Hiding in a Panopticon: Grand Challenges in Internet Anonymity

Bryan Ford, David Isaac Wolinsky, Joan Feigenbaum,
Henry Corrigan-Gibbs, Ewa Syta, John Maheswaran,
Ramakrishna Gummadi **– Yale**

Vitaly Shmatikov, Amir Houmansadr,
Chad Brubaker **– UT Austin**

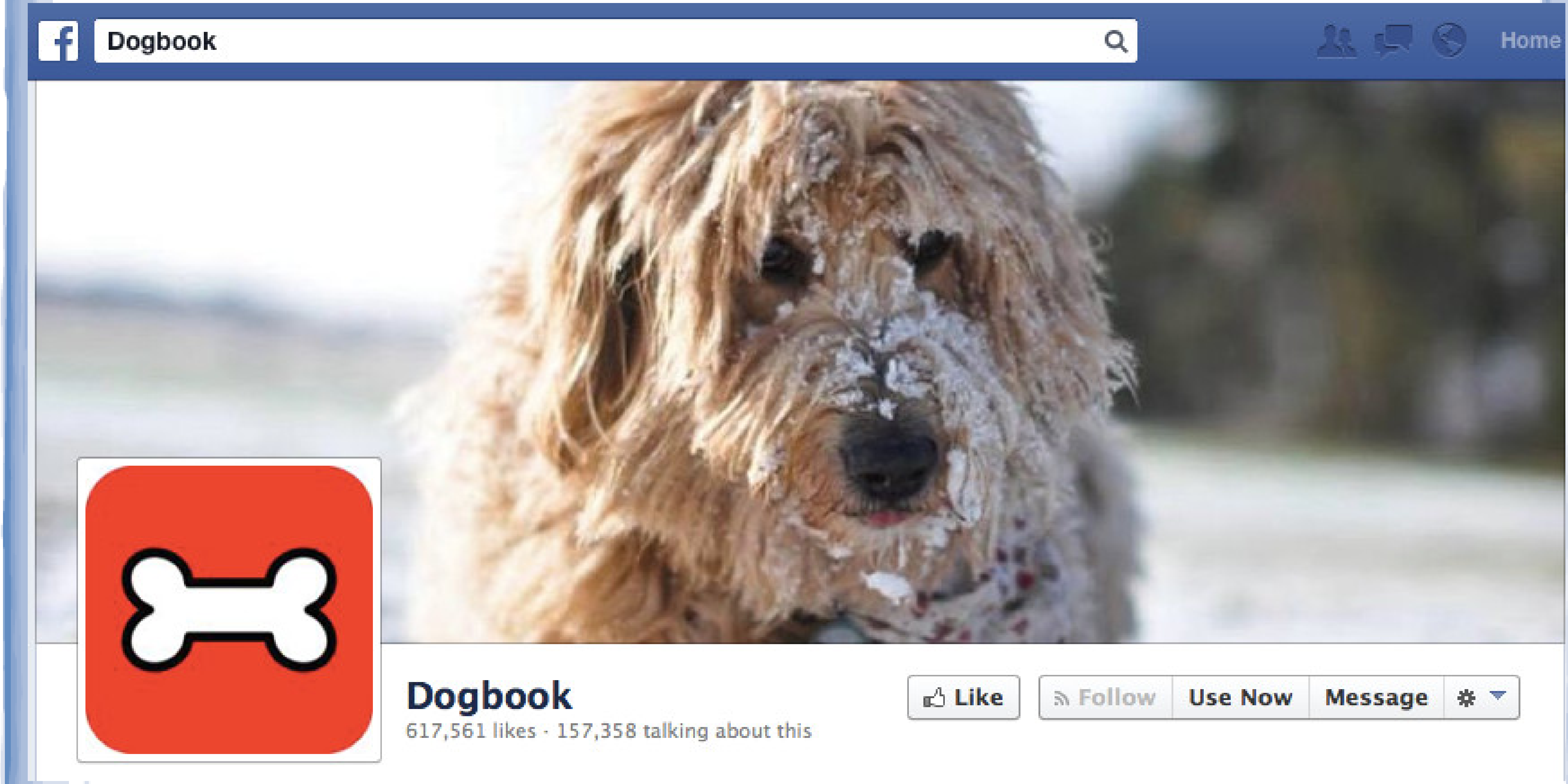
Aaron Johnson **– US Naval Research Lab**

NDSS/SENT Workshop – February 23, 2014

“Nobody knows you're a dog?”

Actually, they know exactly what kind you are

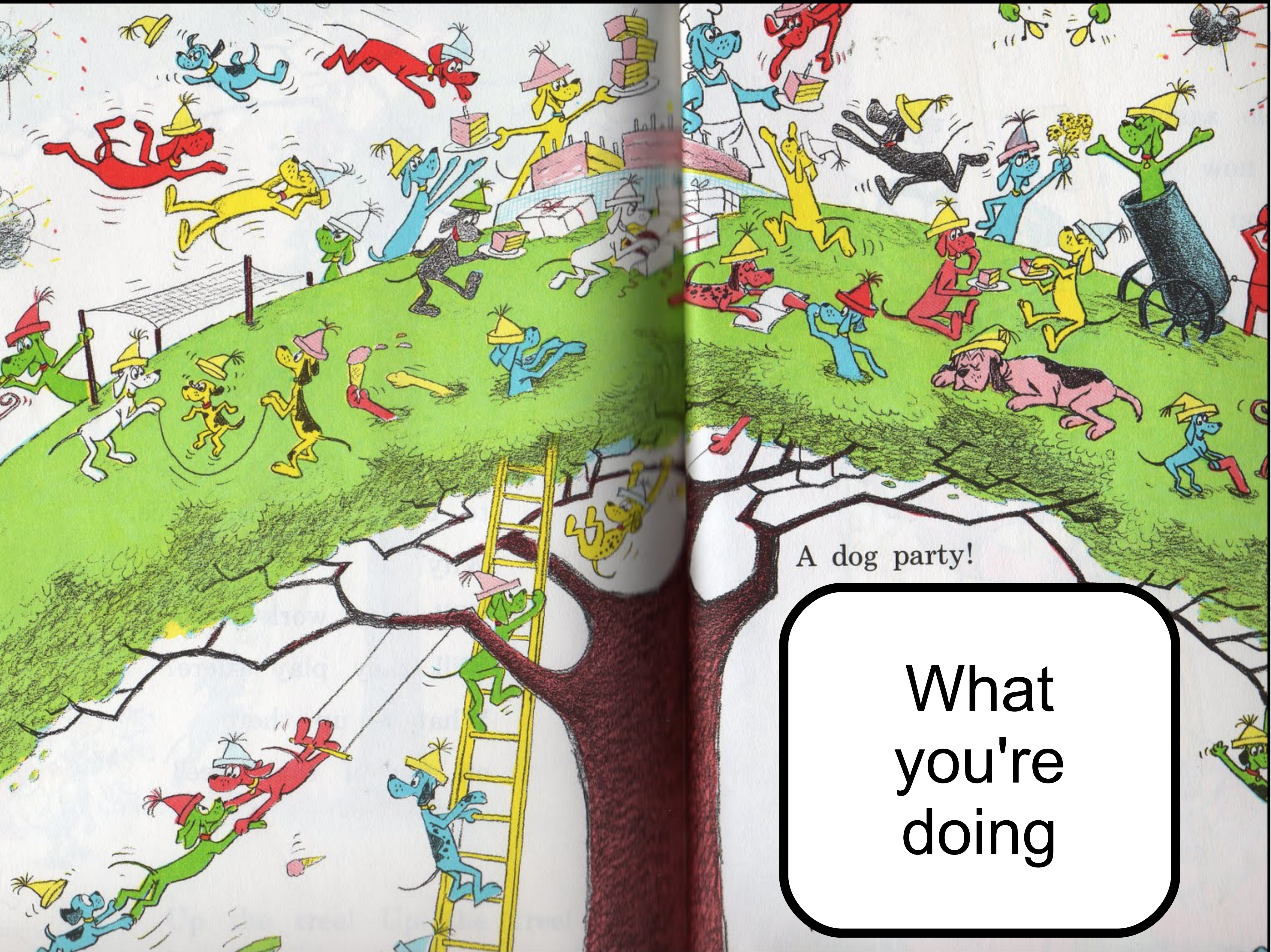
Who your friends are...



Dogbook

Dogbook

617,561 likes · 157,358 talking about this



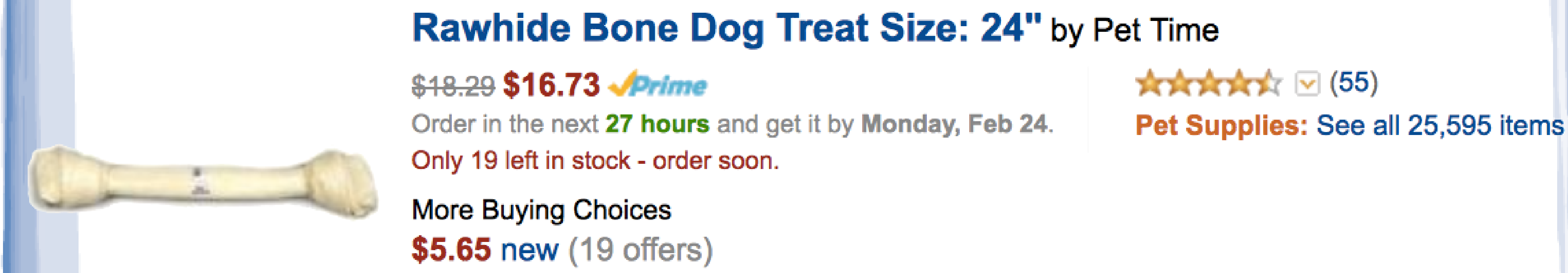
A dog party!

What you're doing

dogazon

What you *and* your friends like to buy

Gift suggestion ...



Rawhide Bone Dog Treat Size: 24" by Pet Time

~~$18.29~~ **$16.73** ✓Prime

★★★★☆ (55)

Order in the next **27 hours** and get it by **Monday, Feb 24**.

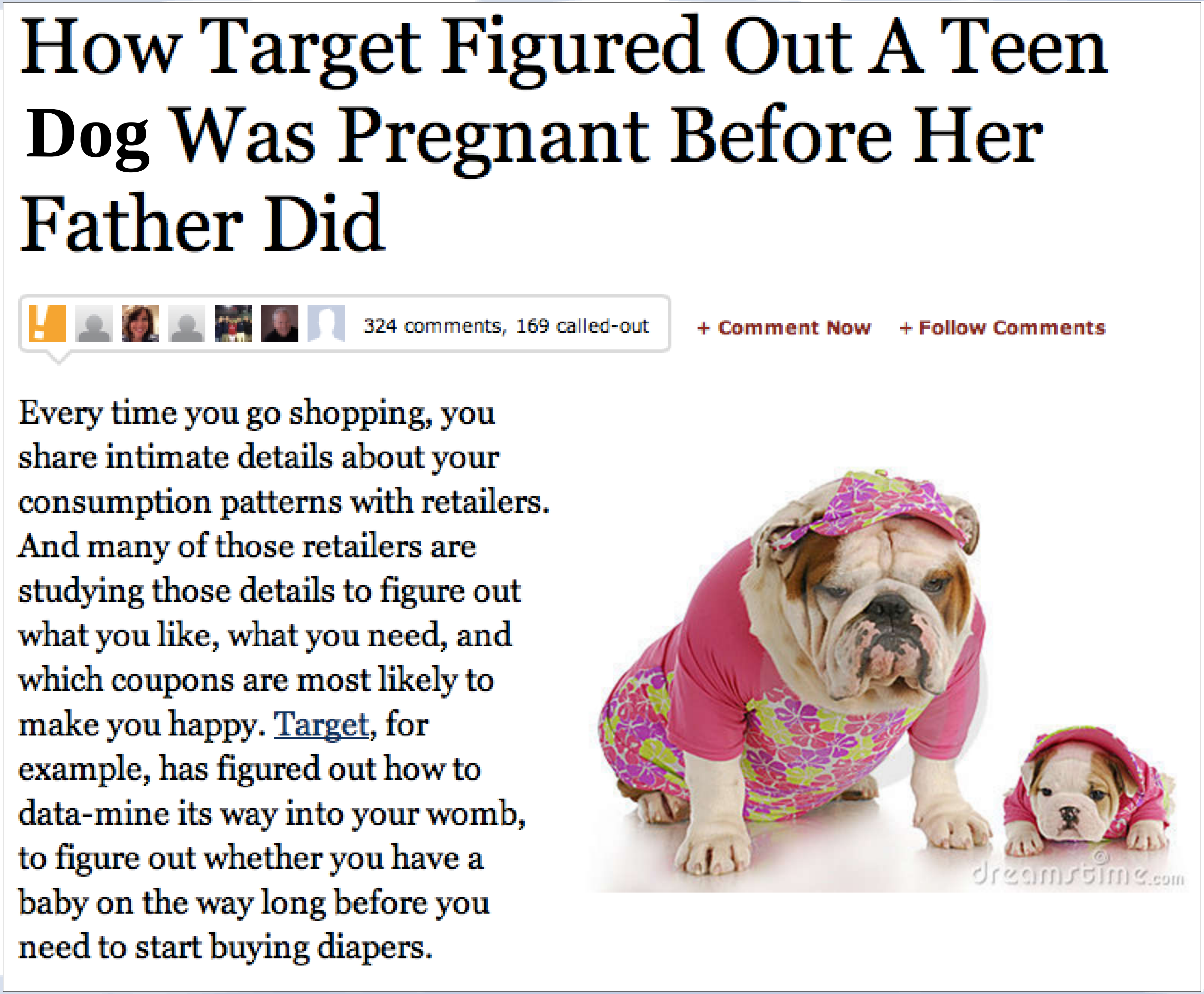
Pet Supplies: See all 25,595 items

Only 19 left in stock - order soon.

More Buying Choices

$5.65 new (19 offers)

... based on Rover's Dogbook likes



How Target Figured Out A Teen Dog Was Pregnant Before Her Father Did

324 comments, 169 called-out

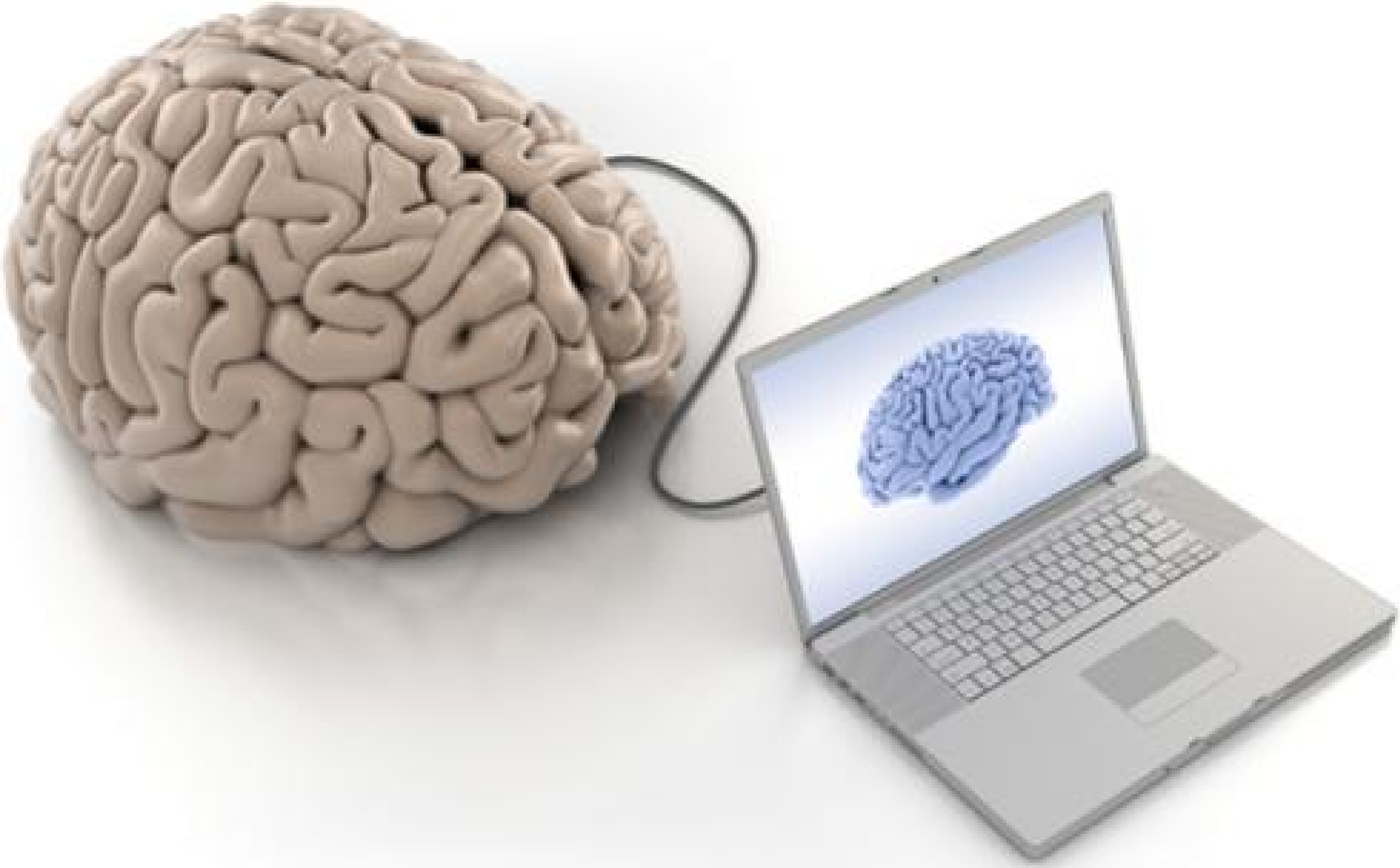
+ Comment Now + Follow Comments

Every time you go shopping, you share intimate details about your consumption patterns with retailers. And many of those retailers are studying those details to figure out what you like, what you need, and which coupons are most likely to make you happy. Target, for example, has figured out how to data-mine its way into your womb, to figure out whether you have a baby on the way long before you need to start buying diapers.

Why should I care about privacy if I have nothing to hide?

Reason 1: Freedom of Thought

- We invented computers to help us think.



