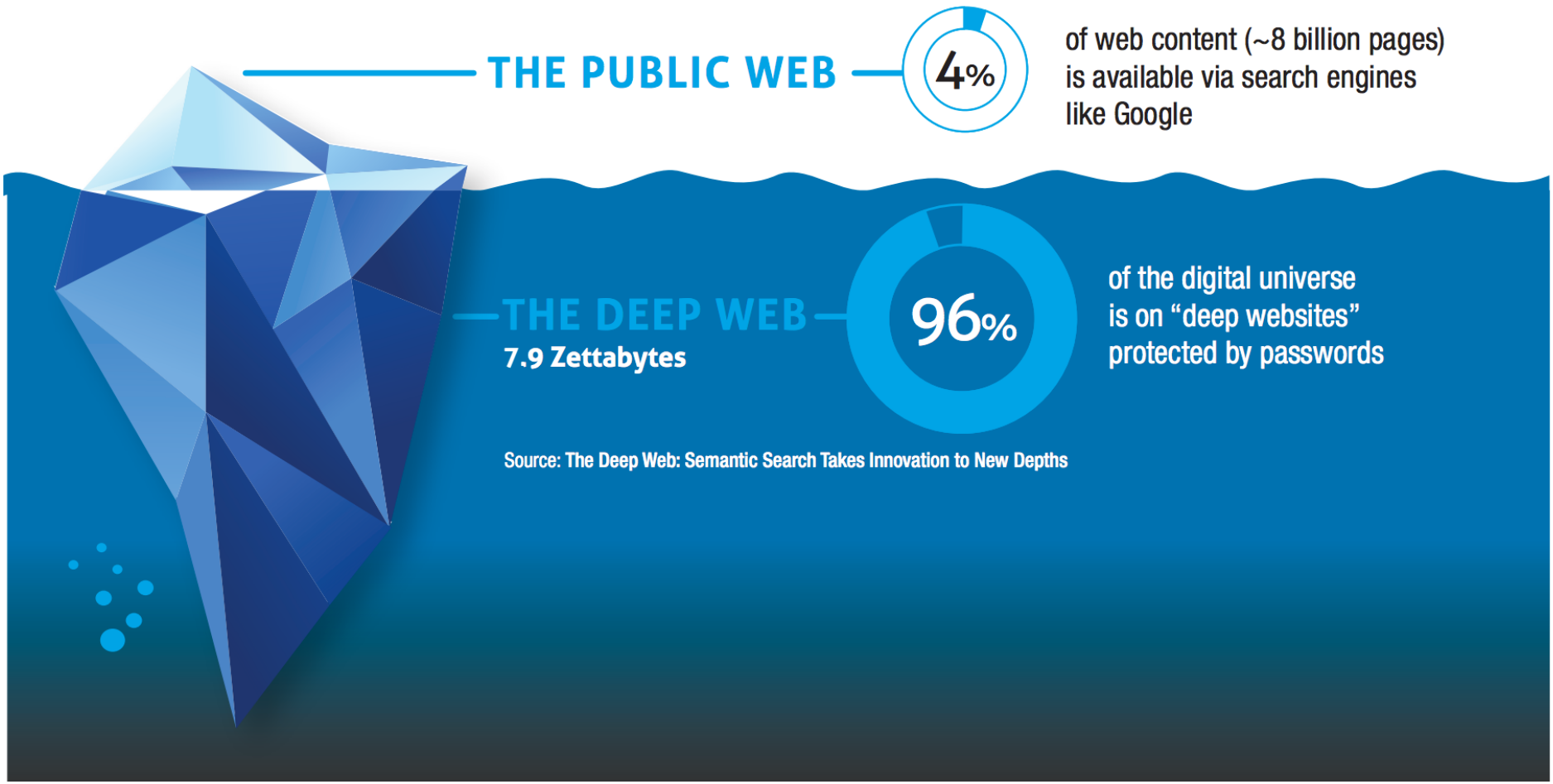
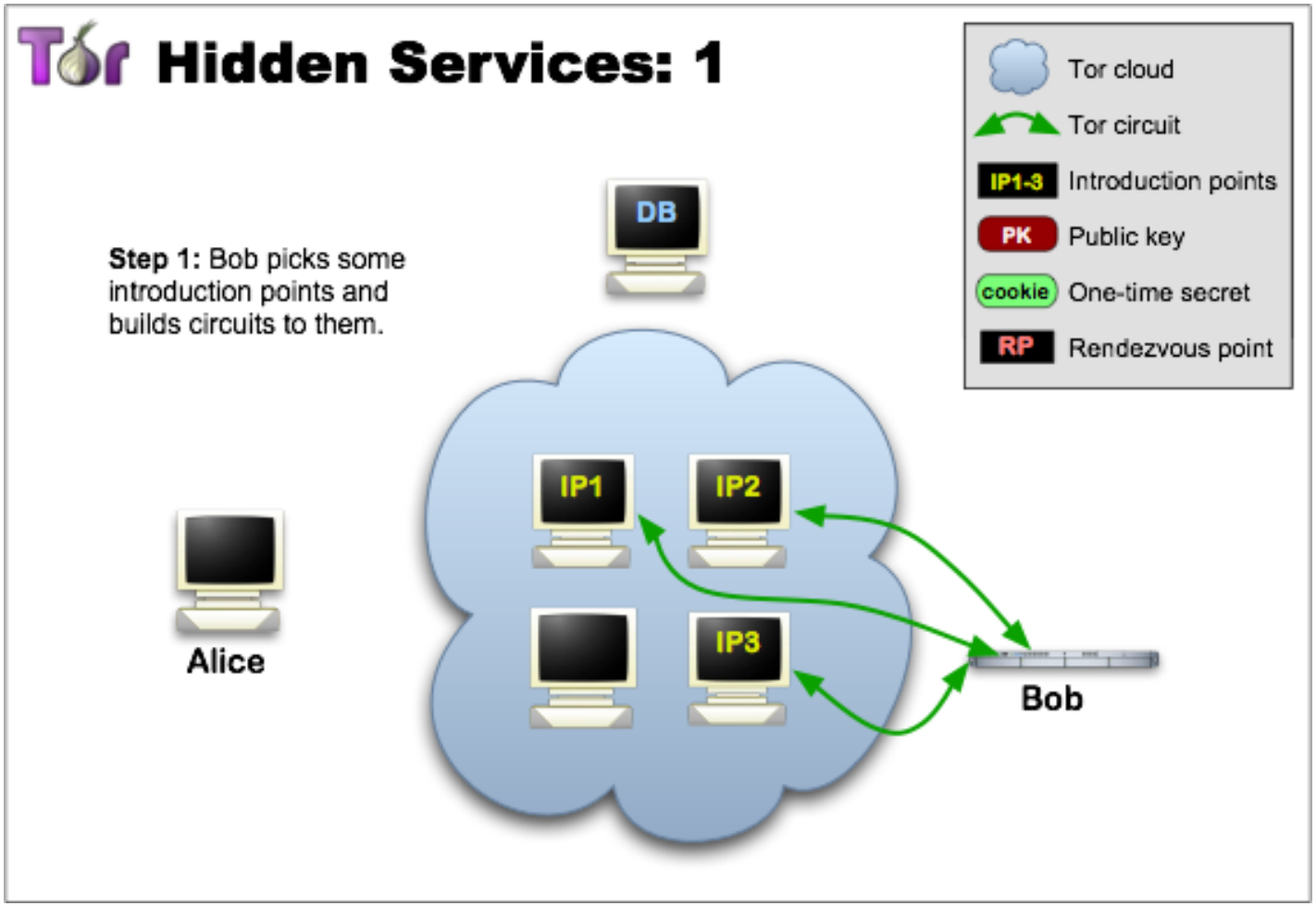
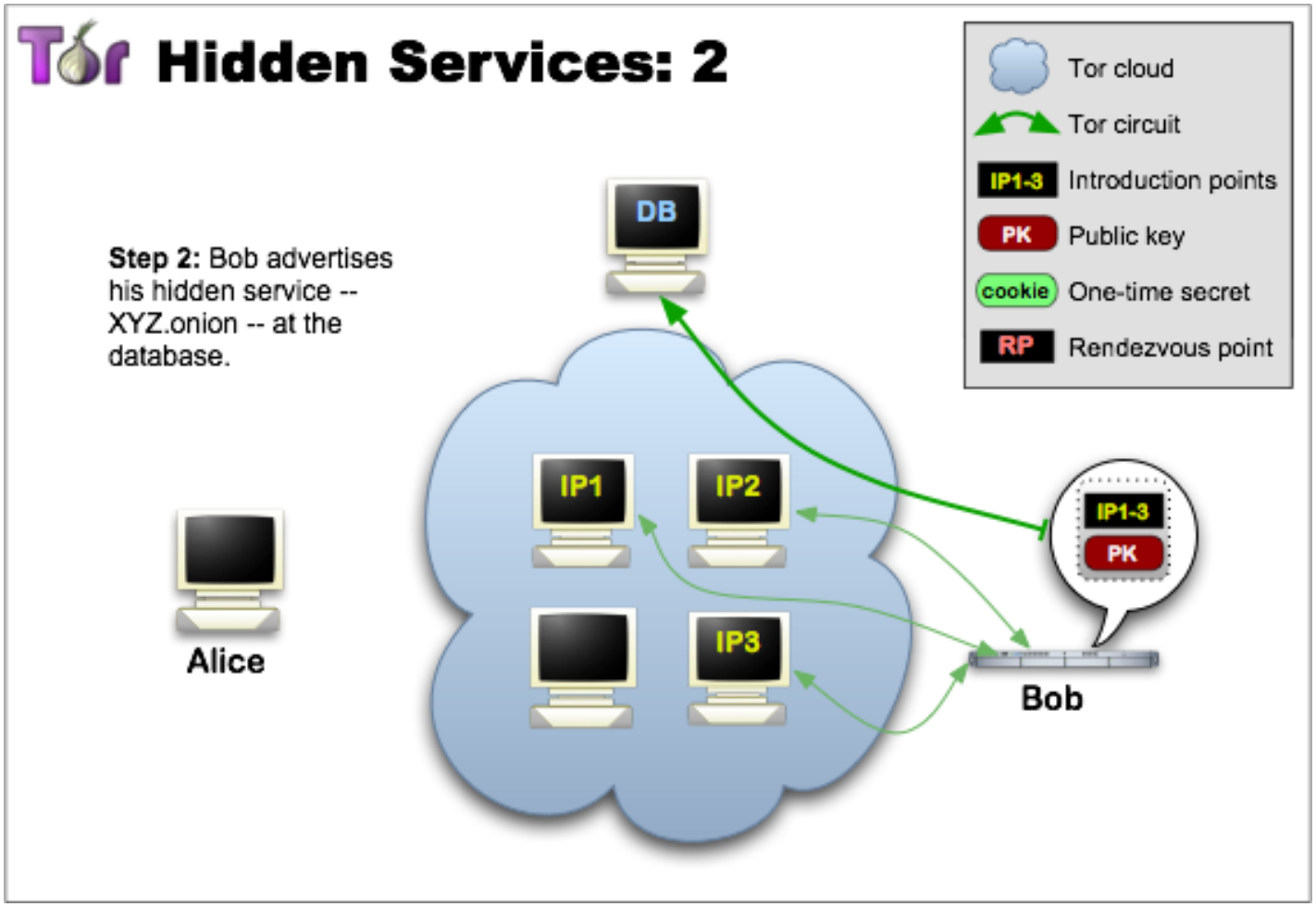
The Deep Web


THE PUBLIC WEB
4%
of web content (~8 billion pages) is available via search engines like Google
THE DEEP WEB
7.9 Zettabytes
96%
of the digital universe is on "deep websites" protected by passwords
Source: The Deep Web: Semantic Search Takes Innovation to New Depths

