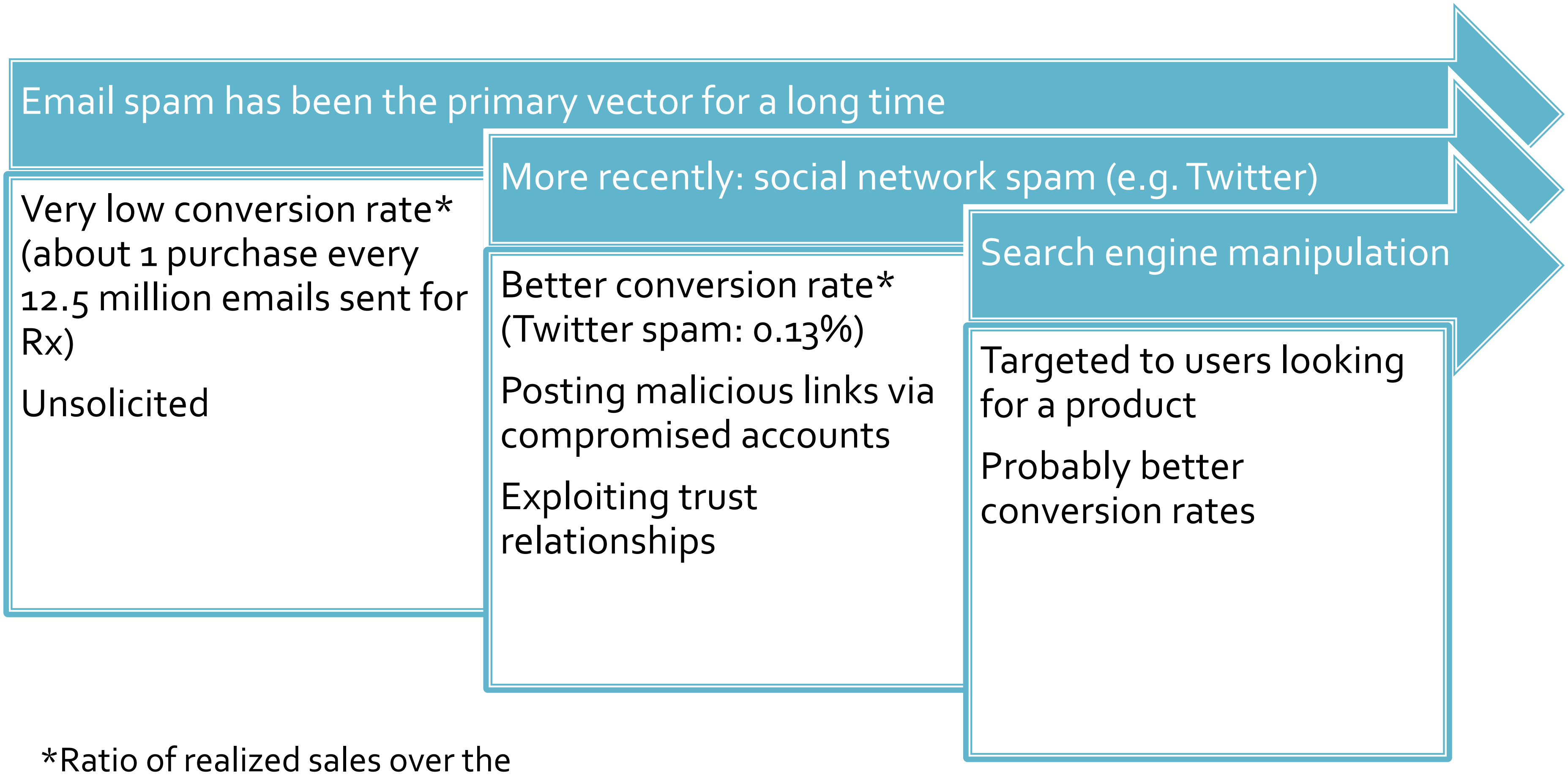
Supply chain: high-level overview



Advertising unlicensed drugs



Evolution of advertising of illicit products



Email spam has been the primary vector for a long time

Very low conversion rate* (about 1 purchase every 12.5 million emails sent for Rx)

Unsolicited

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

Better conversion rate* (Twitter spam: 0.13%)

Posting malicious links via compromised accounts

Exploiting trust relationships

Search engine manipulation

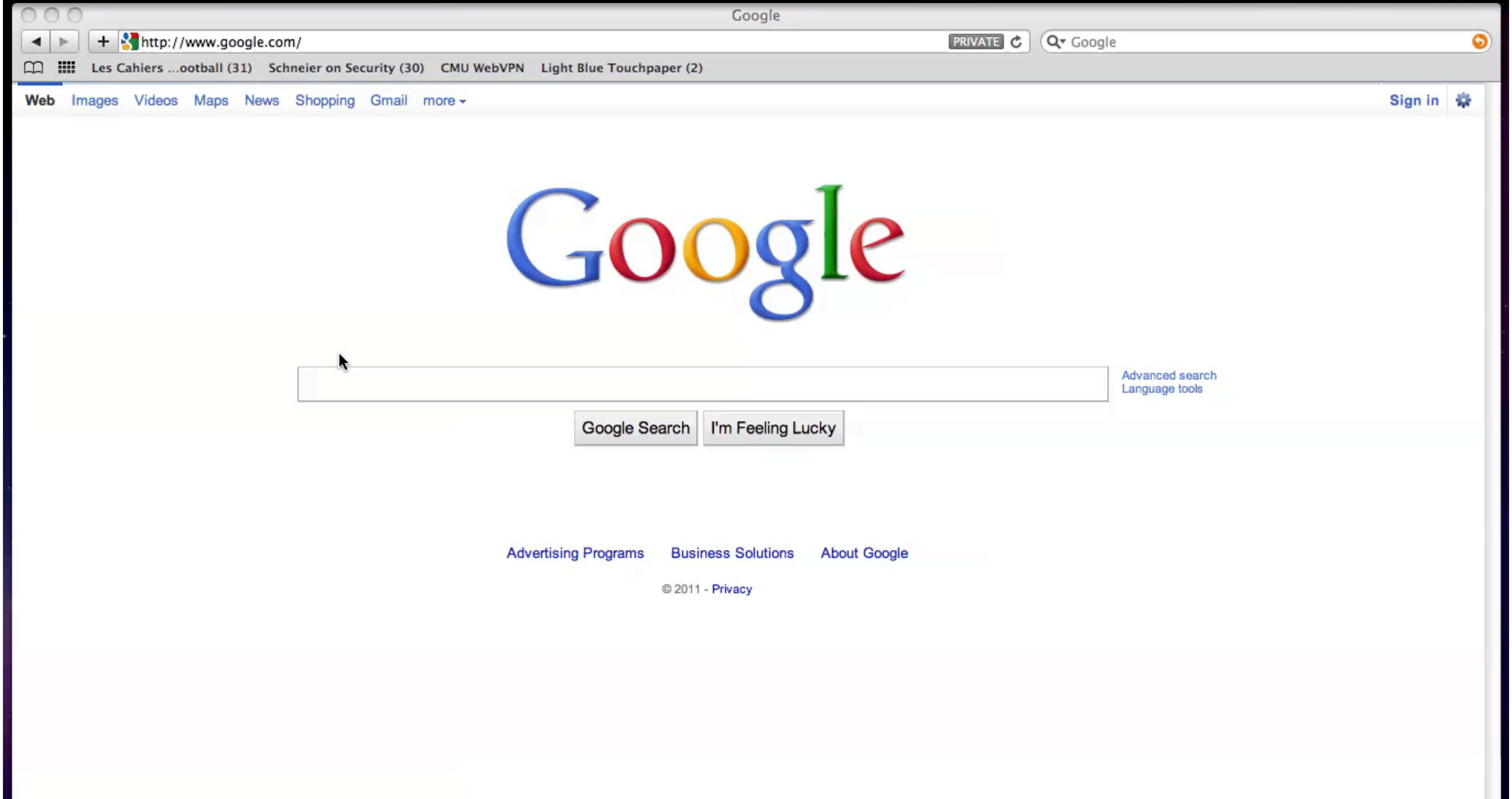
Targeted to users looking for a product

Probably better conversion rates

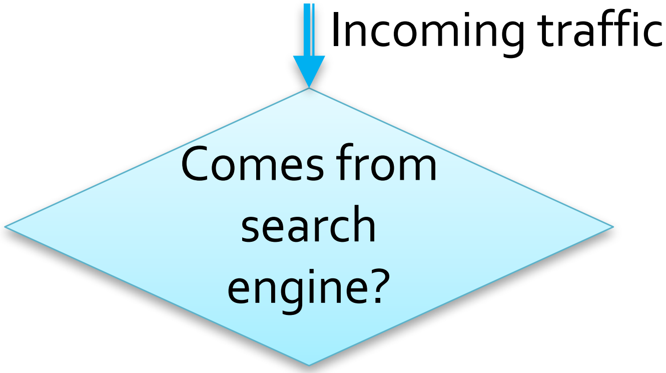
*Ratio of realized sales over the number of emails/clicks

Search-redirection

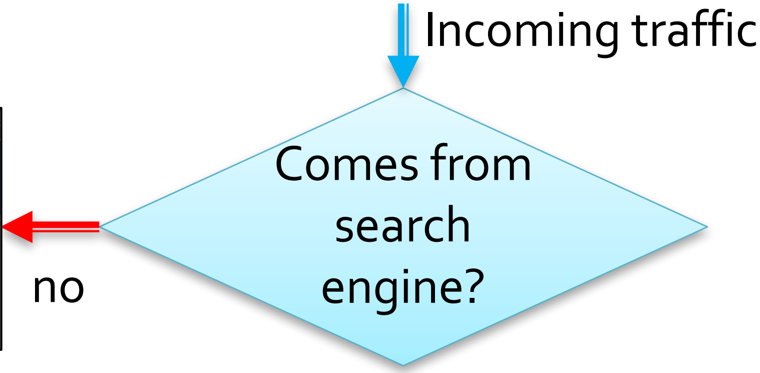
[LMC, USENIX Security 2011]

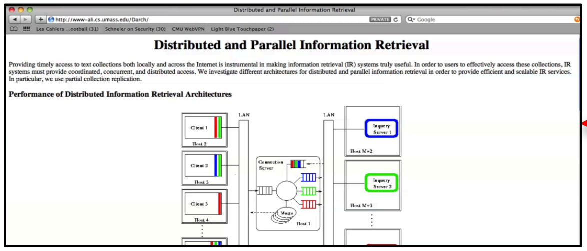


Attack modus operandi: Malware logic



Attack modus operandi: Malware logic



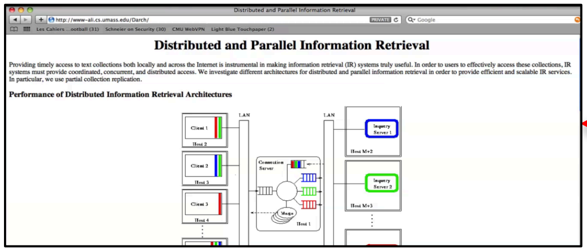

Distributed and Parallel Information Retrieval
http://www-ali.cs.umass.edu/Darch/
PRIVATE
Google
Les Cahiers ...ootball (31)
Schneier on Security (30)
CMU WebVPN
Light Blue Touchpaper (2)

Distributed and Parallel Information Retrieval

Providing timely access to text collections both locally and across the Internet is instrumental in making information retrieval (IR) systems truly useful. In order to users to effectively access these collections, IR systems must provide coordinated, concurrent, and distributed access. We investigate different architectures for distributed and parallel information retrieval in order to provide efficient and scalable IR services. In particular, we use partial collection replication.

Performance of Distributed Information Retrieval Architectures

Attack modus operandi: Malware logic

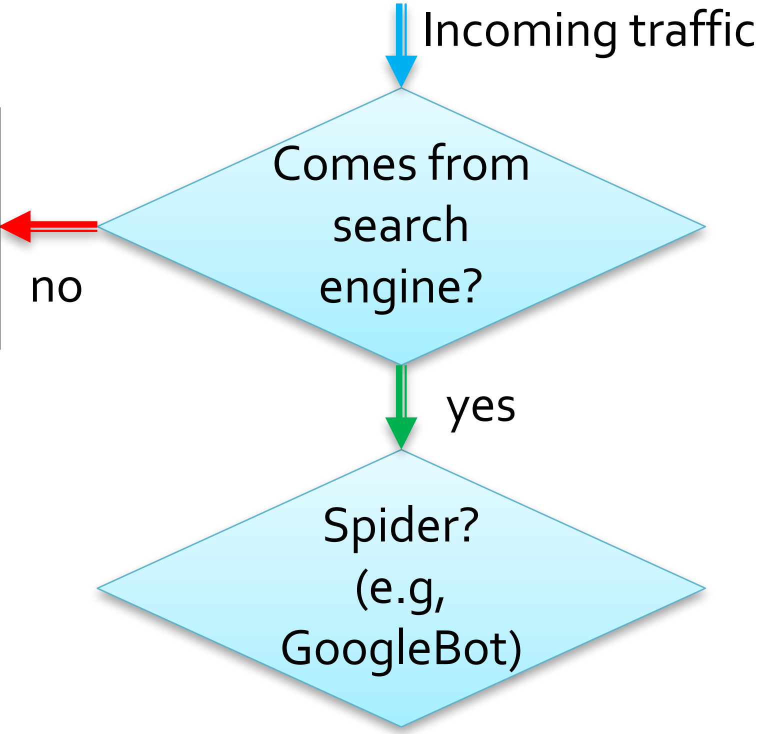


Distributed and Parallel Information Retrieval
http://www-ali.cs.umass.edu/Darch/
PRIVATE
Google
Les Cahiers ...ootball (31)
Schneier on Security (30)
CMU WebVPN
Light Blue Touchpaper (2)

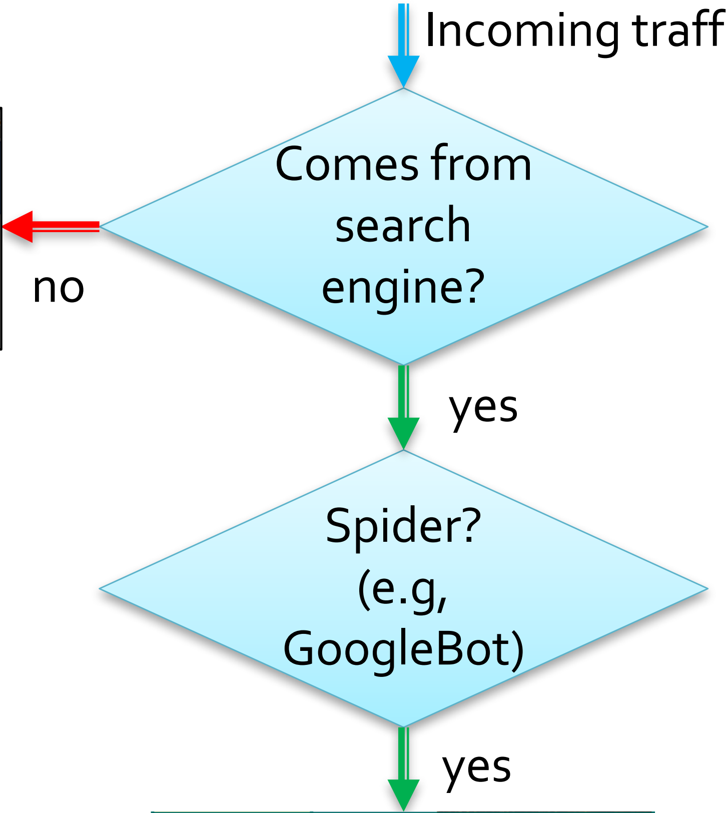
Distributed and Parallel Information Retrieval

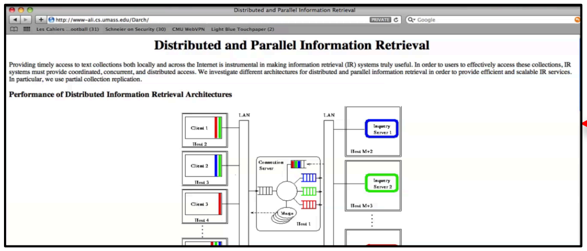
Providing timely access to text collections both locally and across the Internet is instrumental in making information retrieval (IR) systems truly useful. In order to users to effectively access these collections, IR systems must provide coordinated, concurrent, and distributed access. We investigate different architectures for distributed and parallel information retrieval in order to provide efficient and scalable IR services. In particular, we use partial collection replication.

Performance of Distributed Information Retrieval Architectures



Attack modus operandi: Malware logic

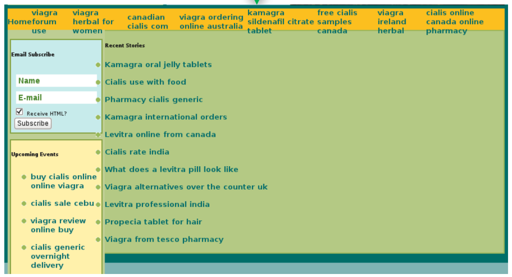



Distributed and Parallel Information Retrieval
http://www-ali.cs.umass.edu/Darch/
PRIVATE
Google
Les Cahiers ...ootball (31)
Schneier on Security (30)
CMU WebVPN
Light Blue Touchpaper (2)

Distributed and Parallel Information Retrieval

Providing timely access to text collections both locally and across the Internet is instrumental in making information retrieval (IR) systems truly useful. In order to users to effectively access these collections, IR systems must provide coordinated, concurrent, and distributed access. We investigate different architectures for distributed and parallel information retrieval in order to provide efficient and scalable IR services. In particular, we use partial collection replication.

Performance of Distributed Information Retrieval Architectures


Homeforum
viagra use
viagra herbal for women
canadian cialis com
viagra ordering online australia
kamagra sildenafil citrate tablet
free cialis samples canada
viagra ireland herbal
cialis online canada online pharmacy

Email Subscribe

Name

E-mail

☑ Receive HTML?