Reason 1: Freedom of Thought

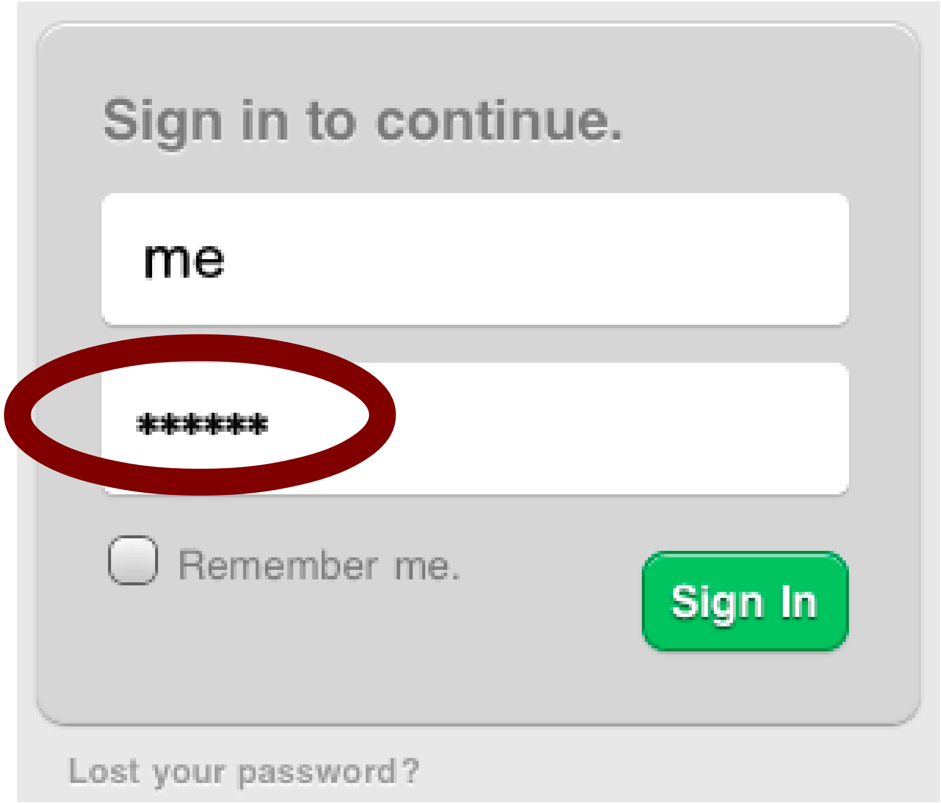
- We invented computers to help us think.
- Ubiquity brings dependence

Reason 1: Freedom of Thought

- We invented computers to help us think.
- Ubiquity brings dependence
- Whoever can read your private data can read your thoughts

Reason 2: Personal Security

You think *this* is your password?



Sign in to continue.

me

******

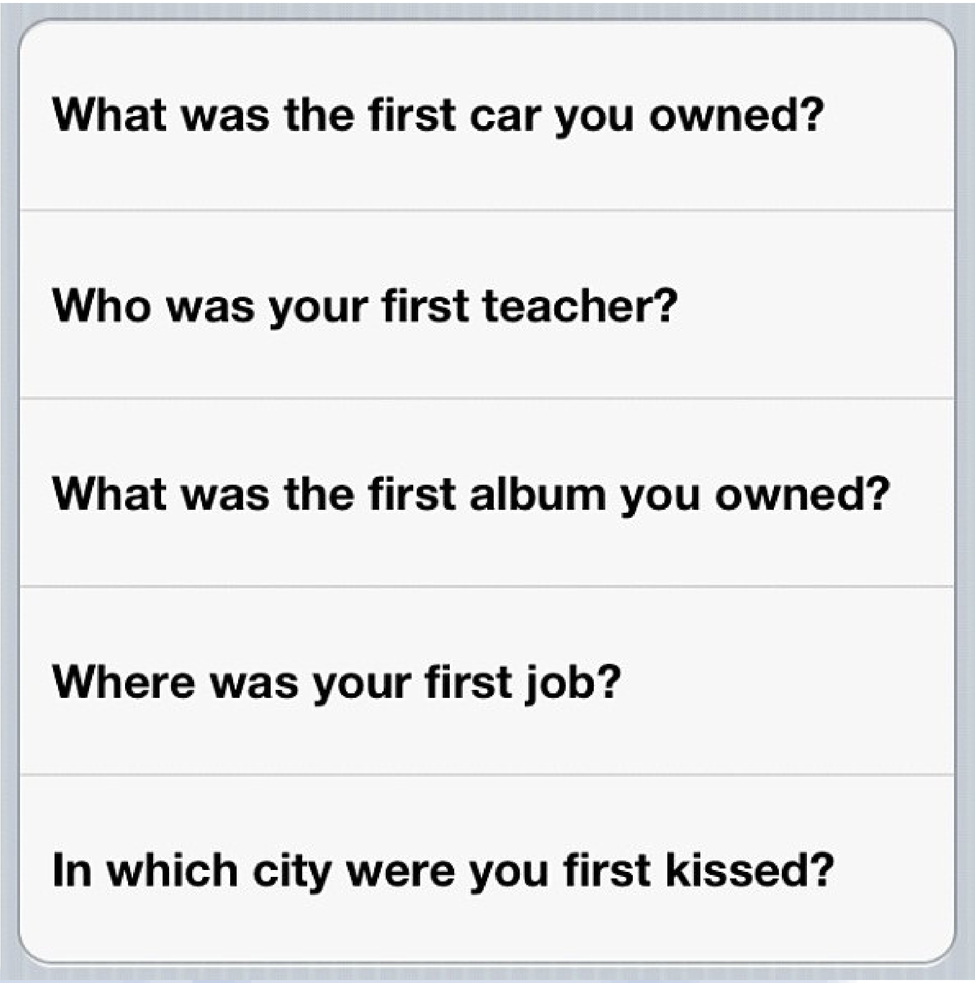
□ Remember me.

Reason 2: Personal Security

No, that's just a temporary access token.

This is your password.

Your life is
your password.



What was the first car you owned?

Who was your first teacher?

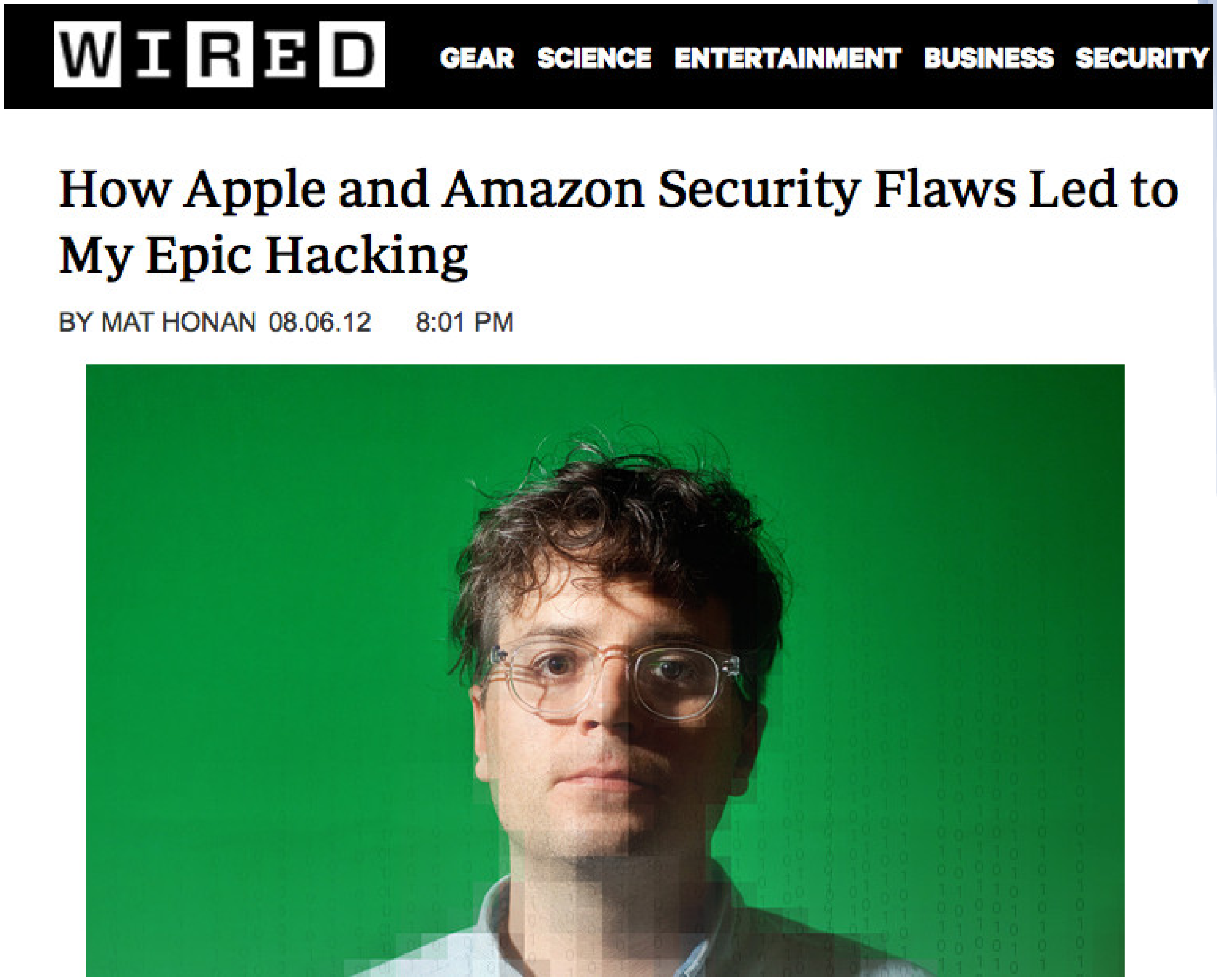
What was the first album you owned?

Where was your first job?

In which city were you first kissed?

Reason 2: Personal Security

Whoever can data-mine your life has your password



How Apple and Amazon Security Flaws Led to My Epic Hacking

BY MAT HONAN 08.06.12 8:01 PM

Who Wants to Track You Online?

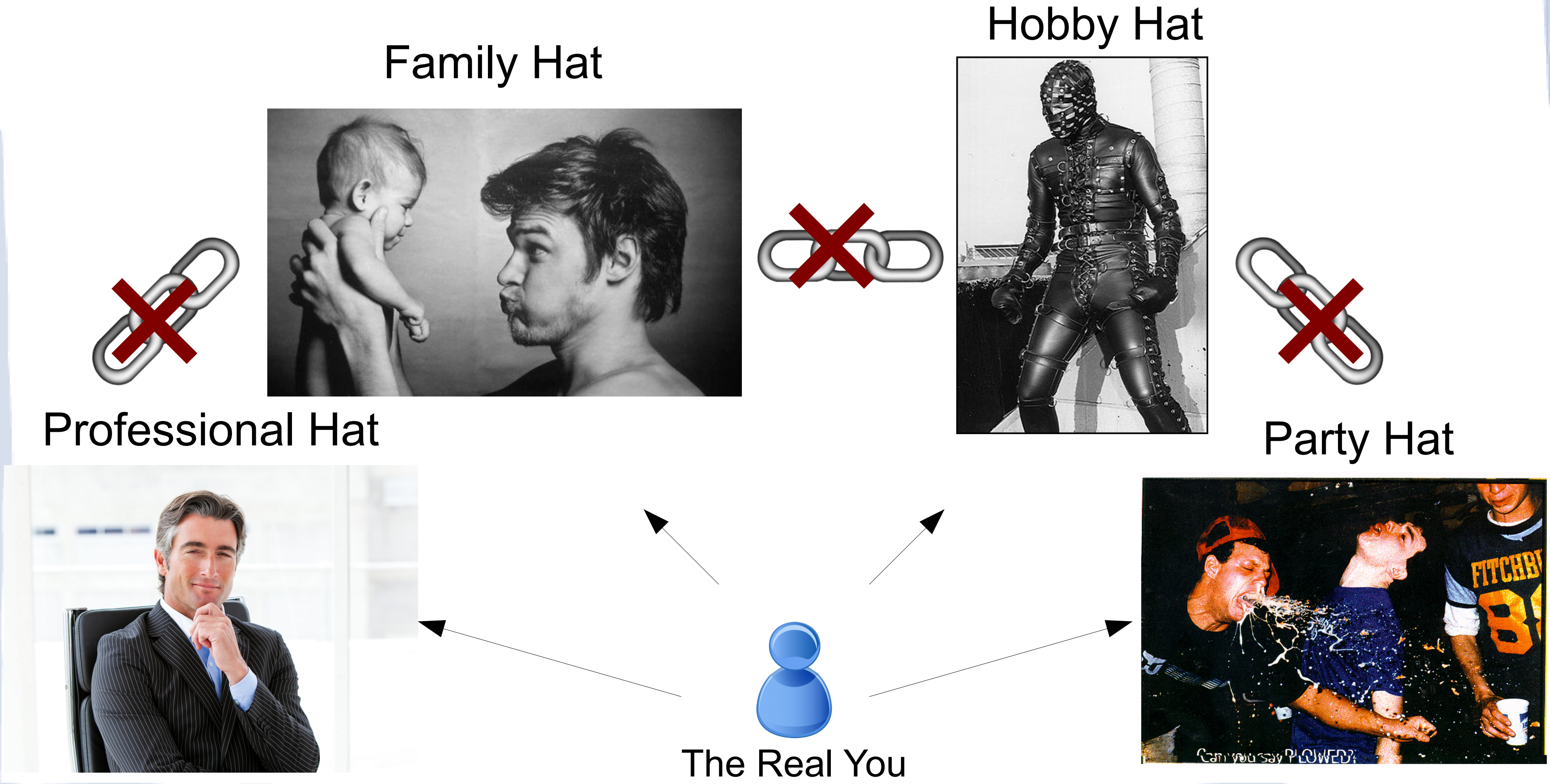
- Advertisers (if you ever spend money)
- Vendors (if you ever buy things)
- Thieves (if you have any money)
- Stalkers (if you're a domestic abuse victim)
- Competitors (if you're a business)
- Extremists (if you're minority/gay/pro-choice...)
- The Police (if you're “of interest” w/in 3 hops)
- The Mob (if you're the police)

What tracking protection do we need?

Some people really need anonymity...

What tracking protection do we need?

Many people just tend to wear multiple hats



Family Hat

Hobby Hat

Professional Hat

Party Hat

The Real You

Talk Outline

✓ Why Anonymity?
- **Current State of the Art**
- Grand Challenges in Anonymity
  - Global traffic analysis
  - Active interference attacks
  - Intersection attacks
  - De-anonymizing exploits
  - Accountability provisions
- Conclusion

What protection can we get now?

Many weak defense options

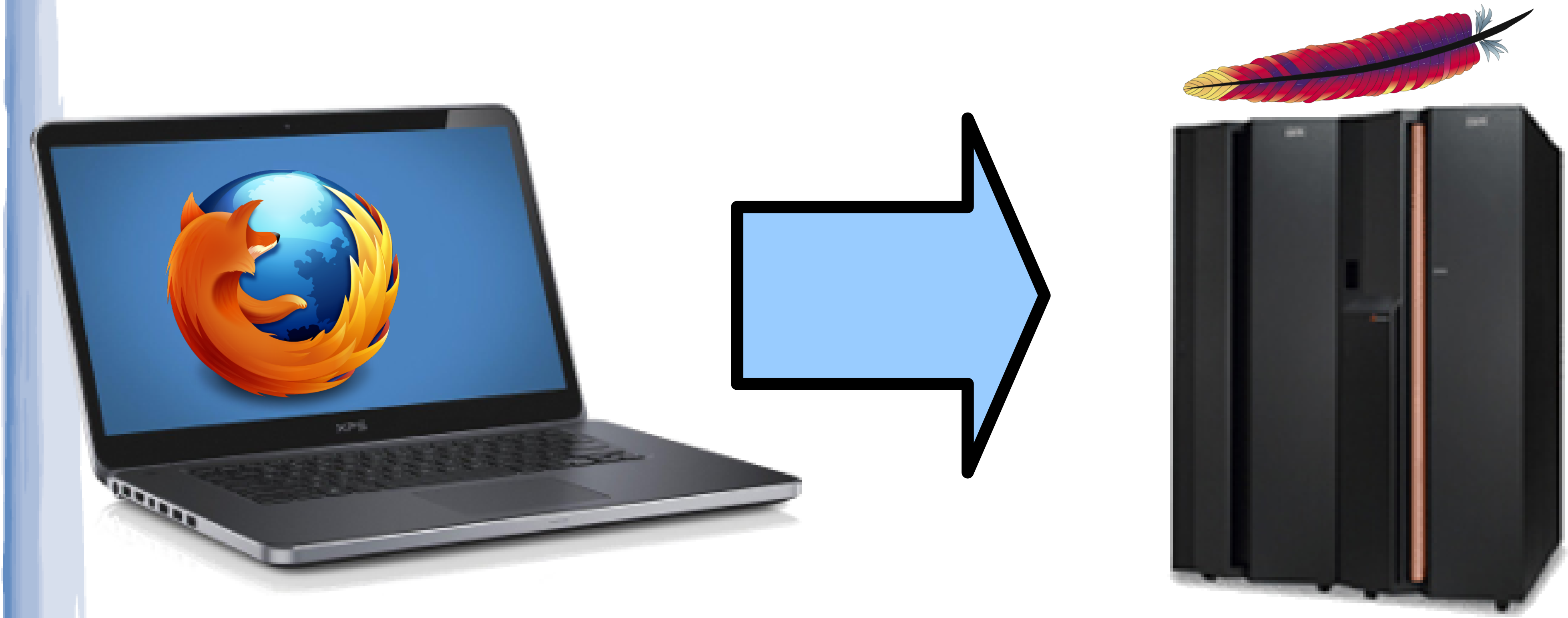
- Disable cookies, browser history, Flash, Java
- “Do-Not-Track” HTTP option
- “Hide” behind NATs, firewalls, corporate VPNs
- Commercial proxy/VPN providers

Current state-of-the-art

- Onion routing systems – e.g., Tor

Do Not Track

Universal Web Tracking Opt Out



```
GET /something/here HTTP/1.1
Host: example.com
DNT: 1
```

Do Not Track

Universal Web Tracking Opt Out

Please don't track me,
pretty please???

Of course we'll
respect your privacy –
promise!

International New York Times

Business Day

Technology

SLIPSTREAM

Do No

By NATASHA

Published: Oc

THE cam

last mont

Topic: Privacy

Compare

Follow via:

Why

miser

Summary: *A*

the big data-co

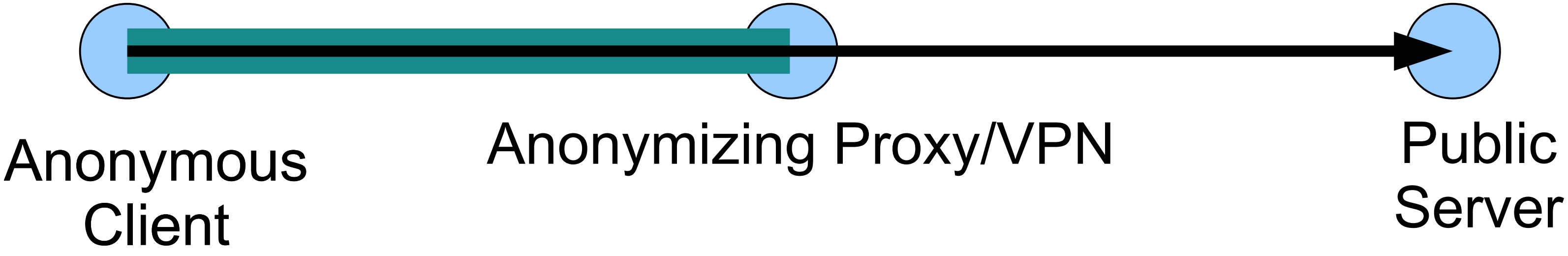
nothing at all.