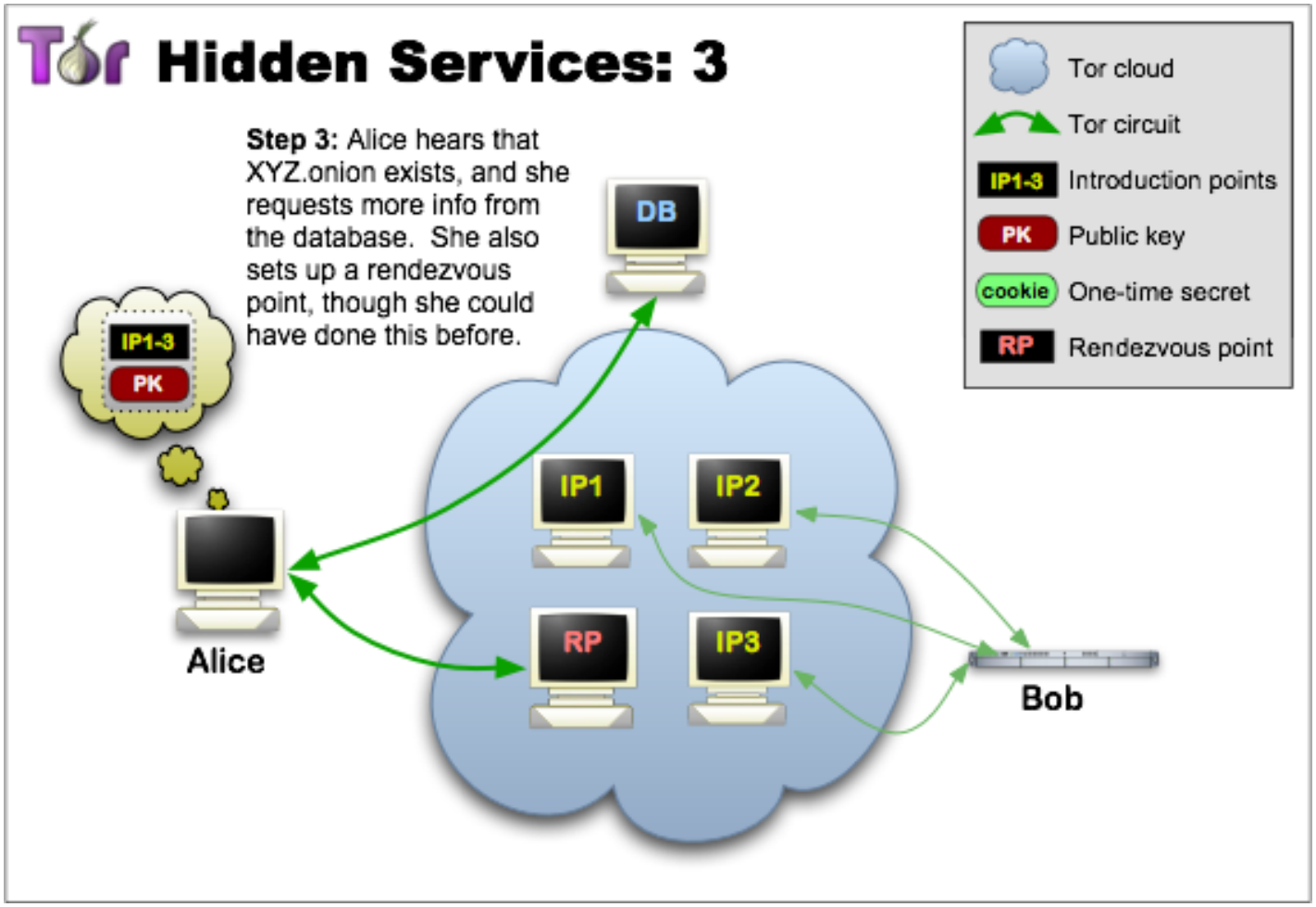
Tor Hidden Services: Overview



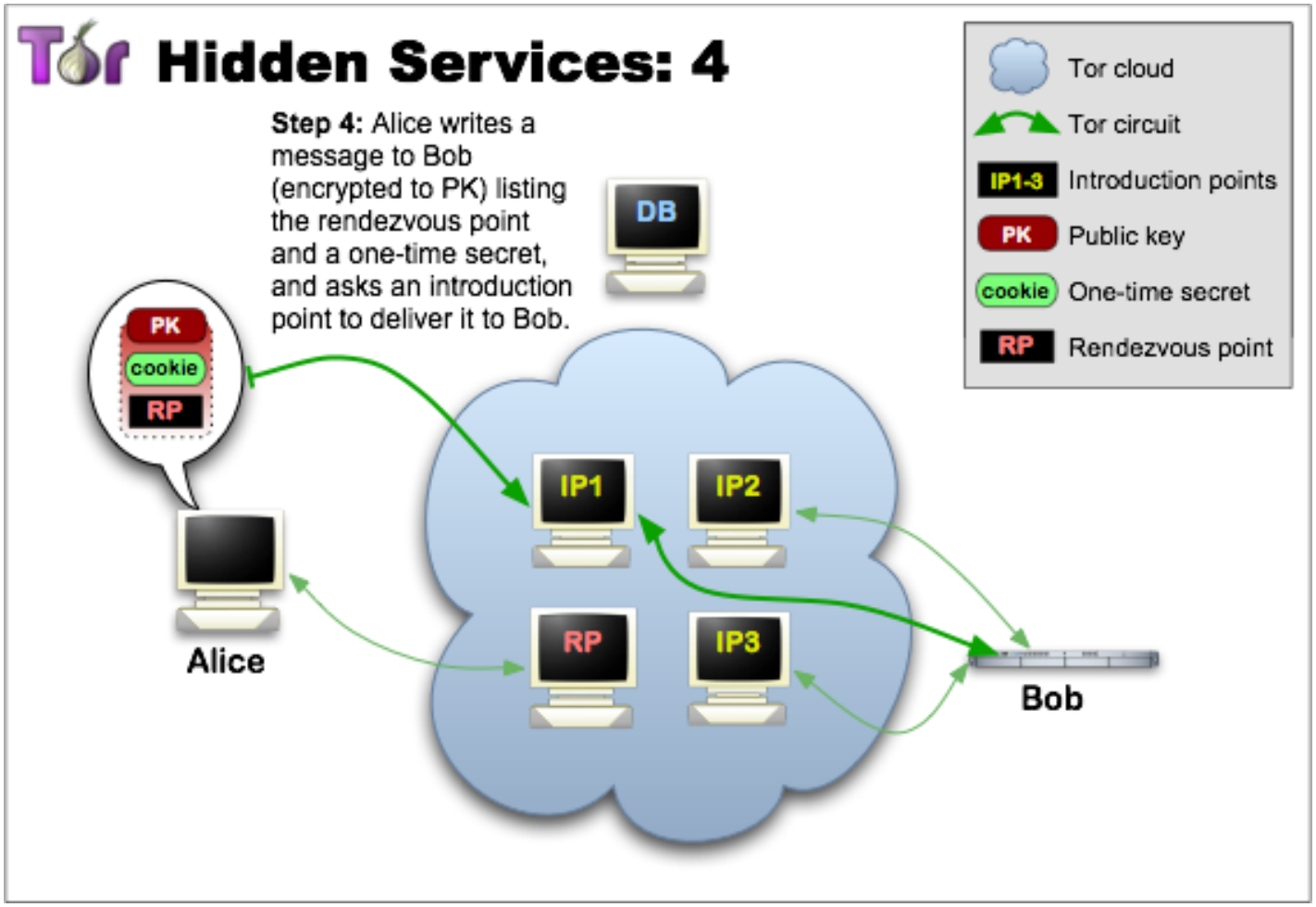
Tor Hidden Services



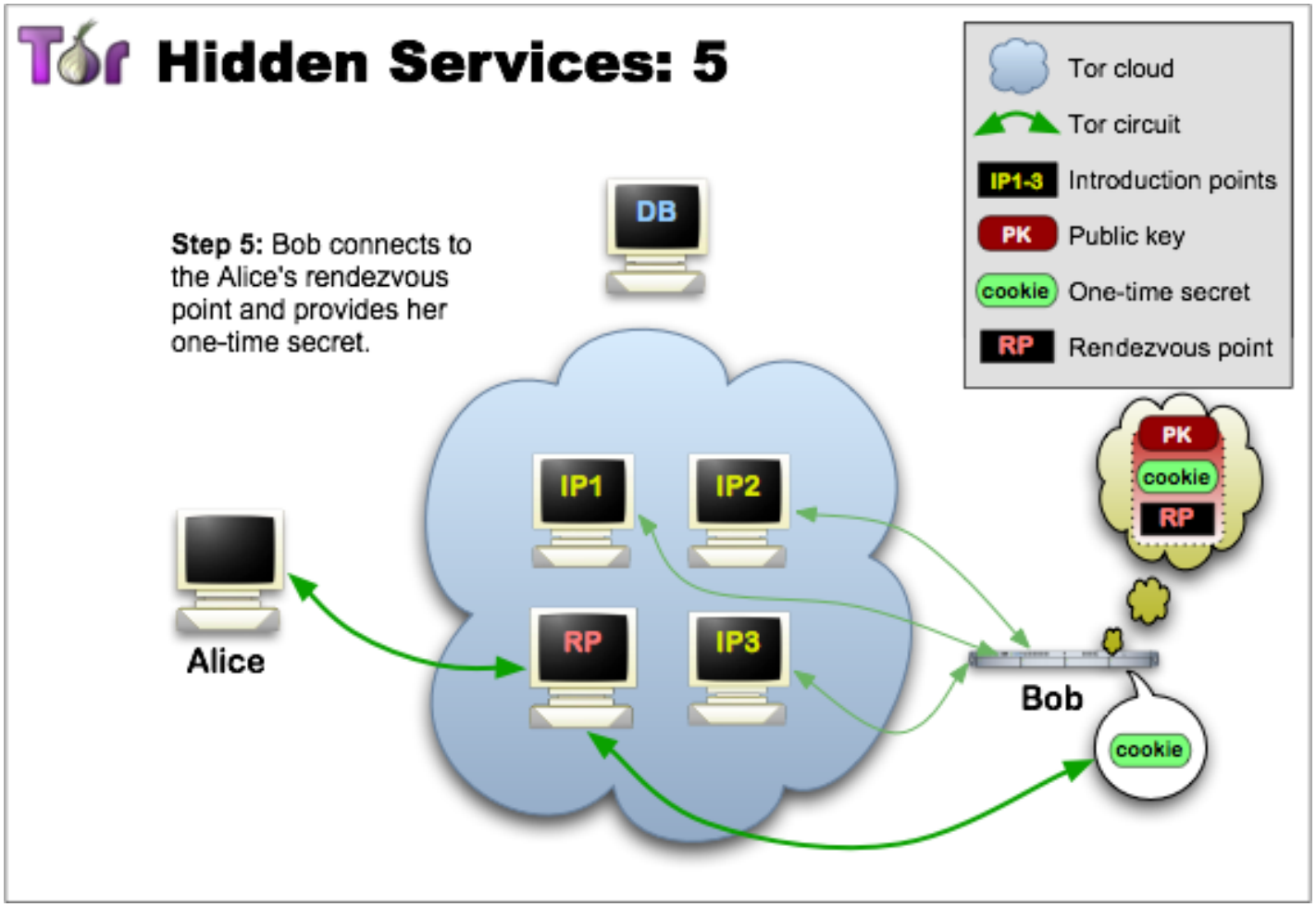
Tor Hidden Services



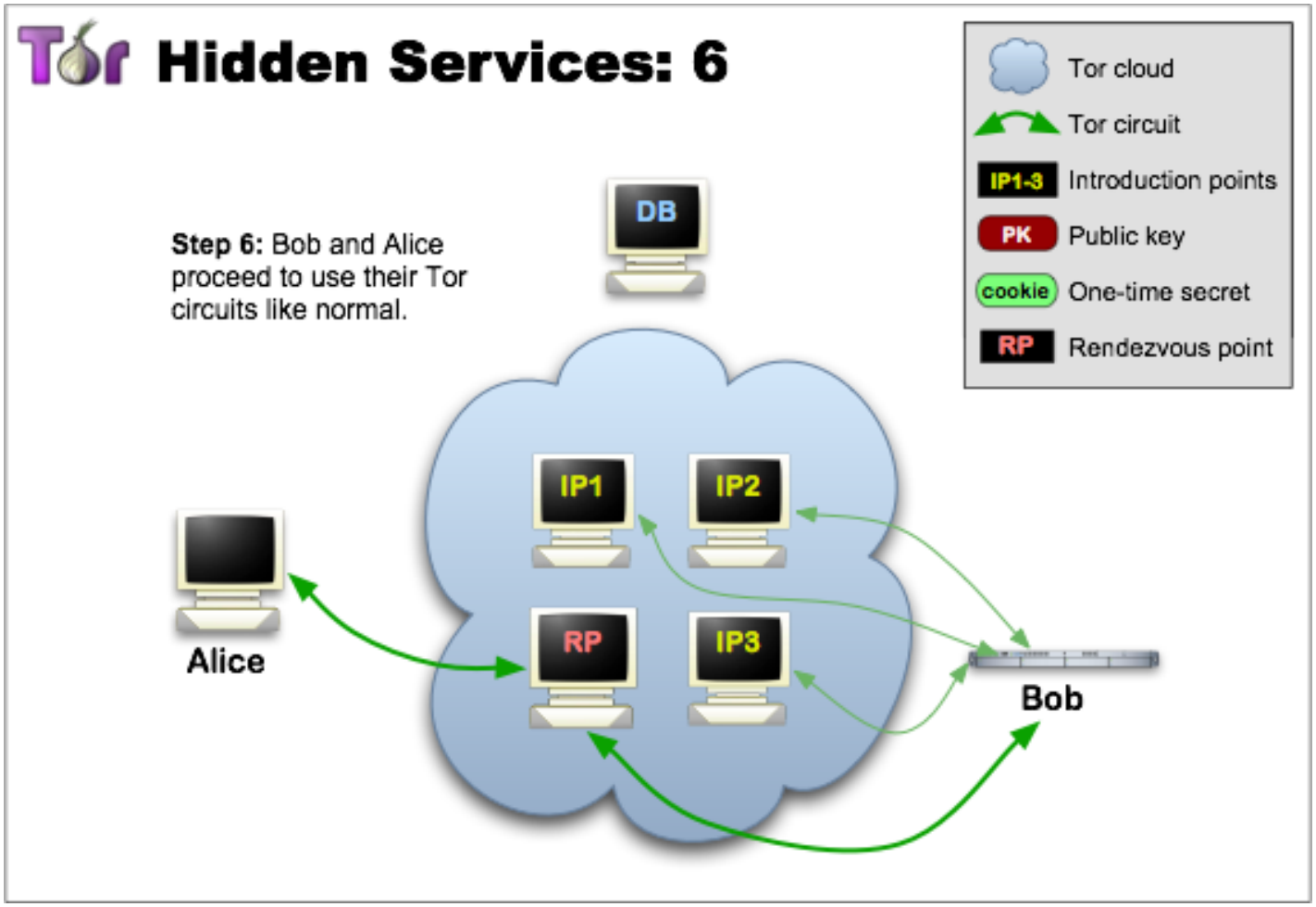
Tor Hidden Services



Tor Hidden Services



Tor Hidden Services



SilkRoad Marketplace

Silk Road
anonymous marketplace

Welcome **now0pen**!
messages(0) | **orders(0)** | **account(฿0)** | **settings** | **log out**

search | (0)

Shop by category:
- Drugs(752)
  - Cannabis(280)
  - Ecstasy(35)
  - Dissociatives(11)
  - Psychedelics(84)
  - Opioids(62)
  - Stimulants(53)
  - Other(107)
  - Benzos(70)
- Lab Supplies(6)
- Digital goods(98)
- Services(48)
- Money(55)
- Weaponry(15)
- Home & Garden(14)
- Food(4)
- Electronics(5)
- Books(49)
- Drug paraphernalia(28)
- XXX(30)
- Medical(3)
- Computer equipment(4)
- Apparel(4)
- Musical instruments(2)
- Tickets(1)
- Forgeries(13)


5 Marijuana Butter Chocolate Chip...
฿8.53

4mg. TIZANIDINE (zanaflex) x25
฿2.09


US customers only Express...
฿2.79

4 x 20MG Original Lily Cialis
฿7.85


(1g) High-grade Crystal Meth
฿11.95


MindFood - Protect your brain!...
฿3.69


to US 1/4 lb (qp) BC Master Kush...
฿121.37


How to Grow Mushrooms
฿0.14


Mushroom Indoor Growing - Easy...
฿0.29

News:
- Escrow hedging **update**
- New feature to help protect **sellers**
- We are **hiring**! Get paid for a referral, too...
- Reclaim lost coins from **MyBitcoin.com**
- Seller ranking and feedback **overhaul**
- Change your Mt. Gox **password**

recent feedback:

SilkRoad Marketplace

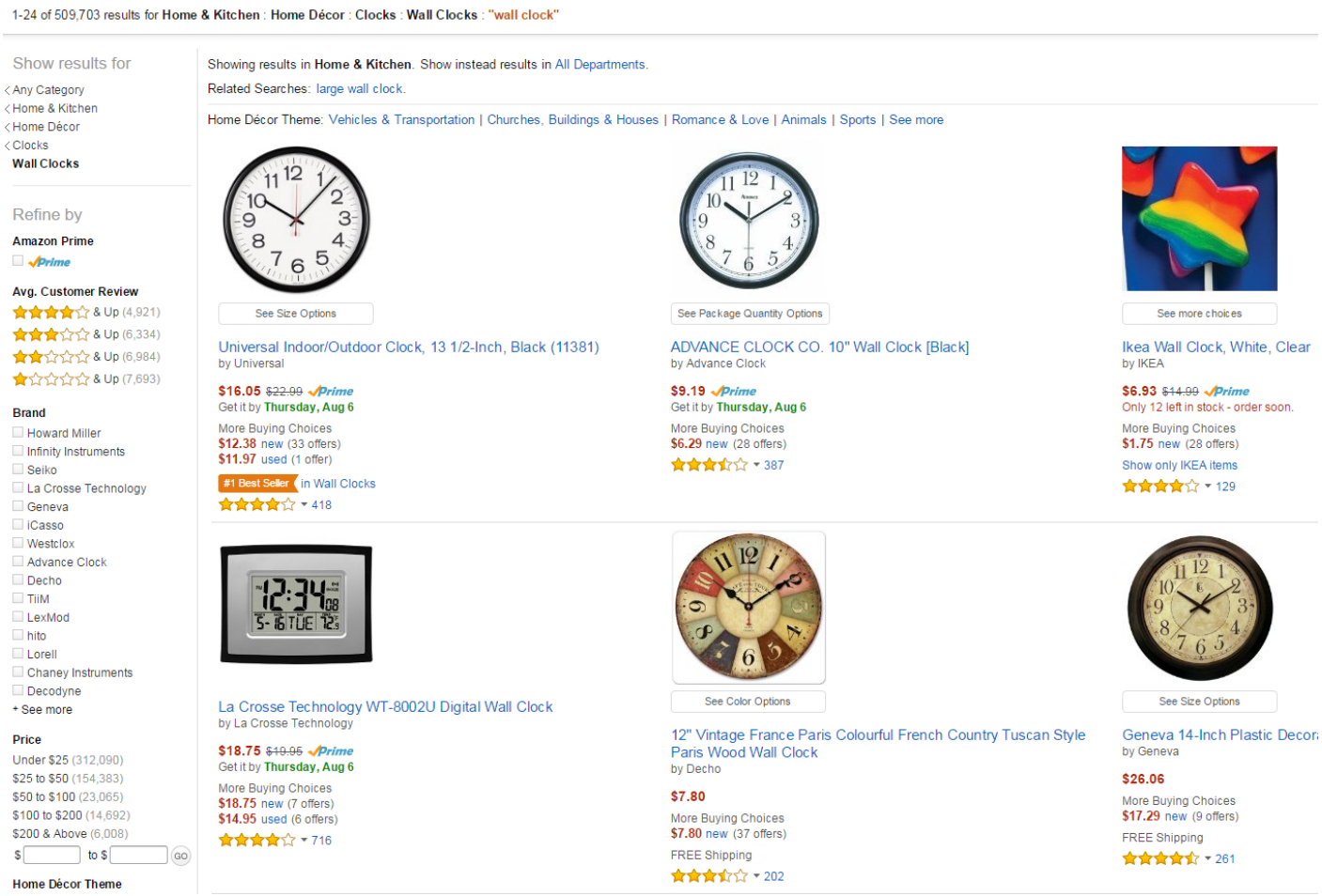
Measuring the Longitudinal Evolution of the Online Anonymous Marketplace Ecosystem

Kyle Soska
Carnegie Mellon University
ECE / Cylab
ksoska@cmu.edu

Nicolas Christin
Carnegie Mellon University
ECE / Cylab
nicolasc@cmu.edu

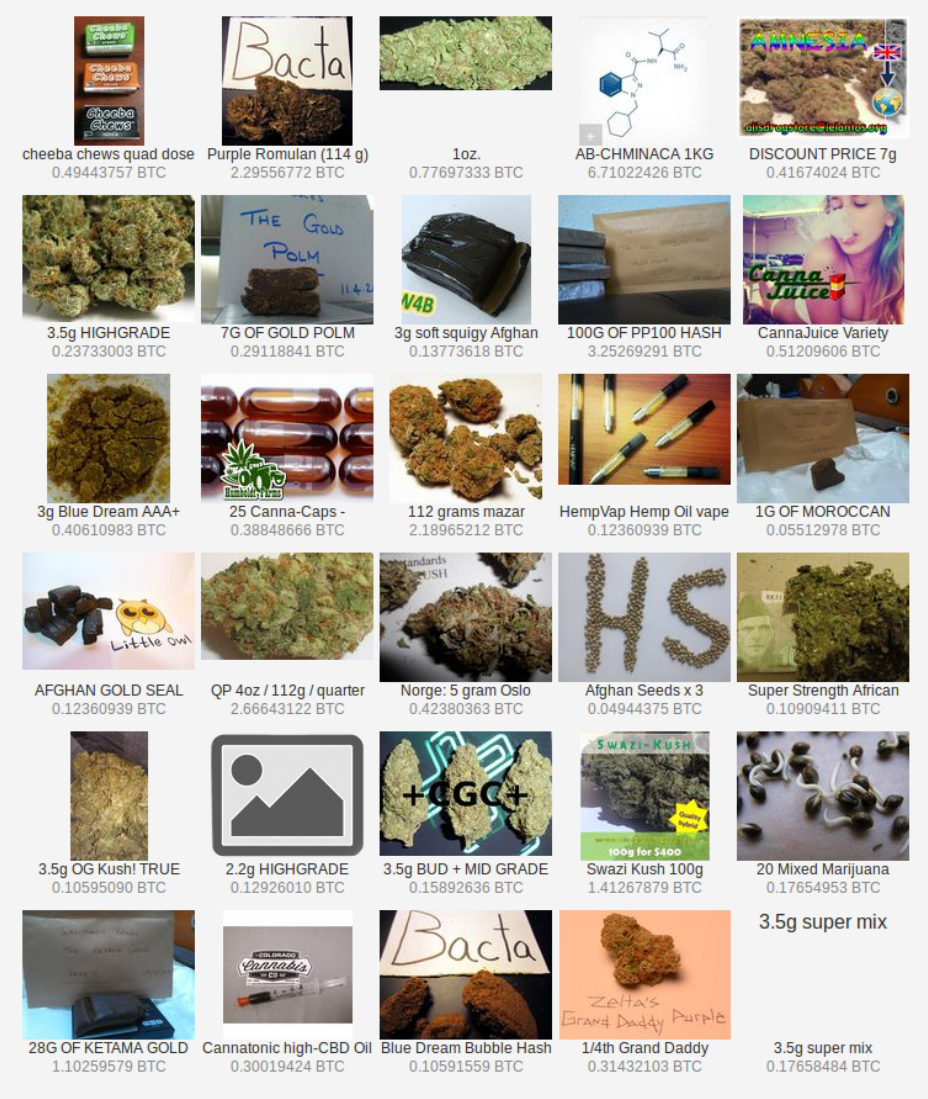
Conventional Commerce

Internet Commerce



Conventional Illicit Commerce

Illicit Internet Commerce



Anonymous Marketplaces

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

Anonymous Marketplace Technology

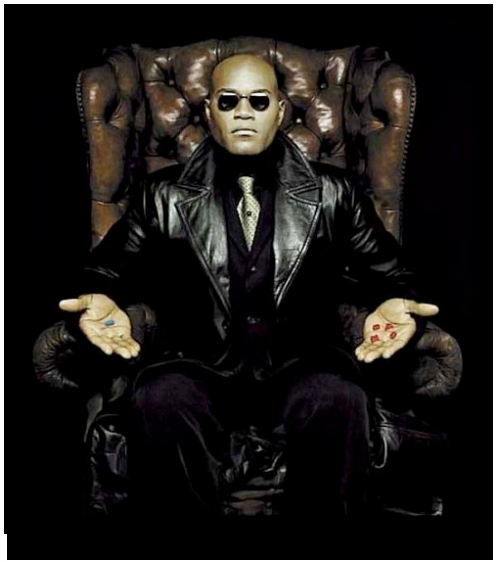
- Hidden Website (Tor Hidden Service, I2P)
  - Customers
    - No cost of creation
    - No information needed
  - Vendors
    - Vendor bonds required
    - Often invite only
    - Public feedback history

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

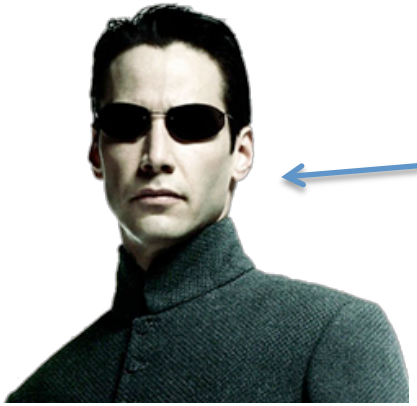
- Hidden Messages(PGP)

Market Transactions

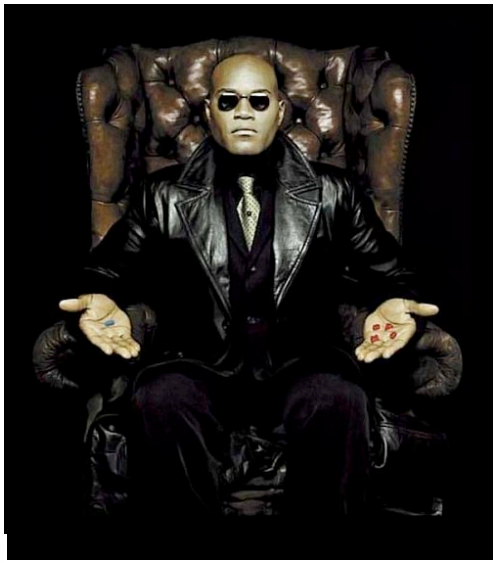
"I'll take the red pill"



Market Transactions

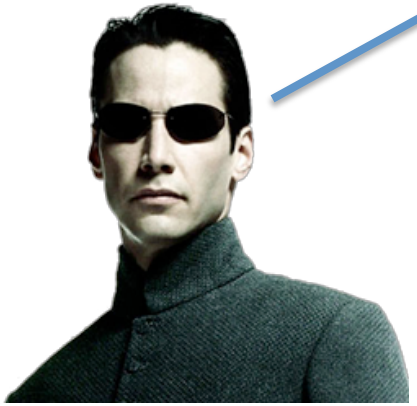


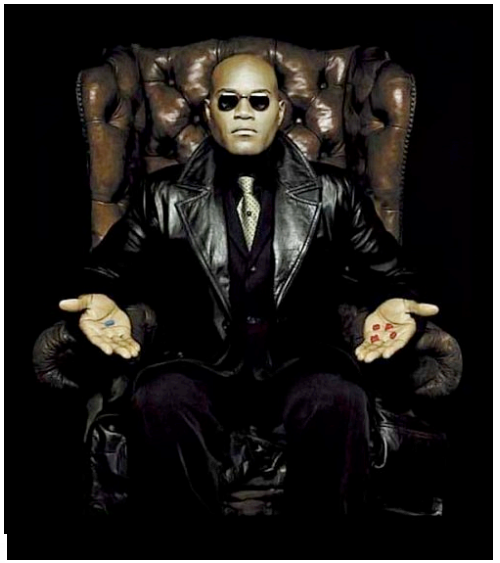
"1 BTC please"