Subscribe

Upcoming Events

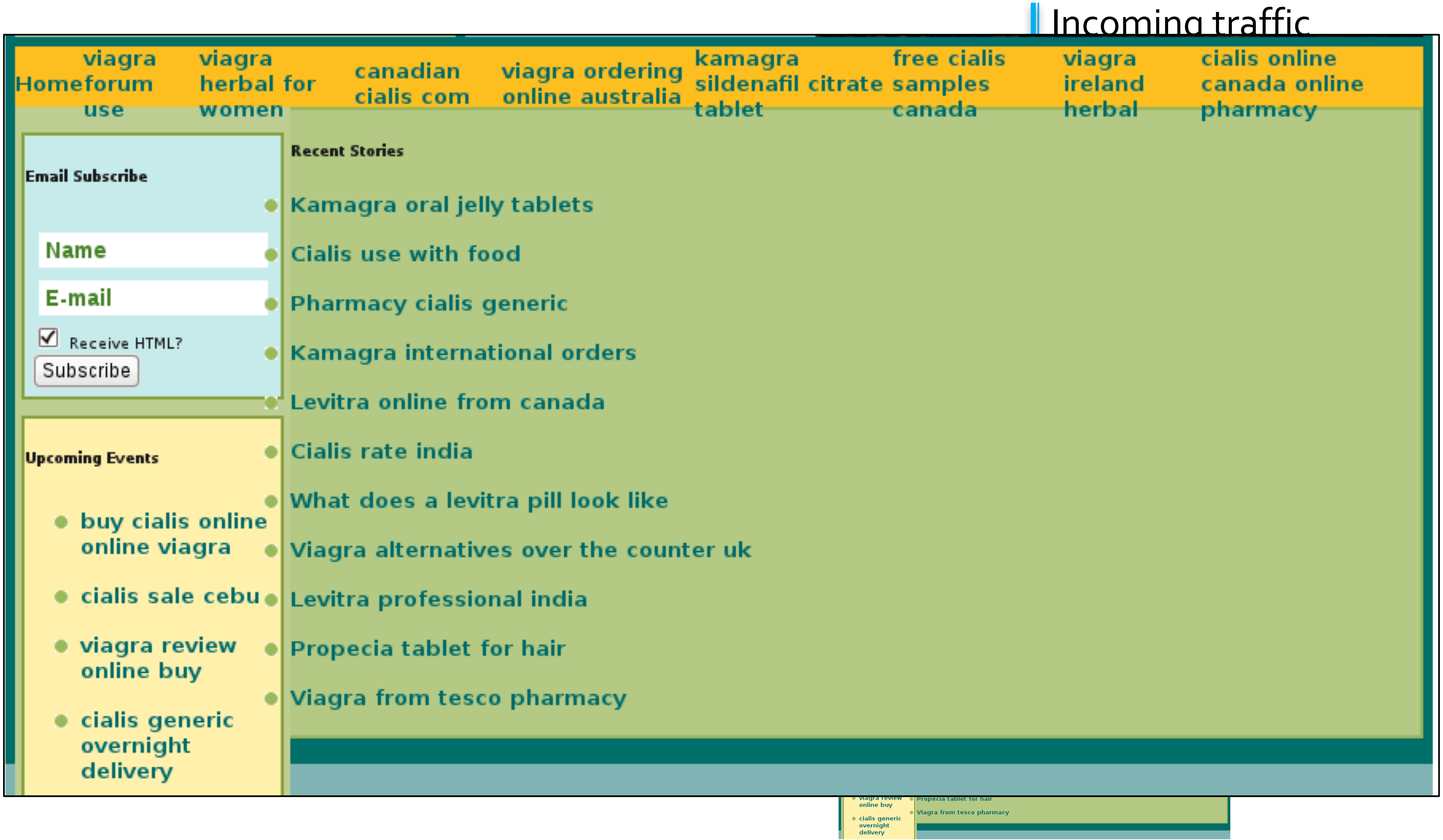
- buy cialis online online viagra
- cialis sale cebu
- viagra review online buy
- cialis generic overnight delivery

Recent Stories

- Kamagra oral jelly tablets
- Cialis use with food
- Pharmacy cialis generic
- Kamagra international orders
- Levitra online from canada
- Cialis rate india
- What does a levitra pill look like
- Viagra alternatives over the counter uk
- Levitra professional india
- Propecia tablet for hair
- Viagra from tesco pharmacy

Attack modus operandi: Malware logic

Incoming traffic



viagra Homeforum use | viagra herbal for women | canadian cialis com | viagra ordering online australia | kamagra sildenafil citrate tablet | free cialis samples canada | viagra ireland herbal | cialis online canada online pharmacy

Email Subscribe

Name

E-mail

☑ Receive HTML?

Subscribe

Upcoming Events

- buy cialis online online viagra
- cialis sale cebu
- viagra review online buy
- cialis generic overnight delivery

Recent Stories

- Kamagra oral jelly tablets
- Cialis use with food
- Pharmacy cialis generic
- Kamagra international orders
- Levitra online from canada
- Cialis rate india
- What does a levitra pill look like
- Viagra alternatives over the counter uk
- Levitra professional india
- Propecia tablet for hair
- Viagra from tesco pharmacy

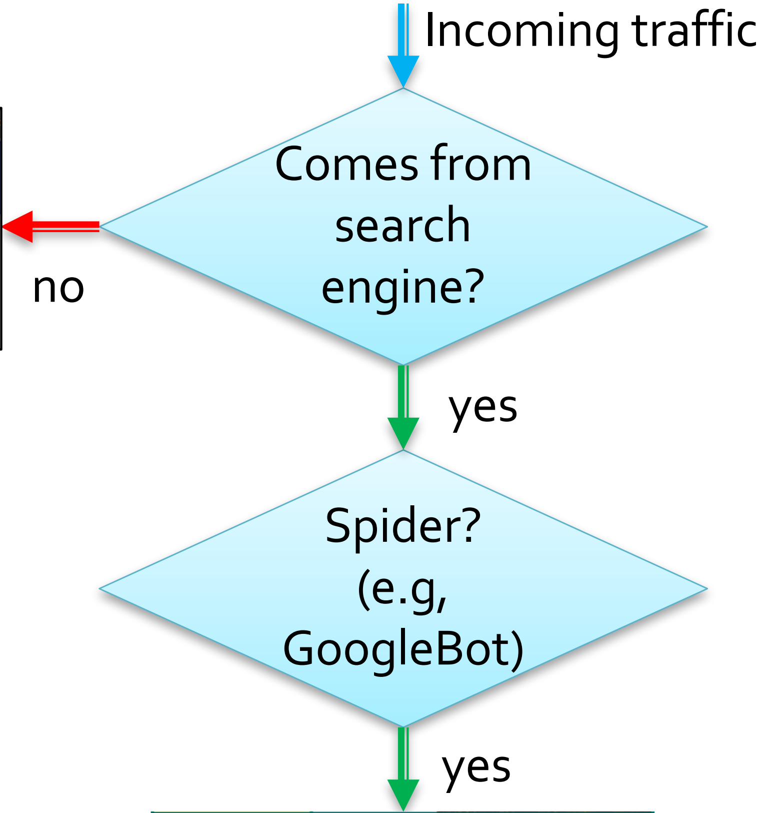
viagra review online buy

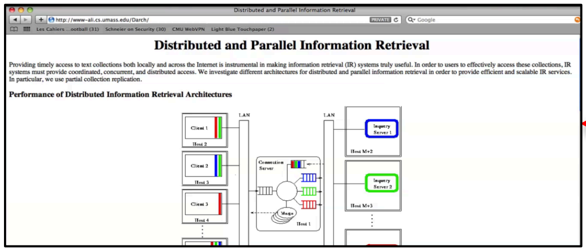
Propecia tablet for hair

cialis generic overnight delivery

Viagra from tesco pharmacy

Attack modus operandi: Malware logic

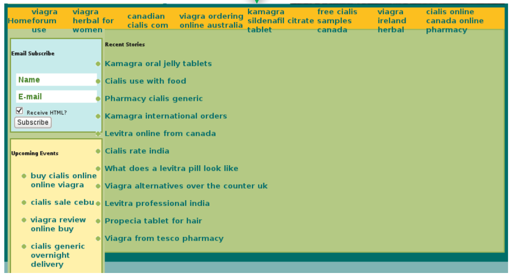



Distributed and Parallel Information Retrieval
http://www-ali.cs.umass.edu/Darch/
PRIVATE
Google
Les Cahiers ...ootball (31)
Schneier on Security (30)
CMU WebVPN
Light Blue Touchpaper (2)

Distributed and Parallel Information Retrieval

Providing timely access to text collections both locally and across the Internet is instrumental in making information retrieval (IR) systems truly useful. In order to users to effectively access these collections, IR systems must provide coordinated, concurrent, and distributed access. We investigate different architectures for distributed and parallel information retrieval in order to provide efficient and scalable IR services. In particular, we use partial collection replication.

Performance of Distributed Information Retrieval Architectures



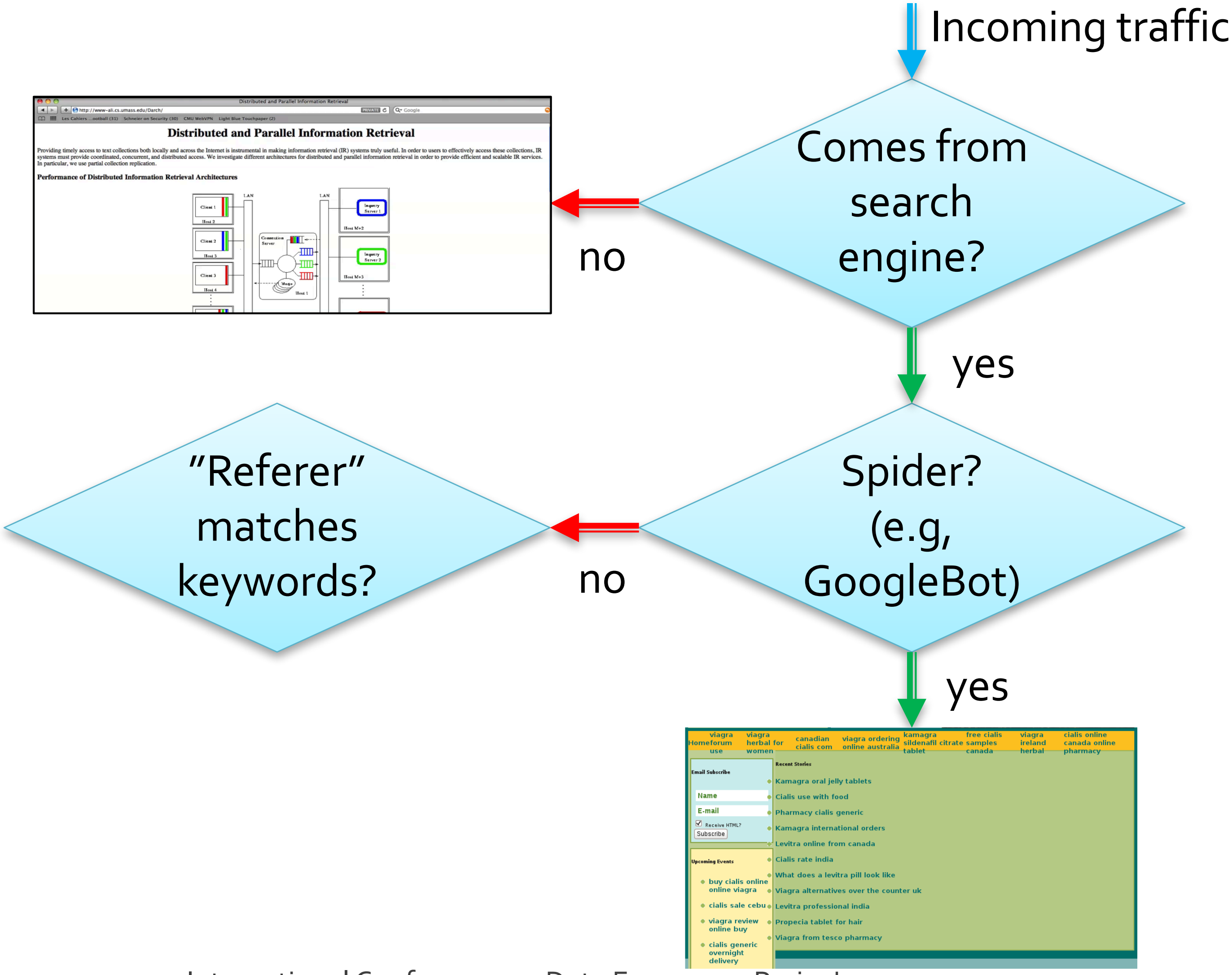
Upcoming Events

- buy cialis online online viagra
- cialis sale cebu
- viagra review online buy
- cialis generic overnight delivery

Recent Stories

- Kamagra oral jelly tablets
- Cialis use with food
- Pharmacy cialis generic
- Kamagra international orders
- Levitra online from canada
- Cialis rate india
- What does a levitra pill look like
- Viagra alternatives over the counter uk
- Levitra professional india
- Propecia tablet for hair
- Viagra from tesco pharmacy

Attack modus operandi: Malware logic



Incoming traffic

Comes from search engine?

no

yes

Spider? (e.g, GoogleBot)

no

yes

"Referer" matches keywords?

Distributed and Parallel Information Retrieval

http://www-ali.cs.umass.edu/Darch/

Les Cahiers ...ootball (31) Schneier on Security (30) CMU WebVPN Light Blue Touchpaper (2)

Distributed and Parallel Information Retrieval

Providing timely access to text collections both locally and across the Internet is instrumental in making information retrieval (IR) systems truly useful. In order to users to effectively access these collections, IR systems must provide coordinated, concurrent, and distributed access. We investigate different architectures for distributed and parallel information retrieval in order to provide efficient and scalable IR services. In particular, we use partial collection replication.

Performance of Distributed Information Retrieval Architectures

viagra Homeforum use | viagra herbal for women | canadian cialis com | viagra ordering online australia | kamagra sildenafil citrate tablet | free cialis samples canada | viagra ireland herbal | cialis online canada online pharmacy

Email Subscribe

Name

E-mail

☑ Receive HTML?

Subscribe

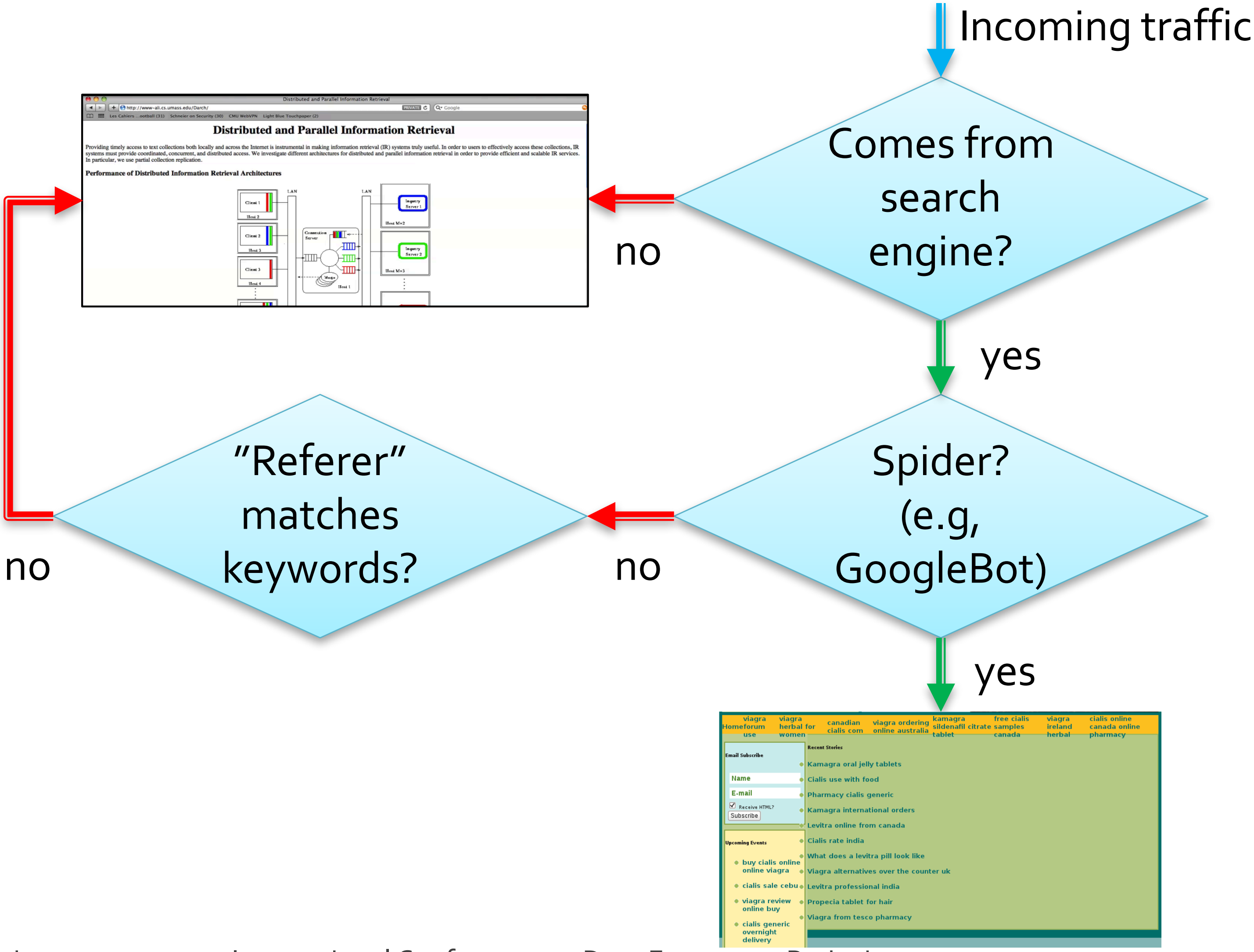
Upcoming Events

- buy cialis online online viagra
- cialis sale cebu
- viagra review online buy
- cialis generic overnight delivery

Recent Stories

- Kamagra oral jelly tablets
- Cialis use with food
- Pharmacy cialis generic
- Kamagra international orders
- Levitra online from canada
- Cialis rate india
- What does a levitra pill look like
- Viagra alternatives over the counter uk
- Levitra professional india
- Propecia tablet for hair
- Viagra from tesco pharmacy

Attack modus operandi: Malware logic



Incoming traffic

Comes from search engine?

no

yes

Spider? (e.g, GoogleBot)

no

yes

"Referer" matches keywords?

no

Distributed and Parallel Information Retrieval

http://www-ali.cs.umass.edu/Darch/

Les Cahiers ...ootball (31) Schneier on Security (30) CMU WebVPN Light Blue Touchpaper (2)

Distributed and Parallel Information Retrieval

Providing timely access to text collections both locally and across the Internet is instrumental in making information retrieval (IR) systems truly useful. In order to users to effectively access these collections, IR systems must provide coordinated, concurrent, and distributed access. We investigate different architectures for distributed and parallel information retrieval in order to provide efficient and scalable IR services. In particular, we use partial collection replication.

Performance of Distributed Information Retrieval Architectures

viagra Homeforum use | viagra herbal for women | canadian cialis com | viagra ordering online australia | kamagra sildenafil citrate tablet | free cialis samples canada | viagra ireland herbal | cialis online canada online pharmacy

Email Subscribe

Name

E-mail

☑ Receive HTML?