By

Commercial VPN services

Popular for circumventing the Great Firewall

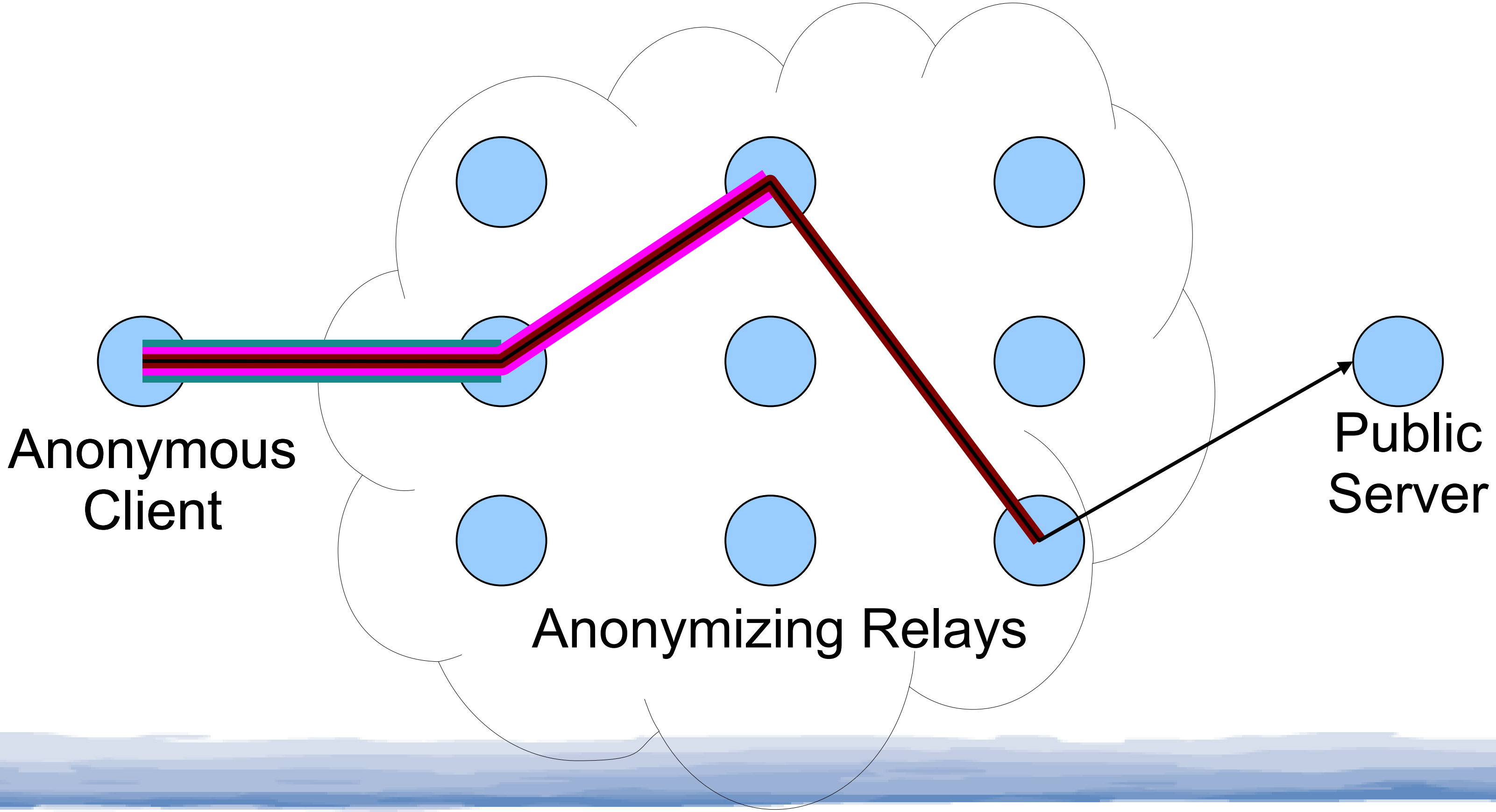
- You build encrypted tunnel with VPN server
- VPN server forwards traffic to destination
- Looks like it's coming from VPN server
- Hope the server operator protects your privacy



The current state-of-the-art

Onion routing tools such as **Tor**

- **https://www.torproject.org**



A more tracking-resistant Internet?

A few choices:

- Extend NAT & proxy protocols to support pseudonyms [Han, SIGCOMM '13]
- Make Tor an Internet standard [Talbot, Nov '13]
- Explore new architectures for anonymity and tracking protection

Rest of this talk focuses on last approach

The Dissent Project

Clean-slate anonymous communications design

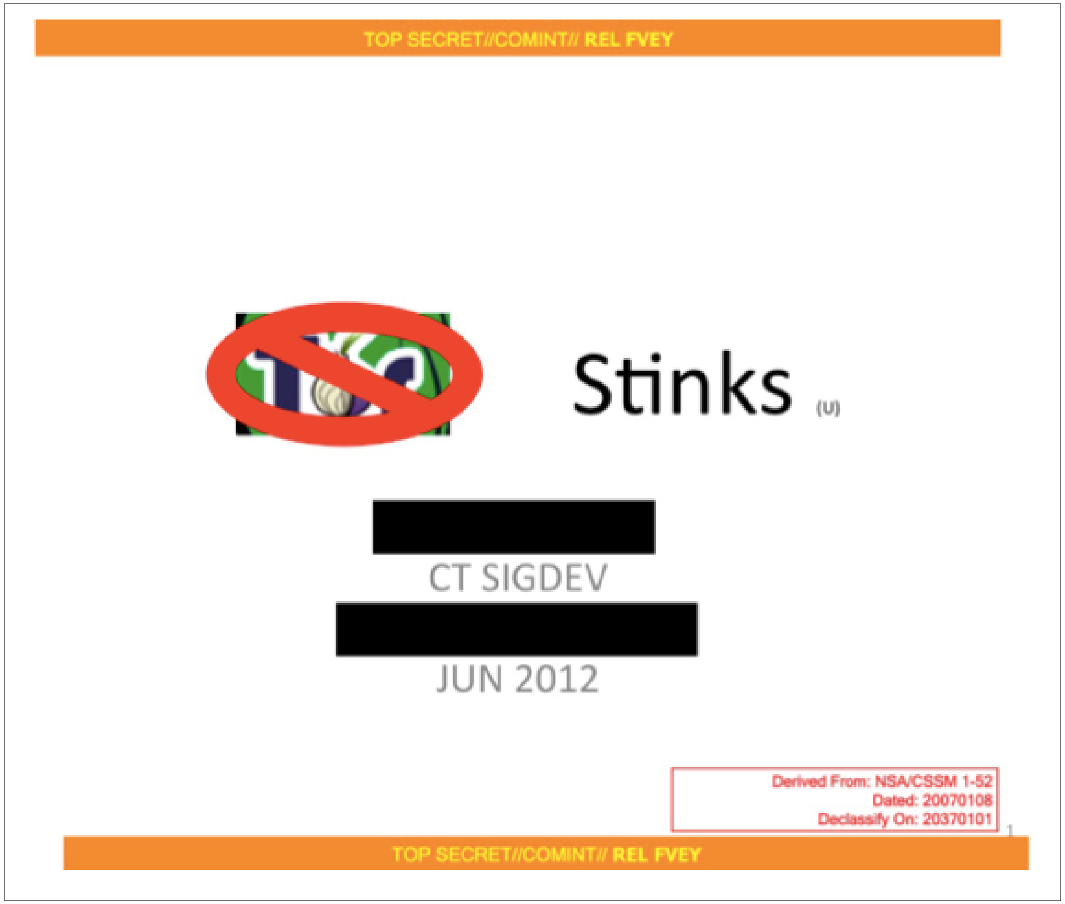
- Offer *quantifiable* and *measurable* anonymity
- Build on primitives offering *provable security*
- Don't just *patch* specific vulnerabilities, but *rearchitect* to address whole *attack classes*

http://dedis.cs.yale.edu/dissent/

[CCS'10, OSDI'12, CCS'13, USENIX Sec'13, ...]

Why rethink online anonymity?

NSA says Tor is the “King of Anonymity” – maybe onion routing is good enough?



TOP SECRET//COMINT// REL FVEY

Stinks (U)

[redacted]

CT SIGDEV

[redacted]

JUN 2012

Derived From: NSA/CSSM 1-52
Dated: 20070108
Declassify On: 20370101

1

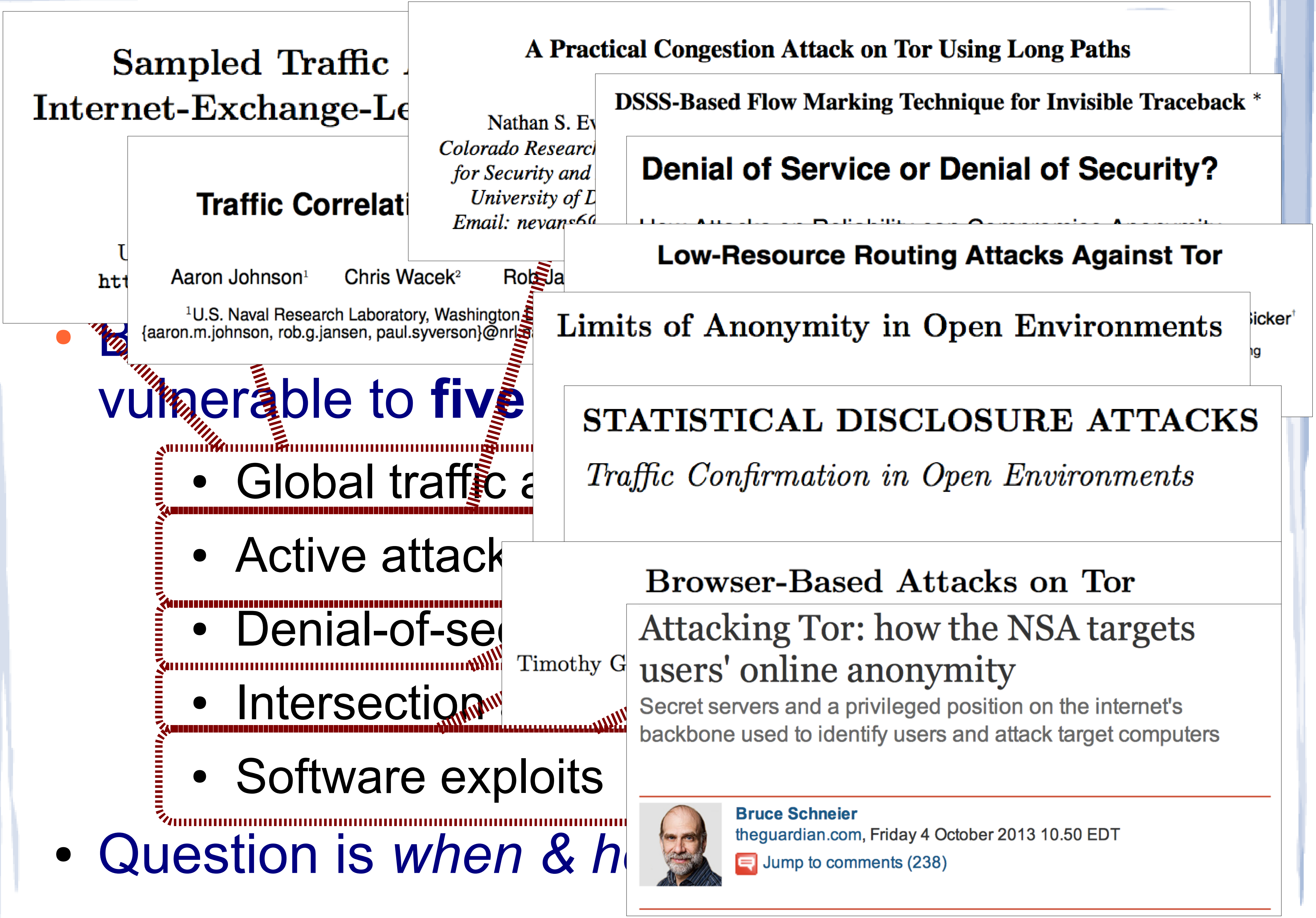
TOP SECRET//COMINT// REL FVEY

- B

  vulnerable to five

  - Global traffic a
  - Active attack
  - Denial-of-se
  - Intersection
  - Software exploits

- Question is *when & h*



Sampled Traffic

Internet-Exchange-Le

U

htt

Traffic Correlati

Aaron Johnson[1] Chris Wacek[2] Rob Ja

[1]U.S. Naval Research Laboratory, Washington

{aaron.m.johnson, rob.g.jansen, paul.syverson}@nrl

A Practical Congestion Attack on Tor Using Long Paths

Nathan S. Ev

Colorado Researc

for Security and

University of D

Email: nevans6

DSSS-Based Flow Marking Technique for Invisible Traceback *

Denial of Service or Denial of Security?

Low-Resource Routing Attacks Against Tor

icker[†]

ng

Limits of Anonymity in Open Environments

STATISTICAL DISCLOSURE ATTACKS

Traffic Confirmation in Open Environments

Browser-Based Attacks on Tor

Timothy G

Attacking Tor: how the NSA targets users' online anonymity

Secret servers and a privileged position on the internet's backbone used to identify users and attack target computers

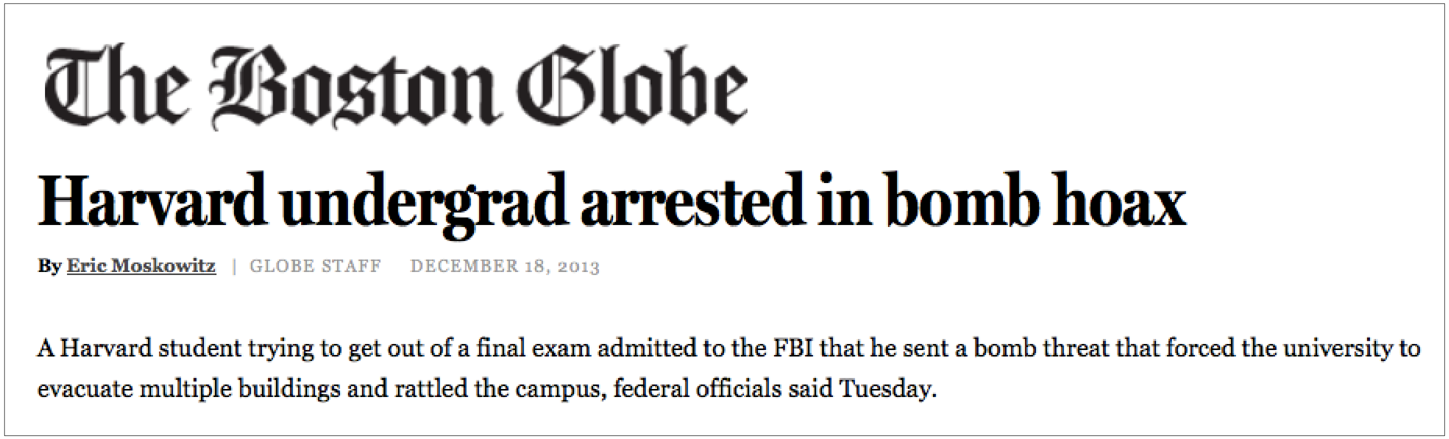
Bruce Schneier

theguardian.com, Friday 4 October 2013 10.50 EDT

Jump to comments (238)

Recent De-anonymization Incidents

Tor is being broken – or *circumvented* – regularly

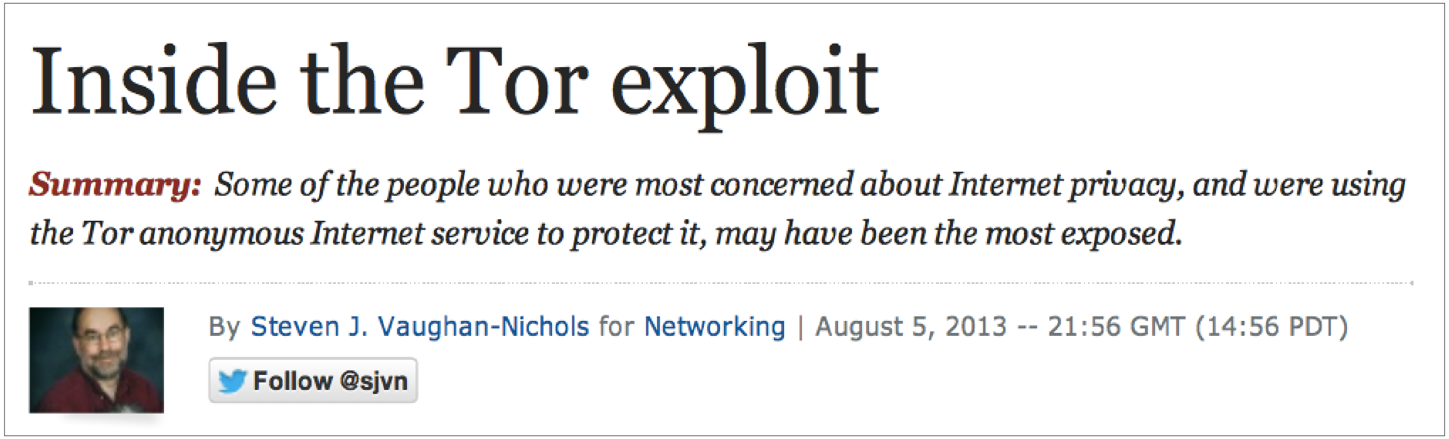


The Boston Globe

Harvard undergrad arrested in bomb hoax

By Eric Moskowitz | GLOBE STAFF DECEMBER 18, 2013

A Harvard student trying to get out of a final exam admitted to the FBI that he sent a bomb threat that forced the university to evacuate multiple buildings and rattled the campus, federal officials said Tuesday.