Market Transactions

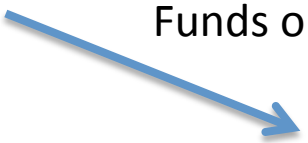
Deposit 1 BTC

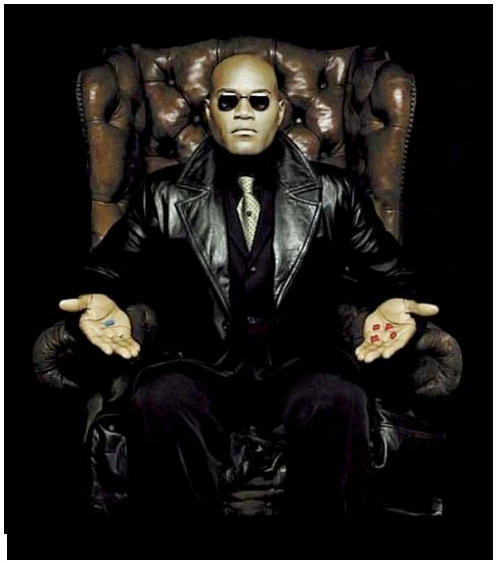




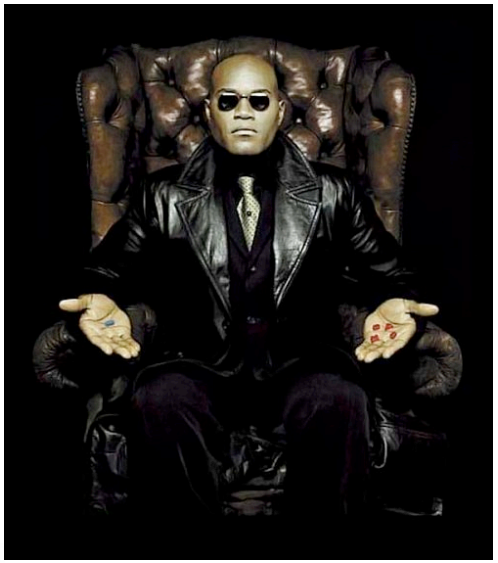
Market Transactions

Funds ok

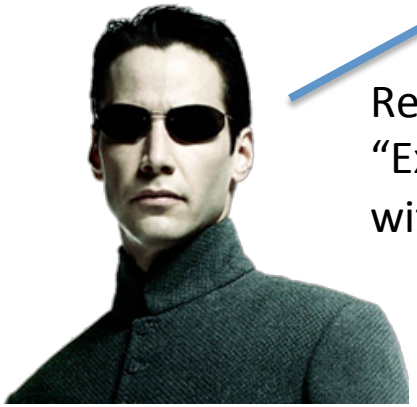




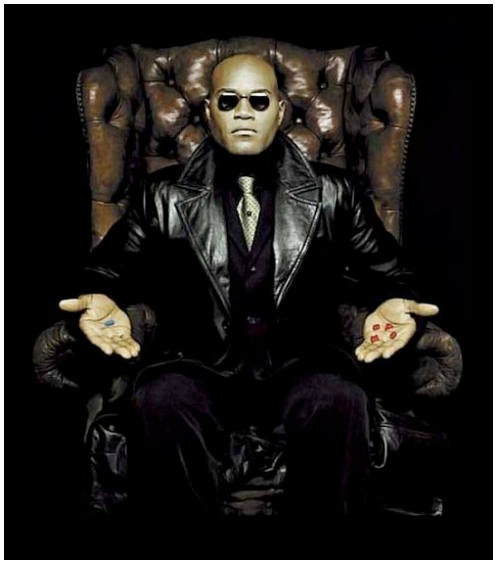
Market Transactions



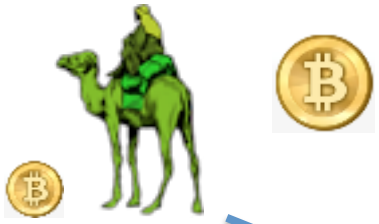
Market Transactions



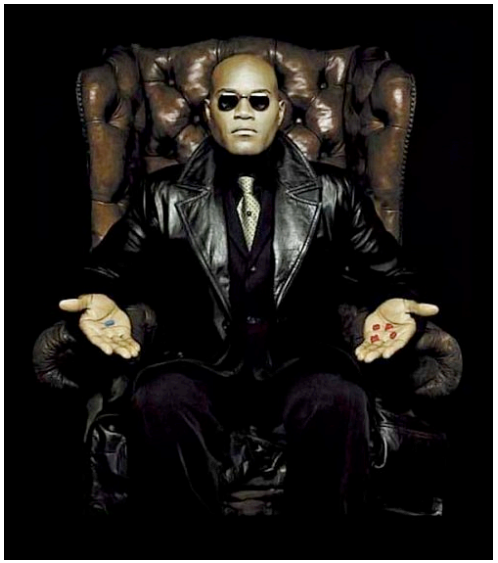
Received
“Excellent seller, would do business with again. A++++”



Market Transactions



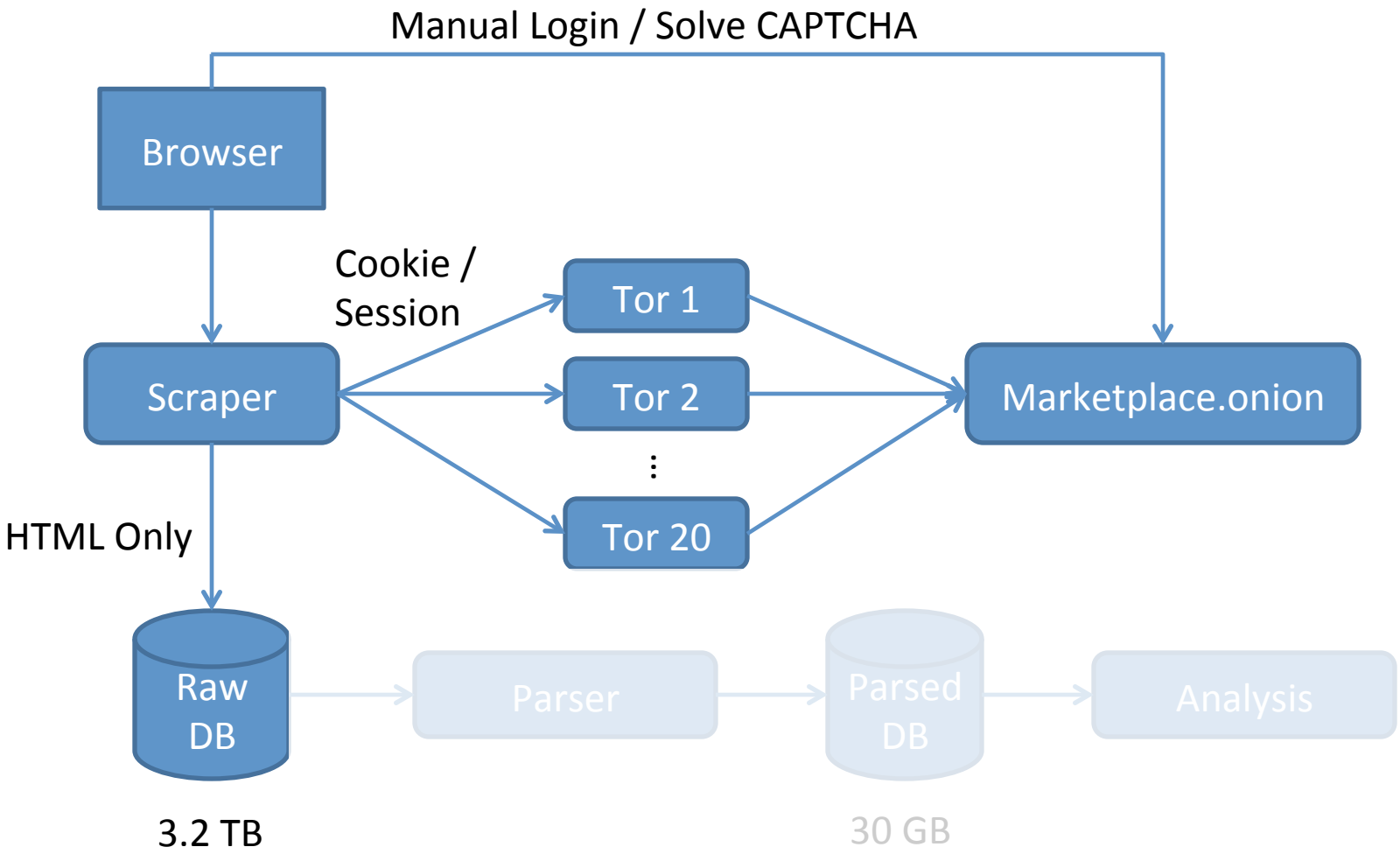
Deposit 0.9 BTC



Questions

- How much is being sold?
- What is being sold?
- How many vendors are relevant?
- What do vendors sell?

Measurement Platform Overview



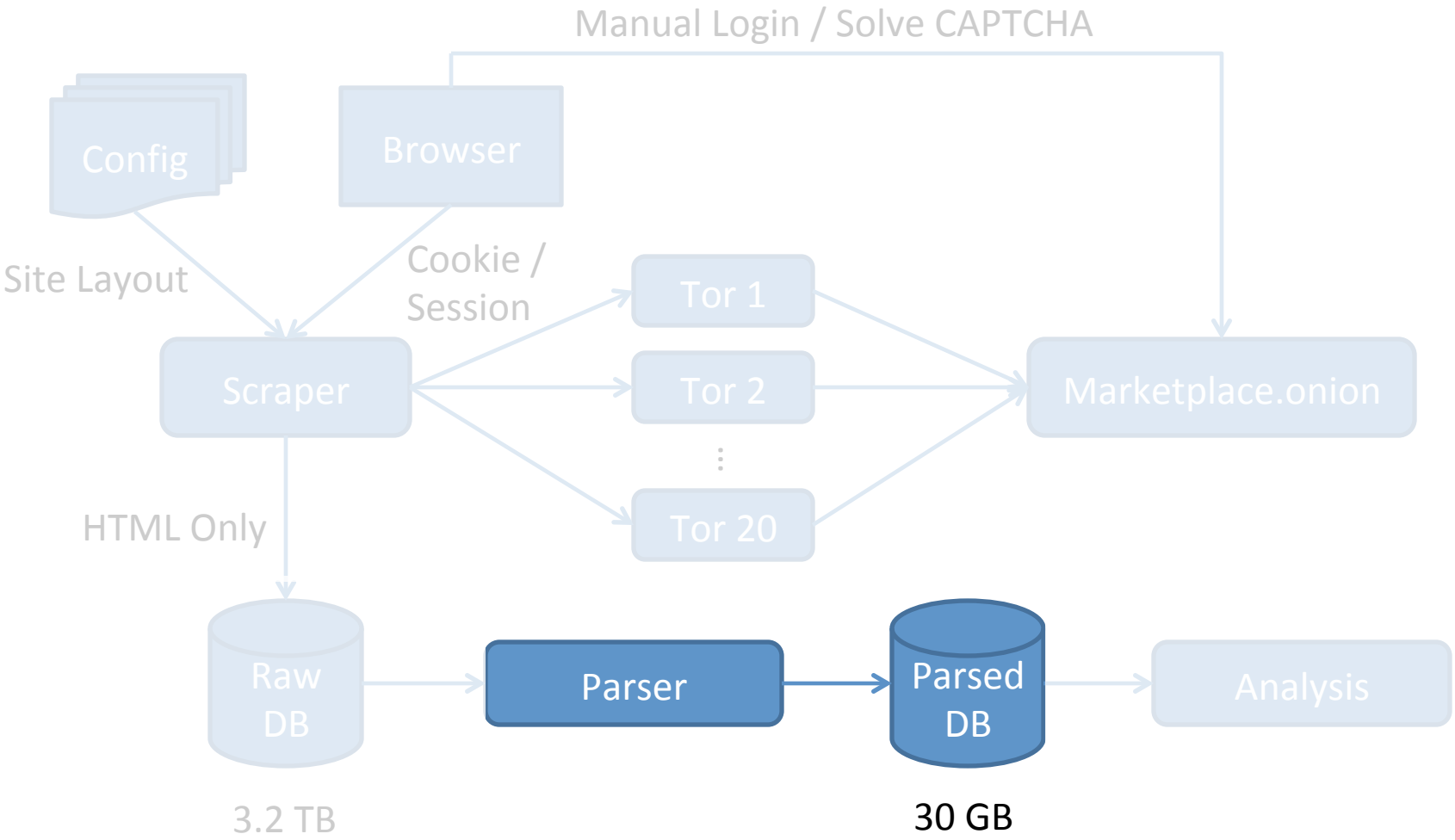
Measurements

- **Stealth**
  - Indistinguishable from real user
  - Random delays, scrape slowly
  - Popular User Agent
  - Browse website “normally”
- **Complete and instantaneous**
  - Dynamic marketplace, moving target
  - Scrape quickly
  - Site availability as low as 70%

Measurements

- **Anti-Scraping Encountered**
  - Rate Limits
  - Cookie Timeout
  - User Account Suspension
- **Totals**
  - 35 Marketplaces 1,908 scrapes total – 3.2 TB
  - 27 – 331,691 pages per scrape
  - 11/22/11 – present

Parsing


Manual Login / Solve CAPTCHA
Config
Browser
Site Layout
Cookie / Session
Tor 1
Tor 2
Tor 20
Scraper
Marketplace.onion
HTML Only
Raw DB
Parser
Parsed DB
Analysis
3.2 TB
30 GB

Silk Road Available Data

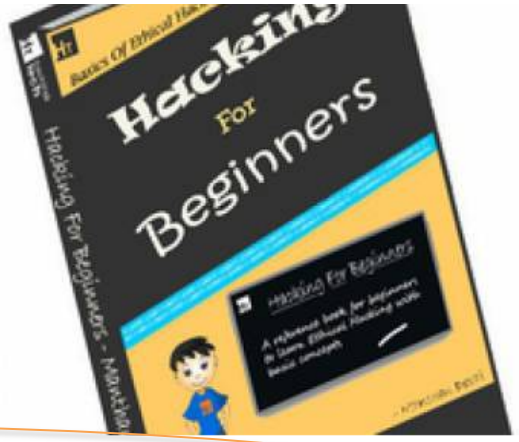
Books

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

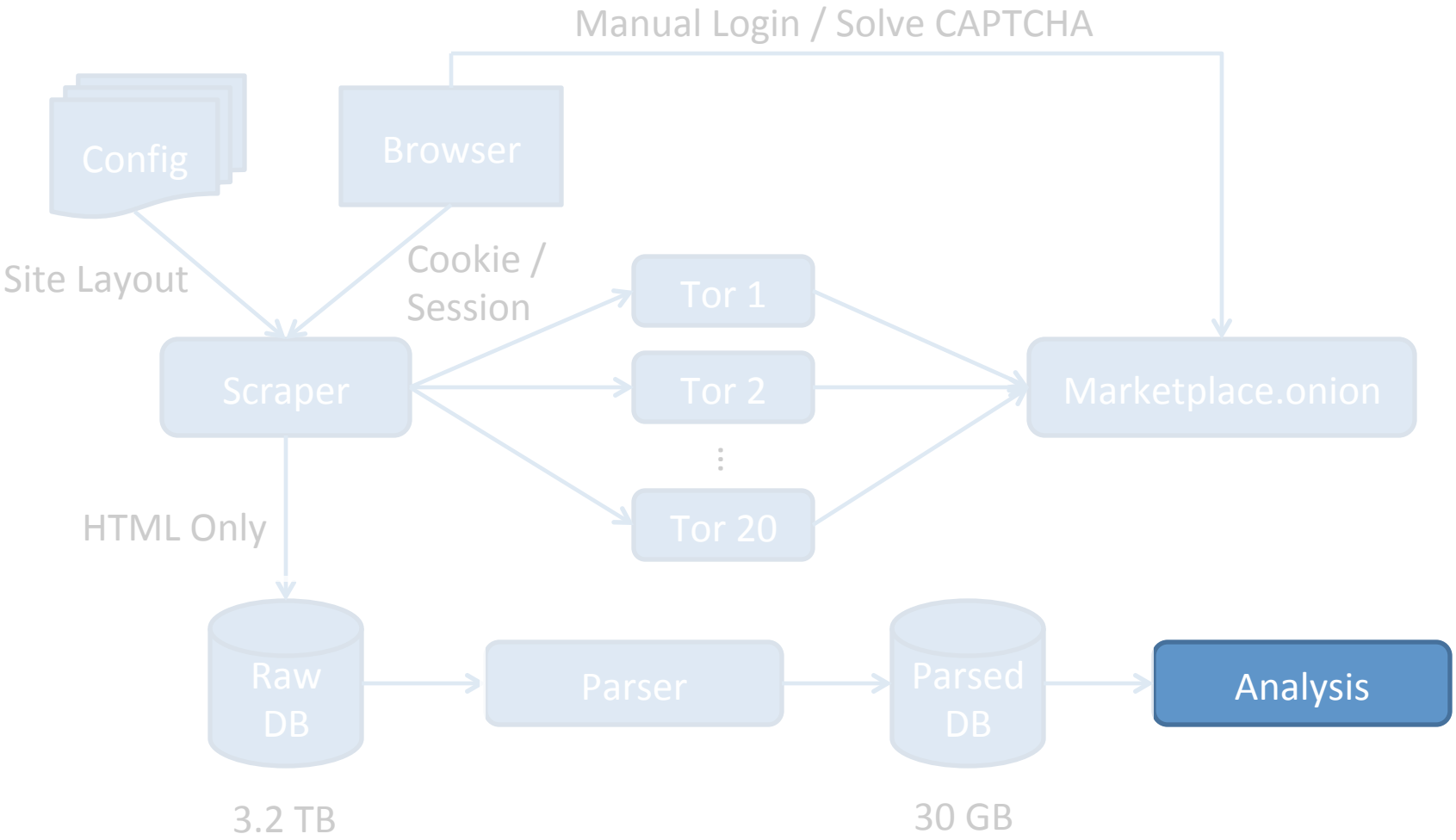
Recent feedback

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

Analysis


Manual Login / Solve CAPTCHA
Config
Browser
Site Layout
Cookie / Session
Tor 1
Tor 2
Tor 20
Scraper
Marketplace.onion
HTML Only
Raw DB
Parser
Parsed DB
Analysis
3.2 TB
30 GB