Subscribe

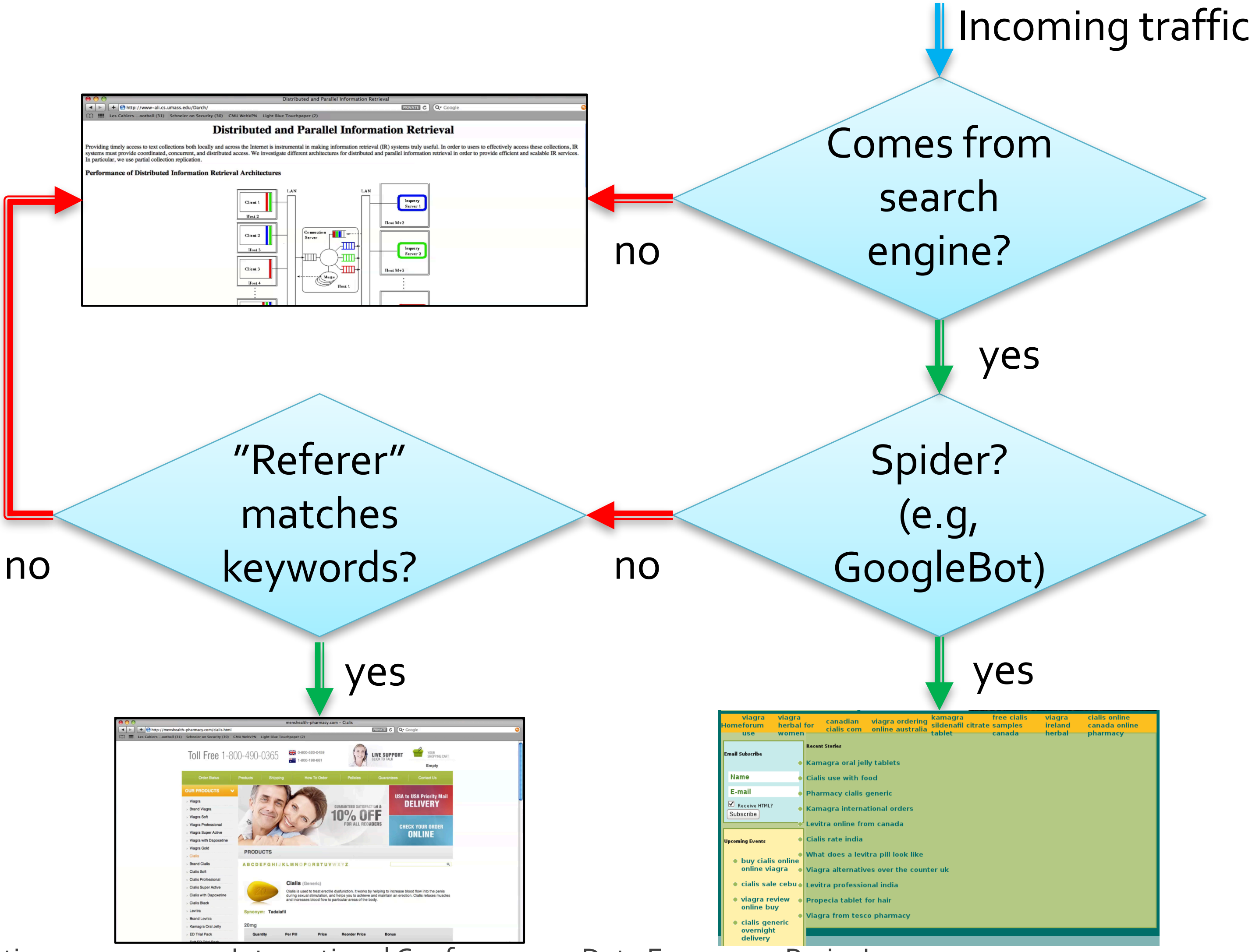
Upcoming Events

- buy cialis online online viagra
- cialis sale cebu
- viagra review online buy
- cialis generic overnight delivery

Recent Stories

- Kamagra oral jelly tablets
- Cialis use with food
- Pharmacy cialis generic
- Kamagra international orders
- Levitra online from canada
- Cialis rate india
- What does a levitra pill look like
- Viagra alternatives over the counter uk
- Levitra professional india
- Propecia tablet for hair
- Viagra from tesco pharmacy

Attack modus operandi: Malware logic



Incoming traffic

Comes from search engine?

no

yes

Spider? (e.g, GoogleBot)

no

yes

"Referer" matches keywords?

no

yes

Distributed and Parallel Information Retrieval
http://www-ali.cs.umass.edu/Darch/
PRIVATE
Google
Les Cahiers ...ootball (31)
Schneier on Security (30)
CMU WebVPN
Light Blue Touchpaper (2)

Distributed and Parallel Information Retrieval

Providing timely access to text collections both locally and across the Internet is instrumental in making information retrieval (IR) systems truly useful. In order to users to effectively access these collections, IR systems must provide coordinated, concurrent, and distributed access. We investigate different architectures for distributed and parallel information retrieval in order to provide efficient and scalable IR services. In particular, we use partial collection replication.

Performance of Distributed Information Retrieval Architectures

menshealth-pharmacy.com - Cialis
http://menshealth-pharmacy.com/cialis.html
PRIVATE
Google
Les Cahiers ...ootball (31)
Schneier on Security (30)
CMU WebVPN
Light Blue Touchpaper (2)

Toll Free 1-800-490-0365

0-800-520-0459

1-800-198-661

- Viagra
- Brand Viagra
- Viagra Soft
- Viagra Professional
- Viagra Super Active
- Viagra with Dapoxetine
- Viagra Gold
- Cialis
- Brand Cialis
- Cialis Soft
- Cialis Professional
- Cialis Super Active
- Cialis with Dapoxetine
- Cialis Black
- Levitra
- Brand Levitra
- Kamagra Oral Jelly
- ED Trial Pack

Cialis (Generic)

Cialis is used to treat erectile dysfunction. It works by helping to increase blood flow into the penis during sexual stimulation, and helps you to achieve and maintain an erection. Cialis relaxes muscles and increases blood flow to particular areas of the body.

Synonym: Tadalafil

20mg

Quantity	Per Pill	Price	Reorder Price	Bonus

Homeforum
viagra use
viagra herbal for women
canadian cialis com
viagra ordering online australia
kamagra sildenafil citrate tablet
free cialis samples canada
viagra ireland herbal
cialis online canada online pharmacy

Email Subscribe

Name

E-mail

☑ Receive HTML?

Subscribe

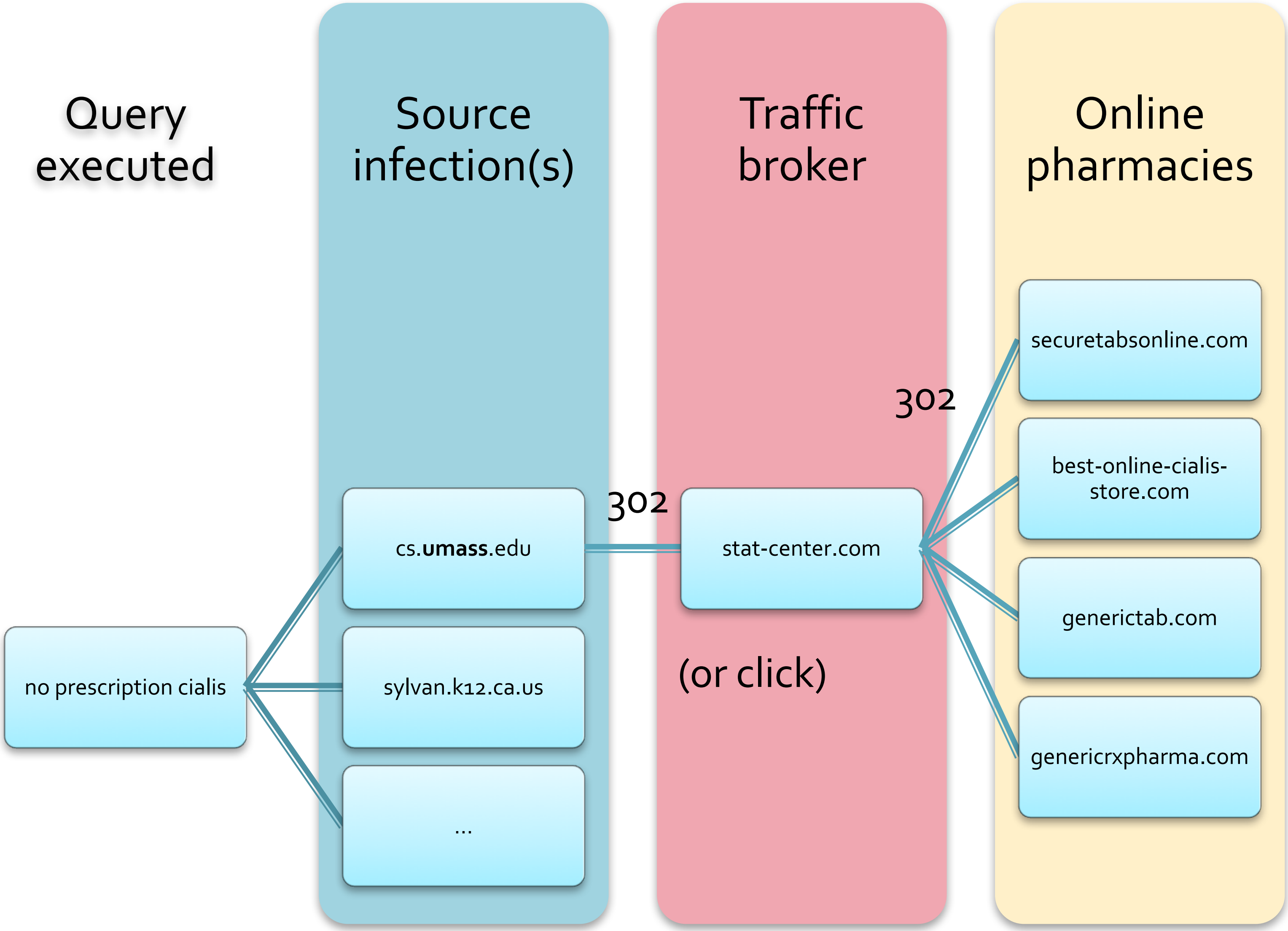
Upcoming Events

- buy cialis online online viagra
- cialis sale cebu
- viagra review online buy
- cialis generic overnight delivery

Recent Stories

- Kamagra oral jelly tablets
- Cialis use with food
- Pharmacy cialis generic
- Kamagra international orders
- Levitra online from canada
- Cialis rate india
- What does a levitra pill look like
- Viagra alternatives over the counter uk
- Levitra professional india
- Propecia tablet for hair
- Viagra from tesco pharmacy

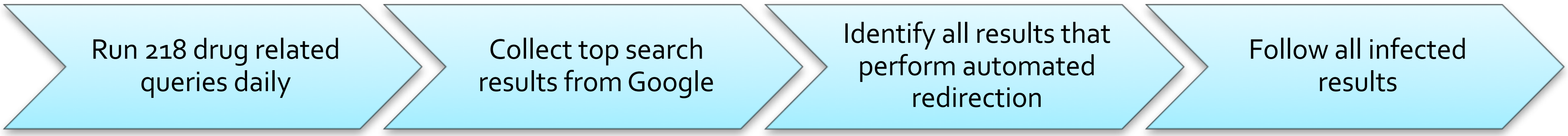
Attack modus operandi: Redirection chains



Questions

- How has this attack evolved?
  - Volume and impact: Does this even matter?
  - Techniques
- Why has the attack evolved?
  - Effectiveness of the defenses?
- Can this be thwarted?
  - Legal intervention vs. technical defenses

Data collection process



Run 218 drug related queries daily

- Daily collection from 4/12/2010 through 9/16/2013

Collect top search results from Google

- Limit due to Google Search API
- Store 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

Datasets collected

[LMC USENIX Security 2011, CCS 2014]

- Dataset 1
  - Aggregate results only
  - Rank of the results unknown
  - Mapping query-results unknown
- Dataset 2
  - Same as Dataset 1, but ranking information known
  - Mapping query-result doesn't include rank
- Dataset 3
  - All information is captured
  - … but new Google API (slightly) limits results returned

Dataset	1	2	3
Period	4/12/10-11/15/10	11/15/10-10/8/11	10/8/11-9/16/13
Search results/query	64	64	16/32
Total results	260,824	3,609,675	1,530,099
Unique URLs	150,955	189,023	122,382
Unique domains	25,182	36,557	30,881

This is 3.5 years worth of data!

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

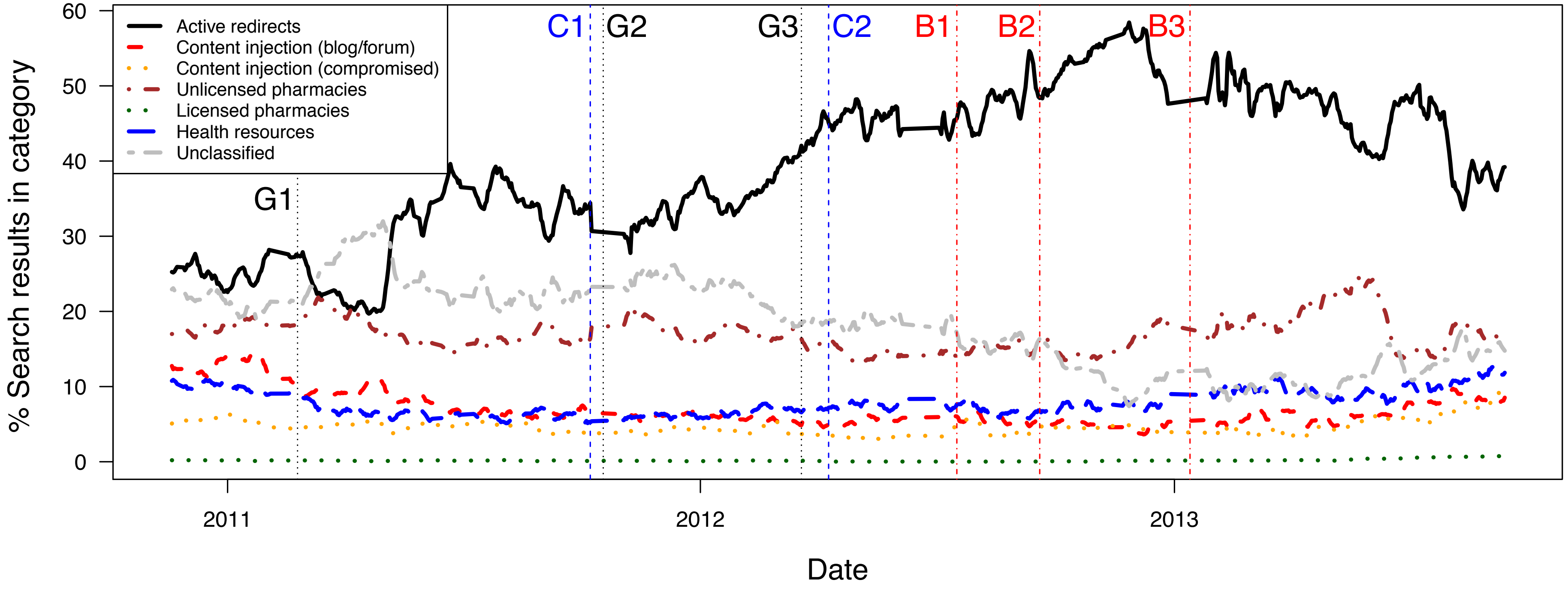
vicodin without prescription
generic cialis free sample
cheap tadalafil
20 mg ambien overdose
prozac side effects
ambien buy online
alprazolam online without prescription buy cheap

Type	Count	Percent
Malicious (Black)	26	22%
Benign (White)	75	34%
Ambiguous (Gray)	117	54%
Total	218	100%

Long-term evolution

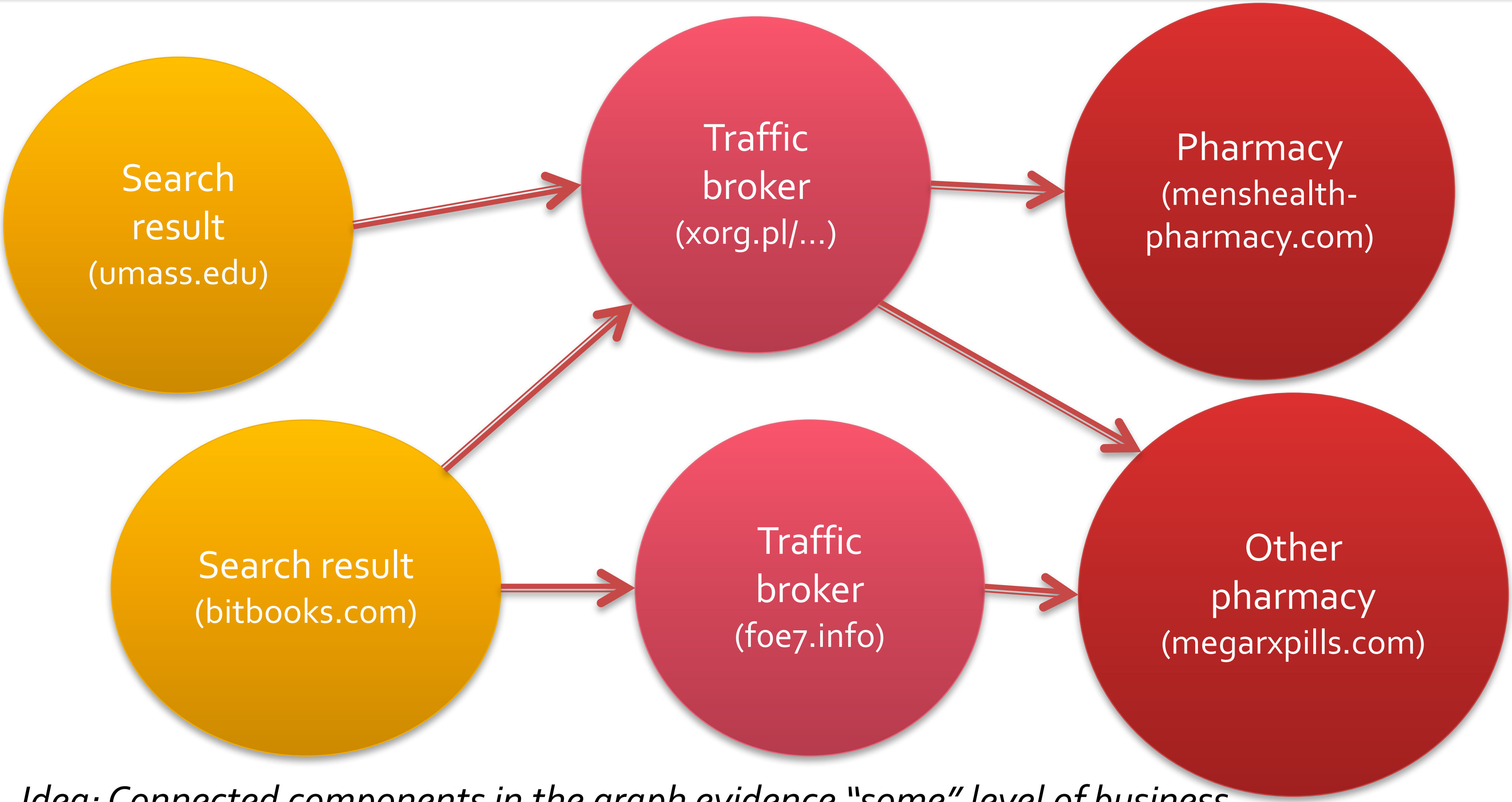
[LMC CCS 2014]