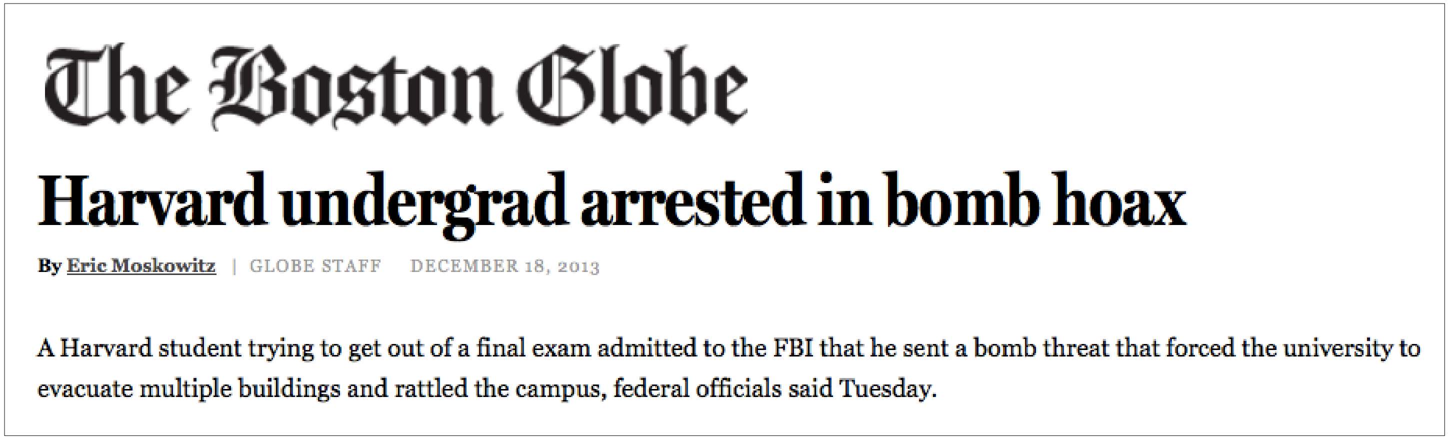


Inside the Tor exploit

Summary: *Some of the people who were most concerned about Internet privacy, and were using the Tor anonymous Internet service to protect it, may have been the most exposed.*

By Steven J. Vaughan-Nichols for Networking | August 5, 2013 -- 21:56 GMT (14:56 PDT)

Recent De-anonymization Incidents



The Boston Globe

Harvard undergrad arrested in bomb hoax

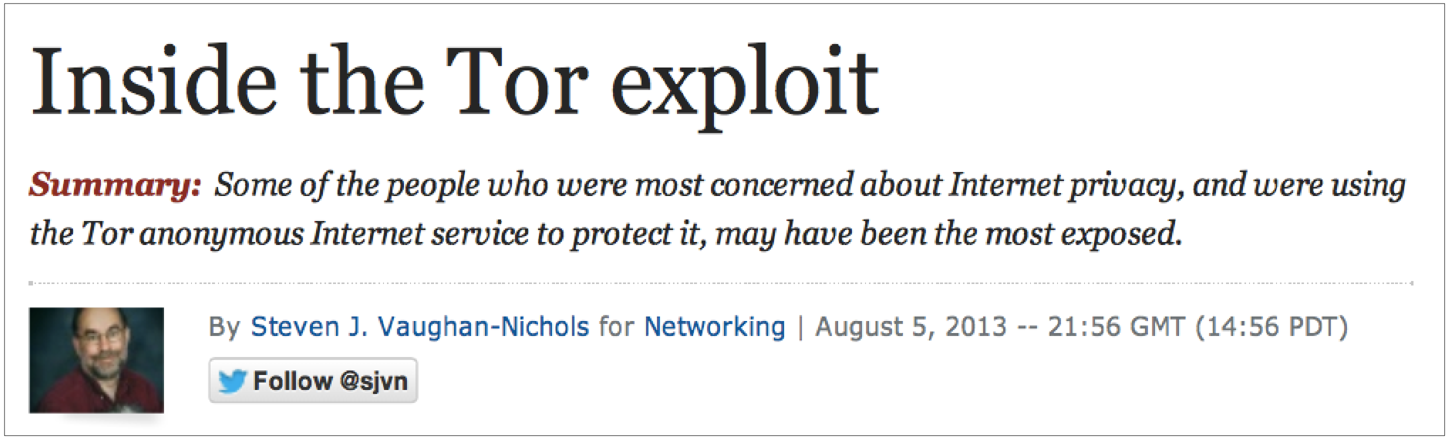
By Eric Moskowitz | GLOBE STAFF DECEMBER 18, 2013

A Harvard student trying to get out of a final exam admitted to the FBI that he sent a bomb threat that forced the university to evacuate multiple buildings and rattled the campus, federal officials said Tuesday.

Lessons from the bomb hoax:

- Traffic analysis attacks are effective
- Intersection attacks are effective
- Anonymity systems need *accountability:* more graceful deterrents to abuse

Recent De-anonymization Incidents



Inside the Tor exploit

Summary: *Some of the people who were most concerned about Internet privacy, and were using the Tor anonymous Internet service to protect it, may have been the most exposed.*

By Steven J. Vaughan-Nichols for Networking | August 5, 2013 -- 21:56 GMT (14:56 PDT)

Lessons from the Tor exploit:

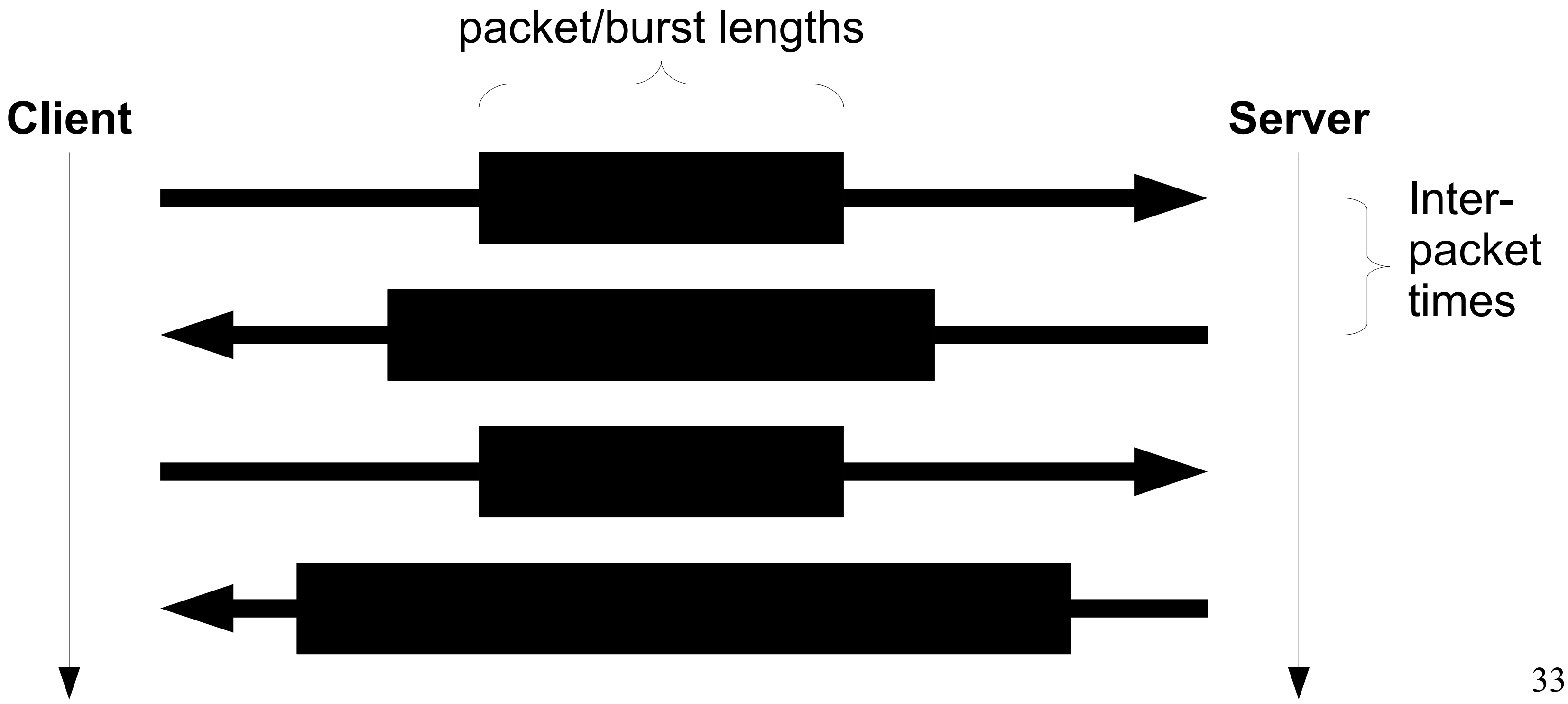
- Client OS isolation model is just as important as tracking-resistance protocols themselves
- Long-term anonymity requires resistance to malware, stains, beacons of all kinds

Talk Outline

- ✓ Why Anonymity?
- ✓ Current State of the Art
- **Grand Challenges in Anonymity**
  - Global traffic analysis
  - Active interference attacks
  - Intersection attacks
  - De-anonymizing exploits
  - Accountability provisions
- Conclusion

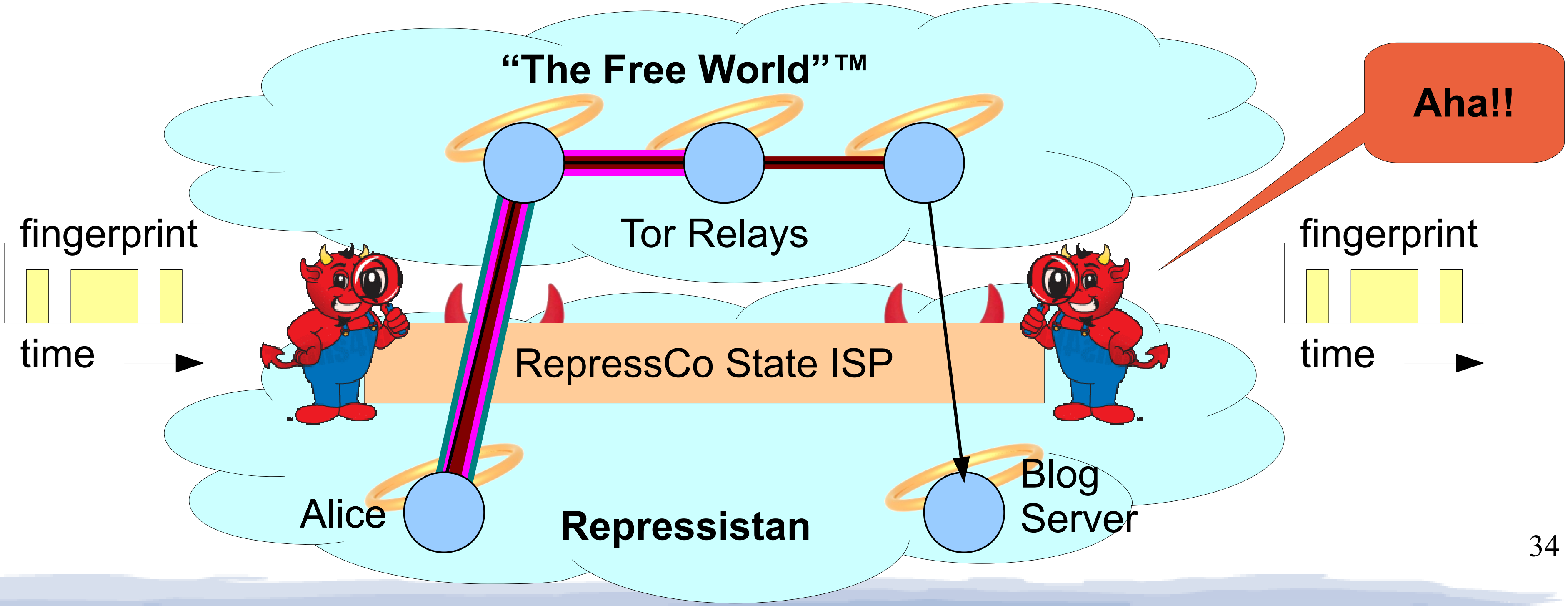
The Traffic Analysis Problem

- Most communication has a *traffic pattern*
  - Lengths and timings of packets in each direction
  - Pattern can be *fingerprinted* without seeing content



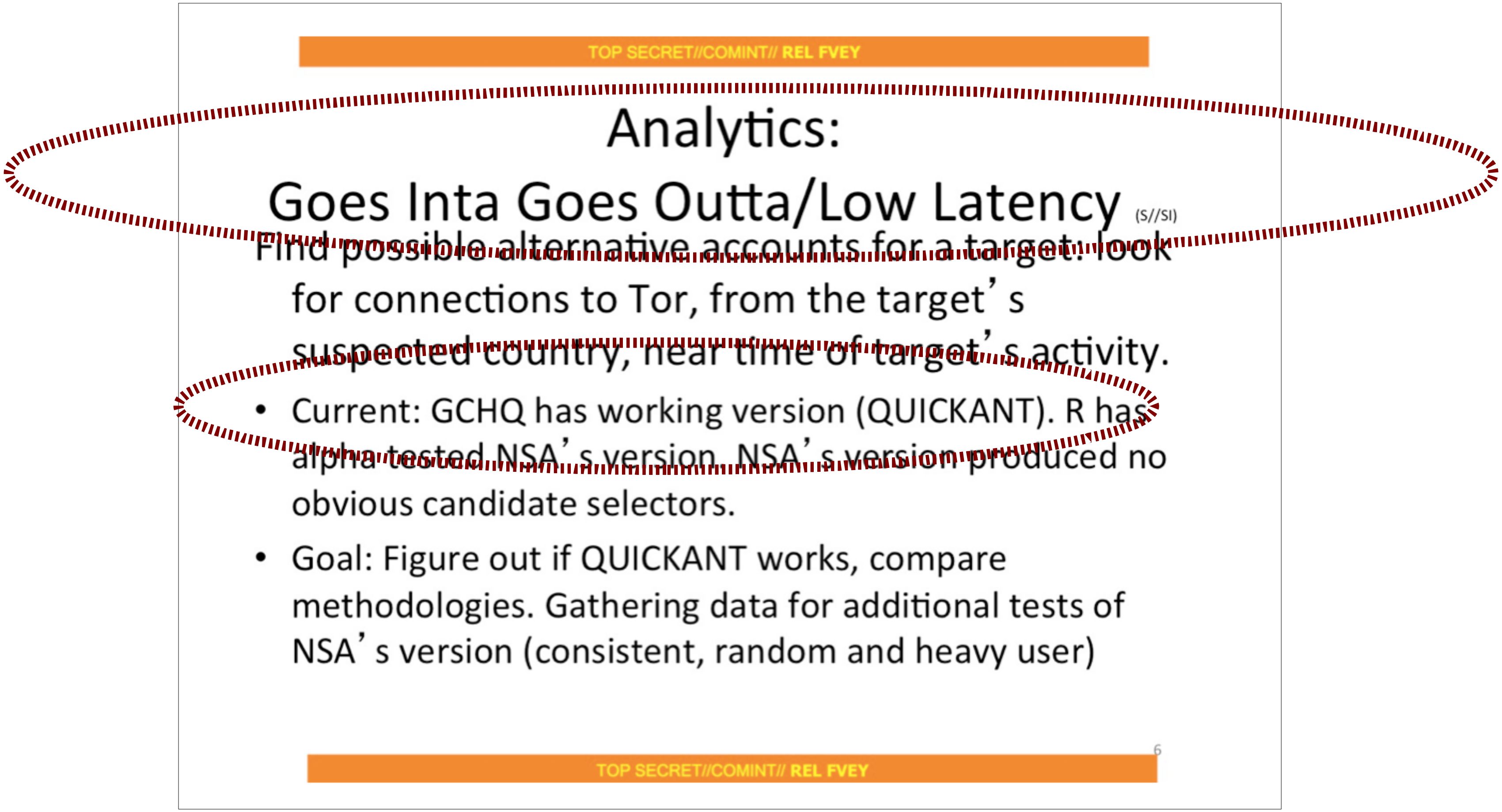
Tor Traffic Analysis Scenario

- Alice in Repressistan uses Tor to post on blog server hosted in Repressistan
- State ISP controls *both* entry and exit hops
- Fingerprint & correlate traffic to **deanonymize**



Do Attackers Actually *Do* This?

Not sure, but some are *working hard on it...*



TOP SECRET//COMINT// REL FVEY

Analytics:

Goes Inta Goes Outta/Low Latency (S//SI)

- Find possible alternative accounts for a target: look for connections to Tor, from the target's suspected country, near time of target's activity.
- Current: GCHQ has working version (QUICKANT). R has alpha-tested NSA's version. NSA's version produced no obvious candidate selectors.
- Goal: Figure out if QUICKANT works, compare methodologies. Gathering data for additional tests of NSA's version (consistent, random and heavy user)

6

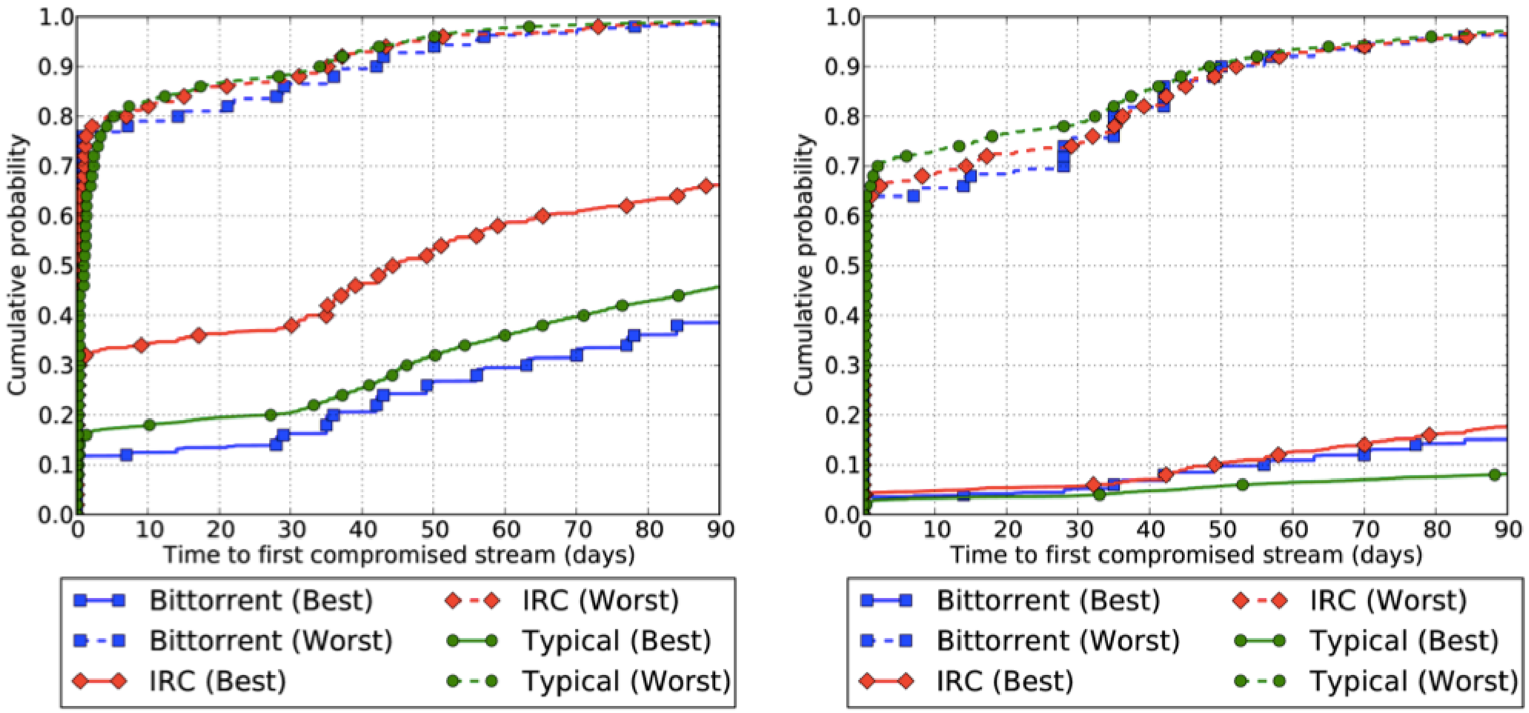
TOP SECRET//COMINT// REL FVEY

("Tor Stinks" slide deck, Guardian 10/4/2013)

Can De-Anonymize "Real" Users?