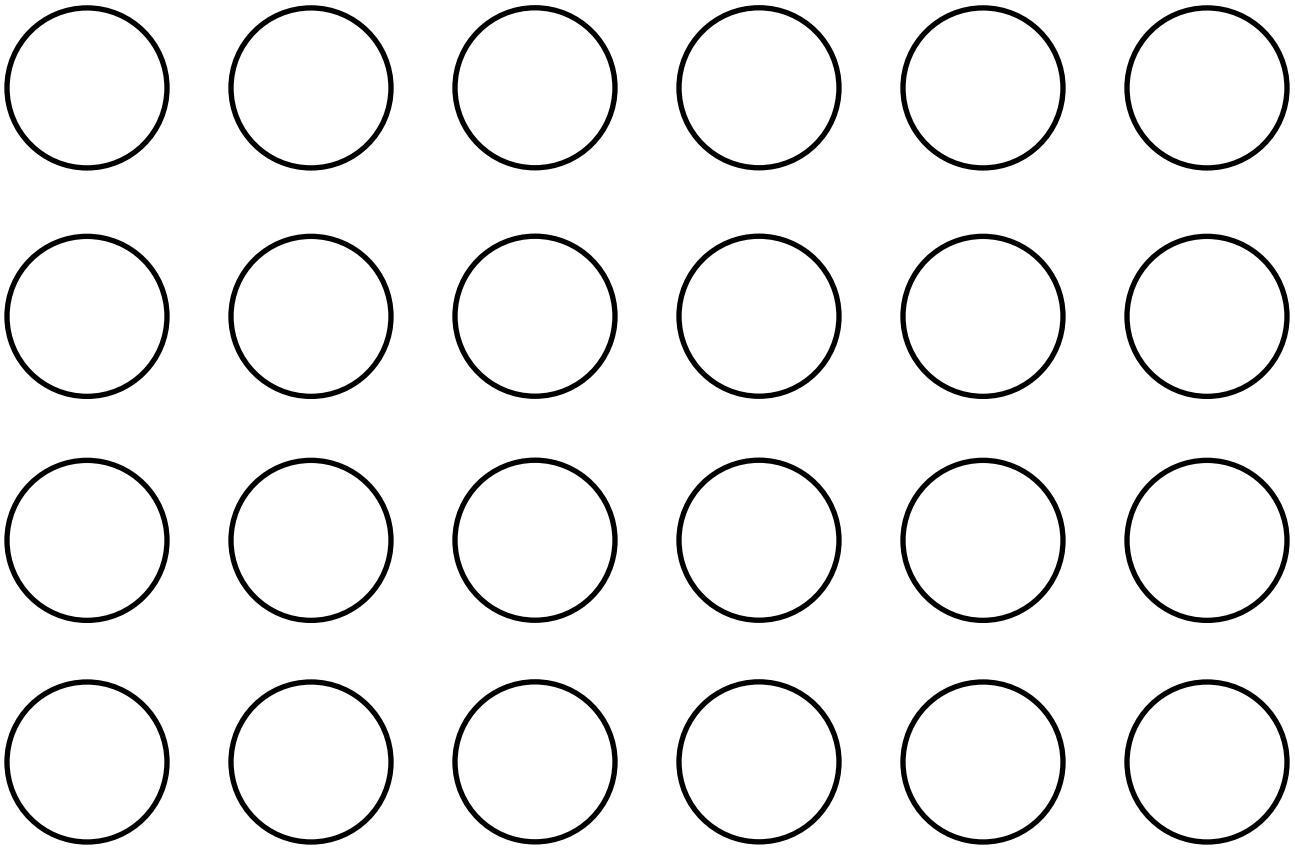
Data Completeness

- **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

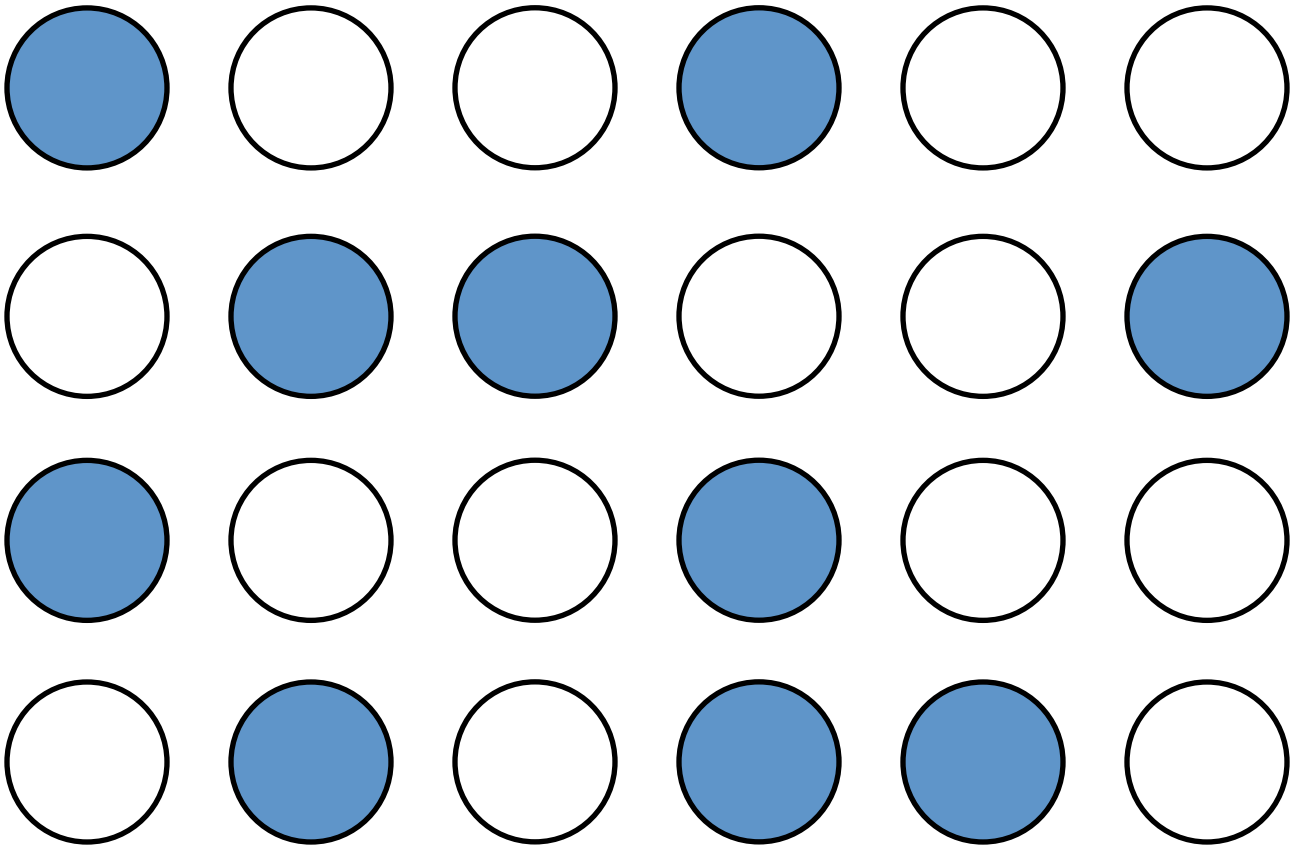
Mark and Recapture

Population Size = 24



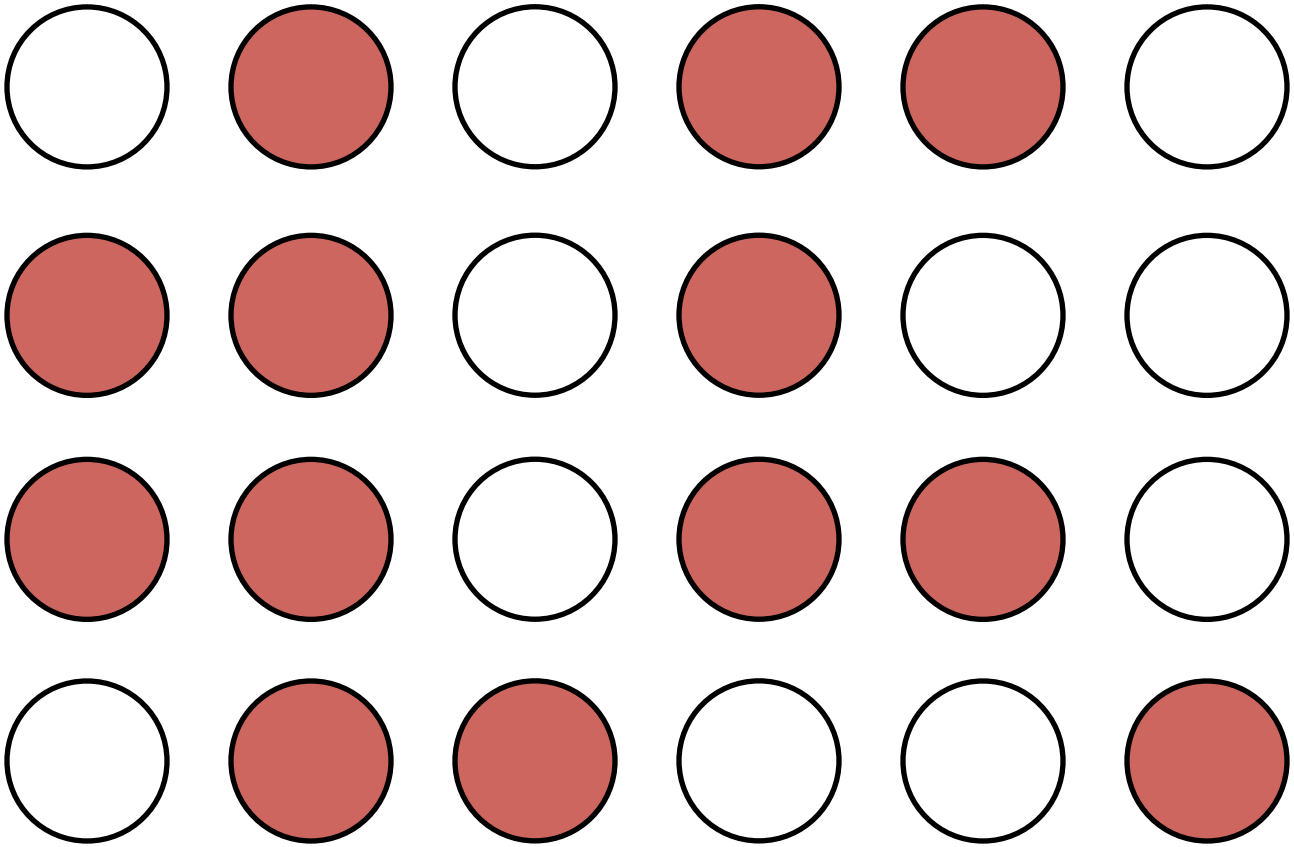
Mark and Recapture

Sample Size = 10



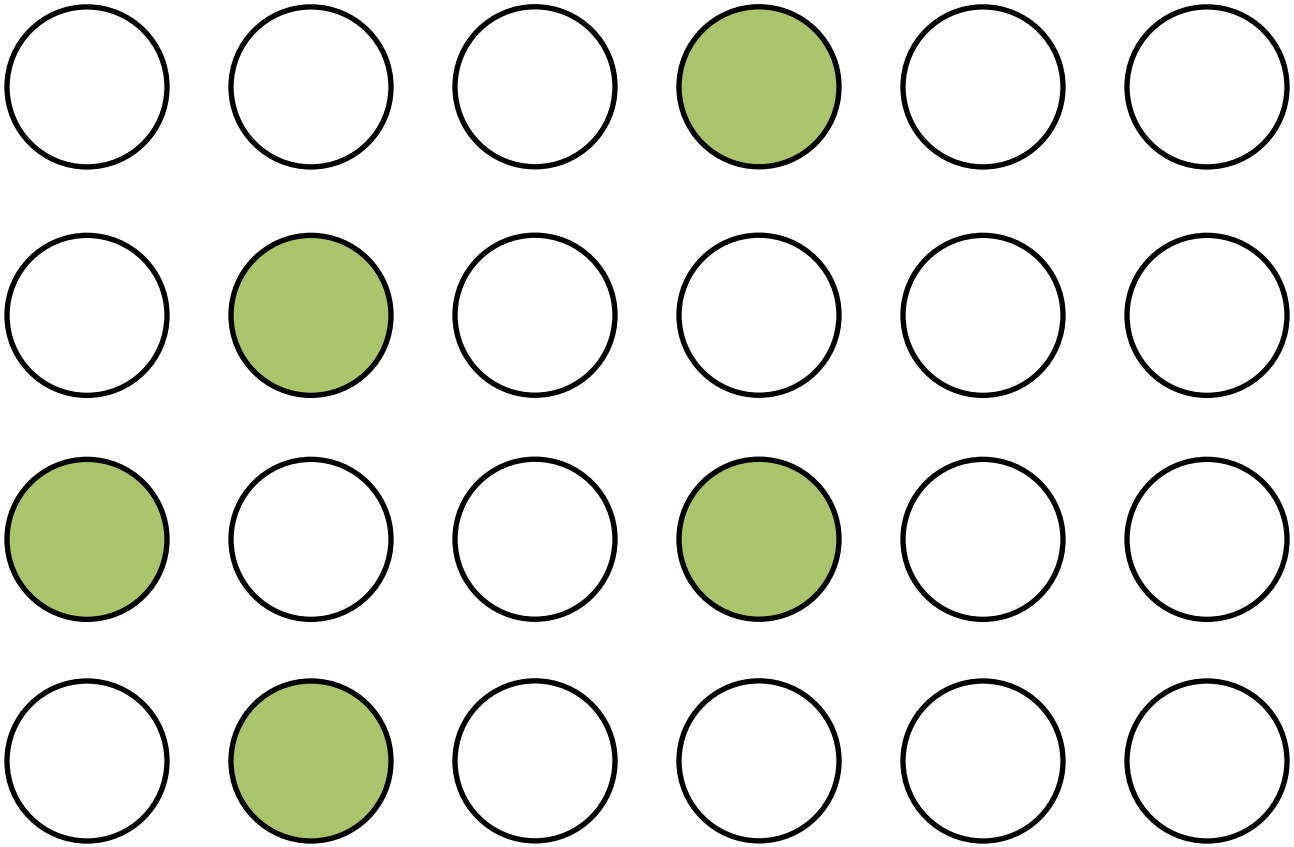
Mark and Recapture

Sample Size = 13

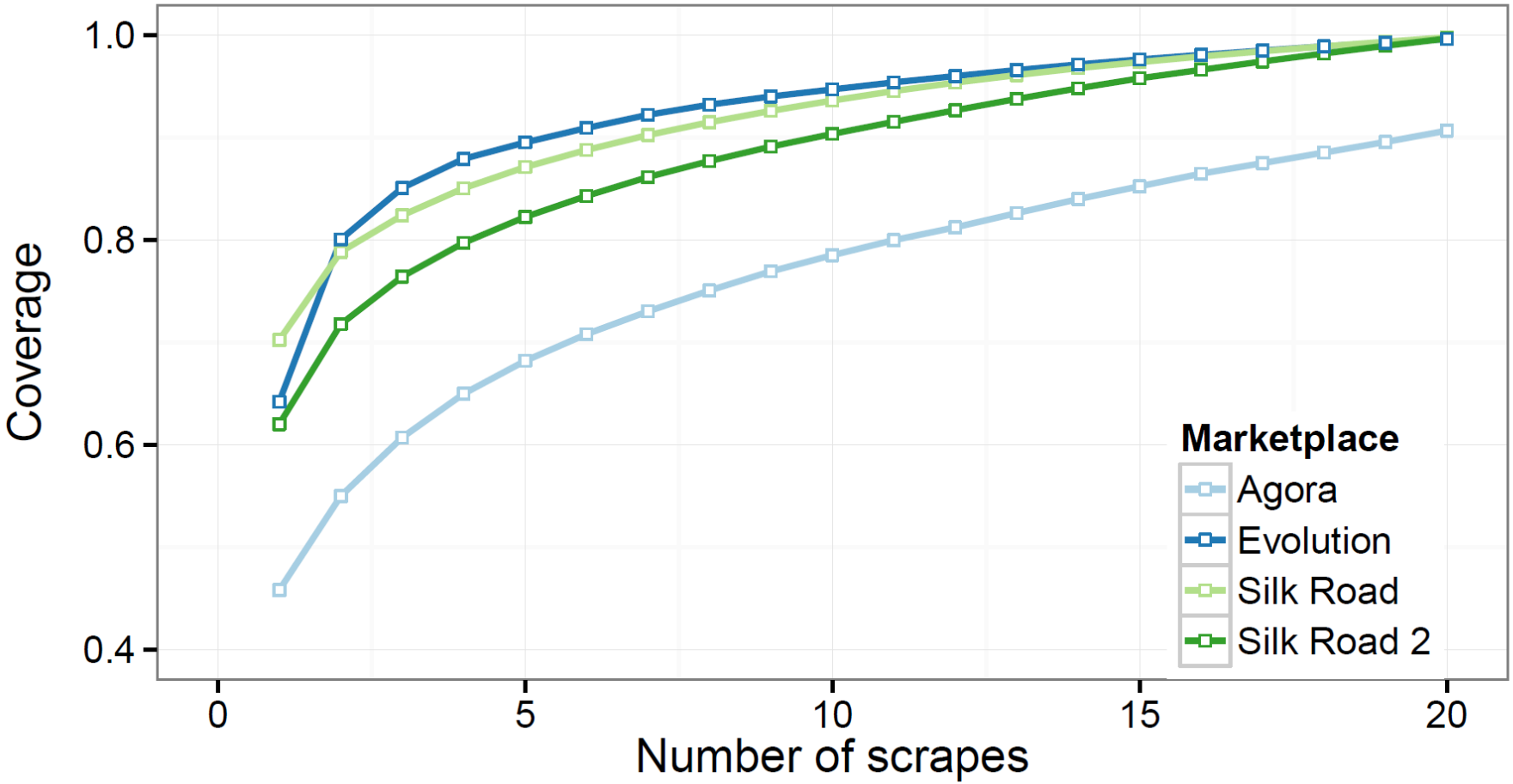


Mark and Recapture

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 1 feedback, conservative estimate

Alternative Transaction Proxies

- **Counting # Item Listings**
  - Very efficient and convenient
  - Assumes that there exists some stable ratio between transaction volume and # listings
  - Daily $\mathit{volume/\#\ Listings}$ for The Evolution Marketplace in July 2014 and September 2014 differ by factor of 4

Uniqueness

- **Problem:**
  - 100s of observations of same feedback
  - Double counting leads to over-estimations
  - Feedback may be updated, deleted

- **Solution:**
  - Automatically detect updated feedbacks
    - Only keep most recent version
  - Hash {timestamp, title, vendor, message, rating}

Holding Prices

- Feedbacks are useful to vendors but are destroyed when the listing is removed
- Vendors raise listing prices prohibitively high

$0.02 -> $1,000.00

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

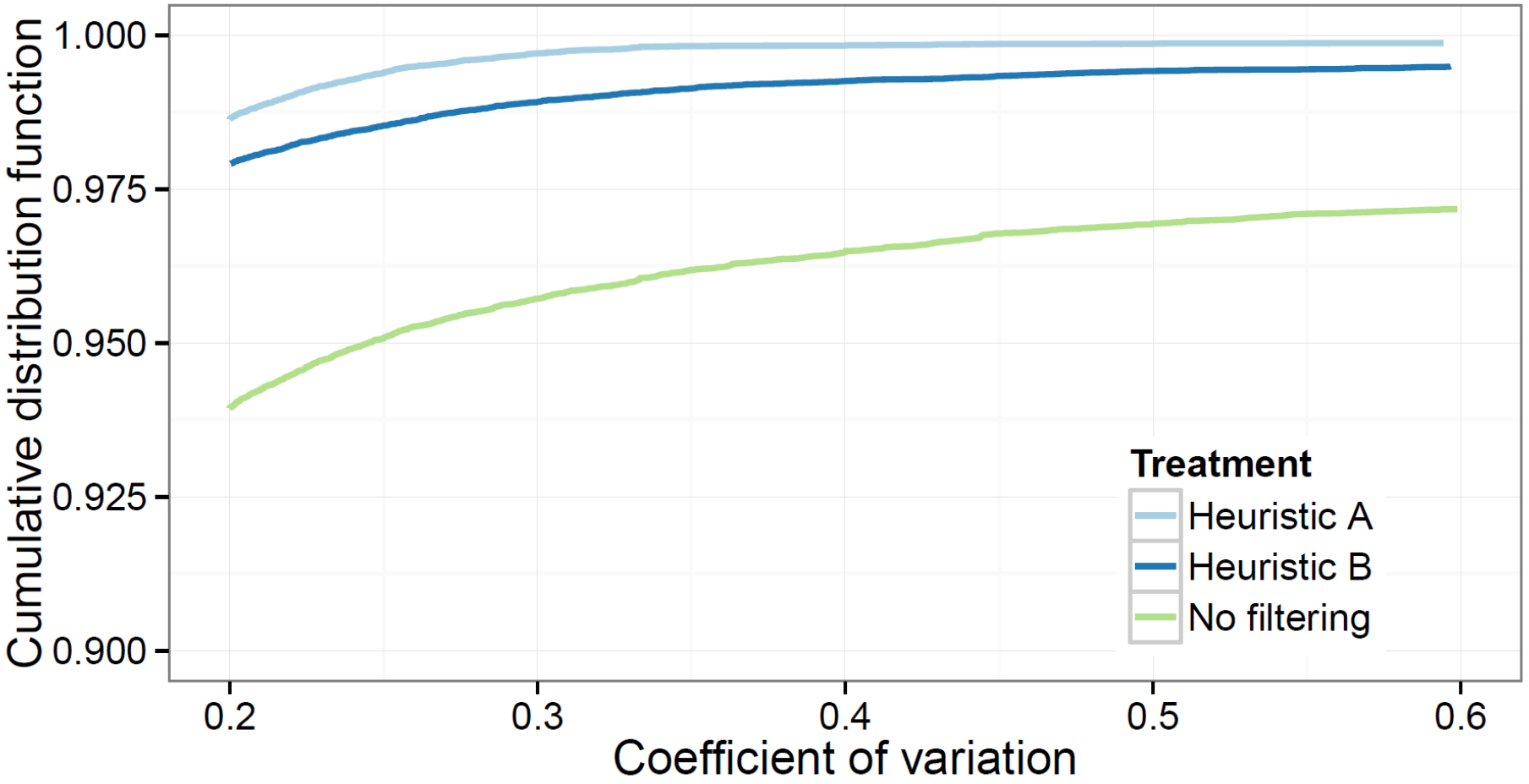
- Need to look at historical price for item