Evolution of search results



G1: Google changes search ranking algorithm
G2: Google starts removing query info from "Referer" field
G3: Google is done deploying Referer modifictations
B1, B2, B3 : Firefox, Safari, Chrome switch to HTTPS-only search
(C1,C2: major changes to our collection infrastructure)

Uncovering relationships in search results


Search result (umass.edu)
Traffic broker (xorg.pl/...)
Pharmacy (menshealth-pharmacy.com)
Search result (bitbooks.com)
Traffic broker (foe7.info)
Other pharmacy (megarxpills.com)

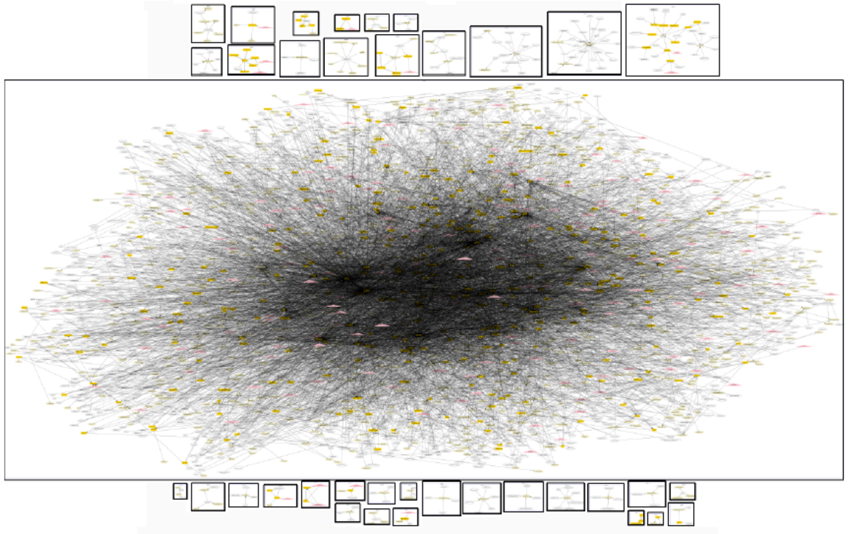
Idea: Connected components in the graph evidence "some" level of business relationships between the nodes they connect

Connected components

- 34 connected components

- One connected component contains
  - 96% of all infected domains
  - 90% of all redirection domains
  - 92% of all pharmacies

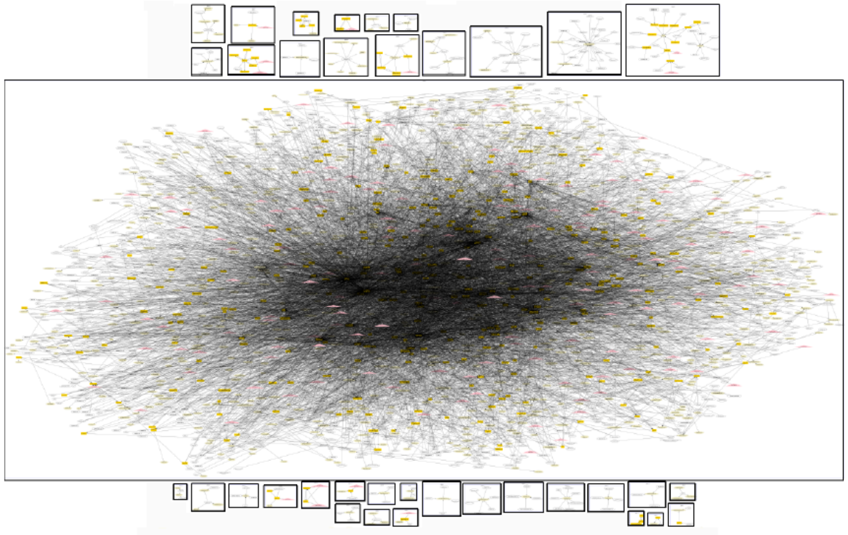
- Is one person responsible for all of this?!



Connected components

- 34 connected components

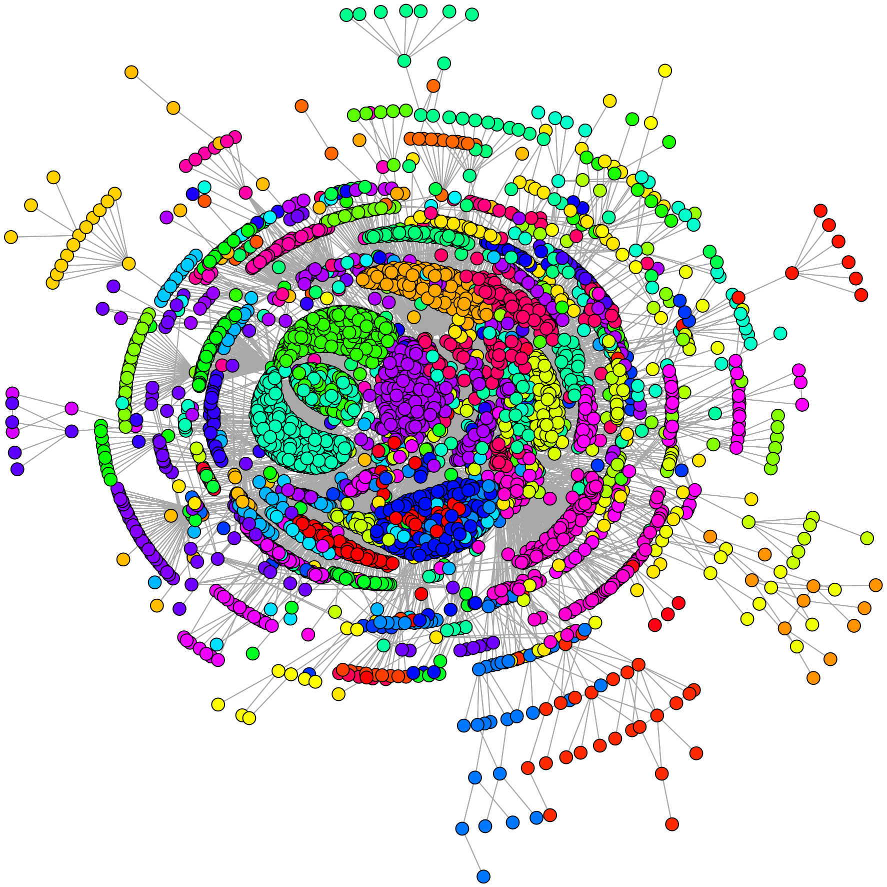
- One connected component contains
  - 96% of all infected domains
  - 90% of all redirection domains
  - 92% of all pharmacies



- Is one person responsible for all of this?!
  - NO!
  - Some advertisers work for several different affiliate networks
  - Certain domains are (ab)used by multiple advertisers

Identifying the main players

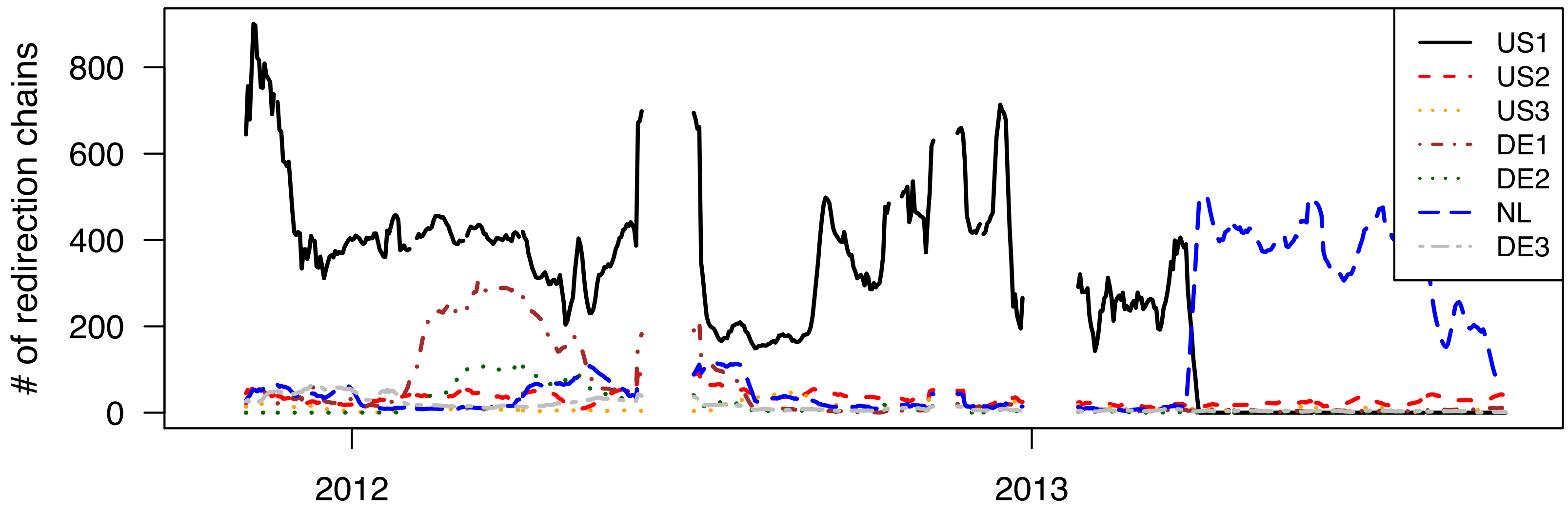
- Run (spinglass) clustering algorithm in big connected component
- Each cluster represented by different color
- Evidence of separate organized groups/campaigns more loosely connected to each other
  - About 10-12 large groups



Illicit advertising infrastructure

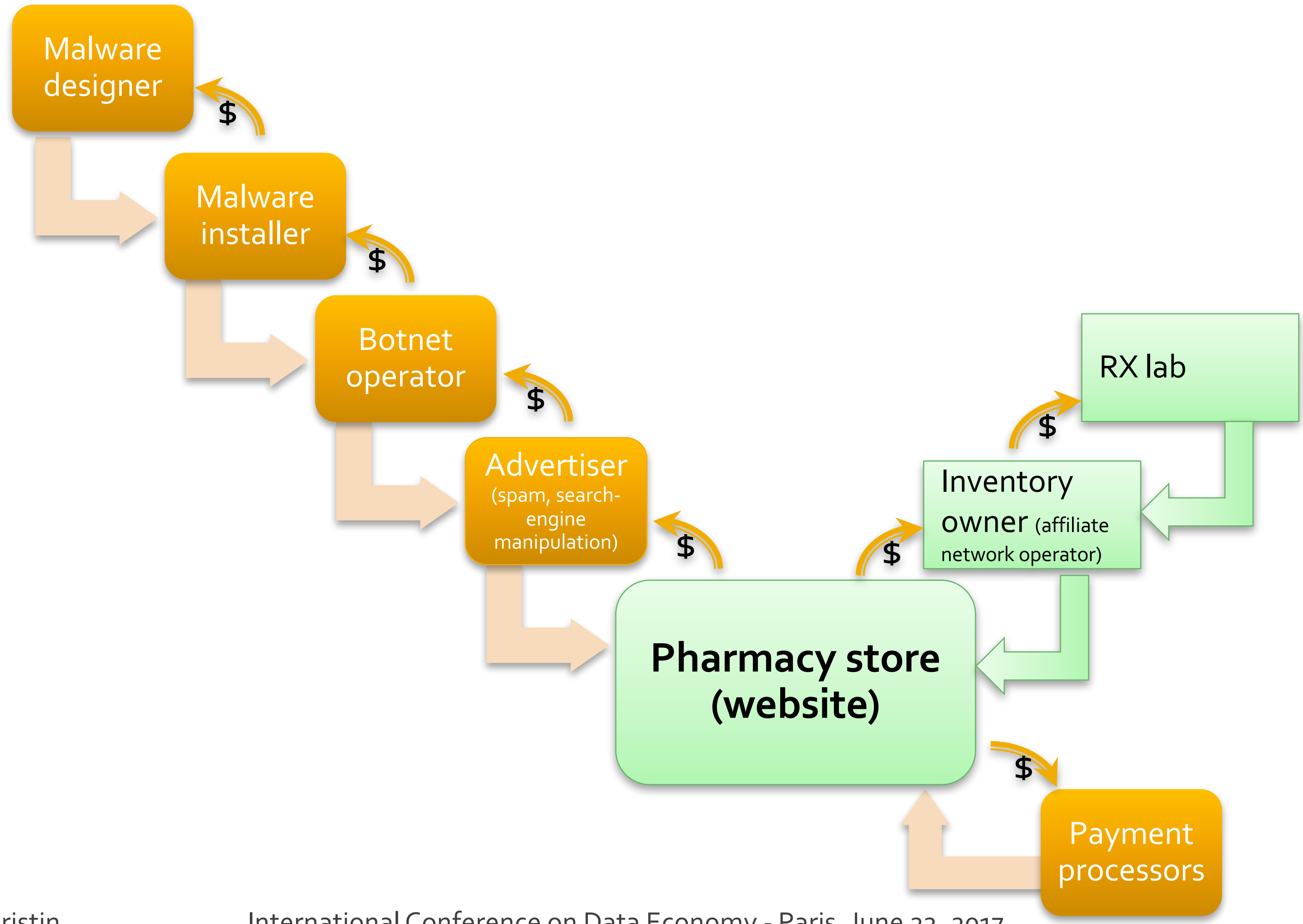
- Traffic brokers are disproportionately hosted on **very few** networks

Traffic brokers observed each day grouped by AS



© Nicolas Christin

Procuring unlicensed drugs

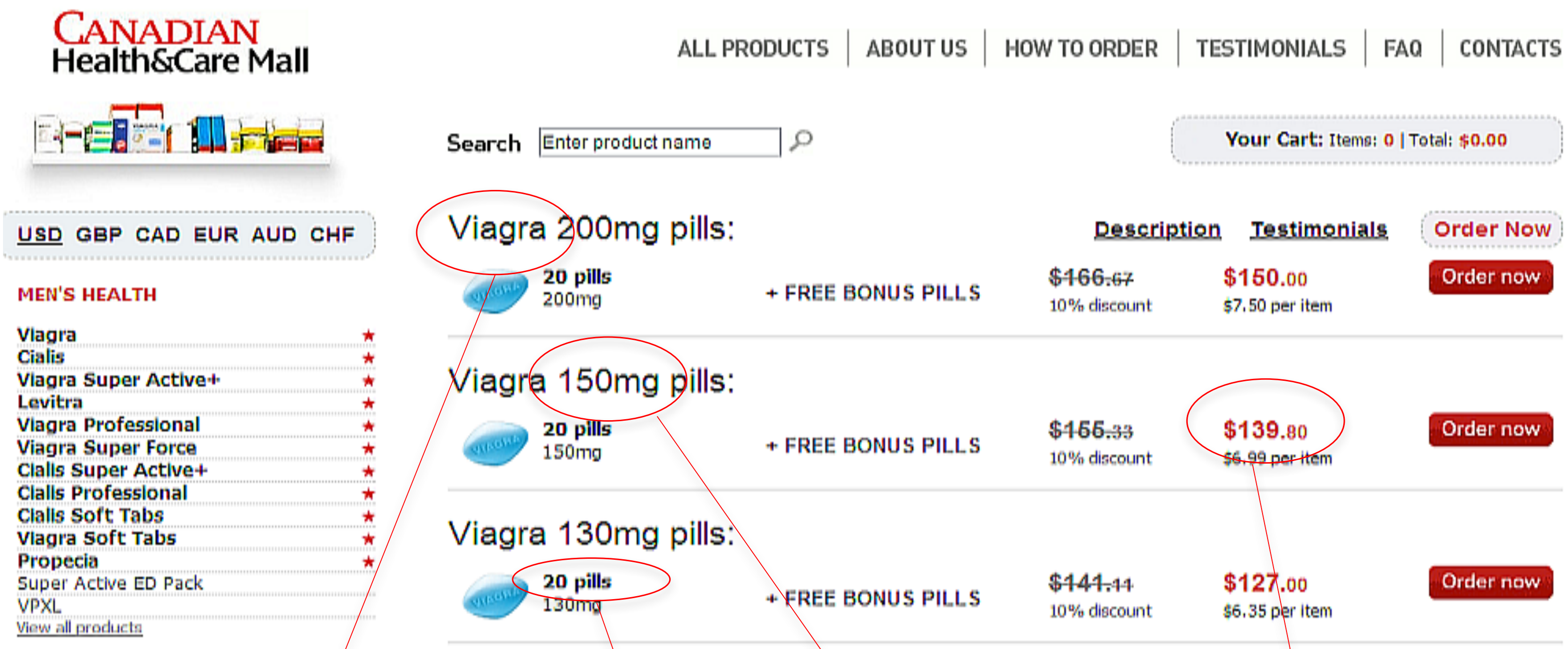


Inventory analysis [LMC, 2013]

Data collected

- Scraped for prices and inventories:
  - 265 unlicensed pharmacies (doing search-redirection attacks) collected between April 3, 2012 and October 16, 2012
  - 265 "blacklisted" pharmacies
    - Randomly sampled out of a corpus of 9000+ NABP "not recommended" pharmacies
    - No overlap with the unlicensed pharmacy corpus

Scraping



CANADIAN Health&Care Mall

ALL PRODUCTS | ABOUT US | HOW TO ORDER | TESTIMONIALS | FAQ | CONTACTS

Search Enter product name

Your Cart: Items: 0 | Total: $0.00

USD GBP CAD EUR AUD CHF

MEN'S HEALTH

- Viagra ★
- Cialis ★
- Viagra Super Active+ ★
- Levitra ★
- Viagra Professional ★
- Viagra Super Force ★
- Cialis Super Active+ ★
- Cialis Professional ★
- Cialis Soft Tabs ★
- Viagra Soft Tabs ★
- Propecia ★
- Super Active ED Pack
- VPXL
- View all products

Description Testimonials Order Now

Viagra 200mg pills:				
20 pills 200mg	+ FREE BONUS PILLS	~~$466.67~~ 10% discount	$150.00 $7.50 per item	Order now
Viagra 150mg pills:				
20 pills 150mg	+ FREE BONUS PILLS	~~$155.33~~ 10% discount	$139.80 $6.99 per item	Order now
Viagra 130mg pills:				
20 pills 130mg	+ FREE BONUS PILLS	~~$141.44~~ 10% discount	$127.00 $6.35 per item	Order now

Drug name/product (subsequently mapped to active ingredient/condition)

Quantity

Dosage

Price

Total = 1,451,587 distinct (drug, active ingredient, dosage, unit) tuples collected

1,661 different drug names

© Nicol

Identifying common suppliers: Inventory overlap

- How much overlap is there between distinct unlicensed pharmacies' inventories?