Yes, if attacker can monitor an Internet AS or IXP

- "Users Get Routed", Johnson et al. CCS 13



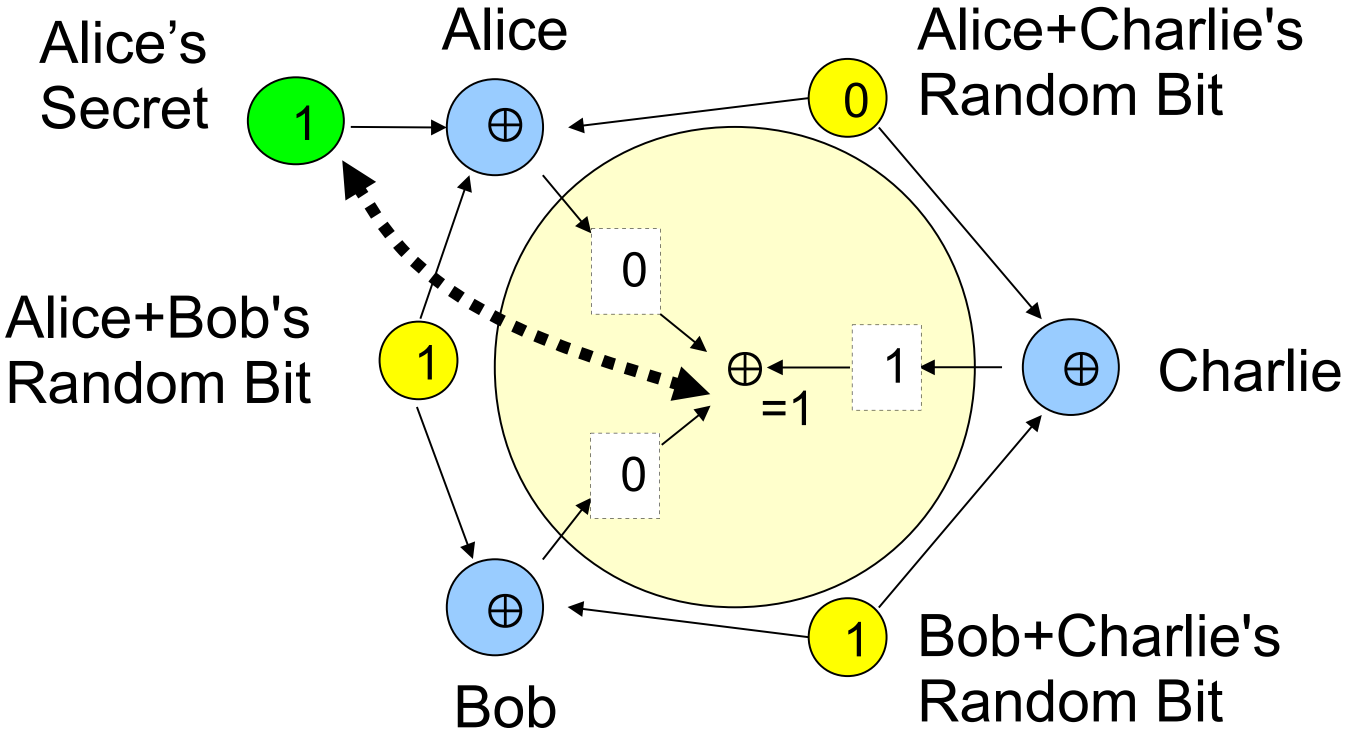
(a) Time to first stream compromised by AS adversary.

(b) Time to first stream compromised by IXP adversary.

Dining Cryptographers (DC-nets)

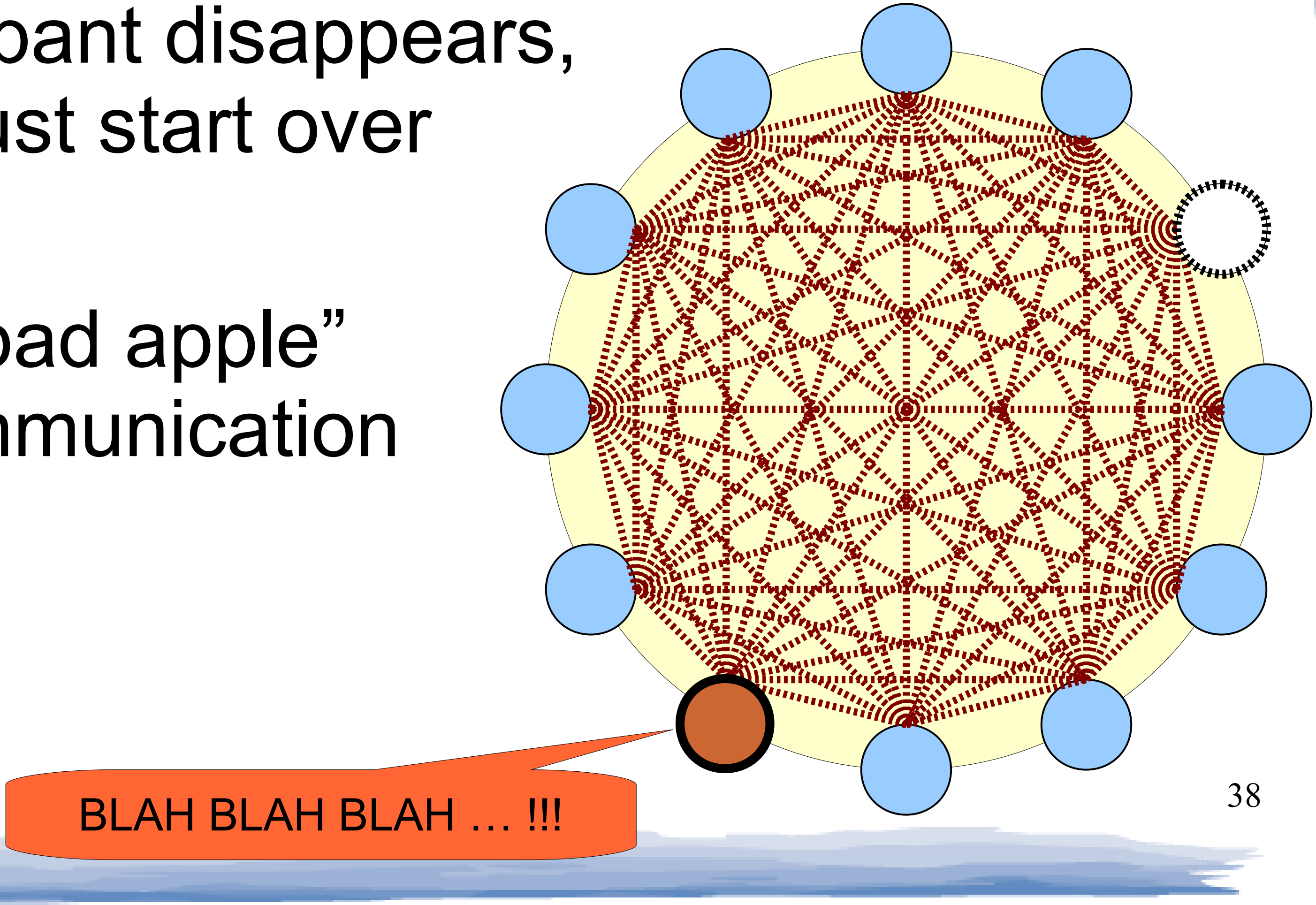
Another fundamental Chaum invention from the 80s...

- Key property: provable anonymity within a group



Why DC-nets Doesn't Scale

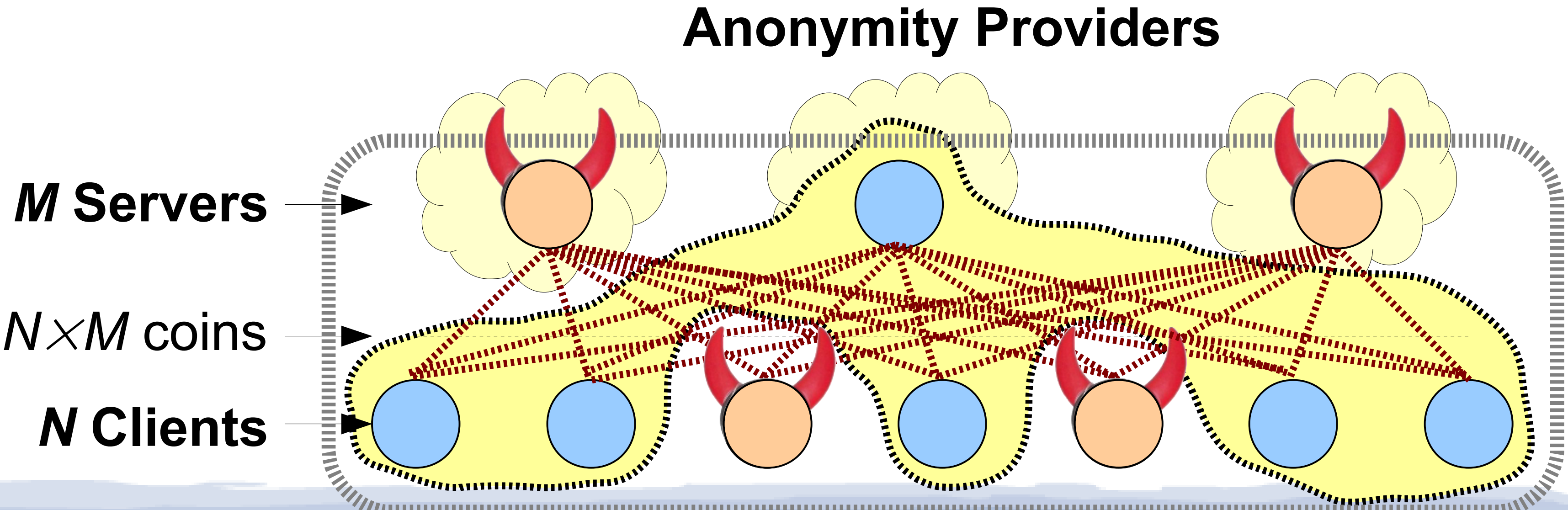
- **Computation cost:** $N \times N$ shared coin matrix
- **Network churn:**
  if *any* participant disappears,
  all nodes must start over
- **Disruption:**
  any single “bad apple”
  can jam communication



“Dissent in Numbers” [OSDI 12]

Scalable DC-nets using client/multi-server model

- Clients share coins *only* with servers
- As long as *at least one* honest server *exists*, yields ideal anonymity among *all honest clients*

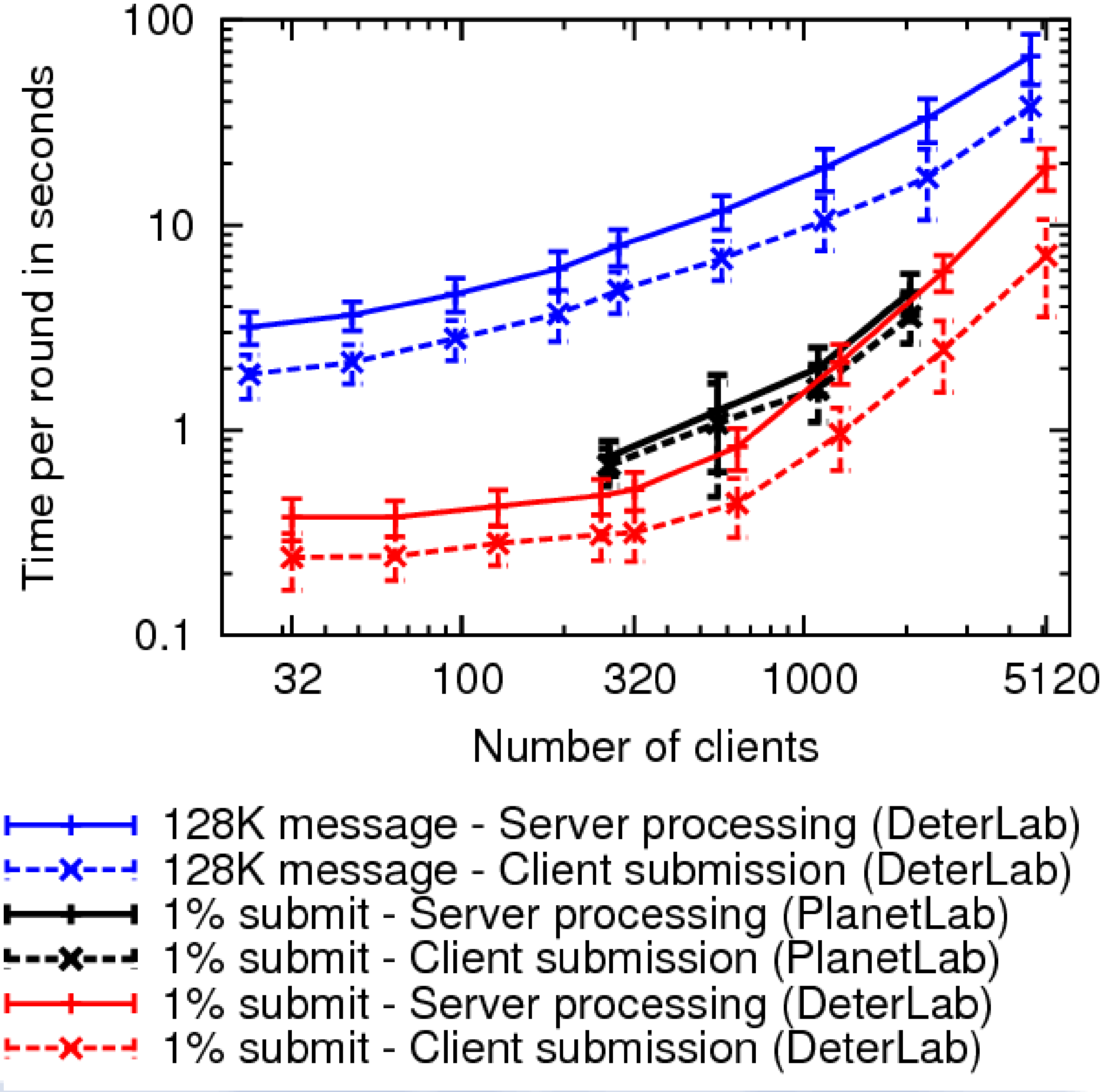


Scaling to Thousands of Clients

100× larger anonymity sets

- (Herbivore, Dissent v1: ~40 clients)

<1 sec latency w/ 1000 clients



Talk Outline

- ✓ Why Anonymity?
- ✓ Current State of the Art
- **Grand Challenges in Anonymity**
  - ✓ Global traffic analysis
  - **Active interference attacks**
  - Intersection attacks
  - De-anonymizing exploits
  - Accountability provisions
- Conclusion

The Ups and Downs of Diversity

Tor has a highly **diverse** worldwide user base

- Diverse types of users, countries, languages
- Diverse reasons for using Tor

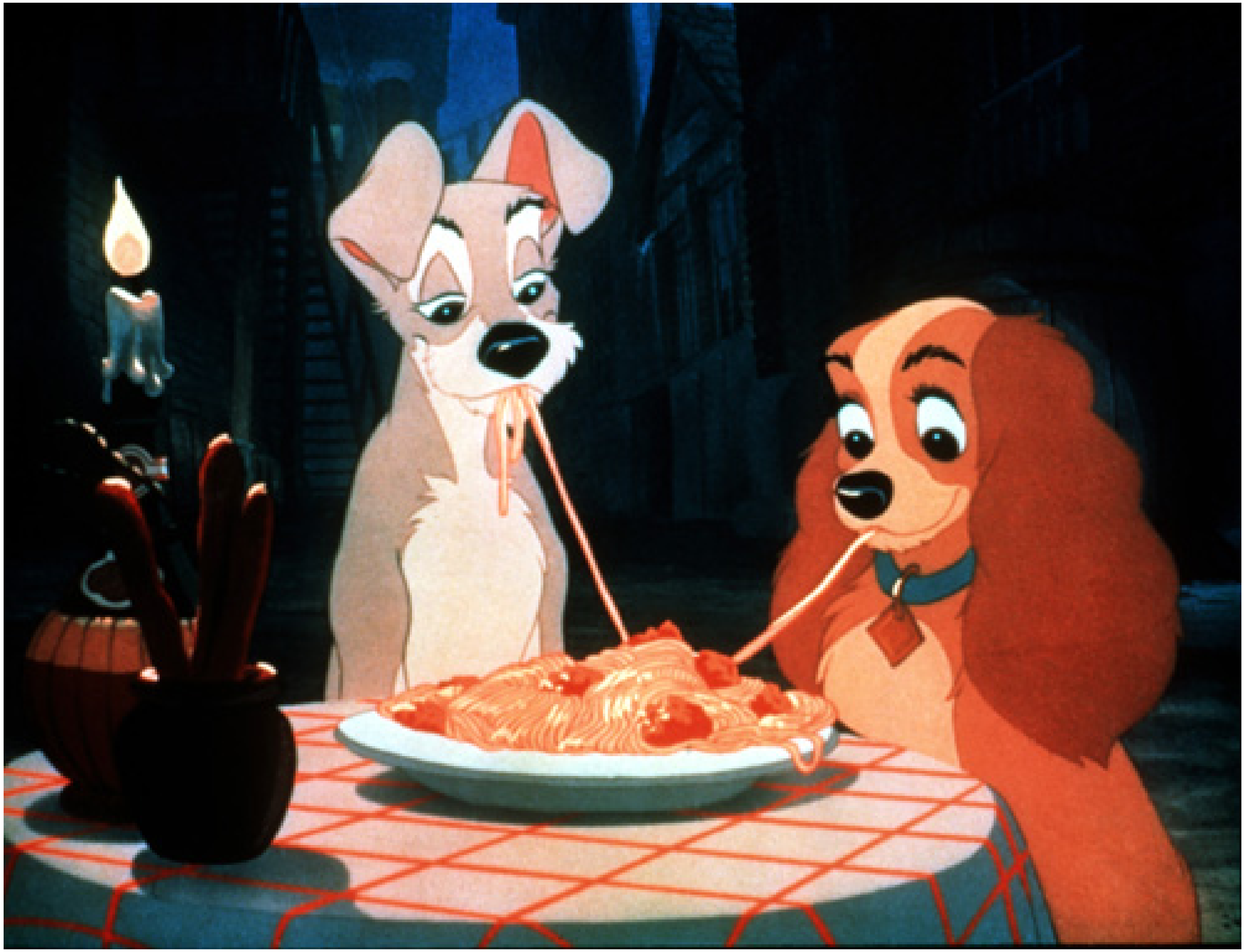
This diversity is crucial for **the Tor system...**
but no **individual user** gets all that “anonymity”

- Most excluded due to location, time, language...
- There is no meaningful anonymity
  except within a meaningful community
  of users who might plausibly behave like you

Tor hides you in a tangle of wires...

...or a plate of spaghetti

But tug on either end of a strand...