Holding Prices

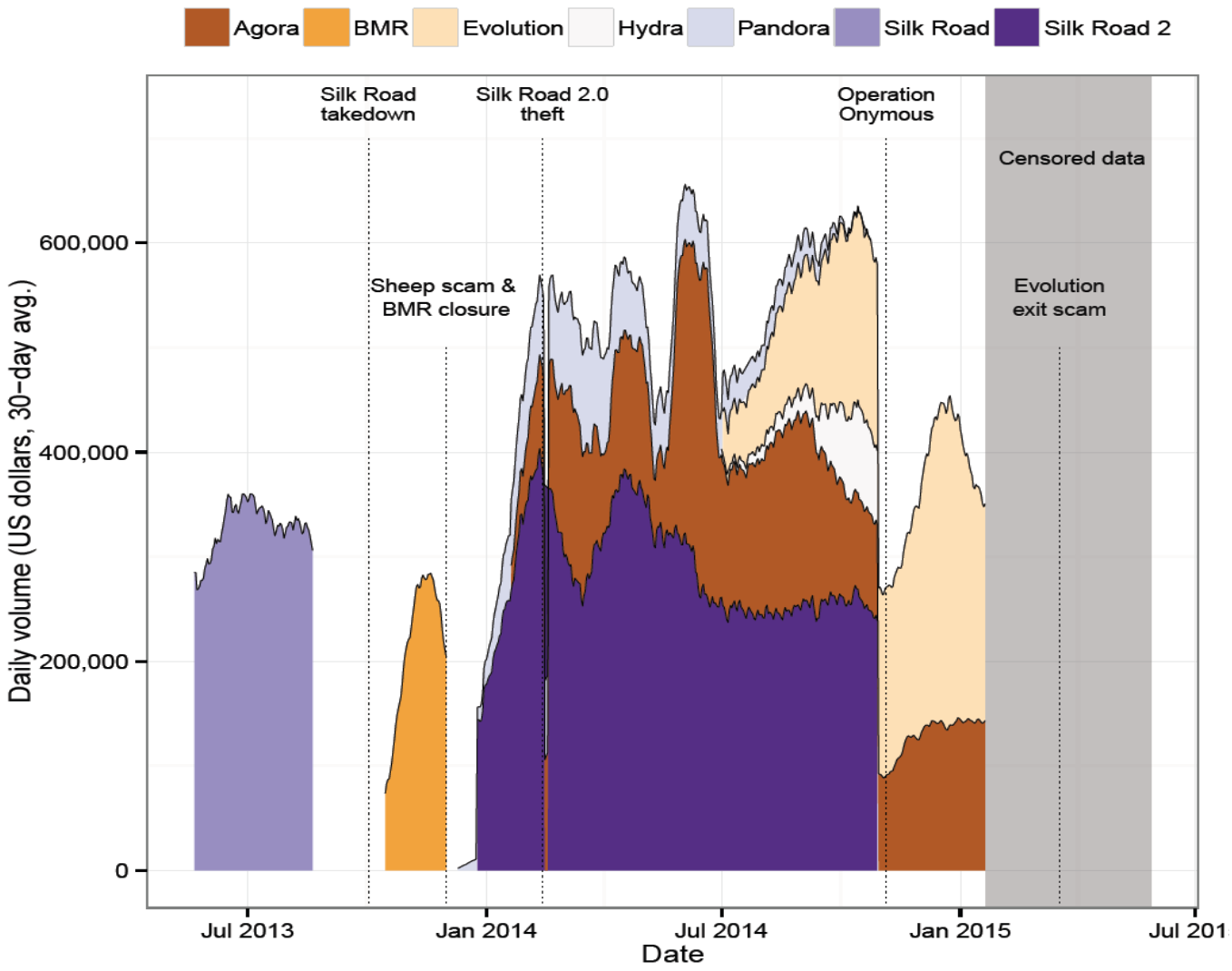
- Heuristic A:
  - Remove all free things
  - Remove all things > $100,000
  - Calculate median of remaining prices
  - Remove everything greater than 5x median
  - Remove things less than 25% of median
- Heuristic B:
  - Remove all things > $100,000
  - Remove upper quartile
  - Remove everything greater than 100x cheapest non-zero price

- Evaluation
  - Coefficient of Variation

Holding Prices CDF

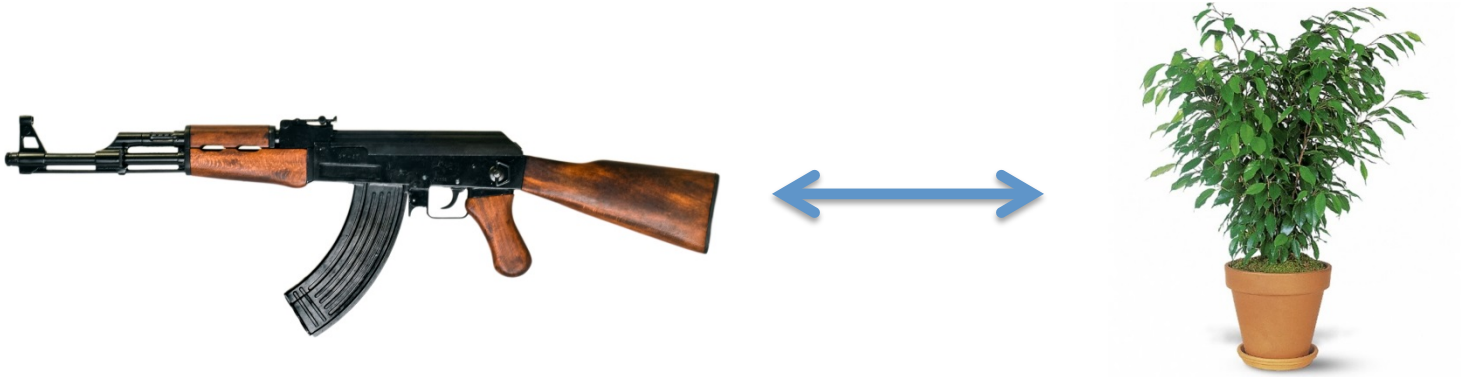


Sales Volume



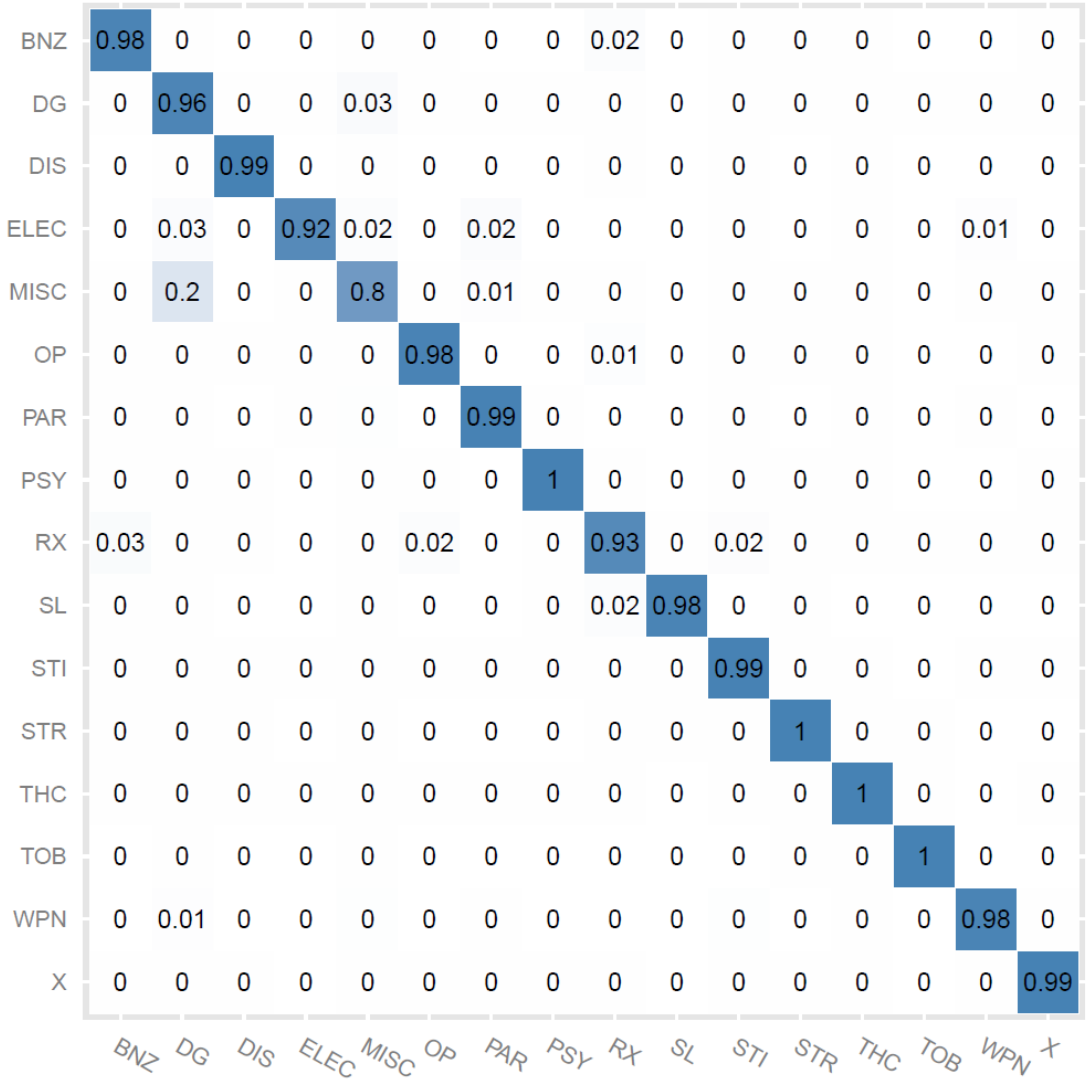
Product Categories

- **What is being sold?**
  - Product labels are often unavailable or inaccurate

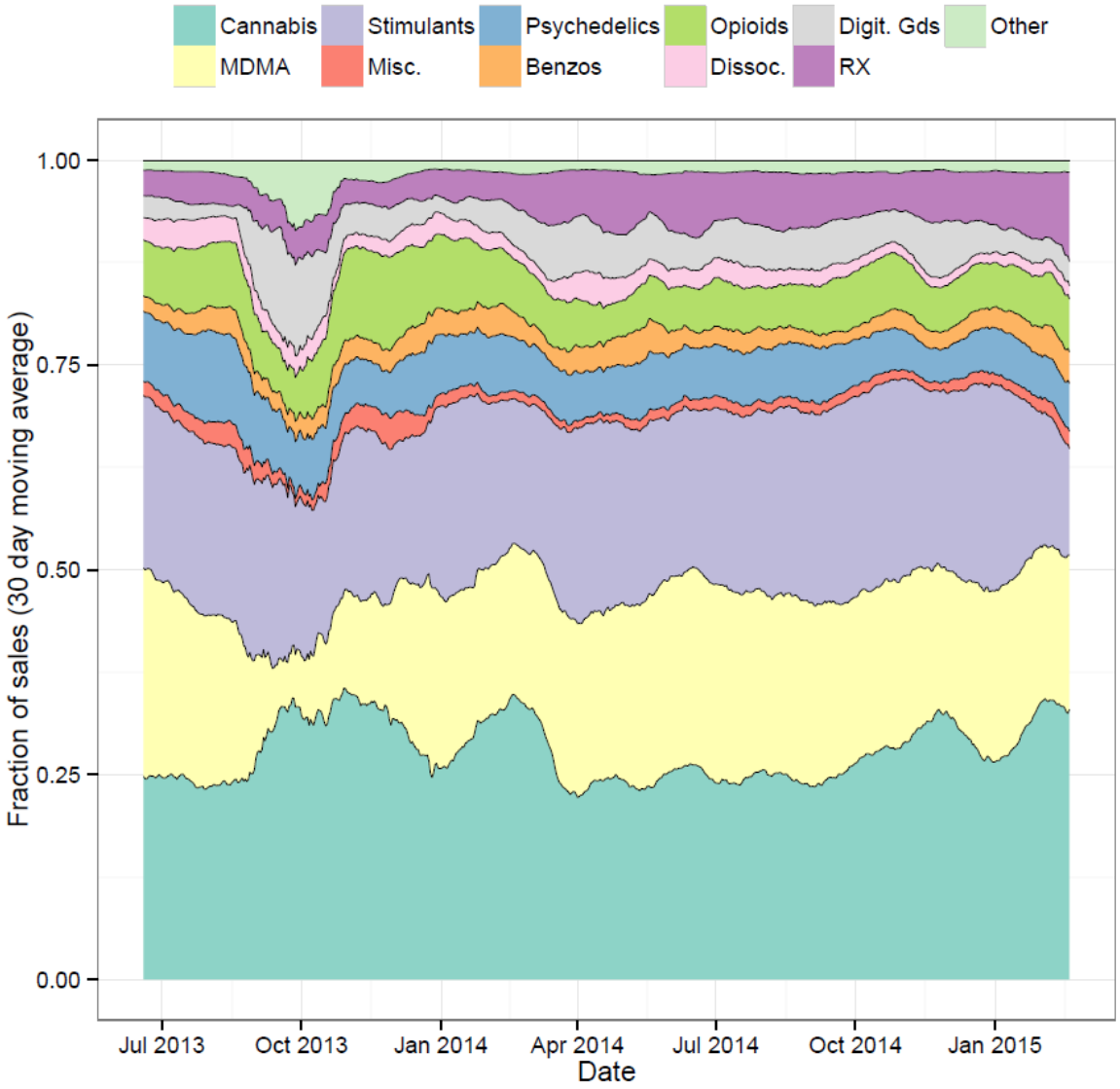


- Classifier trained from Agora and The Evolution Marketplace
  - Listing title and description concatenated and TFIDF
  - 1,941,538 unique samples, 162,198 words tokenized
  - Predicts 16 class labels