- Jaccard distance:

Inventory of pharmacy $\mathscr{A}$

$$J_\delta(A,B) = 1 - \frac{|A \cap B|}{|A \cup B|}$$

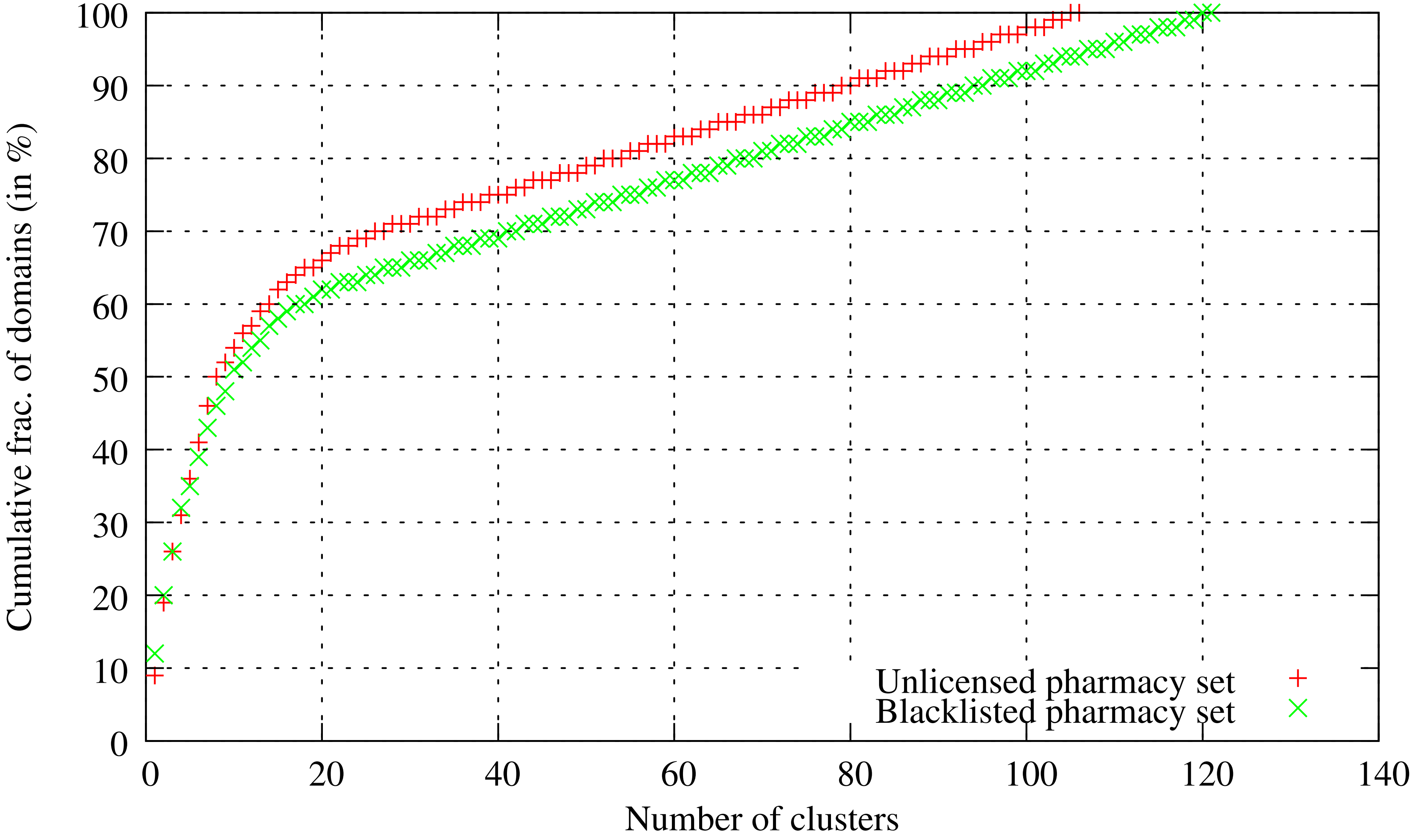
Inventory of pharmacy $\mathscr{B}$

  - Identical inventories $\Leftrightarrow J_\delta(A,B) = 0$
  - No overlap at all $\Leftrightarrow J_\delta(A,B) = 1$

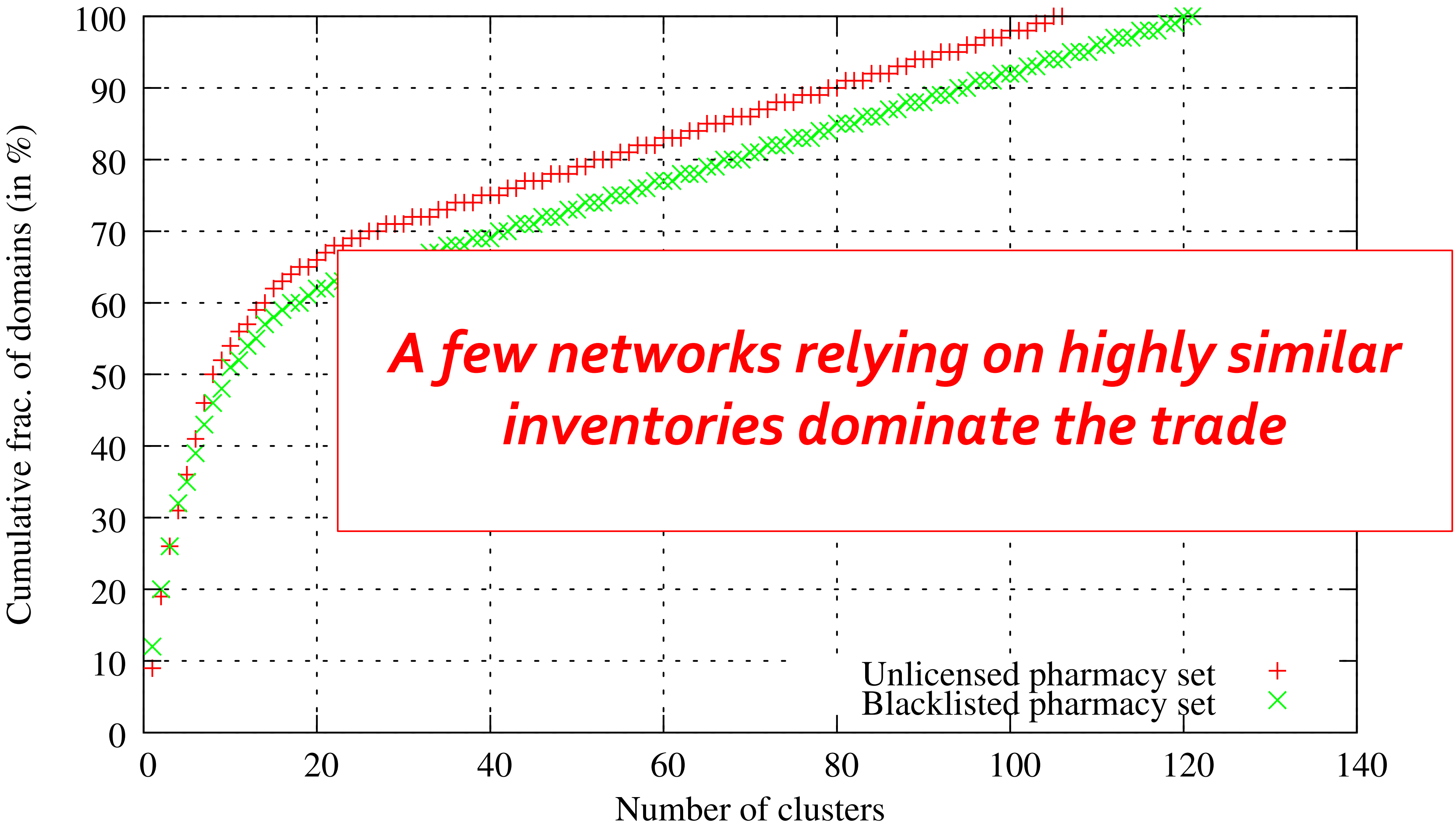
Clustering inventories

- Inventory A and Inventory B belong to the same cluster iff $J_\delta(A, B) < t$
  - t is an arbitrary threshold, $0<t<1$
- Distance between two groups of inventories X, Y:
  - Minimum linkage: $J_\delta(X, Y) = \min\{J_\delta(x, y) : x \in X, y \in Y\}$
  - Maximum linkage: $J_\delta(X, Y) = \max\{J_\delta(x, y) : x \in X, y \in Y\}$
  - Average linkage: $J_\delta(X, Y) = \frac{1}{|X| \cdot |Y|} \sum_{x \in X} \sum_{y \in Y} J_\delta(x, y)$

Clustering inventories: Average linkage, *t=0.31*

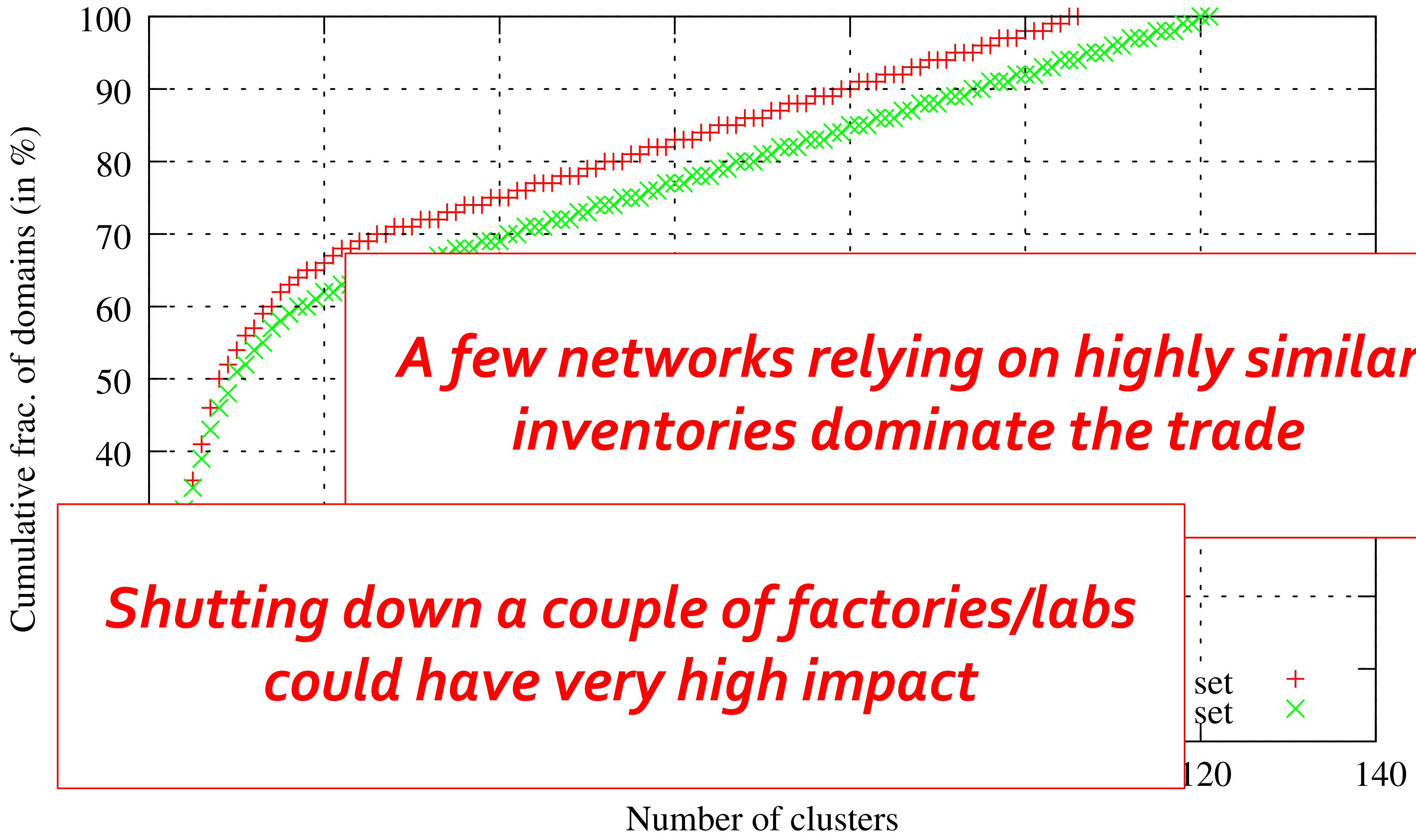


Clustering inventories: Average linkage, *t=0.31*



A few networks relying on highly similar inventories dominate the trade

Clustering inventories: Average linkage, *t=0.31*



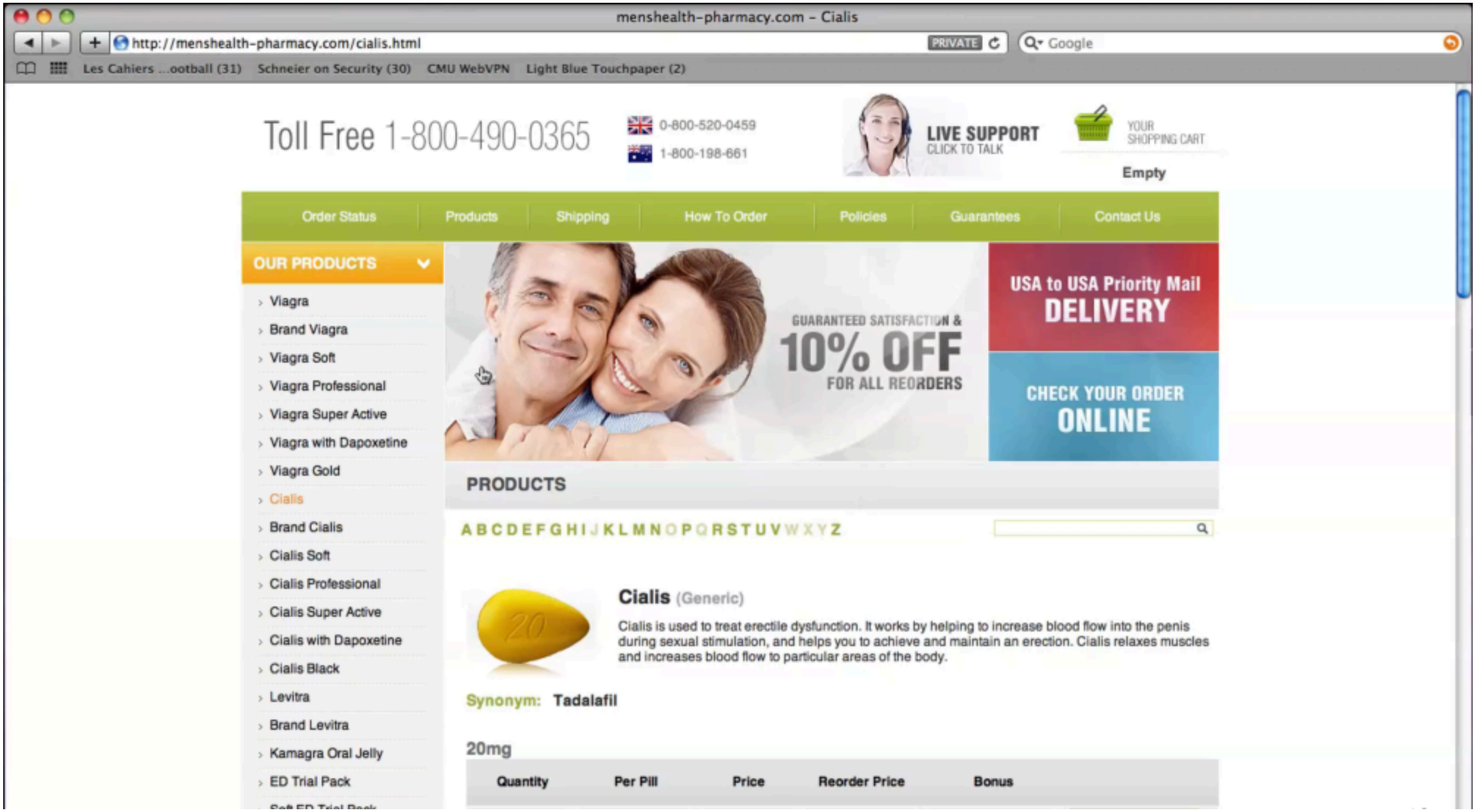
A few networks relying on highly similar inventories dominate the trade

Shutting down a couple of factories/labs could have very high impact

Intervening, or: what does the analysis tell us

- High concentration in traffic brokers
    - Orders of magnitude less numerous than pharmacies and infected hosts
  - Mostly hosted on same networks
  - Structure hasn't changed much over four years
  - Opportunities for takedowns seem ripe
    - Jurisdiction issues?
- High concentration is suppliers (labs)
  - Of strong interest to manufacturers…

Evolution of illicit Internet commerce



Cialis (Generic)