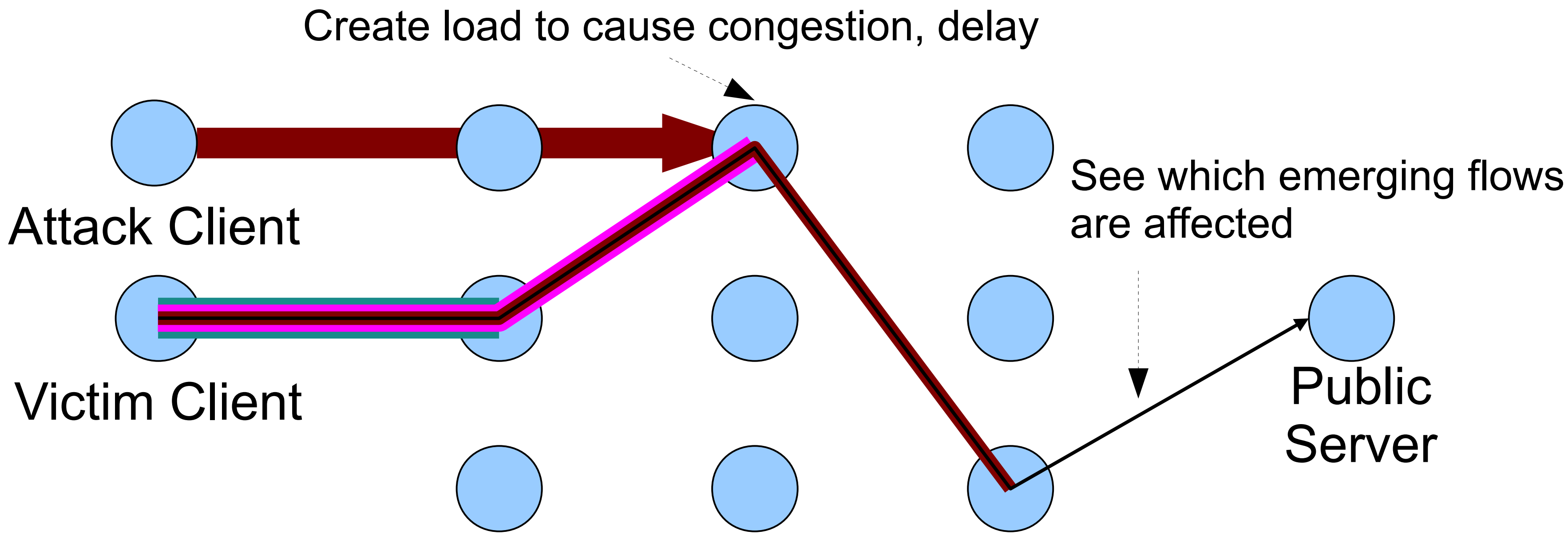


...and you'll find the other

Active Attacks

Attacker perturbs performance to inject traceable side-channel “markers” into flows

- Example: “congestion attacks” against Tor (e.g., Murdoch 05, Evans 09)


Create load to cause congestion, delay
Attack Client
See which emerging flows are affected
Victim Client
Public Server

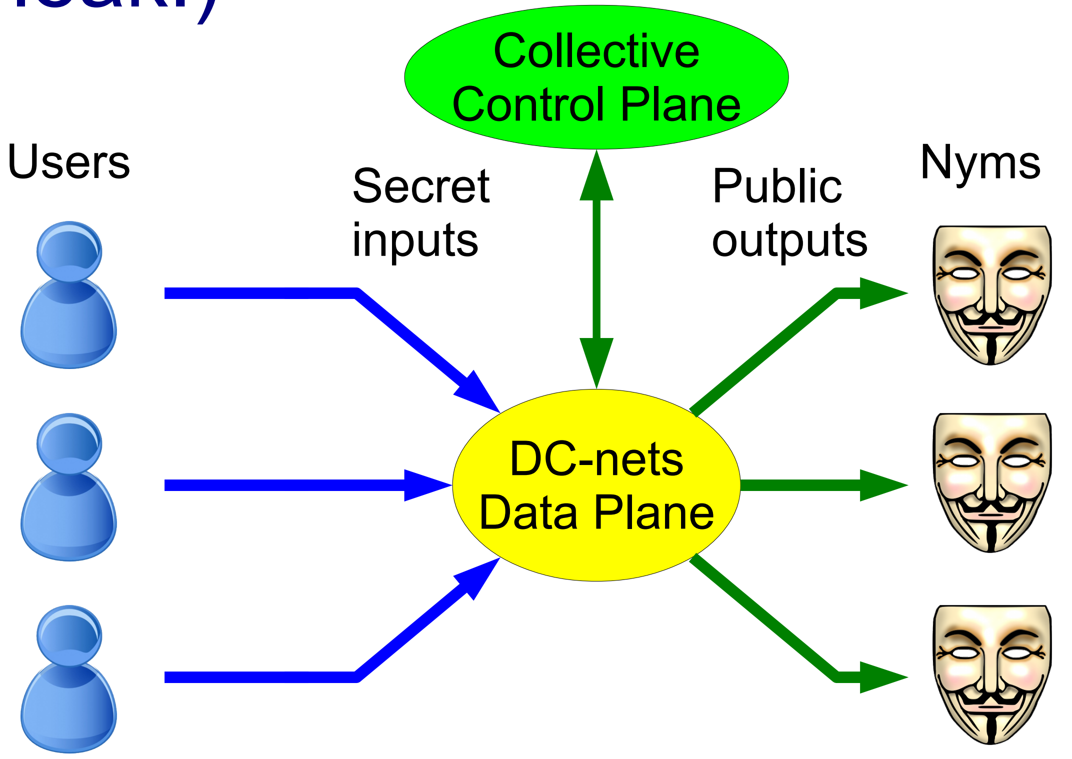
“Community-oriented Anonymity?”

Goal: build strength from groups of *like-minded* users engaging in *collective* activities...

Collective Control Plane (CCP) Model

Policy Oracle controls when/how much to send

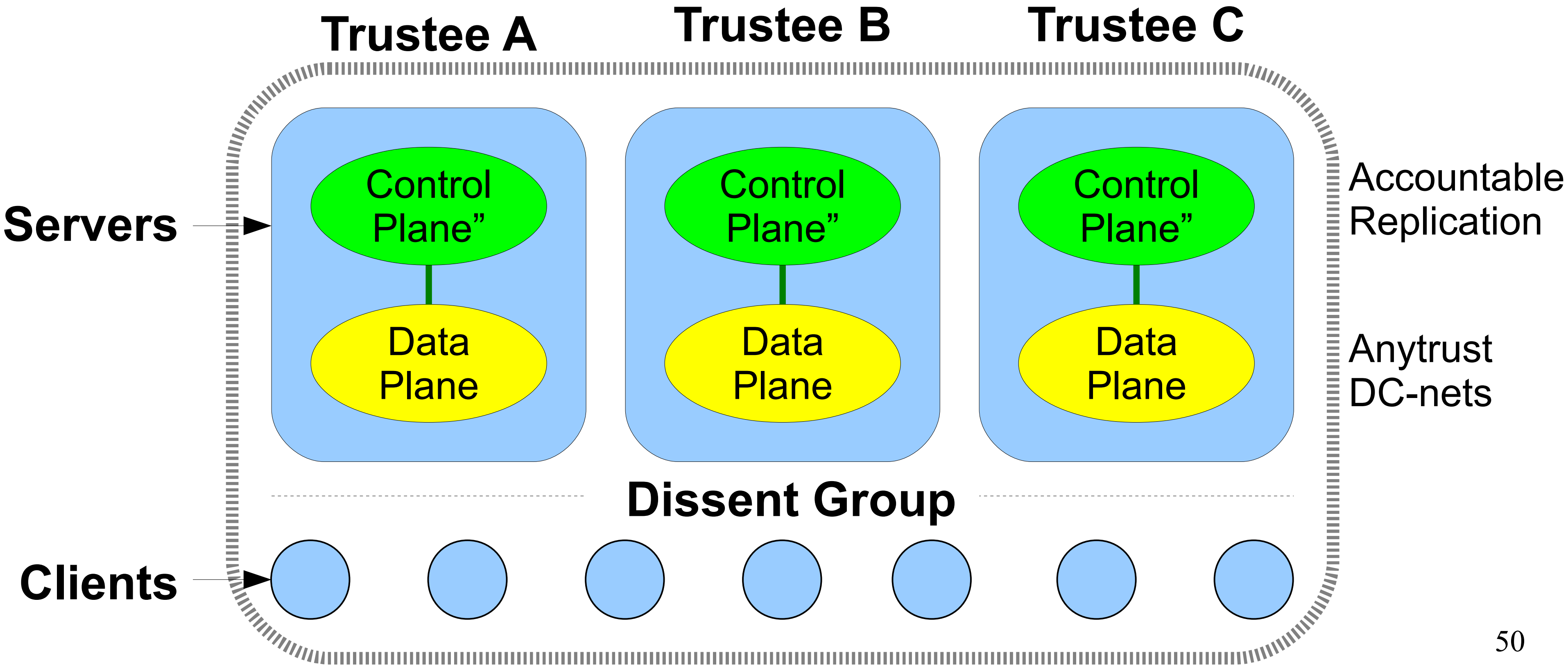
- But *does not know* who owns which nyms (can't leak!)



Implementing the CCP

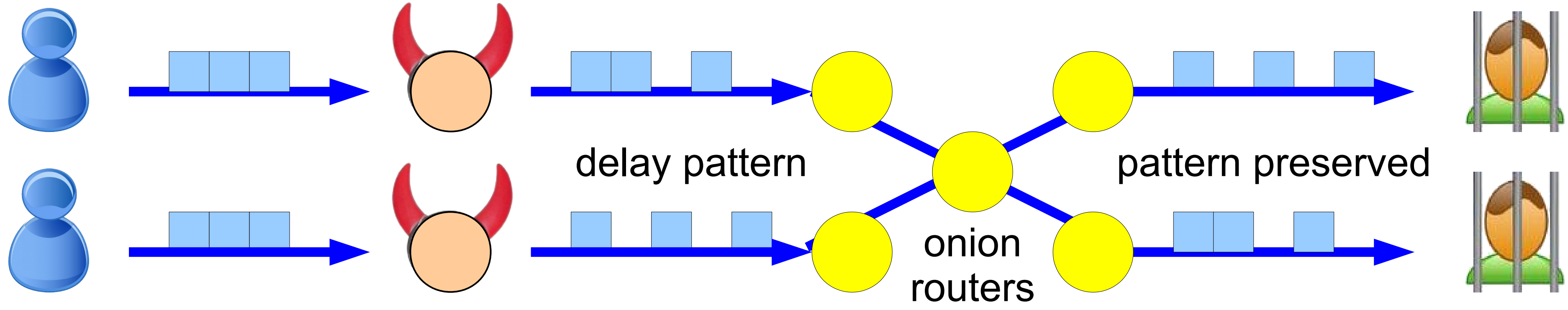
Accountable replication of control plane logic

- Each server implements copy, all must agree

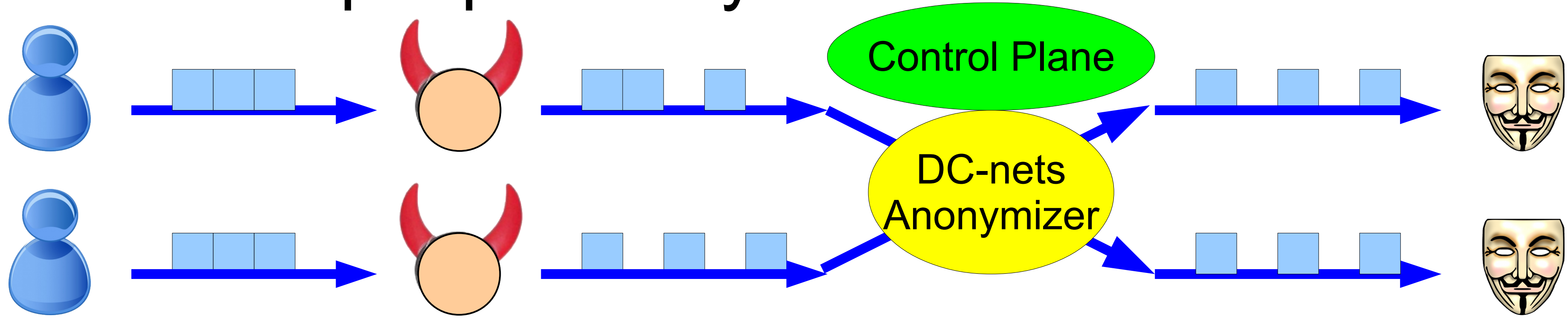


How CCP Counters Active Attacks

Onion routing preserves *individual* flow properties:



Dissent output paced by *collective* control:

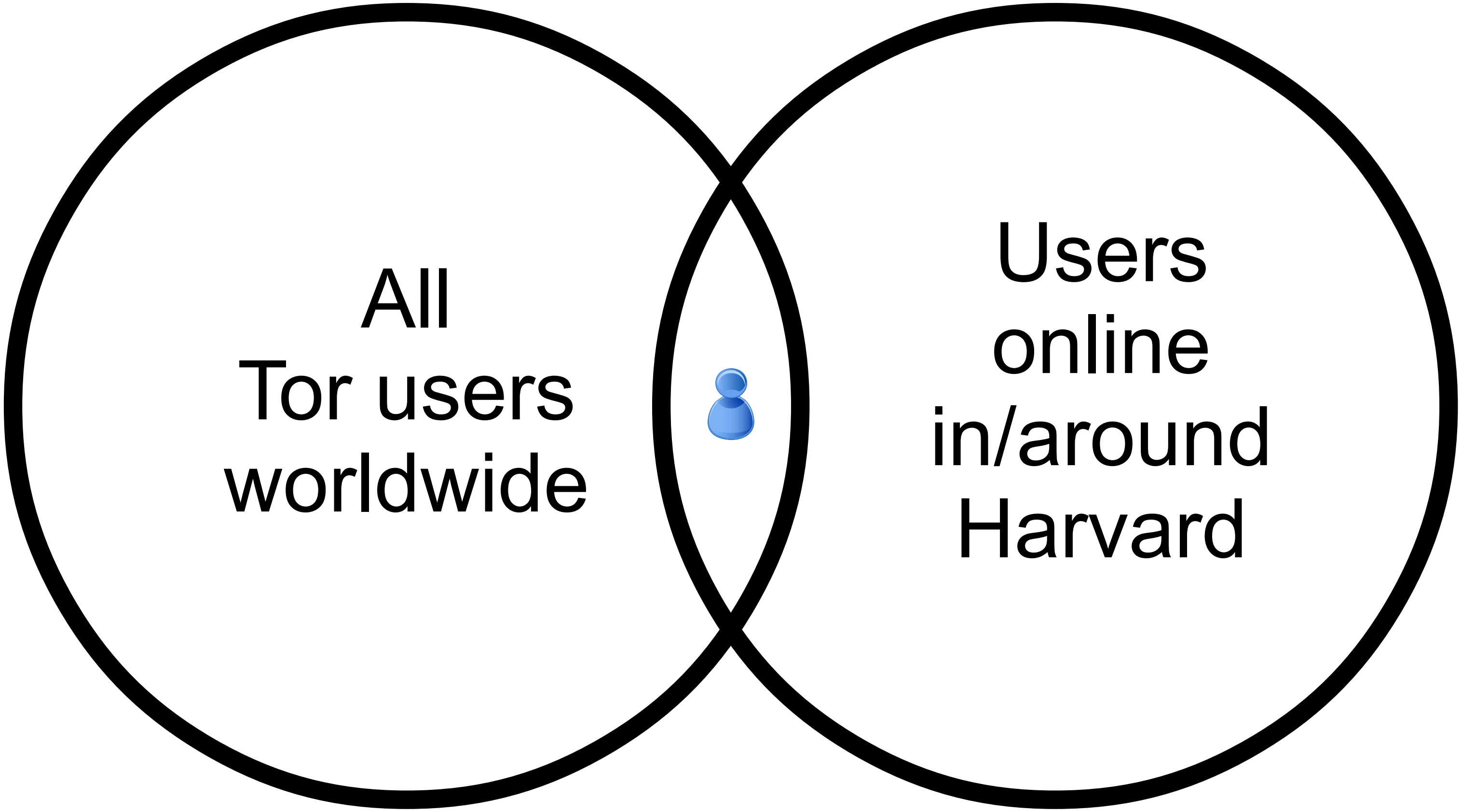


Talk Outline

- ✓ Why Anonymity?
- ✓ Current State of the Art
- **Grand Challenges in Anonymity**
  - ✓ Global traffic analysis
  - ✓ Active interference attacks
  - **Intersection attacks**
  - De-anonymizing exploits
  - Accountability provisions
- Conclusion

The Bomb Hoax Attack

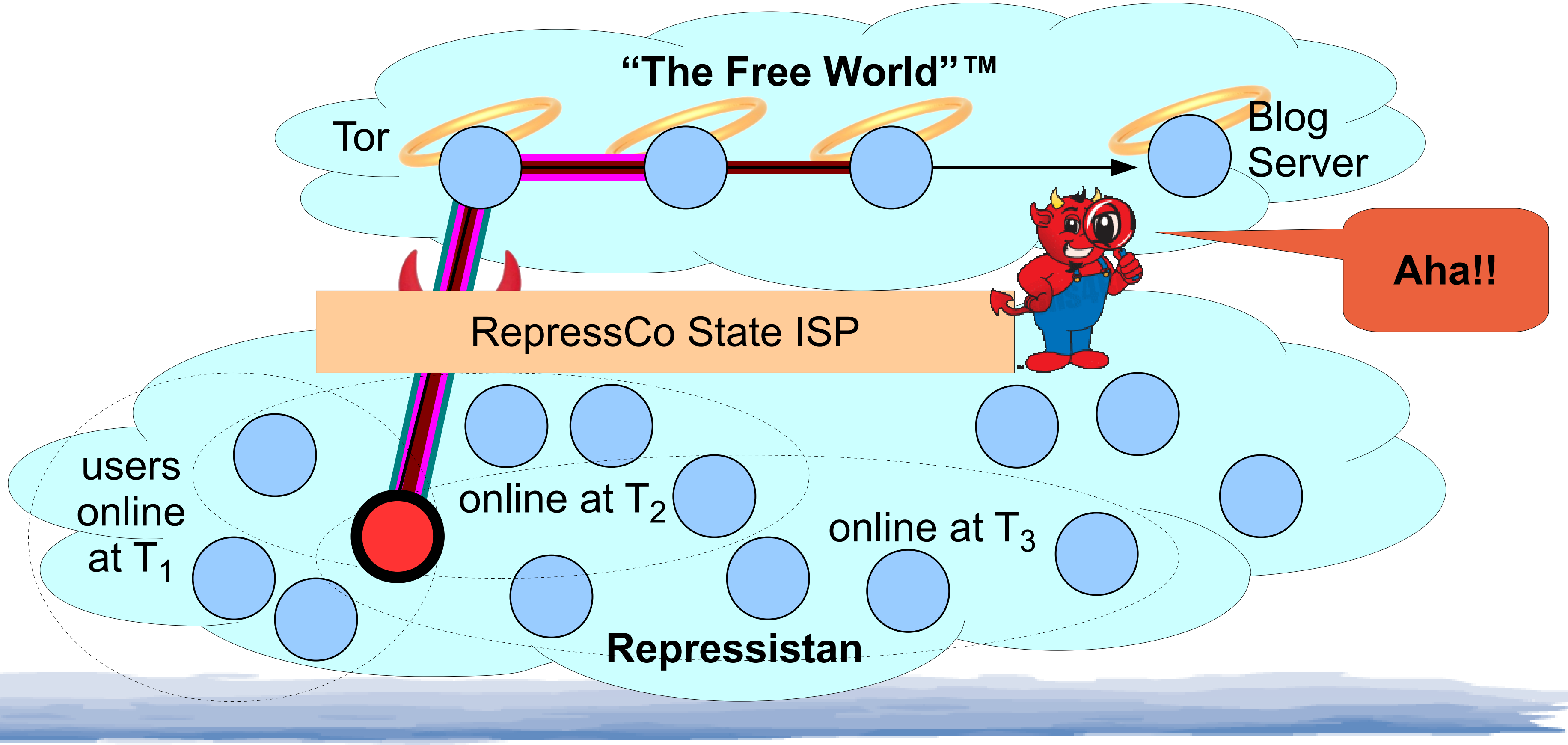
The Harvard bomb hoaxer was de-anonymized by a specially trivial intersection attack