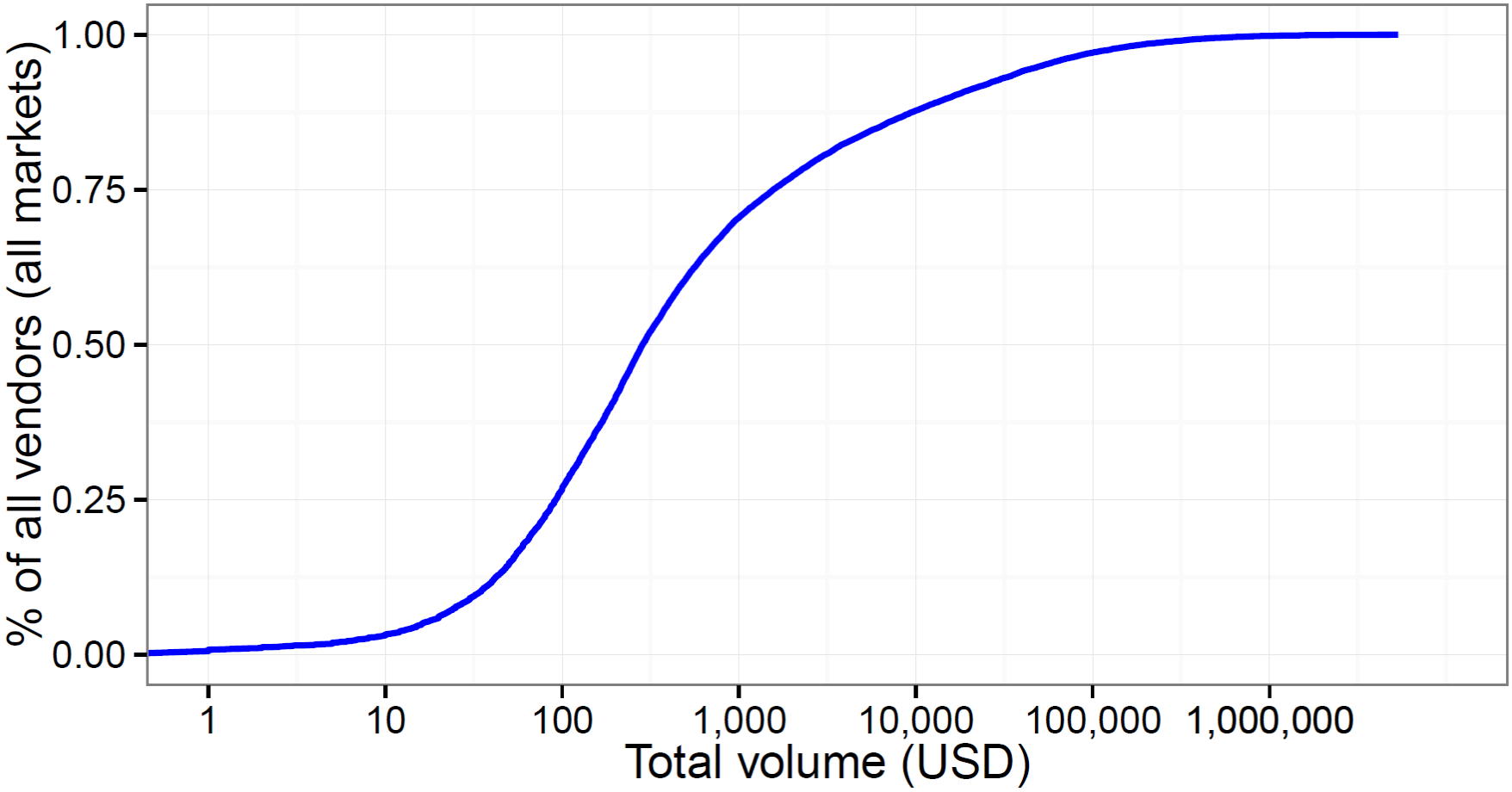
Confusion Matrix



Item Sales Per Category



Vendor Volumes CDF

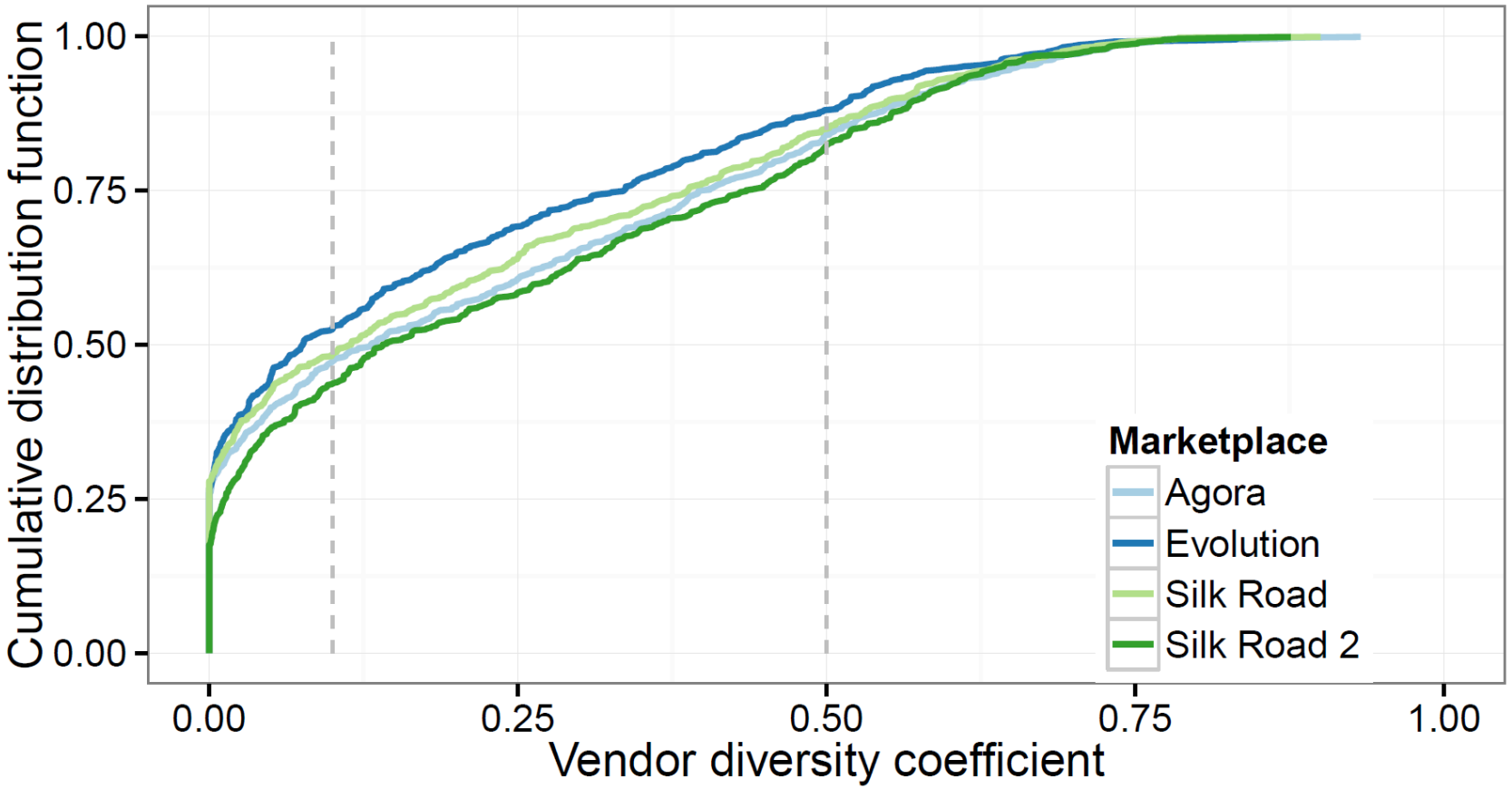


Vendor Diversity

- **Do vendors specialize in what they are selling?**
  - Do vendors sell what they make?
  - Does a single online presence sell goods for several diversified suppliers?

- **Coefficient of Diversity**
  - 0 – all sales from same category
  - 1 – equal sales from each category
  - Only vendors > $10,000 total sales considered

Vendor Diversity CDF



Validation

- Trial evidence GX226A, GX227C places Silk Road 1 weekly volumes at $475,000/week in late March 2012, consistent with our estimates

- Administrator reports Silk Road 2 daily volumes of around $250,000 in September 2014, similar to our estimated $270,000

- Leaked Agora vendor page shows sales total on June 5, 2014 to be $3,460, our observations yielded $3,408

Takeaways

- Anonymous Marketplaces are very easy to setup and use and have wide customer appeal
- Anonymous Marketplace ecosystem transacts in excess of $500,000 / day
- Anonymous Marketplaces are primarily used (~75%) for recreational drugs
- Anonymous Marketplace ecosystem has historically recovered from takedown efforts and scams
- Anonymous Marketplaces are controlled by small set of highly influential vendors

Kyle Soska – ksoska@cmu.edu

Anonymity in Bitcoin

(U) Bitcoin Virtual Currency: Unique Features Present Distinct Challenges for Deterring Illicit Activity

Anonymity in Bitcoin

(U) Bitcoin Virtual Currency: Unique Features Present Distinct Challenges for Deterring Illicit Activity

Ponzi-Scheme Charge Is Good News for Bitcoin

Anonymity in Bitcoin

(U) Bitcoin Virtual Currency: Unique Features Present Distinct Challenges for Deterring Illicit Activity

Ponzi-Scheme Charge Is Good News for Bitcoin

Estimated 18 percent of US drug users bought from Silk Road, says study

Anonymity in Bitcoin

(U) Bitcoin Virtual Currency: Unique Features Present Distinct Challenges for Deterring Illicit Activity

Ponzi-Scheme Charge Is Good News for Bitcoin

Estimated 18 percent of US drug users bought from Silk Road, says study

How much anonymity does Bitcoin really provide?

How Bitcoin works

peer-to-peer network

How Bitcoin works

(pk_A, sk_A)

(pk_B, sk_B)

peer-to-peer network

How Bitcoin works

(pk_A, sk_A)

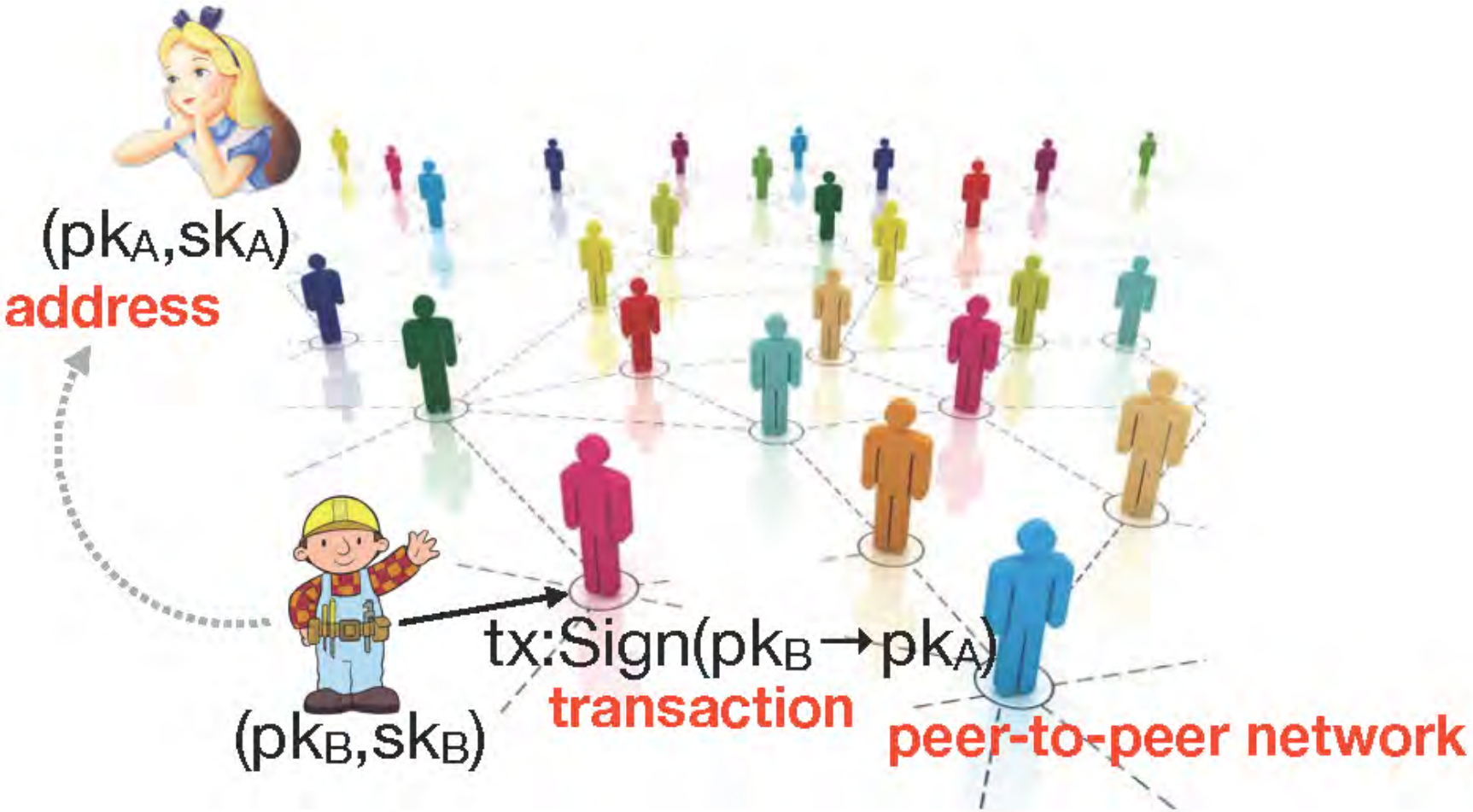
address

(pk_B, sk_B)

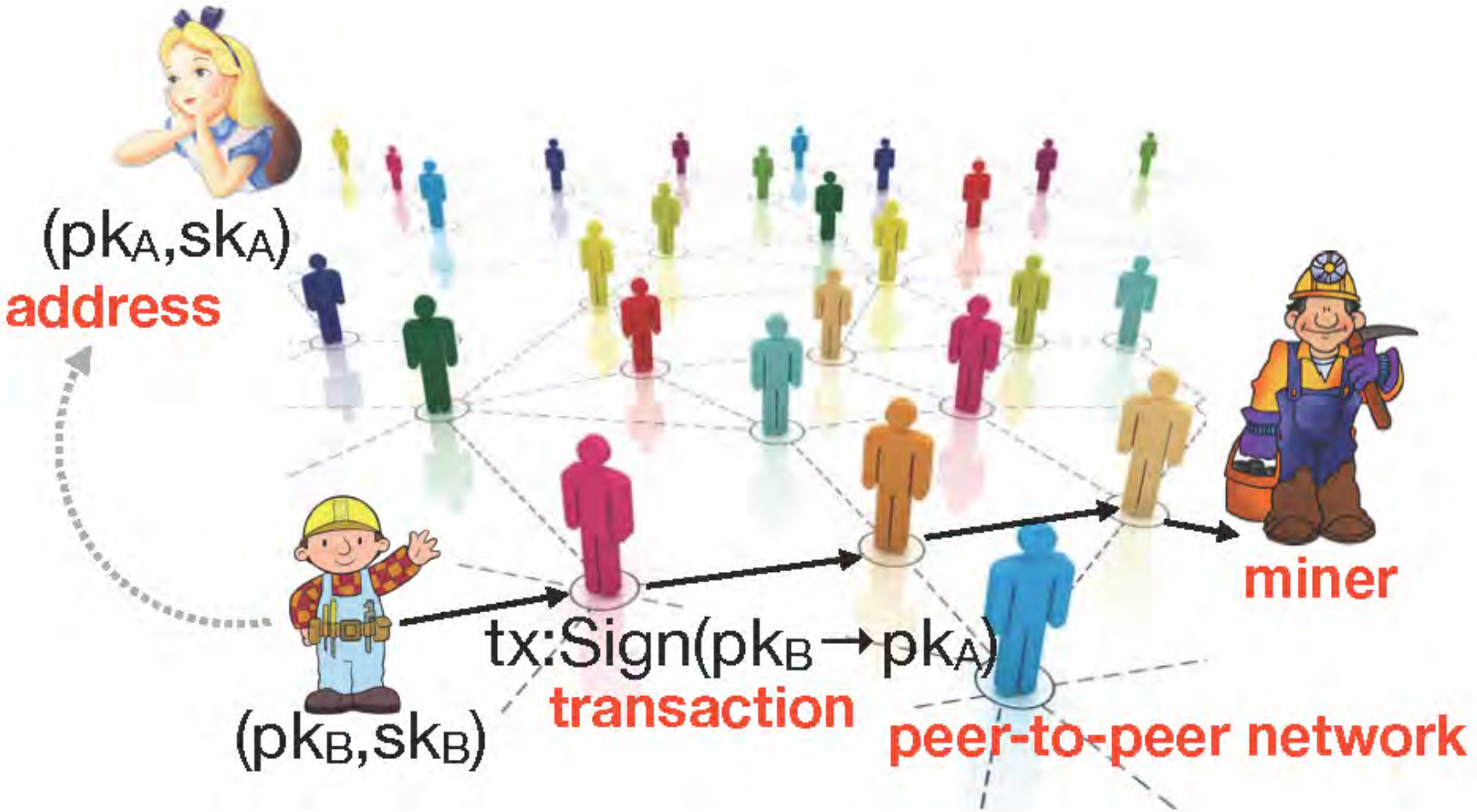
peer-to-peer network

How Bitcoin works

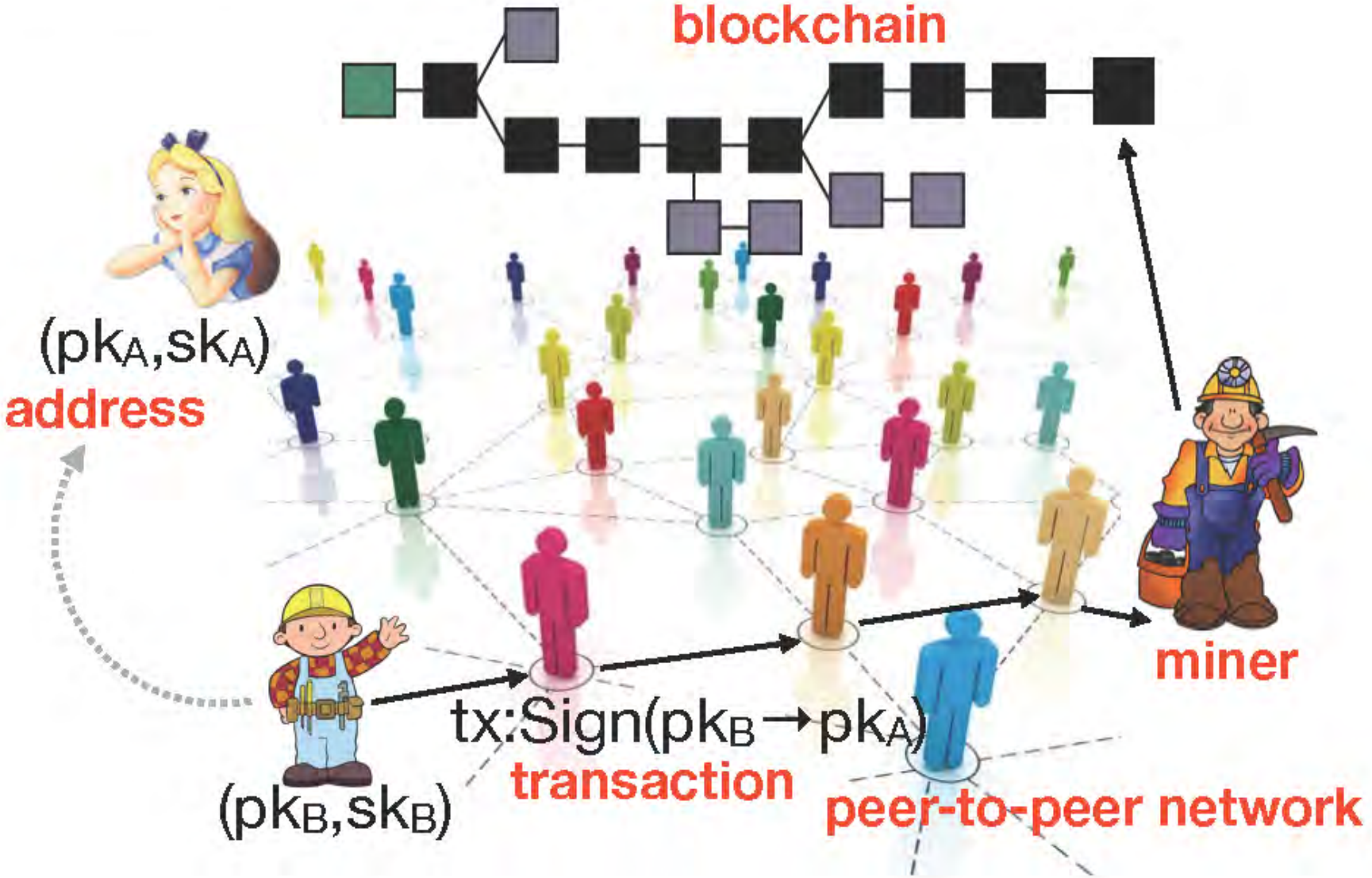
How Bitcoin works



How Bitcoin works



How Bitcoin works



Anonymity in Bitcoin

How much anonymity does Bitcoin really provide?

(pk_A, sk_A)

address

(pk_B, sk_B)

Anonymity in Bitcoin

How much anonymity does Bitcoin really provide?