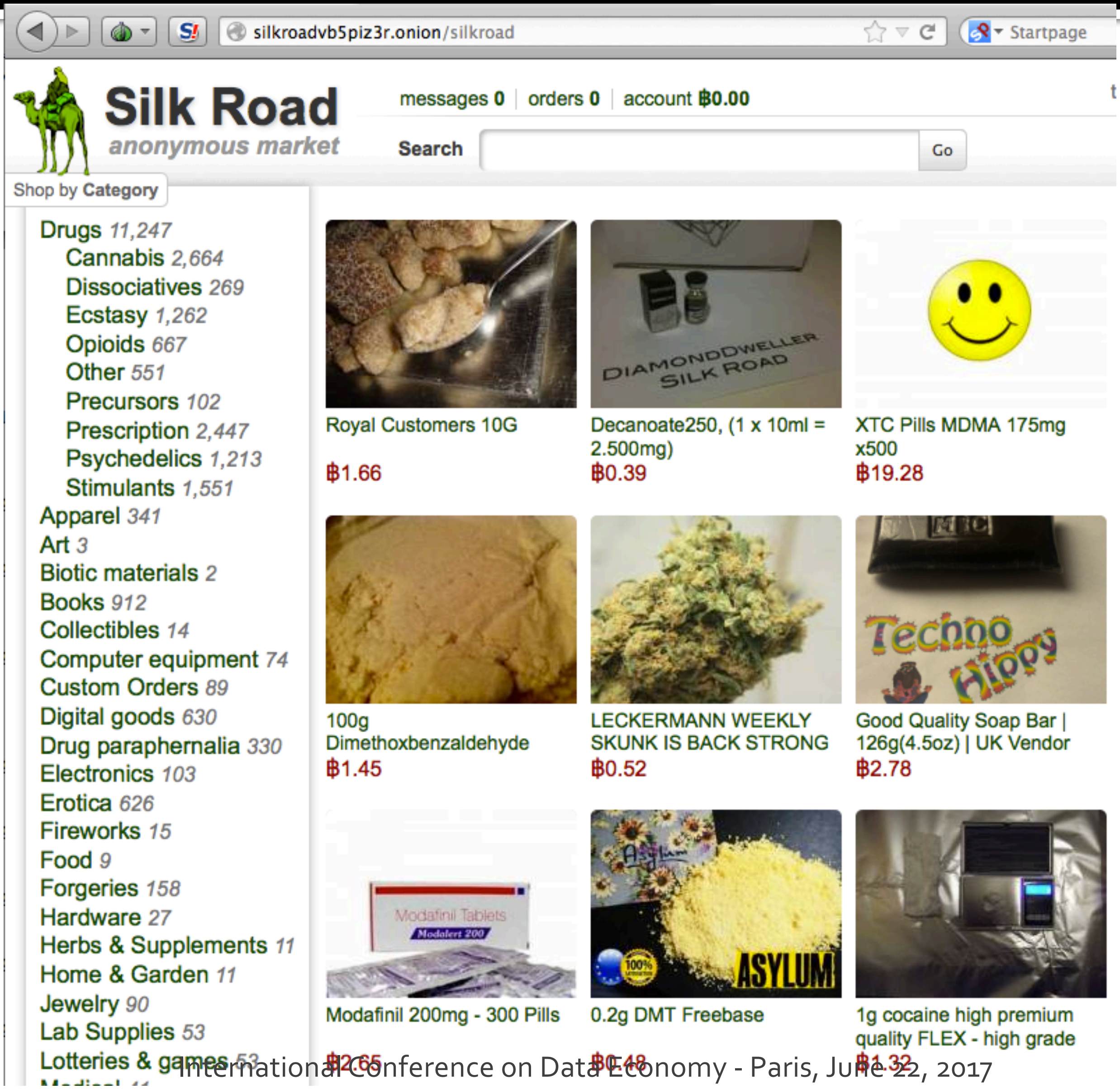
Cialis is used to treat erectile dysfunction. It works by helping to increase blood flow into the penis during sexual stimulation, and helps you to achieve and maintain an erection. Cialis relaxes muscles and increases blood flow to particular areas of the body.

Synonym: Tadalafil

20mg

Quantity	Per Pill	Price	Reorder Price	Bonus

Evolution of illicit Internet commerce



silkroadvb5piz3r.onion/silkroad

Silk Road
anonymous market

Drugs 11,247
- Cannabis 2,664
- Dissociatives 269
- Ecstasy 1,262
- Opioids 667
- Other 551
- Precursors 102
- Prescription 2,447
- Psychedelics 1,213
- Stimulants 1,551

Apparel 341
Art 3
Biotic materials 2
Books 912
Collectibles 14
Computer equipment 74
Custom Orders 89
Digital goods 630
Drug paraphernalia 330
Electronics 103
Erotica 626
Fireworks 15
Food 9
Forgeries 158
Hardware 27
Herbs & Supplements 11
Home & Garden 11
Jewelry 90
Lab Supplies 53
Lotteries & games 53
Medical 11

Royal Customers 10G
฿1.66

Decanoate250, (1 x 10ml = 2.500mg)
฿0.39

XTC Pills MDMA 175mg x500
฿19.28

100g Dimethoxbenzaldehyde
฿1.45

LECKERMANN WEEKLY SKUNK IS BACK STRONG
฿0.52

Good Quality Soap Bar | 126g(4.5oz) | UK Vendor
฿2.78

Modafinil 200mg - 300 Pills
฿2.65

0.2g DMT Freebase
฿0.48

1g cocaine high premium quality FLEX - high grade
฿1.32

Case study:
Online anonymous marketplaces

- Amazon.com of illegal goods
  - Drugs, CC's & Fake IDs, Weapons, etc.
  - No child pornography
- Safety
- Convenience
- Variety
- Accountability
- Competition

Online anonymous marketplace technology

- Hidden Website (Tor Hidden Service, I2P)
  - Customers
    - No cost of creation
    - No information needed
  - Vendors
    - Vendor bonds required
    - Often invite only
    - Public feedback history

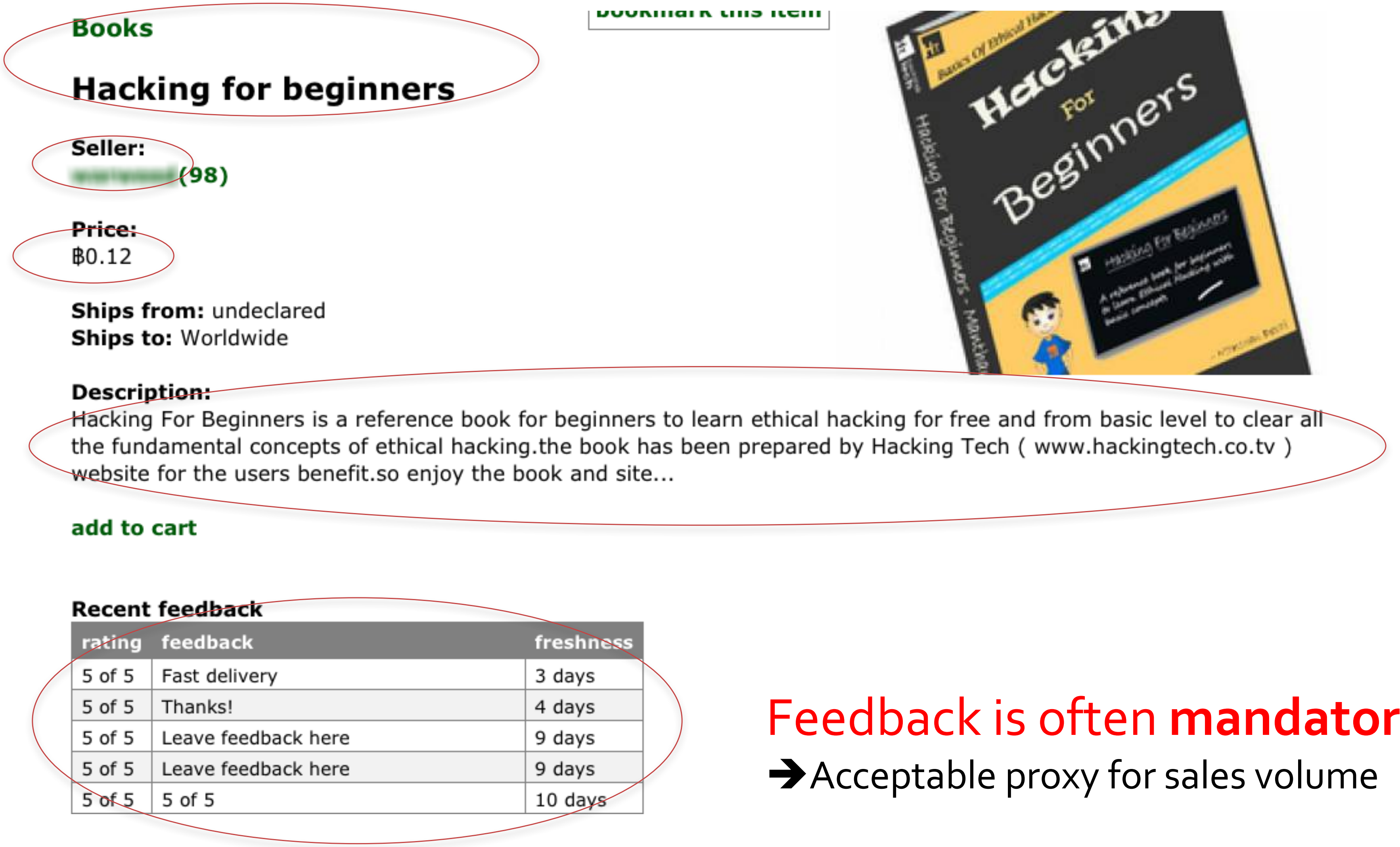
- Payments (Bitcoin)
  - Marketplaces often act as escrow agent
  - Escrow sometimes acts as a mixing service

- Encrypted Messages(PGP)

Questions

- How much is being sold?
- What is being sold?
- How many vendors are relevant?
- What are potentially successful interventions?

Typical listing page



bookmark this item

Books

Hacking for beginners

Seller:
[redacted](98)

Price:
฿0.12

Ships from: undeclared
Ships to: Worldwide

Description:
Hacking For Beginners is a reference book for beginners to learn ethical hacking for free and from basic level to clear all the fundamental concepts of ethical hacking.the book has been prepared by Hacking Tech (www.hackingtech.co.tv) website for the users benefit.so enjoy the book and site...

add to cart

Recent feedback

rating	feedback	freshness
5 of 5	Fast delivery	3 days
5 of 5	Thanks!	4 days
5 of 5	Leave feedback here	9 days
5 of 5	Leave feedback here	9 days
5 of 5	5 of 5	10 days

Feedback is often **mandatory!**

➔Acceptable proxy for sales volume

Measurements

- **Started collection in November 2011**
- **As of August 2015, we had collected**
  - 35 marketplaces
  - 1,908 scrapes total – 3.2 TB
  - 27 – 331,691 pages per scrape
- **Still collecting…**

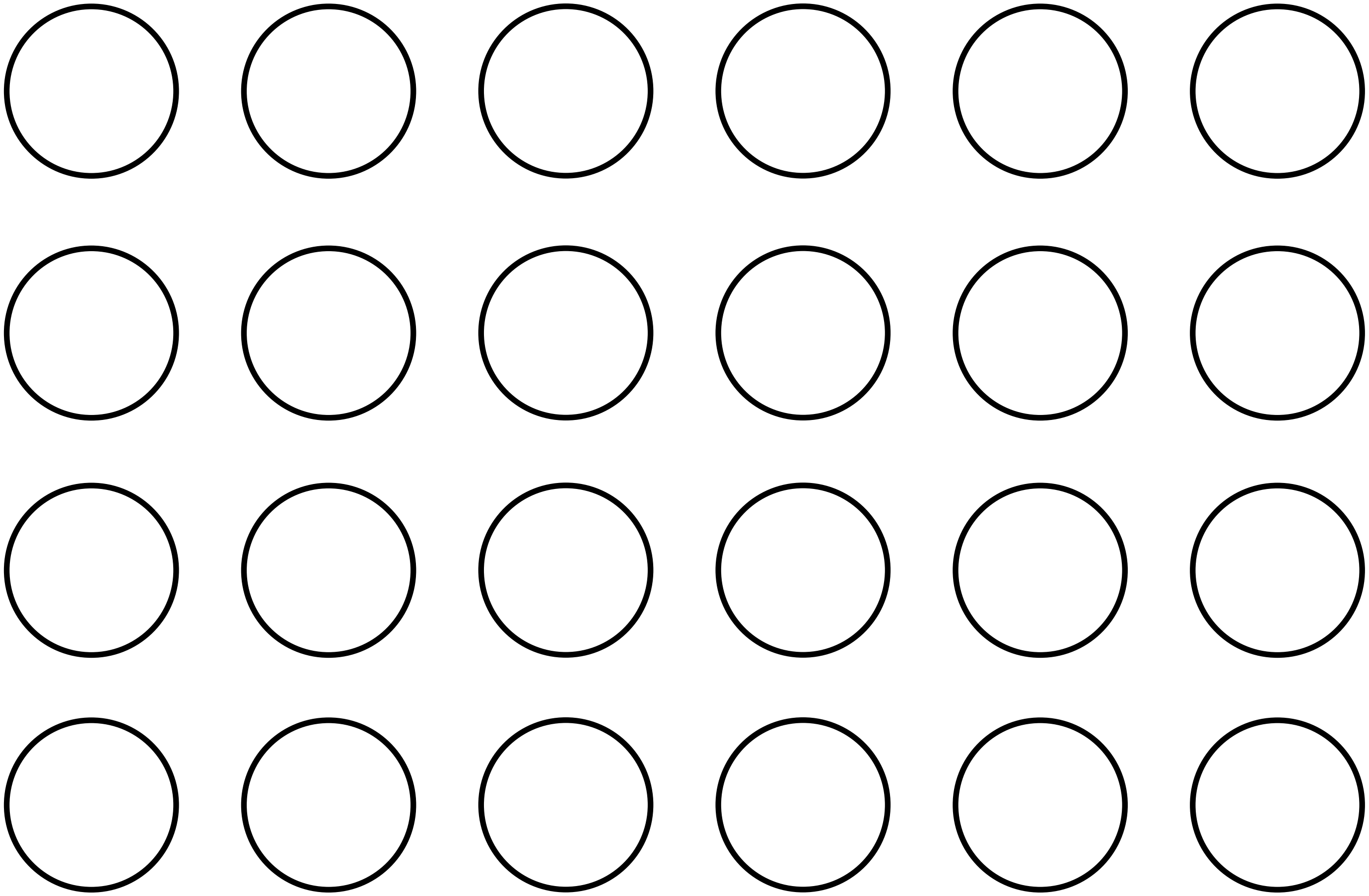
Data completeness

- **How complete is the data?**
  - Unreliable dynamic marketplaces that take days to scrape
  - Empirical observations – lower bound

- **Idea:** Estimate population via mark and recapture
  - Schnabel estimator allows multiple recapture