The Intersection Attack Problem

Kate signs posts with pseudonym “Bob”

- Posts signed messages at times T_1, T_2, T_3
- Police **intersects** user sets online each time



Buddies [CCS '13]

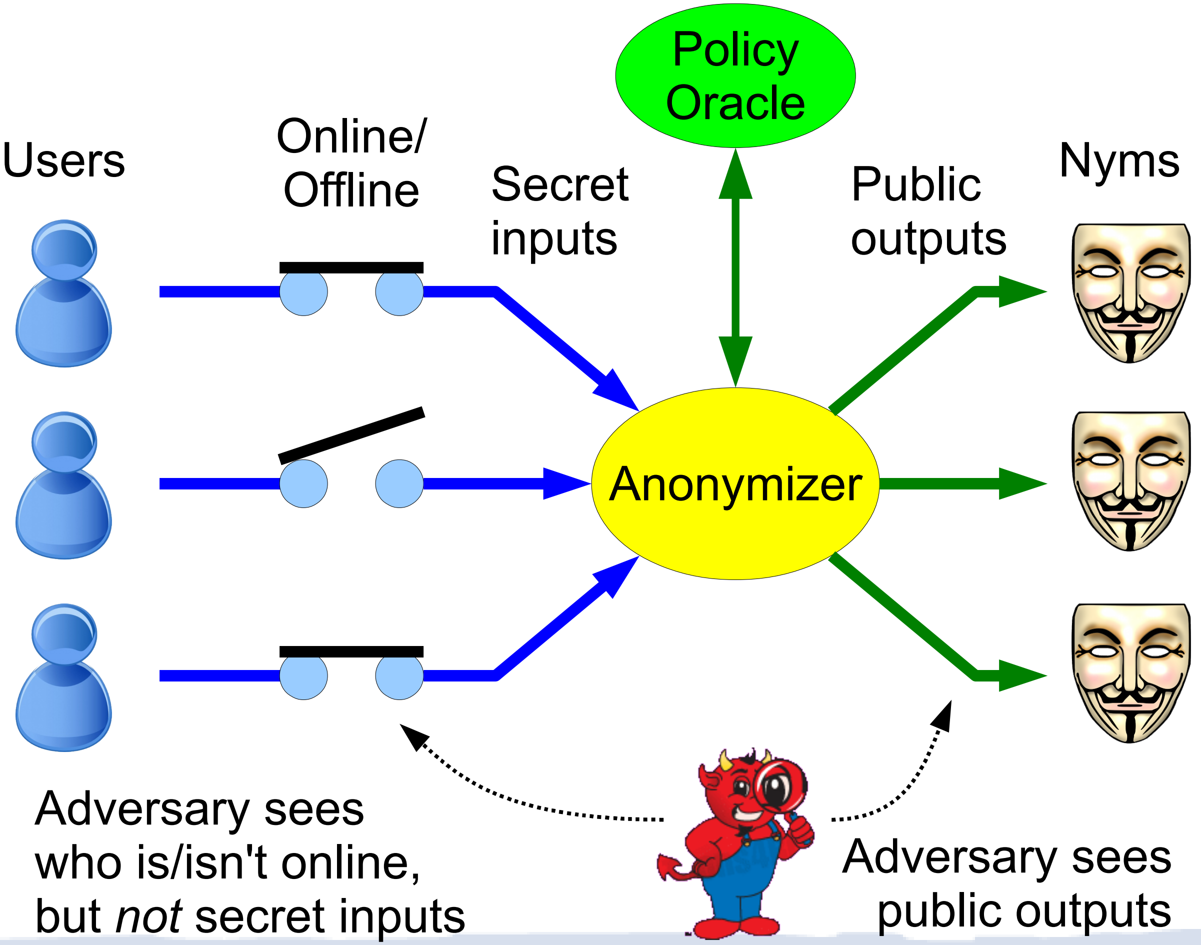
First attempt at building intersection attack resistance into a practical anonymity system

Goals:

- *Measure* anonymity under intersection attack
- Actively *mitigate* anonymity loss
- Enforce *lower bounds* by trading availability

Buddies Conceptual Model

Focus: what adversary learns from *online status*



Policy Oracle
Users
Online/ Offline
Secret inputs
Public outputs
Nyms
Anonymizer
Adversary sees who is/isn't online, but *not* secret inputs
Adversary sees public outputs

Computing Anonymity Metrics

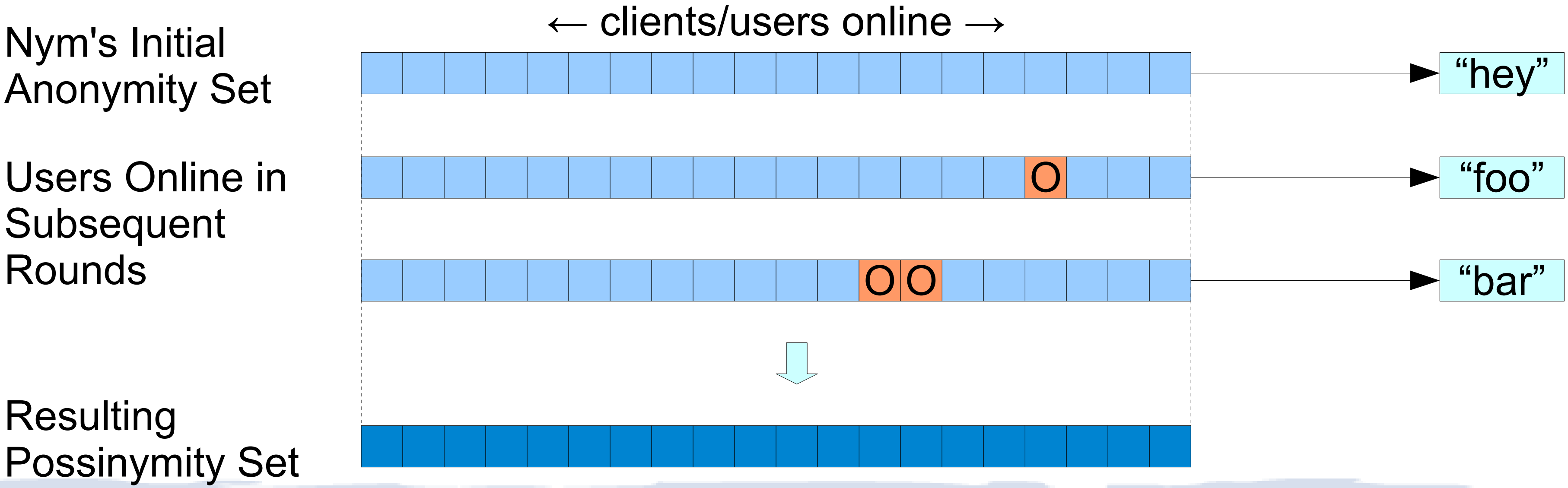
Policy Oracle *simulates an adversary's view*

- Knows who's online each round (via “tags”)
- Performs “intersection attacks” against Nyms
- Computes anonymity metrics
  - **Possinymity:** “possibilistic deniability”
  - **Indinymity:** “probabilistic indistinguishability”
- Reports metrics, uses them in policy decisions

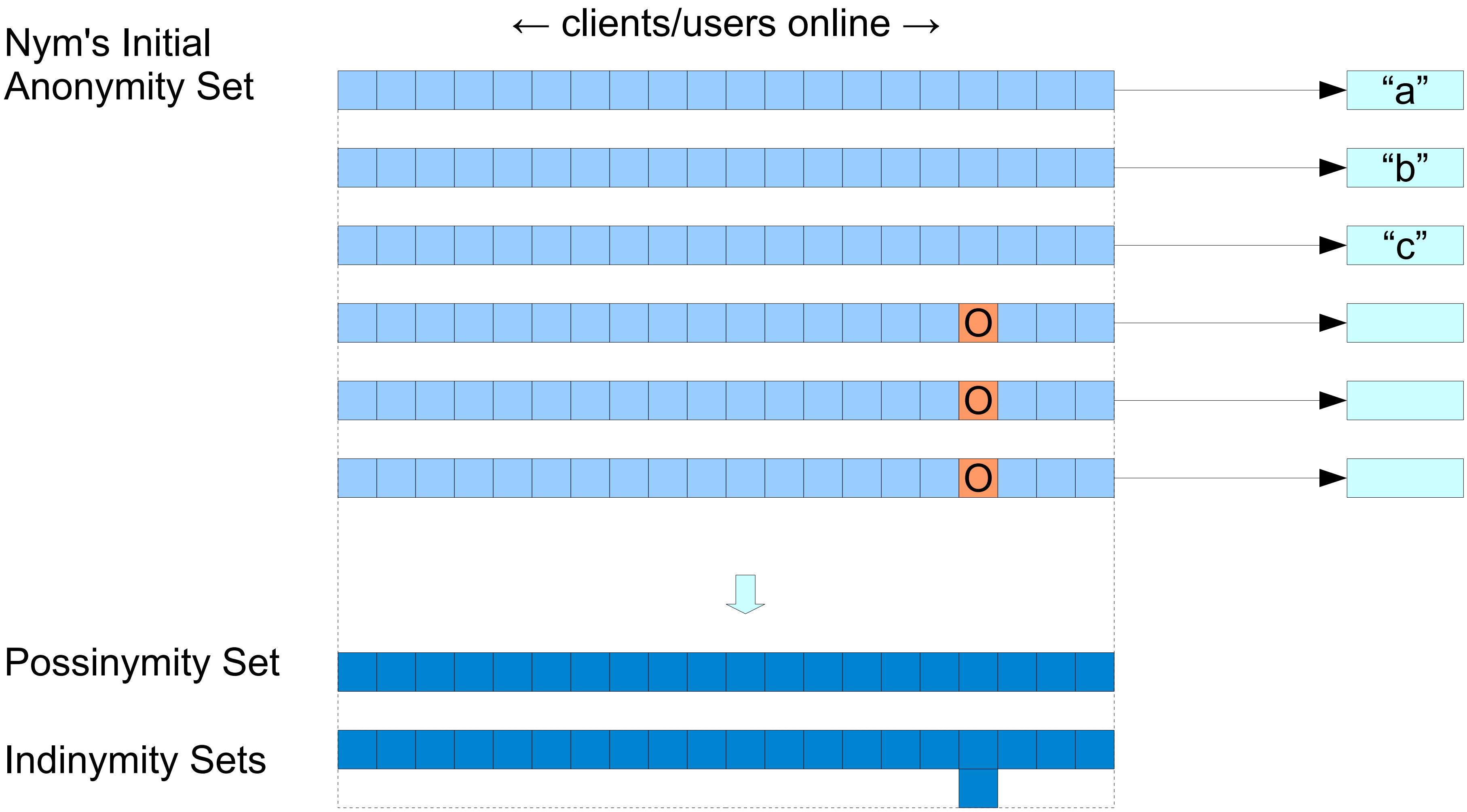
Possinymity: Possibilistic Deniability

Set of users who *could conceivably* own Nym

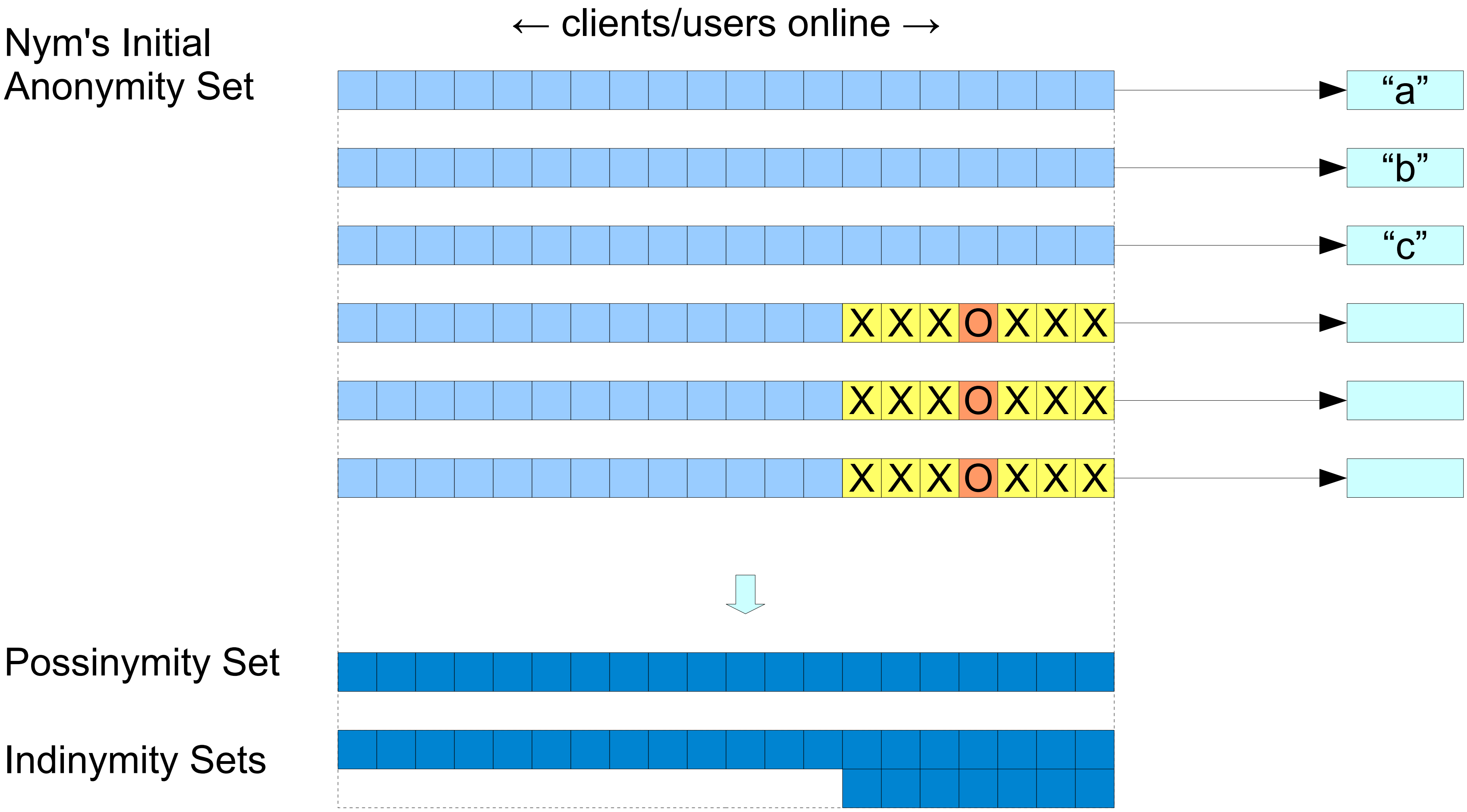
- Intersection of sets of all users *online and unfiltered* in rounds where *a message appears*
- Simplistic, but may build “reasonable doubt”


← clients/users online →
Nym's Initial Anonymity Set
“hey”
Users Online in Subsequent Rounds
O
“foo”
O O
“bar”
Resulting Possinymity Set

The "Statistical Disclosure" Problem

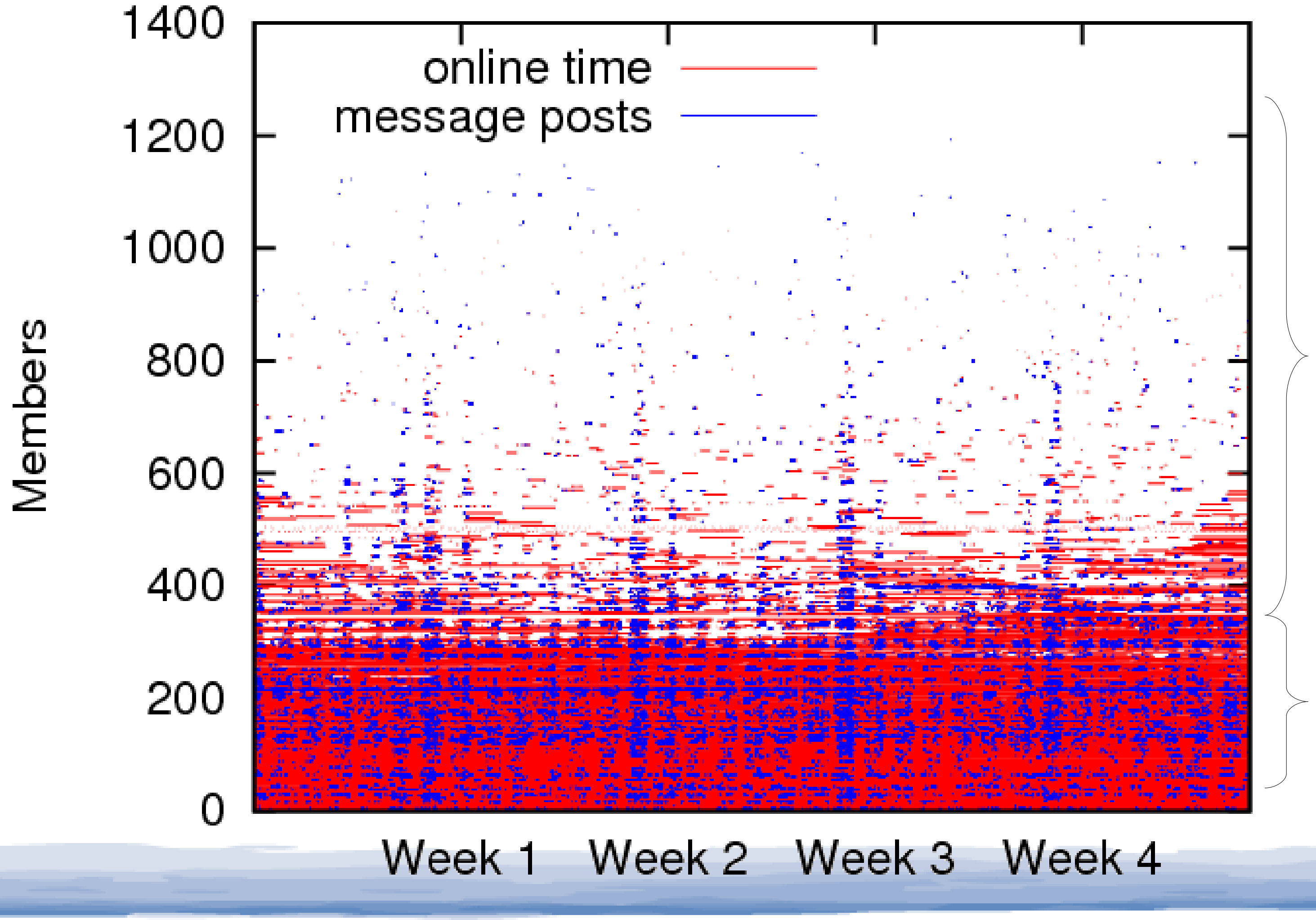


How Dissent Preserves Indinymity



How effective? Depends on users...