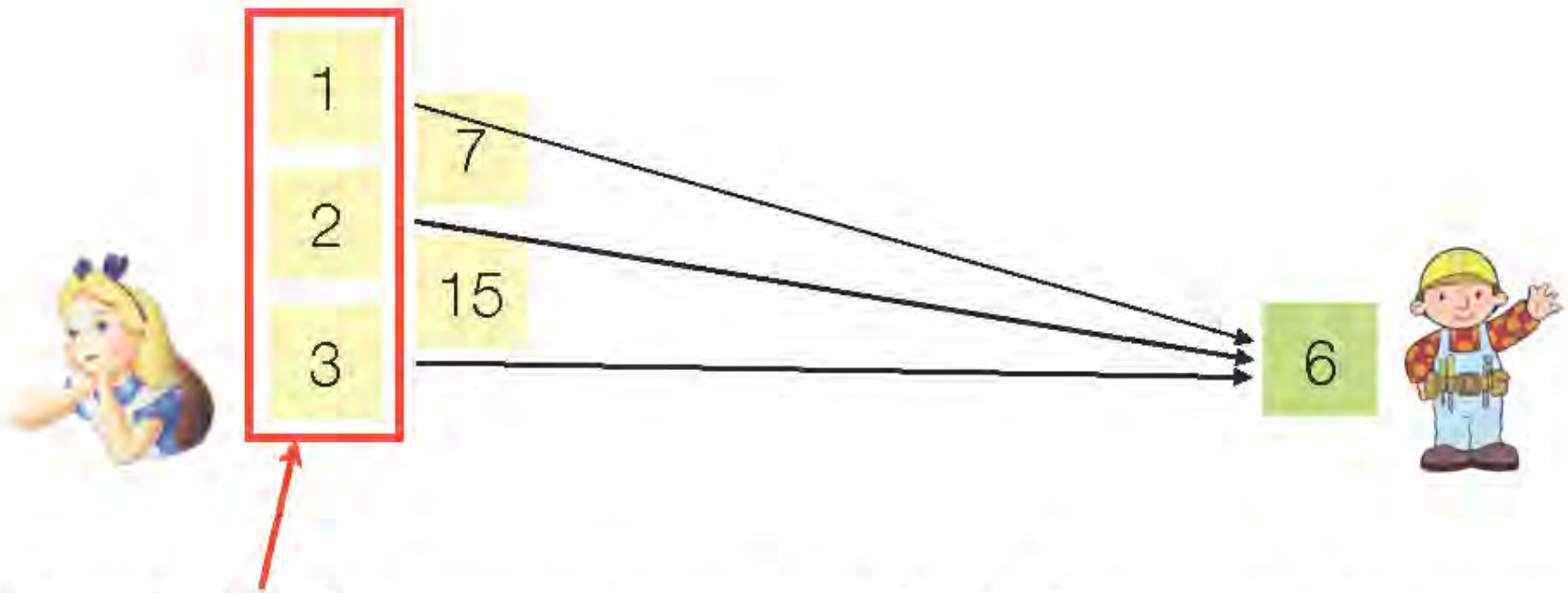
in theory, a lot! addresses are not linked to identity

(pk_A, sk_A)

address

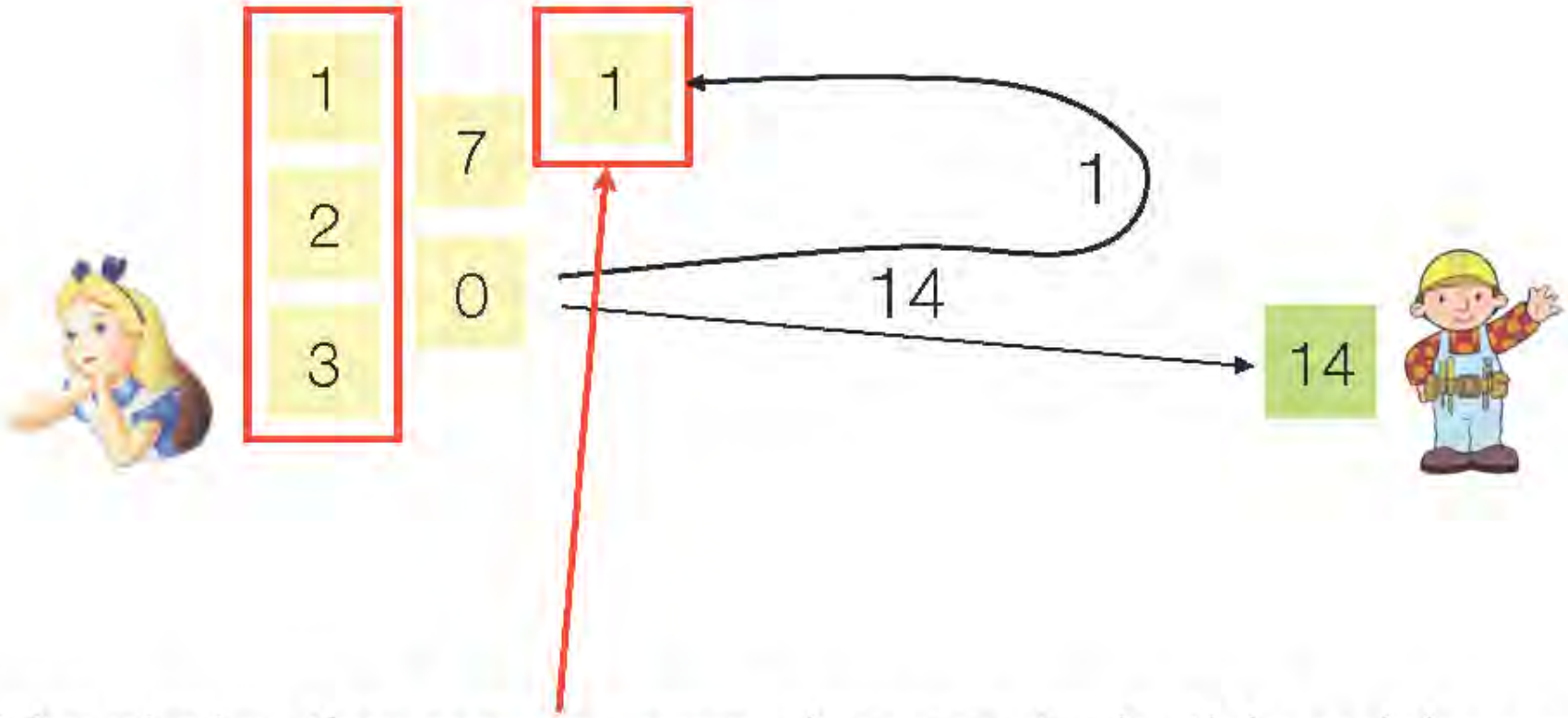
(pk_B, sk_B)

Input clustering [RH13,RS13,A+13,**M**+13,SMZ14]



Heuristic: the same user controls these addresses

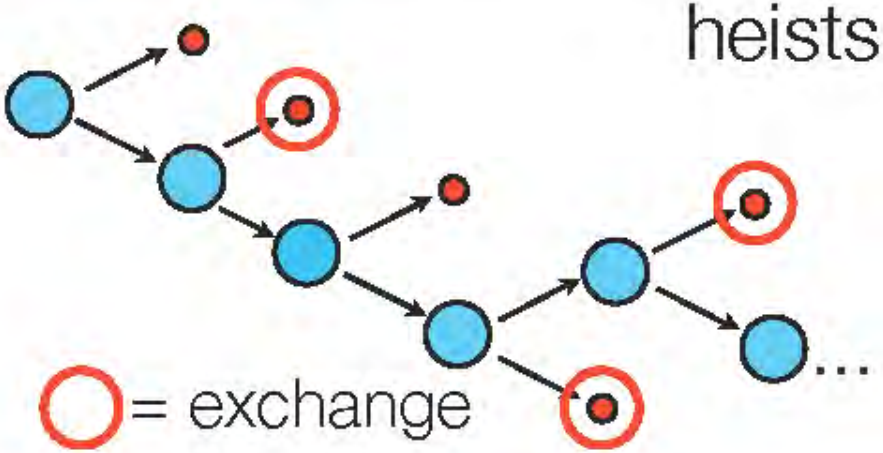
Change clustering [A+13,**M**+13,SMZ14]



Heuristic: the same user also controls this address

Tracking technique [**M**+13,HD**M**+14]

heists



cycle theft

individual thefts

service interaction

Andy Greenberg, Forbes Staff

Covering the worlds of data security, privacy and hacker culture.

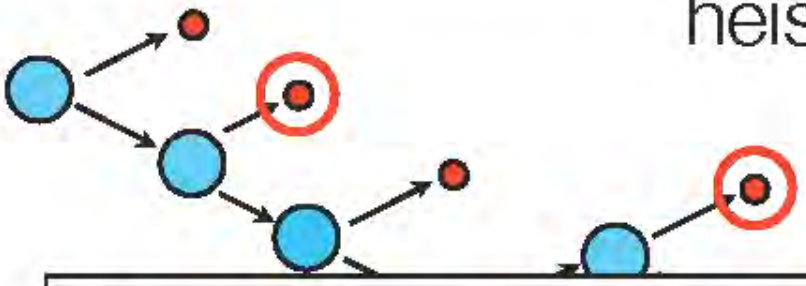
+ Follow (1,192)

SECURITY 9/05/2013 @ 10:36AM | 131,094 views

Follow The Bitcoins: How We Got Busted Buying Drugs On Silk Road's Black Market

Tracking technique [**M**+13,HD**M**+14]

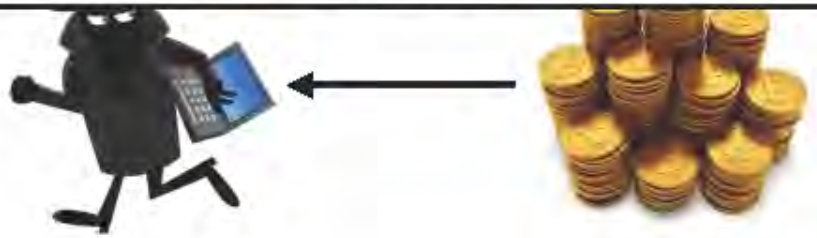
heists



cycle theft

Prosecutors Trace $13.4M in Bitcoins From the Silk Road to Ulbricht's Laptop

If anyone still believes that bitcoin is magically anonymous internet money, the US government just offered what may be the clearest demonstration yet that it's not.



SECURITY

Follow The Bitcoins: How We Got Busted Buying Drugs On Silk Road's Black Market

Anonymity in Bitcoin

How much anonymity does Bitcoin really provide?

in theory, a lot! addresses are not linked to identity

in practice, maybe not so much

Privacy-enhancing overlays

Mixcoin

Anonymity for Bitcoin with accountable mixes

(Full version)

Josep

Destination Address Anonymization in Bitcoin

CoinSwap: Transaction graph disjoint trustless trading

October 30, 2

CoinShuffle: Practical Decentralized

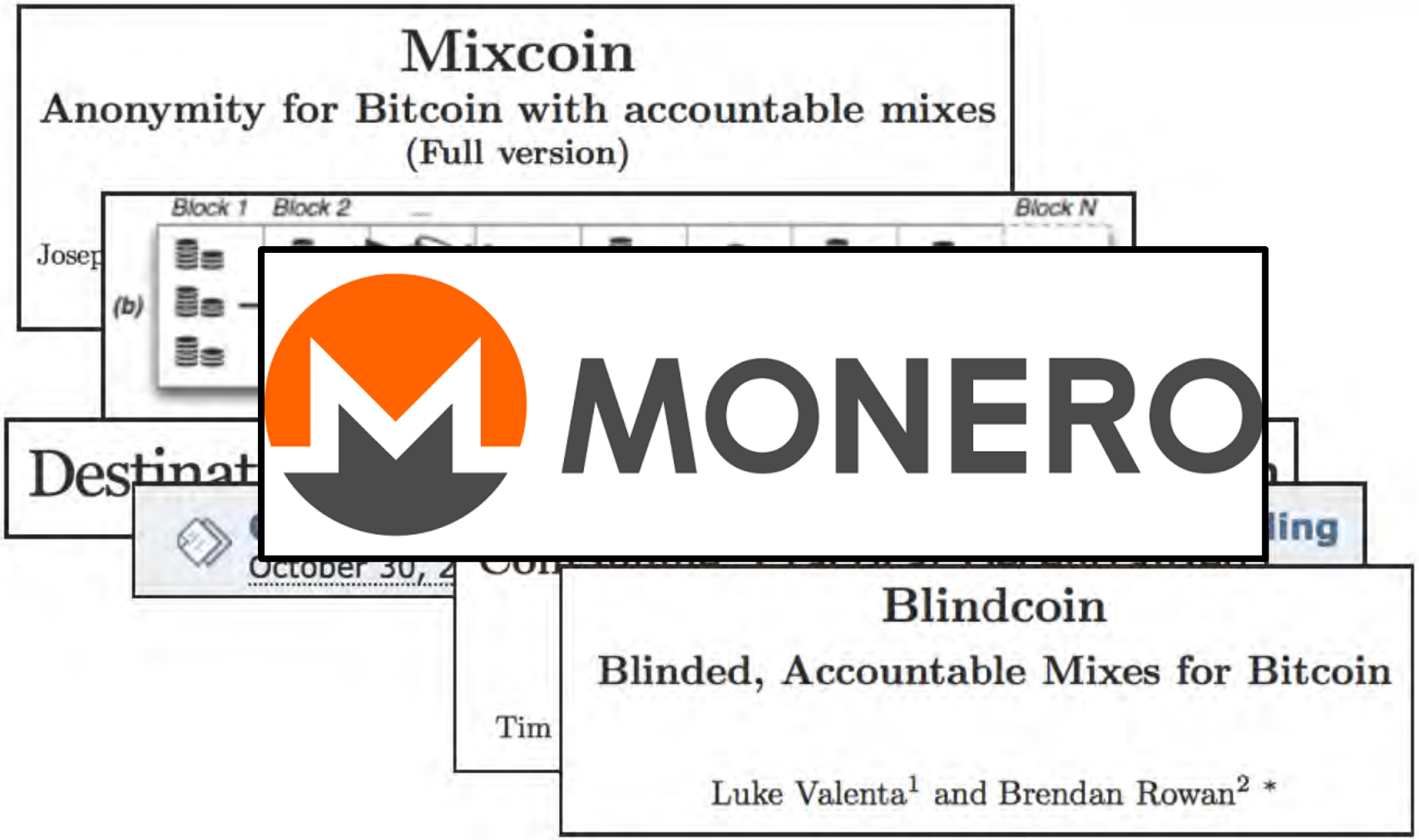
Tim

Blindcoin

Blinded, Accountable Mixes for Bitcoin

Luke Valenta[1] and Brendan Rowan[2] *

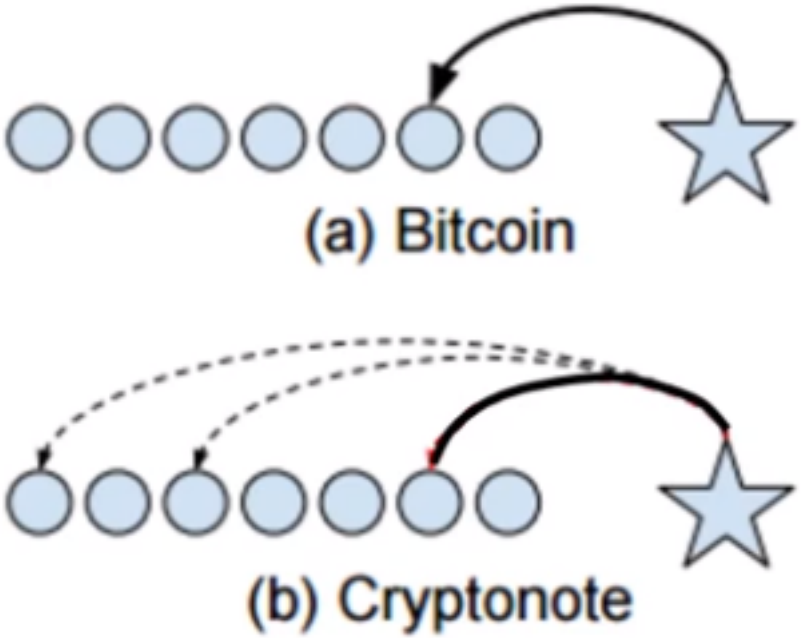
Privacy-enhancing overlays



Monero Anonymity

Malte Möser*, Kyle Soska, Ethan Heilman, Kevin Lee, Henry Heffan, Shashvat Srivastava, Kyle Hogan, Jason Hennessey, Andrew Miller, Arvind Narayanan, and Nicolas Christin

An Empirical Analysis of Traceability in the Monero Blockchain



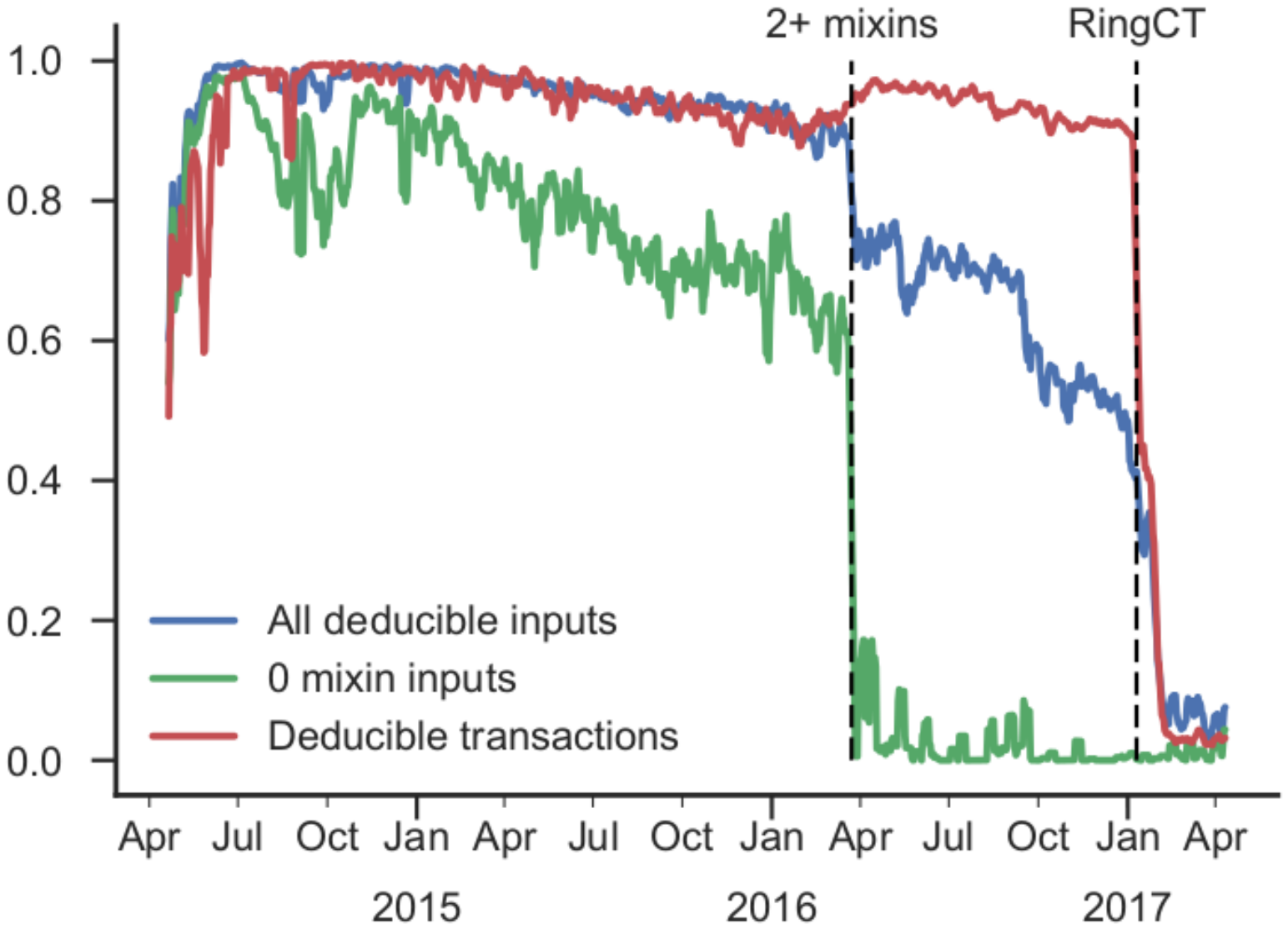


Fig. 5: Fraction of transaction inputs that can be deduced and transactions including at least one deducible input (averaged over intervals of 7 days)

Problem 1: 0-mixins

- The 0-mixin transactions should never be used later as “mix-ins” since they dont help. However, the Monero client does not discriminate

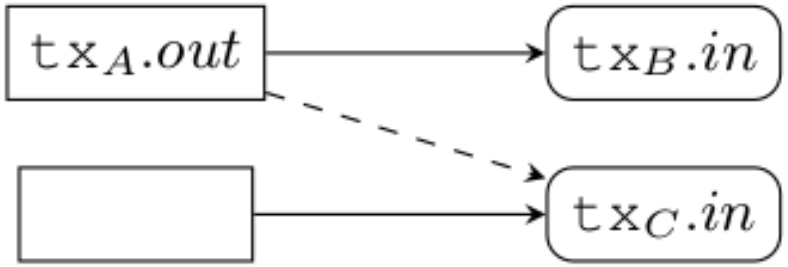
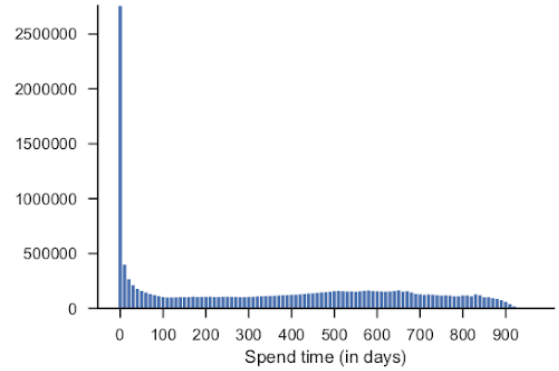
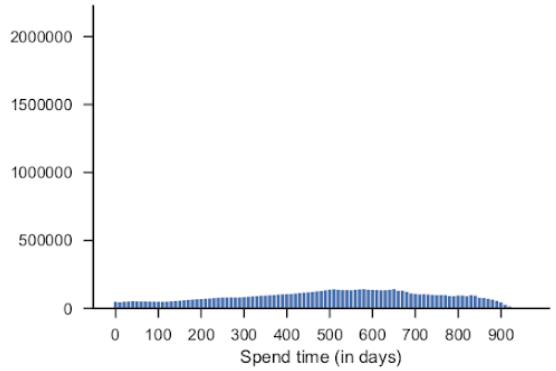


Fig. 4: 0-mixins effectively reduce the untraceability of other transactions: the dashed reference can be ruled out since $\mathtt{tx}_A.out$ must have been spent in $\mathtt{tx}_B.in$.