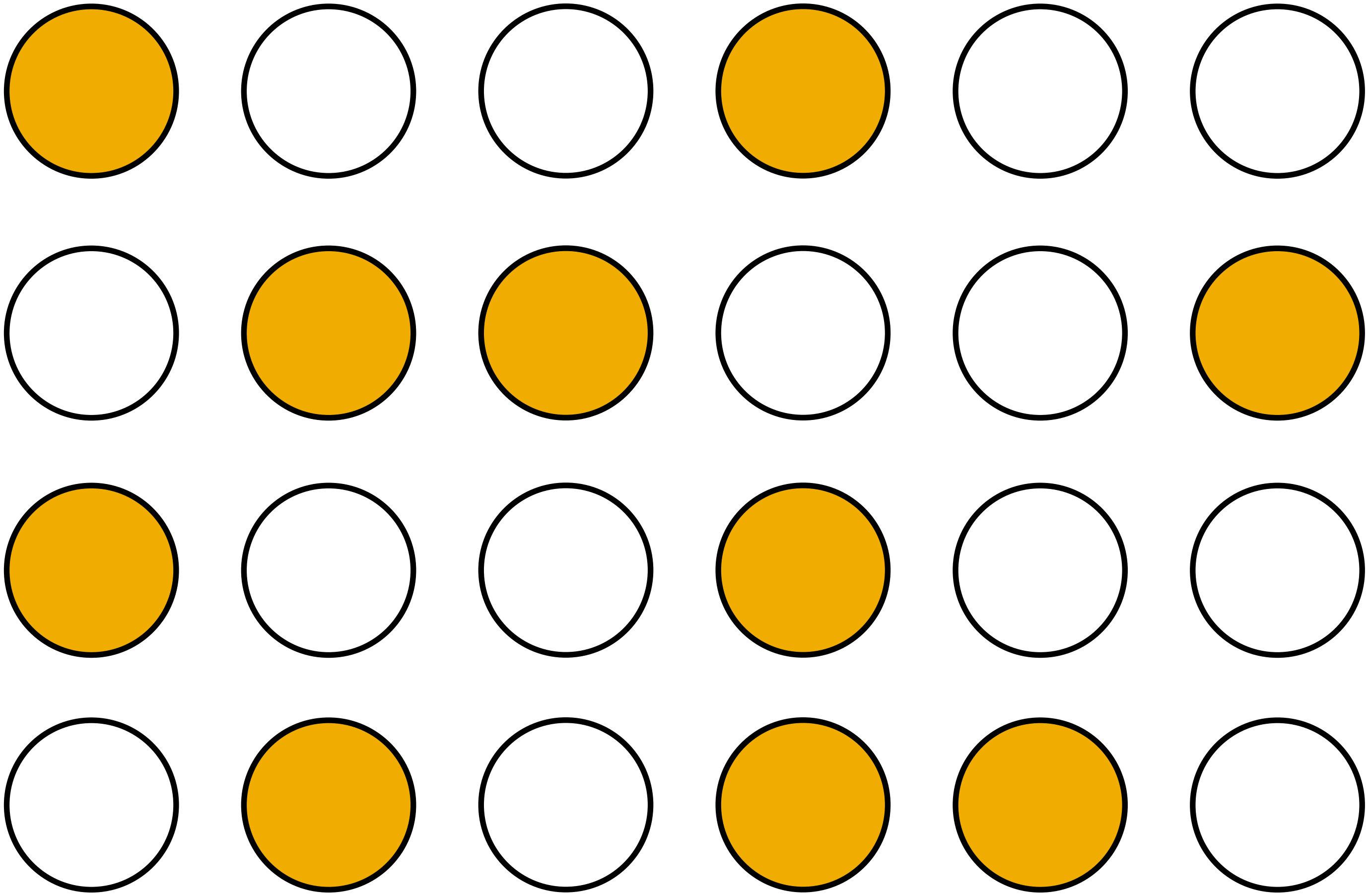
Mark and recapture

Population Size = 24



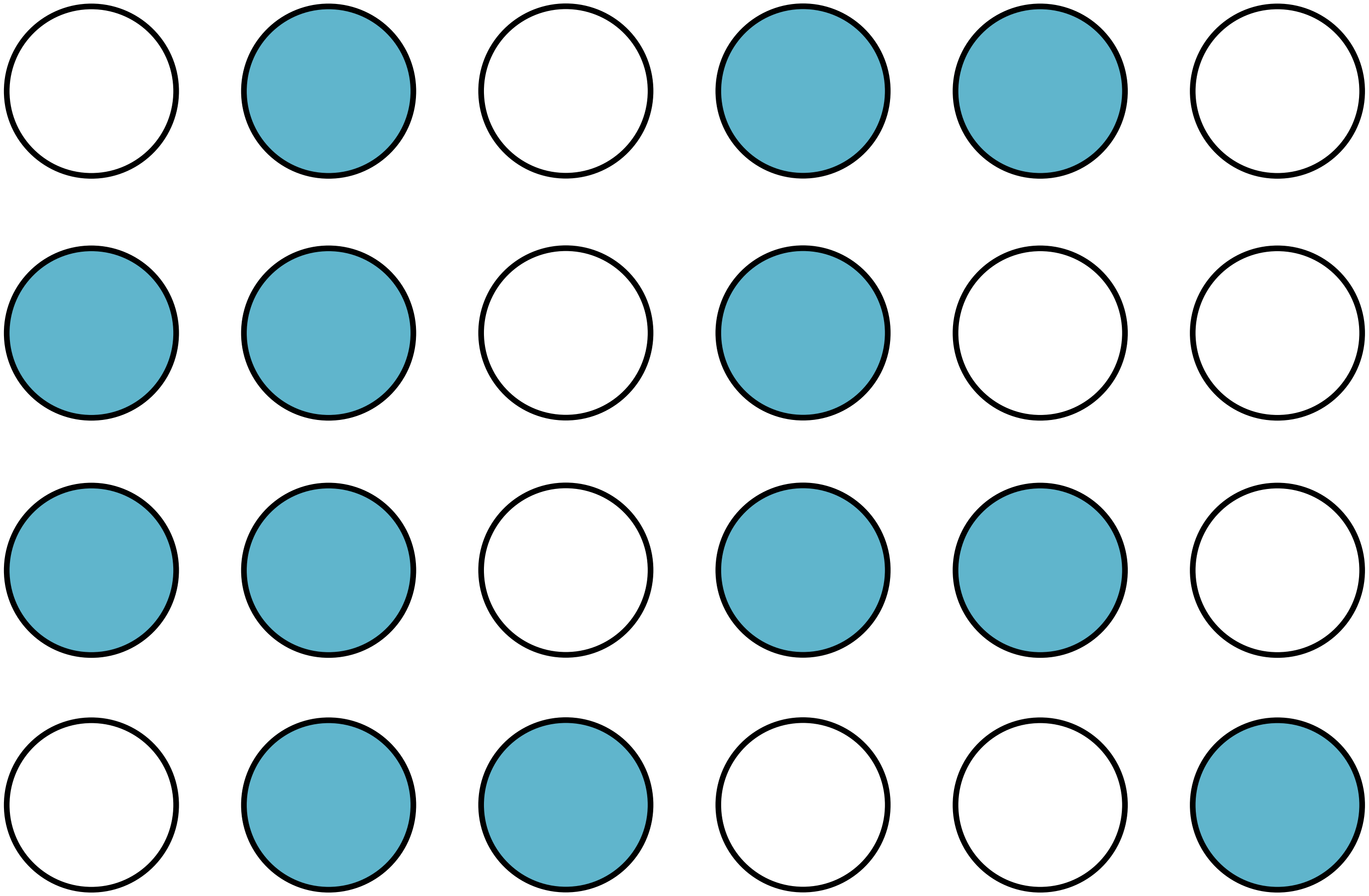
Mark and recapture

Sample Size = 10



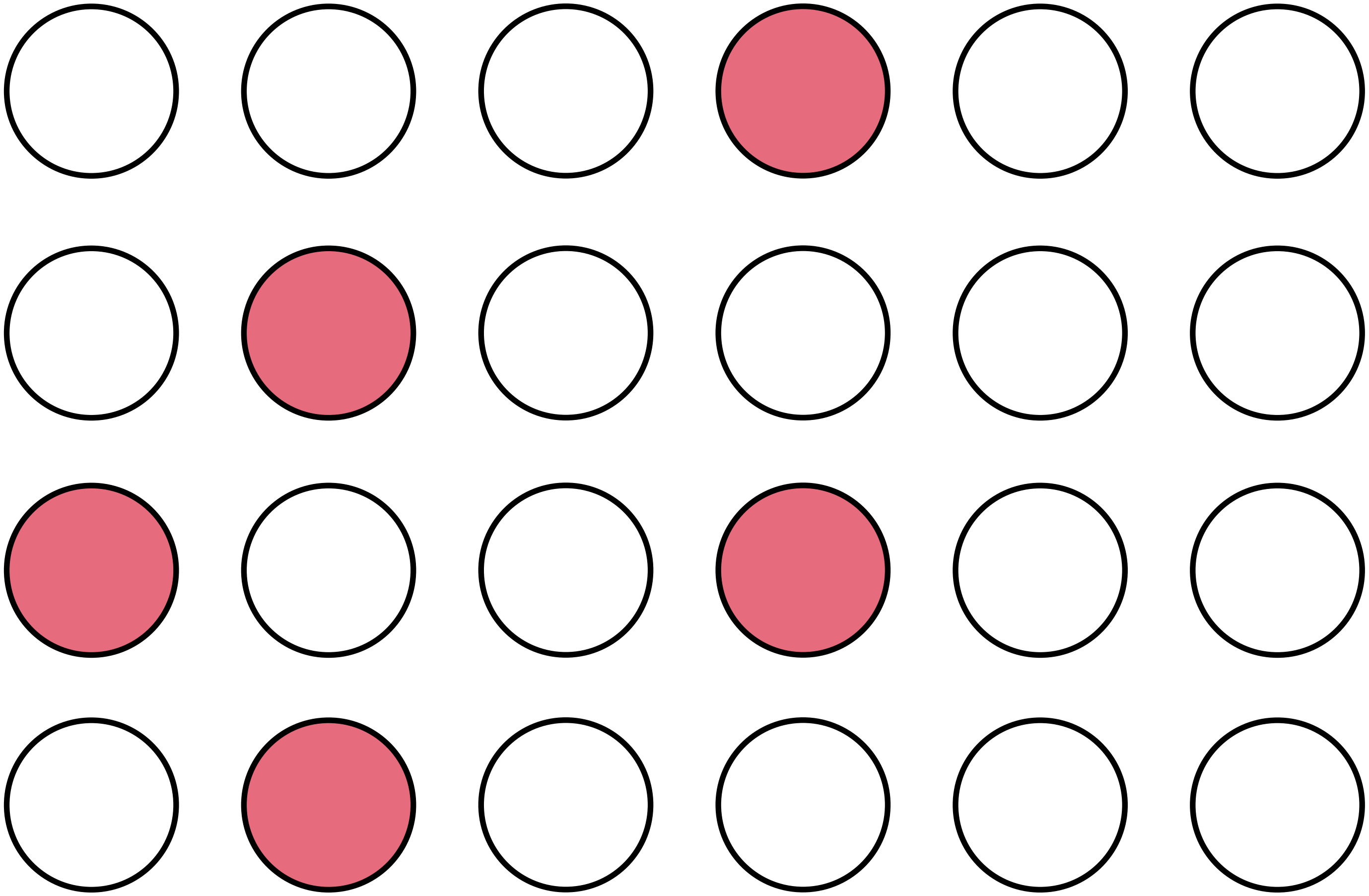
Mark and recapture

Sample Size = 13

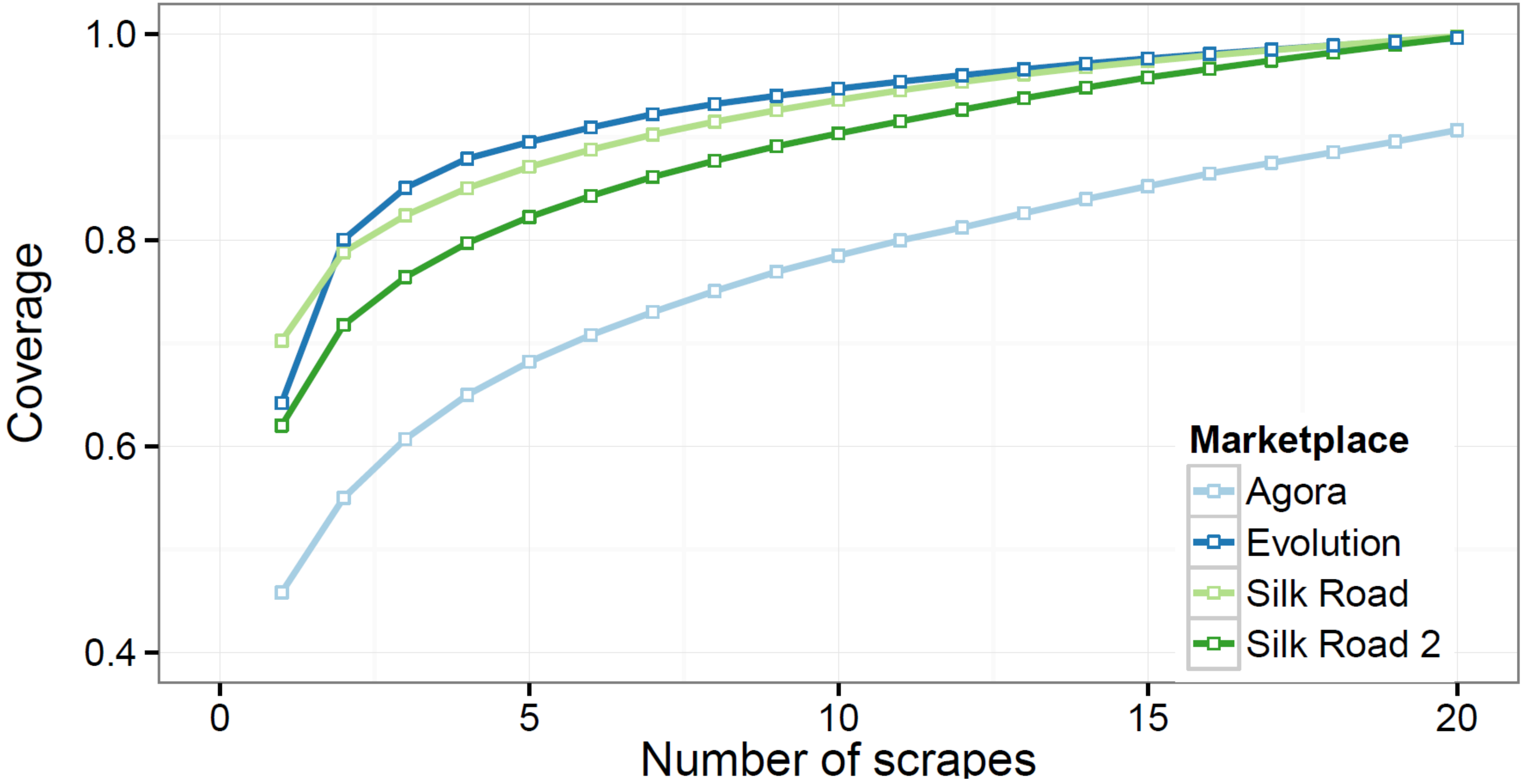


Mark and recapture

Overlap = 5, Population Estimate = 26



Data completeness



Analysis

- **Assumption:** Each feedback corresponds to precisely one transaction
  - Anonymity requires strictly enforced feedback system to establish reputation
  - Possible on many marketplaces to purchase several quantities of item and leave one feedback, conservative estimate

Analysis challenges

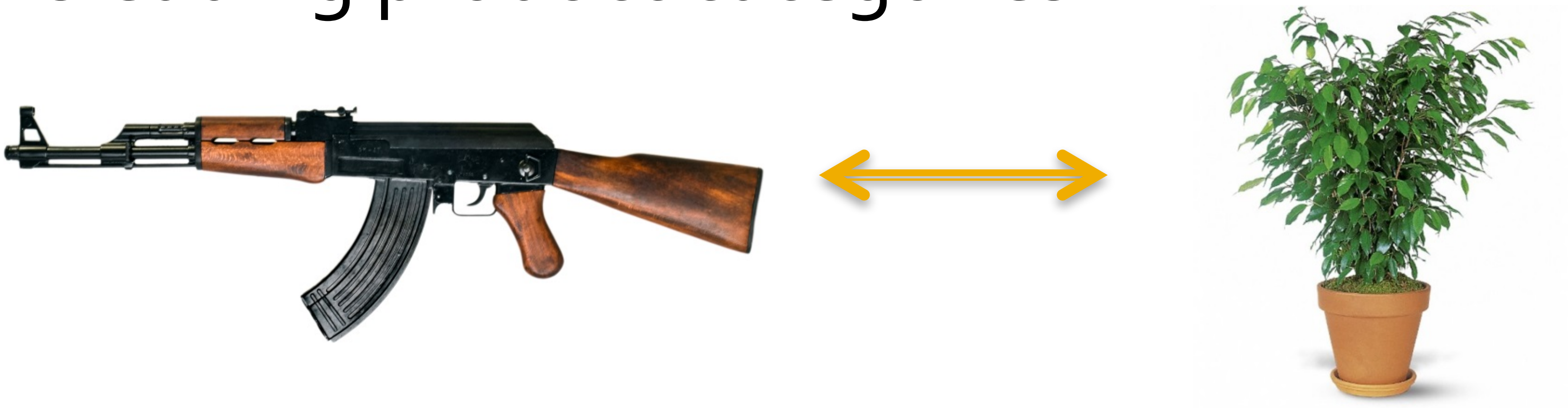
- “Holding prices”
  - Came up with automated statistical filtering of outliers

$0.02 -> $1,000.00

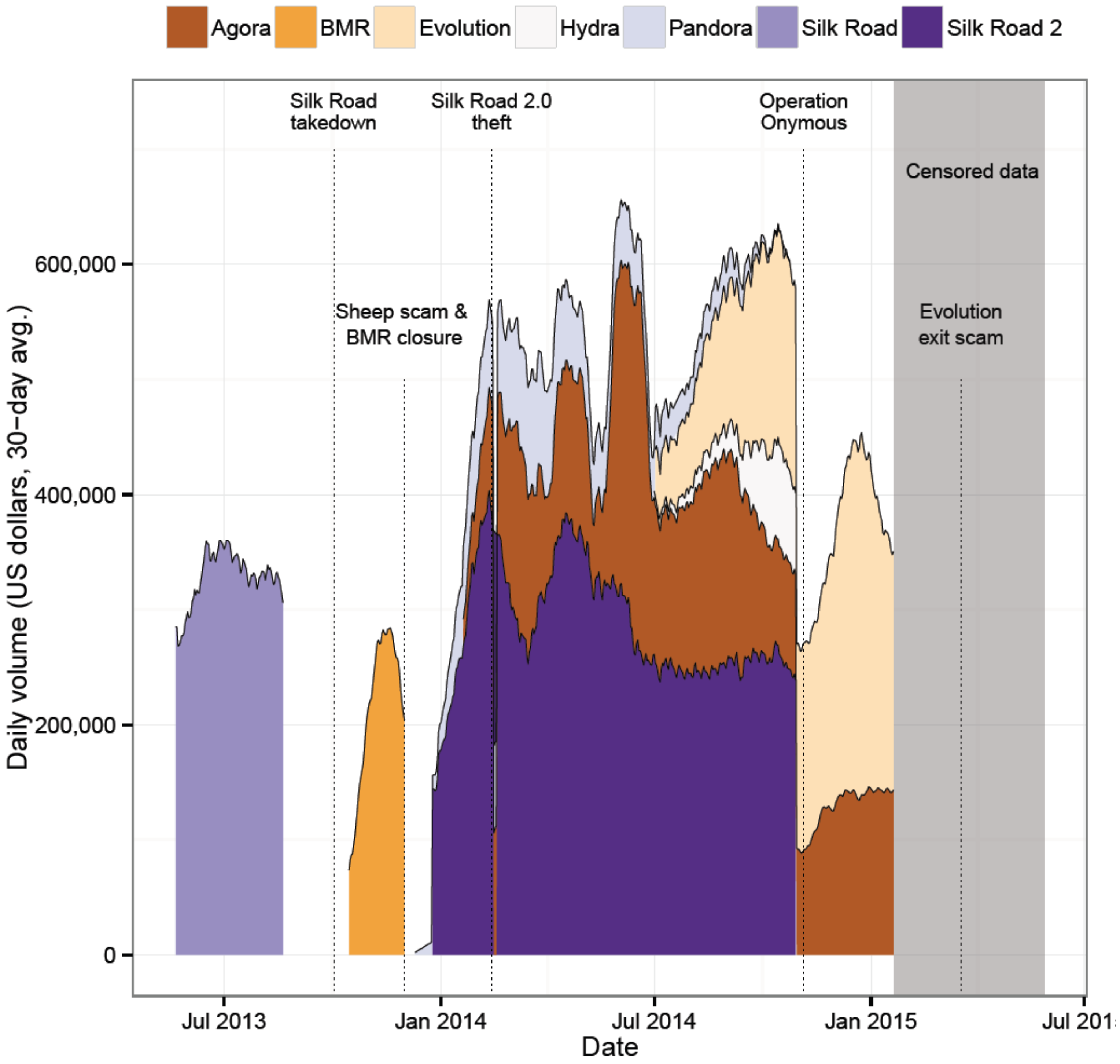
$1,100.00 -> $1,000,000.00

Analysis challenges

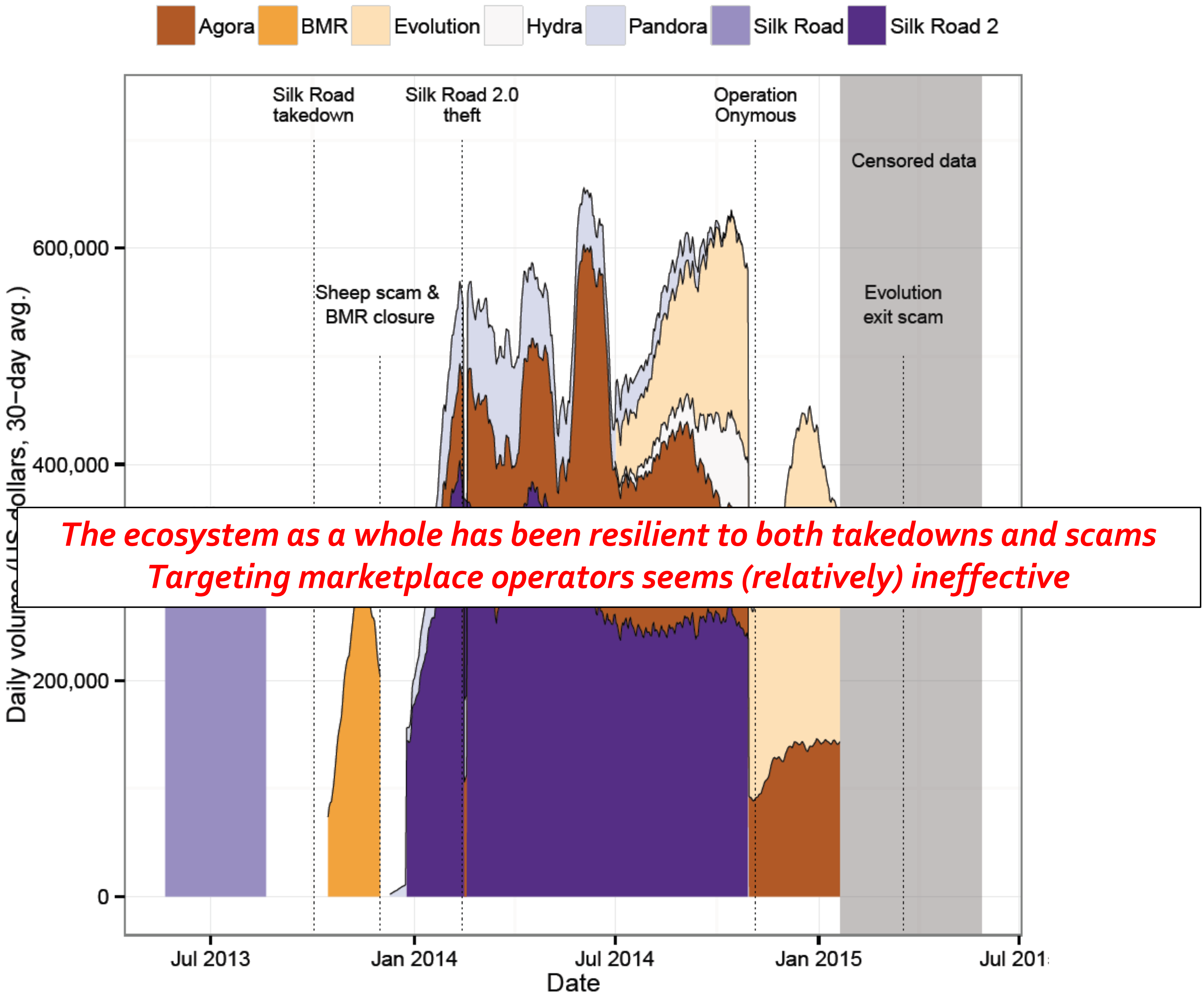
- Misleading product categories



  - Define sixteen categories
  - Designed special purpose classifier to infer which category each listing belongs to
    - Extract from tf-idf