Analysis based on IRC online status traces



Ephemeral users

Where intersection attack resistant anonymity sets may plausibly be found

Key policy and usage model issues

In what contexts might Buddies be realistic?

- **Quickie browsing:** get online long enough to do your thing, then erase *all* linkable state
- **Blogging:** delay-tolerant anonymity among users who sign on at least once a day
- **Always-on apps:** BitTorrent-like background activities that encourage users to stay online

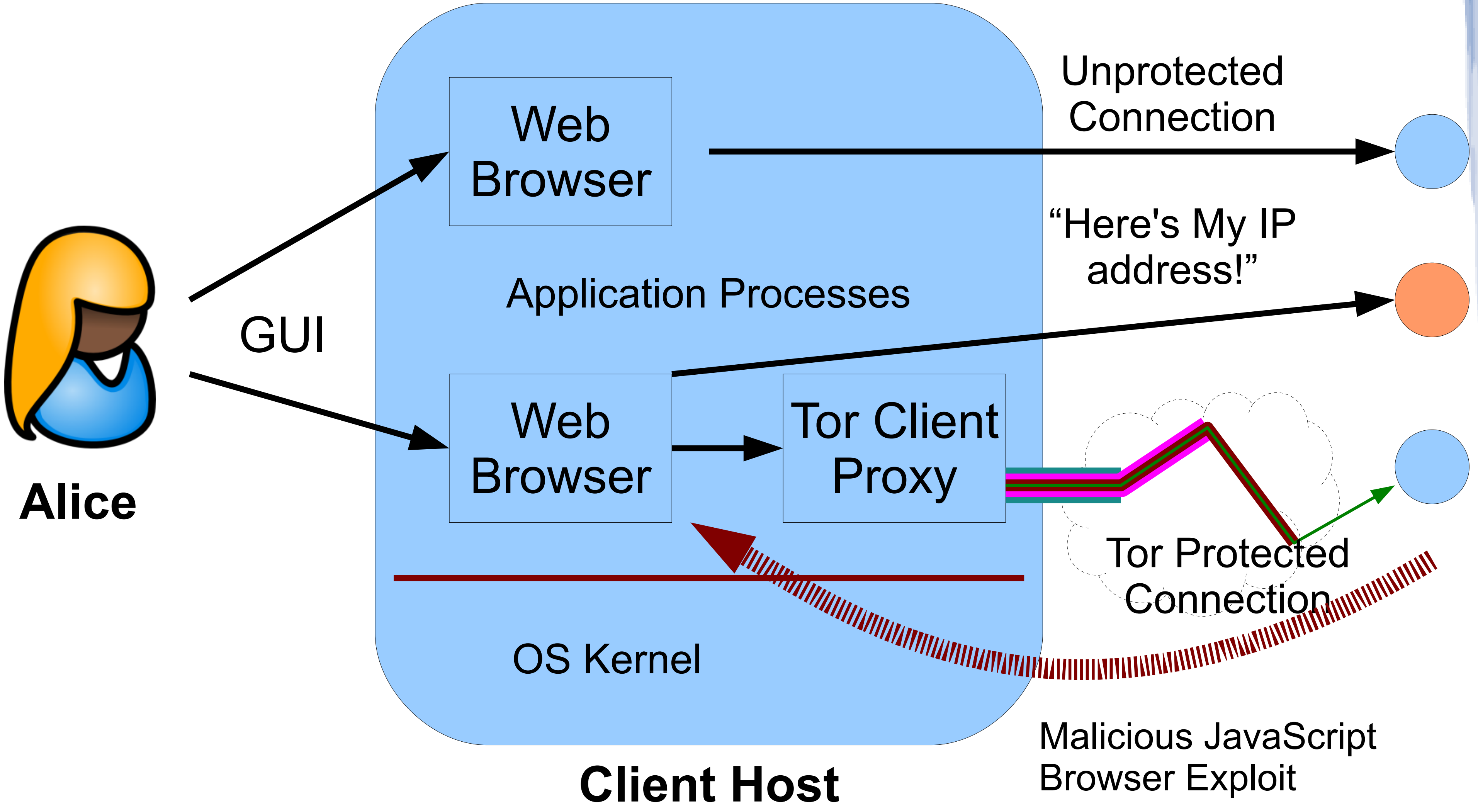
What if your buddy set is stacked with bad-guys?

- **Policy choices:** e.g., “random” vs “reliable”

Talk Outline

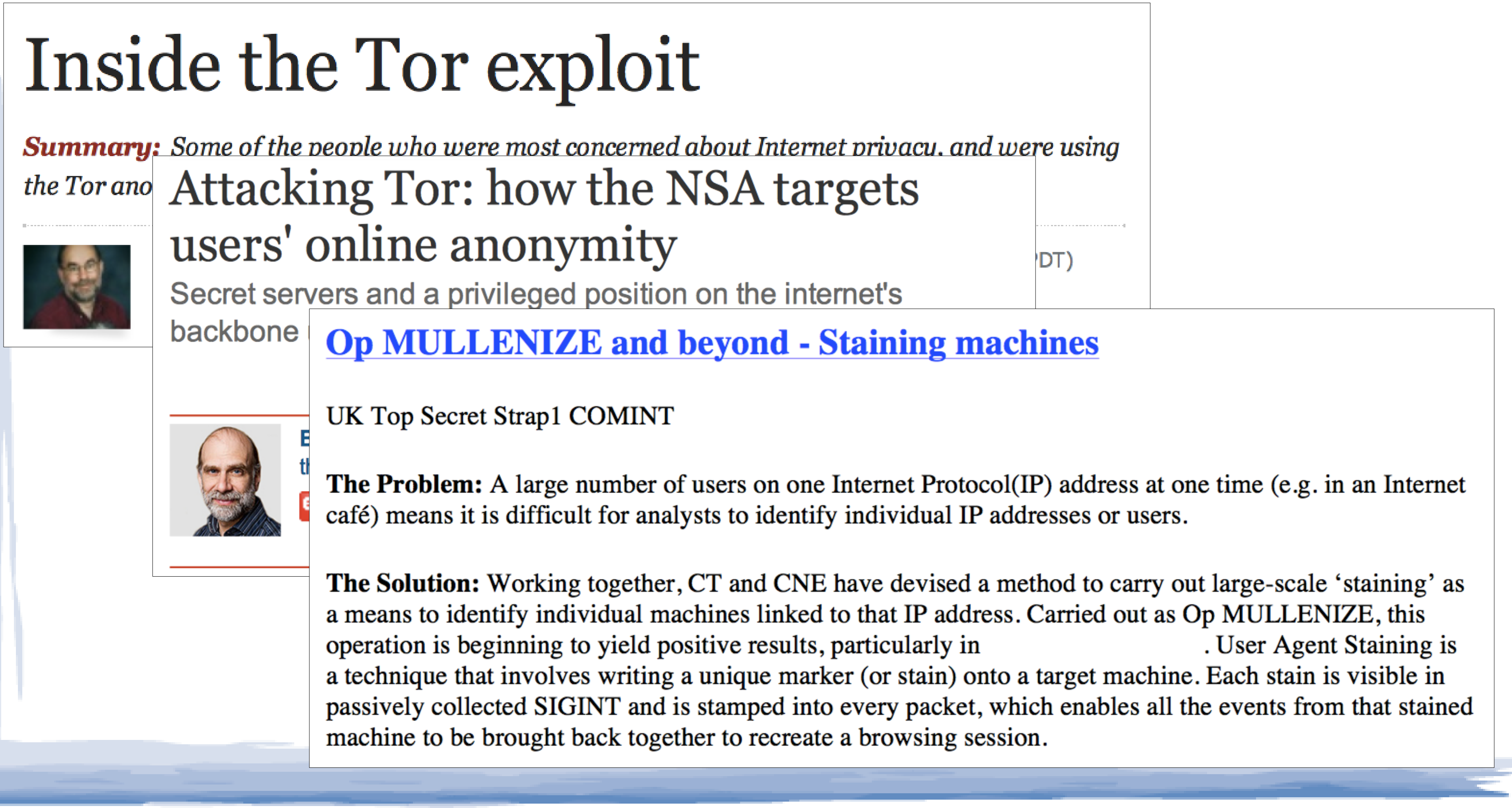
- ✓ Why Anonymity?
- ✓ Current State of the Art
- **Grand Challenges in Anonymity**
  - ✓ Global traffic analysis
  - ✓ Active interference attacks
  - ✓ Intersection attacks
  - **De-anonymizing exploits**
  - Accountability provisions
- Conclusion

Typical System Model



Exploits: The Low-Hanging Fruit

Circumvent the Anonymizer, Attack the Browser



Inside the Tor exploit

Summary: *Some of the people who were most concerned about Internet privacy, and were using the Tor ano*

DT)

Attacking Tor: how the NSA targets users' online anonymity

Secret servers and a privileged position on the internet's backbone

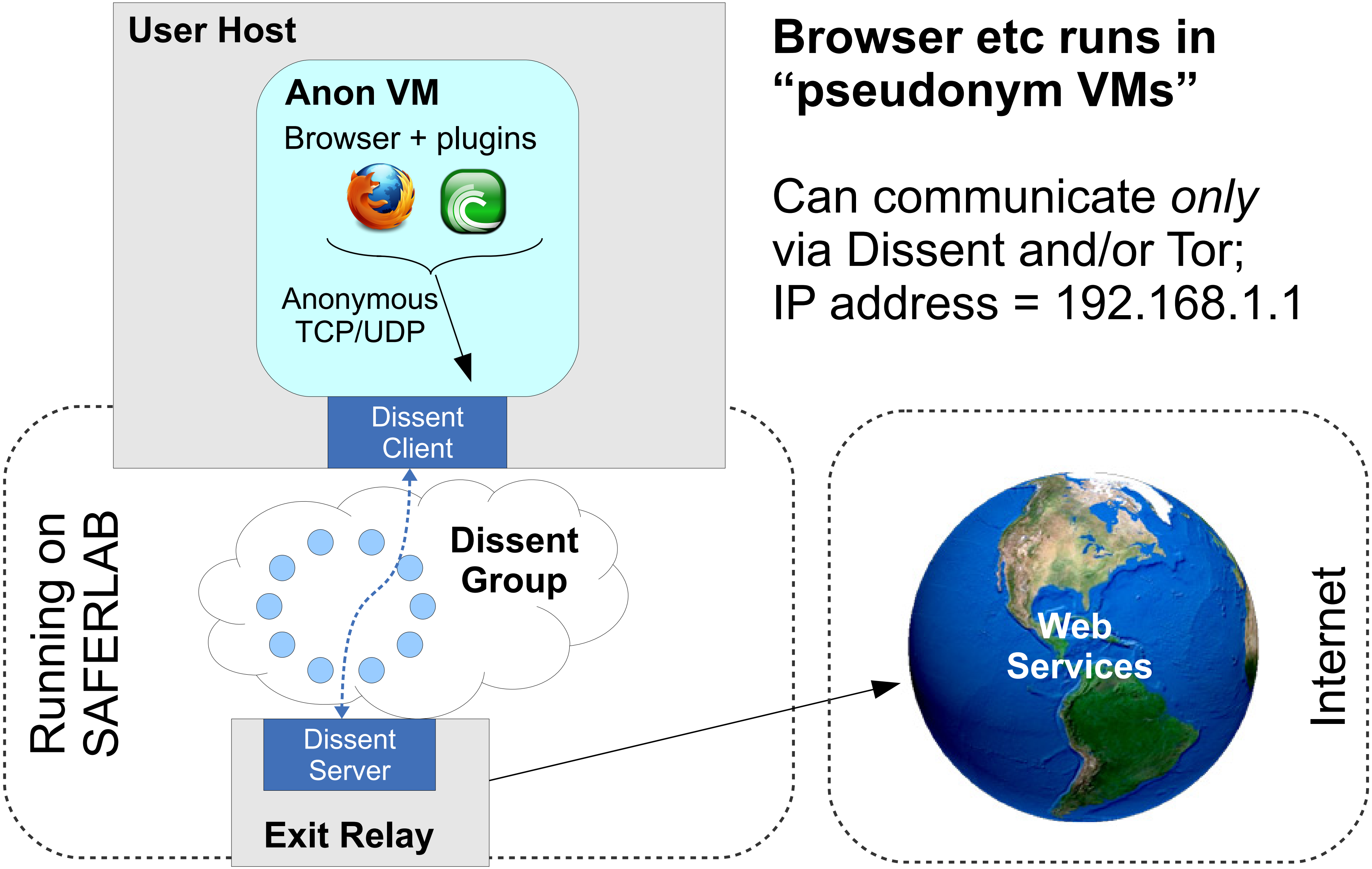
Op MULLENIZE and beyond - Staining machines

UK Top Secret Strap1 COMINT

The Problem: A large number of users on one Internet Protocol(IP) address at one time (e.g. in an Internet café) means it is difficult for analysts to identify individual IP addresses or users.

The Solution: Working together, CT and CNE have devised a method to carry out large-scale 'staining' as a means to identify individual machines linked to that IP address. Carried out as Op MULLENIZE, this operation is beginning to yield positive results, particularly in . User Agent Staining is a technique that involves writing a unique marker (or stain) onto a target machine. Each stain is visible in passively collected SIGINT and is stamped into every packet, which enables all the events from that stained machine to be brought back together to recreate a browsing session.

WiNon: VM-hardened Anonymity



Browser etc runs in "pseudonym VMs"

Can communicate *only* via Dissent and/or Tor; IP address = 192.168.1.1

Talk Outline

- ✓ Why Anonymity?
- ✓ Current State of the Art
- **Grand Challenges in Anonymity**
  - ✓ Global traffic analysis
  - ✓ Active interference attacks
  - ✓ Intersection attacks
  - ✓ De-anonymizing exploits
  - **Accountability provisions**
- Conclusion

Returning to the bomb hoax

Bomb threats are an abuse of anonymity. But:

- Kids do stupid things
- It's our job to educate them
- Is unmasking (& 5 years jail) the only way to deter abuse?

Accountable Anonymity

Accountability can mean many things

- "Accountability & Deterrence" [Feigenbaum'11]

We need deterrents that **escalate gracefully**

1. **Threat of censure** by peers in online forum
2. **Opportunity to retract** without unmasking
3. **Expulsion** from group without unmasking
4. **Unmasking** only as a last resort, via transparent procedures – *not* secret spy tech

Accountability in Dissent

Dissent model can provide:

- **Authenticated pseduonyms**
  - If you post apology and reaction,
peers (and cops) know it's same you
- **1-to-1 mapping** of users to pseudonyms
  - If you get banned, you can't just pop up again
- **Decentralized authority**
  - If *all* Dissent server operators agree you're a hardened criminal, they can de-anonymize you

Accountability Schemes in Dissent

1. **Dissent v1** [CCS'10]:
   use Brickell/Shmatikov shuffle to distribute hash-checked *assignments* before round
   - Simple, but requires expensive shuffle *each* round
2. **Scalable Dissent** [OSDI '12]:
   retroactive disruption-tracing “blame” protocol
   - Complex, efficient when *not* disrupted
3. **Verifiable Dissent** [USENIX Sec 13]:
   proactive verifiability via zero-knowledge proofs
   - Offline possible, lower blame cost *when* disrupted

Talk Outline

- ✓ Why Anonymity?
- ✓ Current State of the Art
- ✓ Grand Challenges in Anonymity
  - ✓ Global traffic analysis
  - ✓ Active interference attacks
  - ✓ Intersection attacks
  - ✓ De-anonymizing exploits
  - ✓ Accountability provisions
- **Conclusion**

Dissent: Current Status

- Proof-of-concept works, available on github
  - **Preliminary:** not at all feature-rich, user-friendly
  - **Don't** use it [yet] for security-critical activities!
- Takes a few steps, but many questions remain
  - How well can we make it perform, scale?
  - Broadcast limits scalability for “point-to-point” use
  - *Might* be very efficient for multicast applications
    - Anonymous chat/microblogging, “town hall” meetings
- Time (and further development) will tell!

Key Future Work Questions

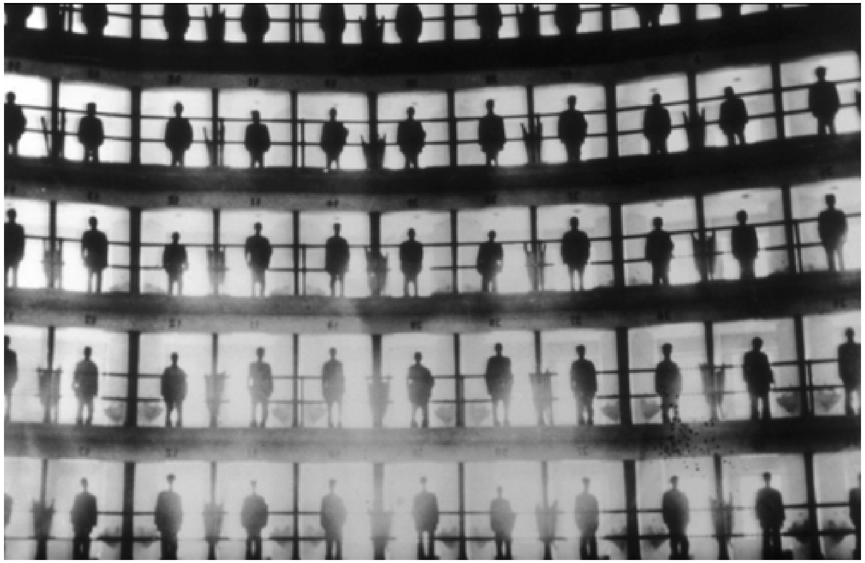
Can we refit the Internet with tracking resistance?

- Make traffic analysis resistant protocols even more scalable, get *everyone* running them
- Community-oriented applications giving people strength in numbers via *relevant* anonymity sets
- Create usage models enabling and incentivizing intersection attack resistant user behaviors
- Build pseudonym isolation, stain resistance into popular client-side operating systems
- Graceful abuse response through accountability

Conclusion

Can you hide in an Internet panopticon?
It's hard! – due to five grand anonymity challenges

- Global traffic analysis
- Active attacks
- Intersection attacks
- Software exploits
- Accountability



Dissent takes a few baby steps toward solutions, but only a starting point for trustworthy anonymity

http://dedis.cs.yale.edu/dissent/