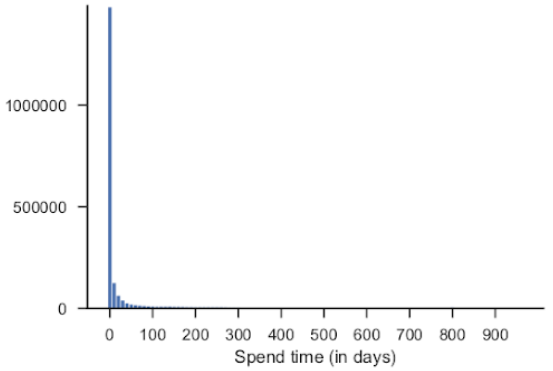
Problem 2: Temporal Analysis



(a) Age distribution among all inputs (mixins and real) from blocks 0.9M-1.2M.



(b) Age distribution among all ruled-out mixins from blocks 0.9M-1.2M.



(c) Estimated age distribution of real inputs (recovered from deducible transactions, among blocks 0.9M-1.2M).

The real transaction input is typically the “first” one

Non-Hidden Hidden Services Considered Harmful

Filippo Valsorda
George Tankersley

What is Tor?

- **T**he **O**nion **R**outer
- Provides client anonymity
- Works by routing your connection though other machines

Hidden Services

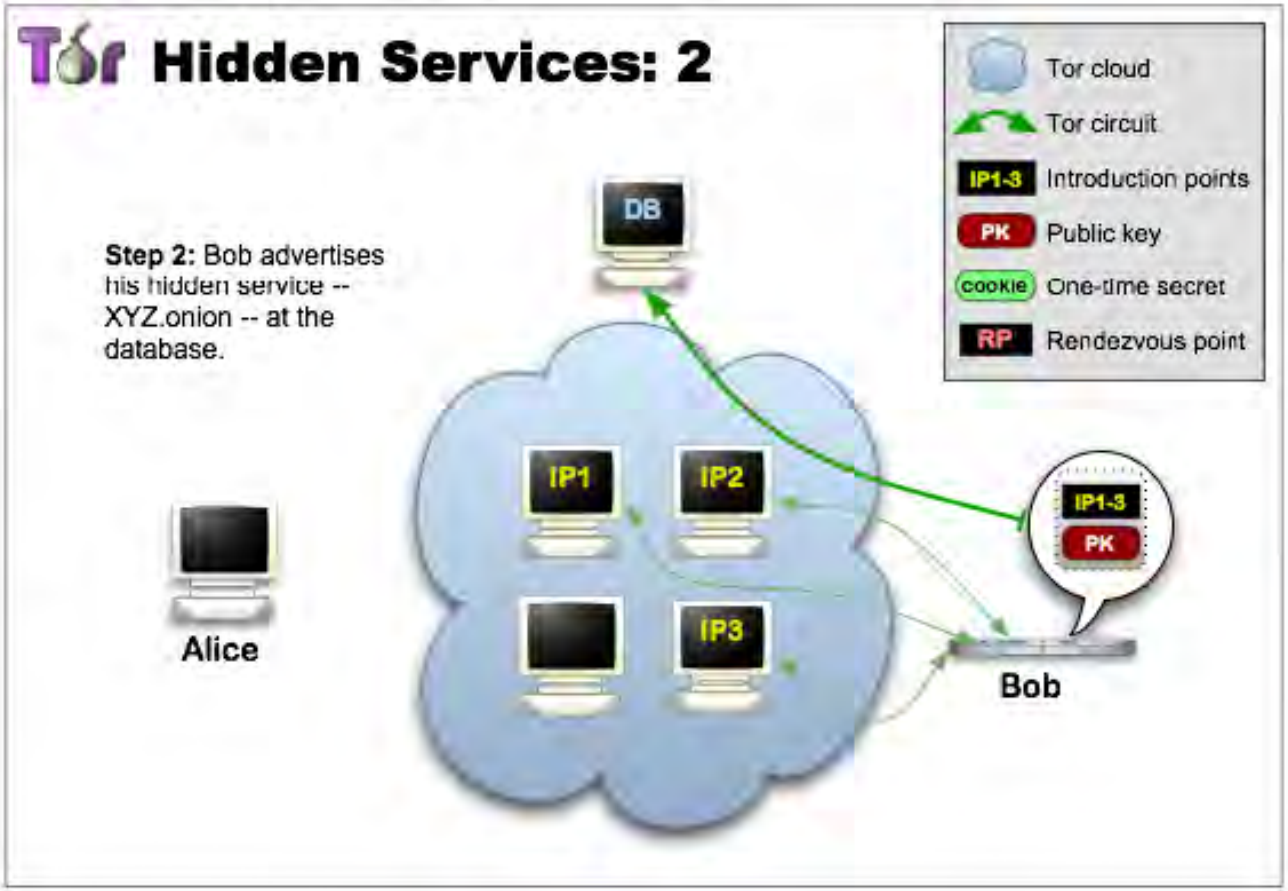
- Provide *bidirectional* anonymity
- Supports generic TCP services
- Famous for drug markets
  - Silk Road
  - Silk Road 2

Hidden Services

But they're actually used for good

- Whistleblowing (SecureDrop)
- Private chat (Ricochet, XMPP-over-HS)
- Anonymous publishing (of course!)

Hidden Services



Hidden Services

The “database” is a DHT made up of stable relays

- directory authorities grant *HSDir* flag
- not related to *Stable* flag

How do we choose where to publish?

HSDir selection

Choose two sets of 3 relays with *HSDir* flag

Think “consistent hashing”

- relays arranged in a ring sorted by identity

Based on a predictable formula (#8244)

HSDir selection

hs-descriptor-id =

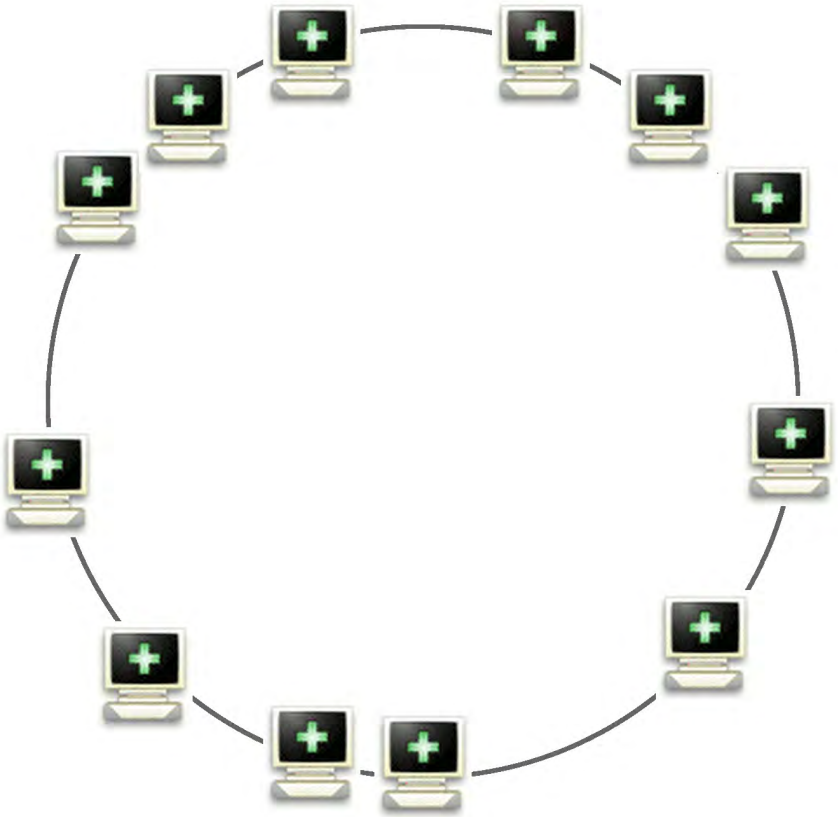
SHA1(id || SHA1(time-period || replica))

id: first 80 bits of SHA1(public key)

time-period: days since epoch (+offset)

replica: which set of HSDirs

HSDir selection



HSDir selection

facebookcorewwwi.onion

descriptor-id =

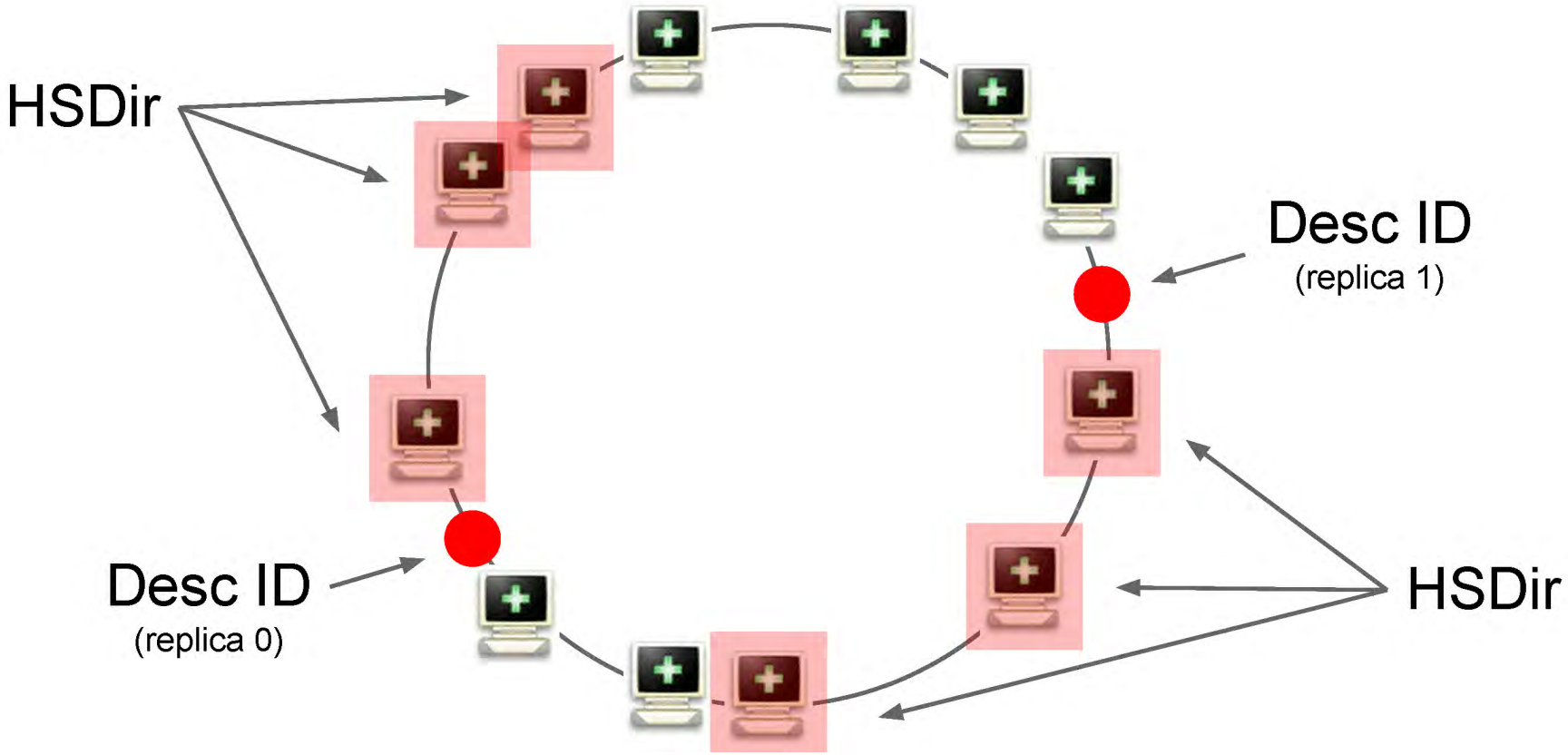
SHA1(facebookcorewwwi || SHA1(16583 || 0))

SHA1(facebookcorewwwi || SHA1(16583 || 1))

replica 0: ys5pml4c6txpw5hnq5v4zn2htytfejf2

replica 1: fq7r4ki5uwcxdxibdl7b7ndvf2mvw2k2

HSDir selection



Why did he just explain all this?

Point of the talk!

Hidden service users face a greater risk of targeted deanonymization than normal Tor users.

Vulnerability of Tor

Low-latency implies correlation attacks

Correlation attacks

in Tor, “both ends” means we’re usually just worried about entry nodes and exit nodes

- **entry nodes** see when a connection starts
- **exit nodes** see when it terminates

Correlation attacks

worried about entry nodes and exit nodes

- ***entry nodes*** *see when a connection starts*
- ***exit nodes*** *see when it terminates*

Tor has protections for entry/exit positions

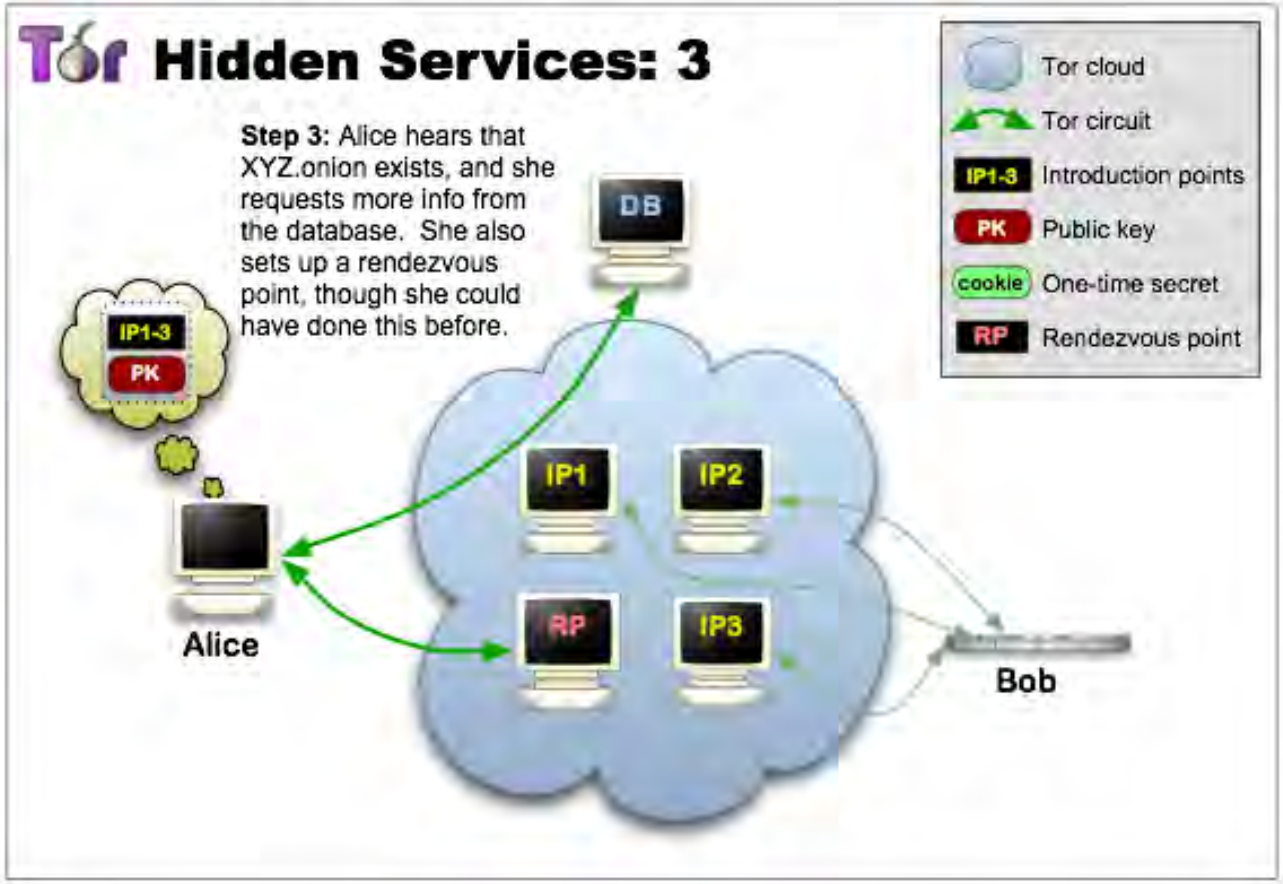
- entry guards, bad relay monitoring, size of network

Correlation attacks

It is hard to become both ends of a circuit.

What else can see when connections happen?

Hidden Services



Hidden Services

An HSDir for a hidden service gets a lookup on ⅙ of requests for information about the hidden service

A lookup indicates a user trying to connect to the hidden service

Correlation attacks

worried about entry nodes and exit nodes

- ***entry nodes*** *see when a connection starts*
- ***exit nodes*** *see when it terminates*

For a hidden service, the HSDir can see when a connection happens

Correlation attacks

*worried about entry nodes and **HSDir***

- ***entry nodes** see when a connection starts*
- ***HSDir** see when it terminates*

For a hidden service, the HSDir can see when a connection happens

Correlation attacks

If your target uses a hidden service, don't need exit relay to see when the connection happens.

Instead, be an HSDir.

Hidden Services

It is very easy to become HSDir

- You just need 4 days uptime
- It should be harder than it is (#8243)

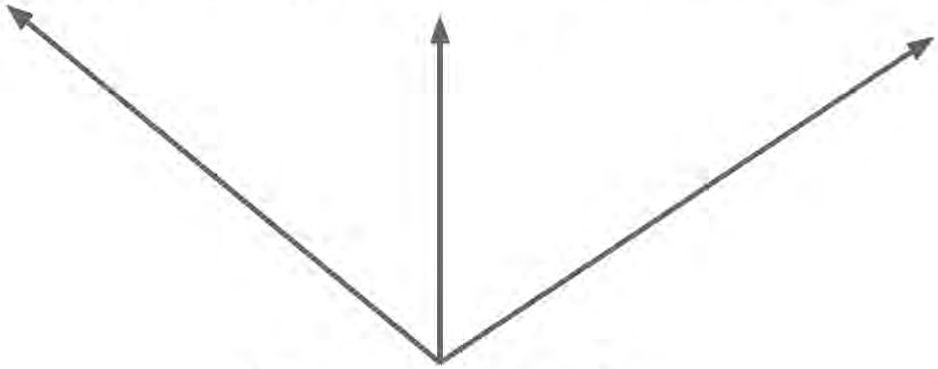
In fact, very easy to become *specific* HSDir

Positioning attack

SHA1(id || SHA1(time-period || replica))

Positioning attack

SHA1(**id** || SHA1(**time-period** || **replica**))



PREDICTABLE

Positioning attack

Predictable and fast? Bruteforce it!

1) Calculate descriptor IDs for the service
2) Generate random 1024-bit RSA key
3) Check if hash precedes the first real descriptor ID in the DHT
4) If not, goto 2

Correlation attacks

If your target uses a hidden service, don't need exit relay to see when the connection happens.

Instead, be **their** HSDir.

Correlation attacks

If your target uses a hidden service, don't need exit relay to see when the connection happens.

Instead, be **every** HSDir.

Vulnerability of Tor

worried about entry nodes and HSDir

- ***entry nodes** see when a connection starts*
- *HSDir see when it terminates*

Vulnerability of Tor

worried about entry nodes and HSDir

- ***many people*** *see when a connection starts*
- *HSDir see when it terminates*

Vulnerability of Tor

worried about entry nodes and HSDir

- ***many people** see when a connection starts*
- *HSDir see when it terminates*

"entry" does not just mean your entry node

- ISP, malicious access point, pen register…

Summarizing all of that

1) HSDirs can serve the same purpose against a hidden service as a malicious exit relay would in a basic correlation attack

2) The “entry side” of a Tor connection can be monitored by means other than compromising guards

Summarizing all of that

It's actually ***worse***, because it's way easier to be the user's HSDir.

Hidden service users face a greater risk of targeted deanonymization than normal Tor users.