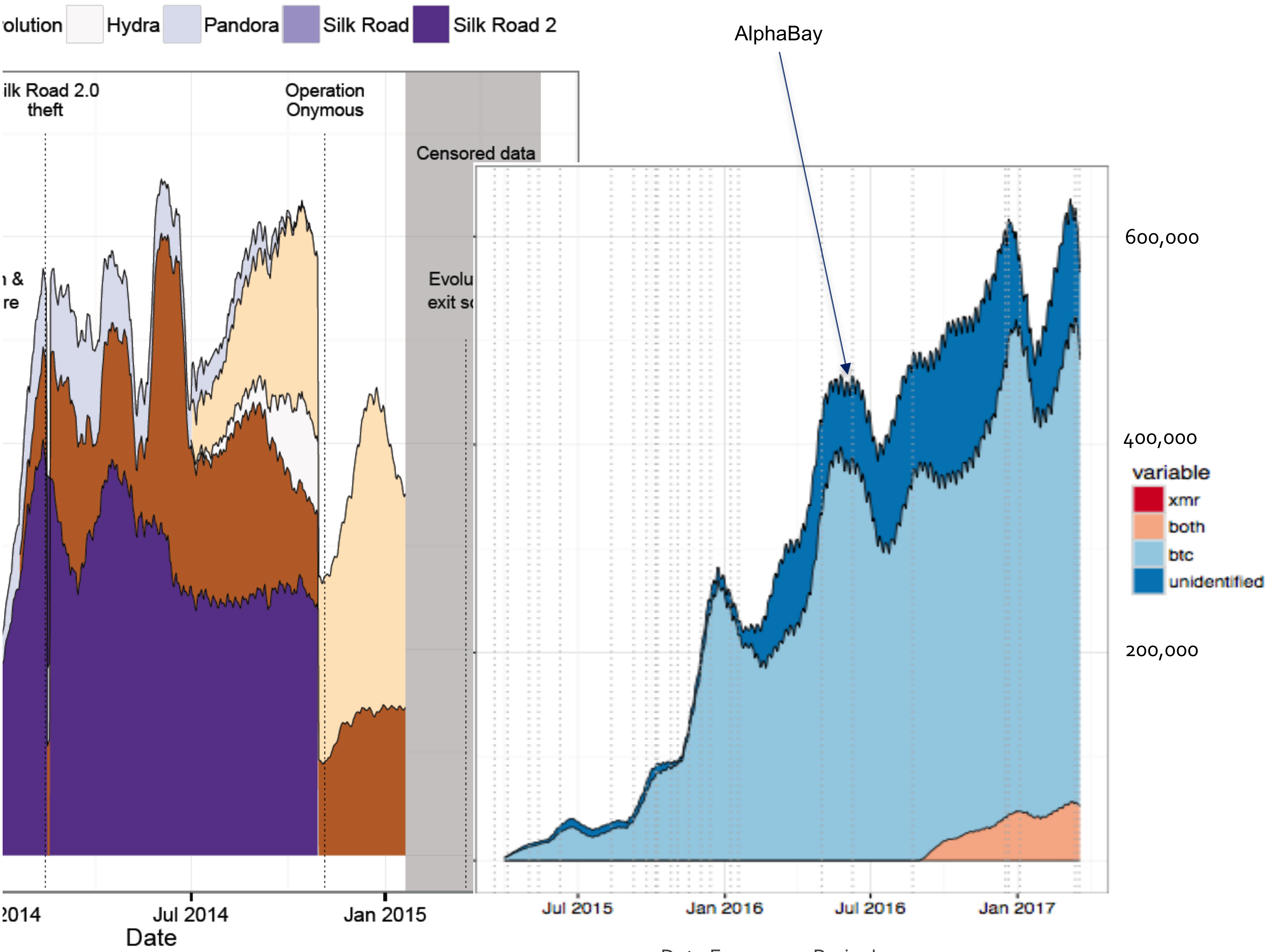


Agora
BMR
Evolution
Hydra
Pandora
Silk Road
Silk Road 2
Silk Road takedown
Silk Road 2.0 theft
Sheep scam & BMR closure
Operation Onymous
Censored data
Evolution exit scam
Daily volume (US dollars, 30−day avg.)
600,000
400,000
200,000
0
Jul 2013
Jan 2014
Jul 2014
Jan 2015
Jul 201
Date

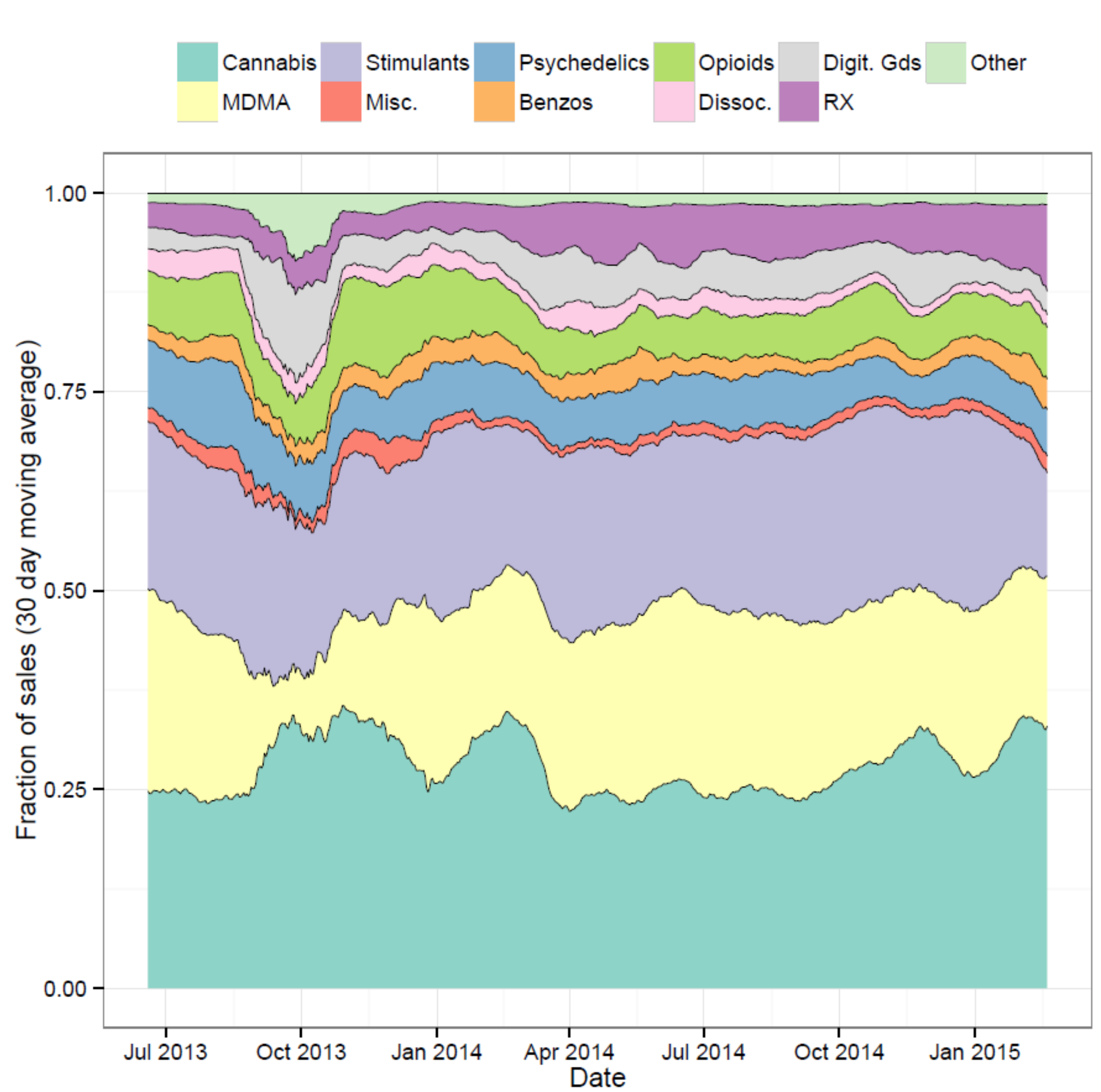


The ecosystem as a whole has been resilient to both takedowns and scams

Targeting marketplace operators seems (relatively) ineffective

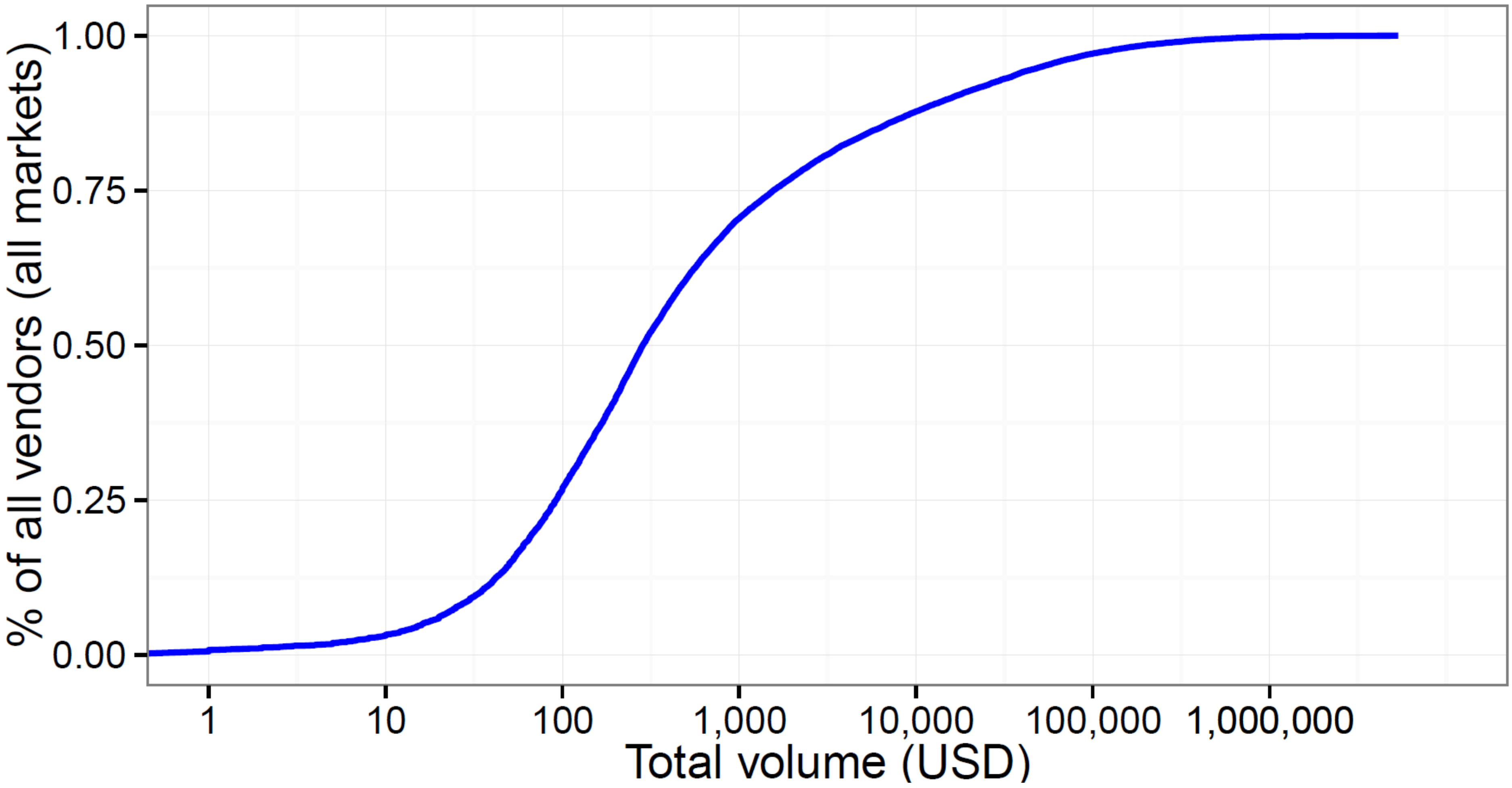

olution
Hydra
Pandora
Silk Road
Silk Road 2
ilk Road 2.0
theft
Operation
Onymous
Censored data
n &
re
Evolu
exit sc
2014
Jul 2014
Jan 2015
Date
AlphaBay
600,000
400,000
200,000
variable
xmr
both
btc
unidentified
Jul 2015
Jan 2016
Jul 2016
Jan 2017

Item sales per category

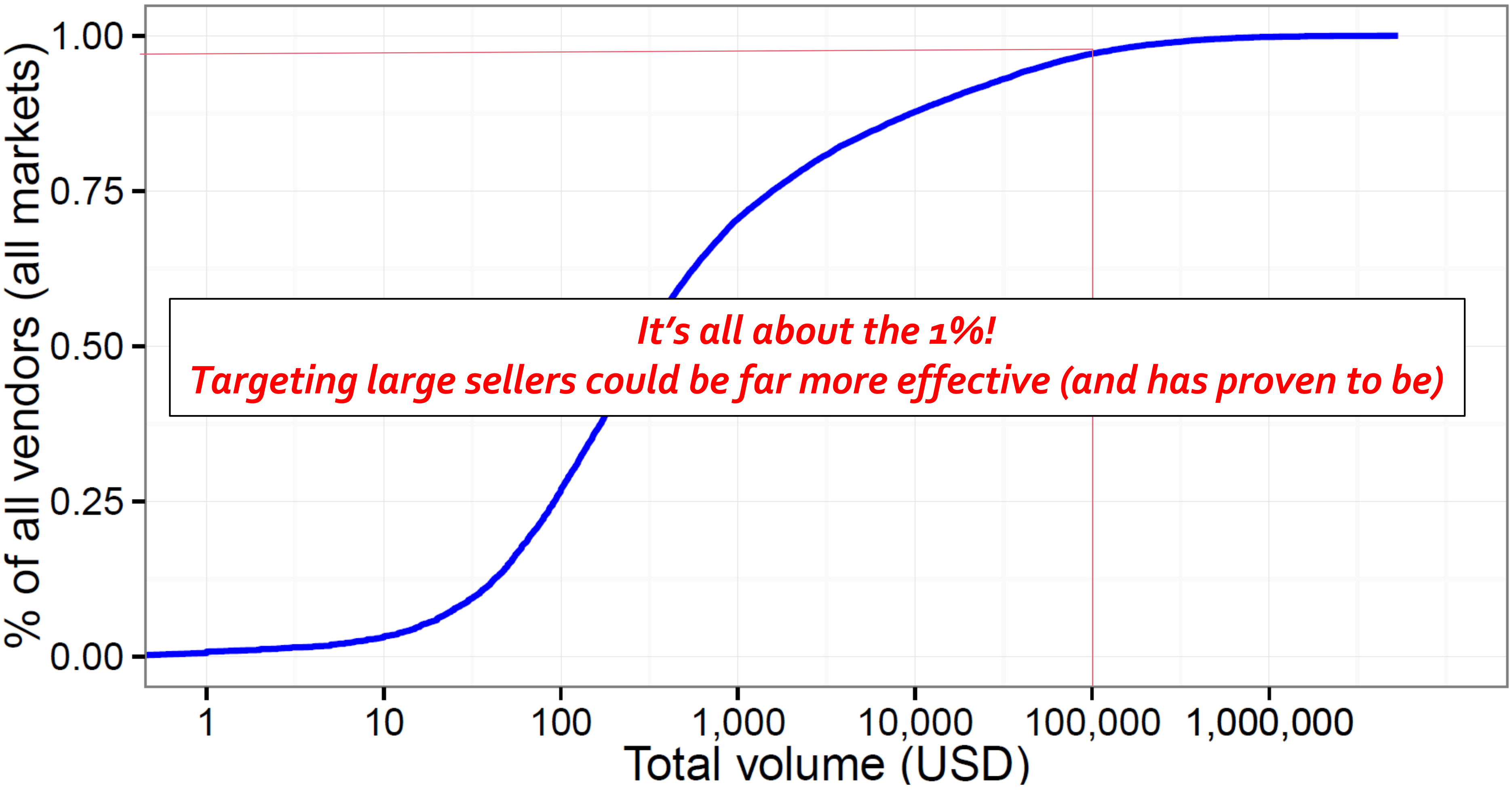


© Nicolas Christin

Vendor volumes



Vendor volumes



It's all about the 1%!

Targeting large sellers could be far more effective (and has proven to be)

Summary

- Collect and analyze data to understand attacker ecosystem and develop better defenses
  - **Development of a science of measurement**
  - Emergence of concentrations
    - Traffic brokers & production labs in pharma, large sellers in narcotics…
    - **Driven by economic properties**
    - Possible intervention points

- Ongoing/future work
  - Using our data to build descriptive (mathematical) models of interactions that can then be used to predict future behavior

Nicolas Christin
nicolasc@cmu.edu / @nc2y
https://www.andrew.cmu.edu/user/nicolasc