Dissent: Design and Experimental Lessons from a Clean-Slate Anonymity System

Bryan Ford
Yale University

Fifth GENI Research and Educational Experiment Camp
May 26, 2015

Dissent Project
Members and Collaborators

Henry Corrigan-Gibbs, Joan Feigenbaum, Bryan Ford, Ramakrishna Gummadi, Daniel Jackowitz, John Maheswaran, Michael F. Nowlan, Ewa Syta, Shu-Chun Weng, David Isaac Wolinsky
– Yale

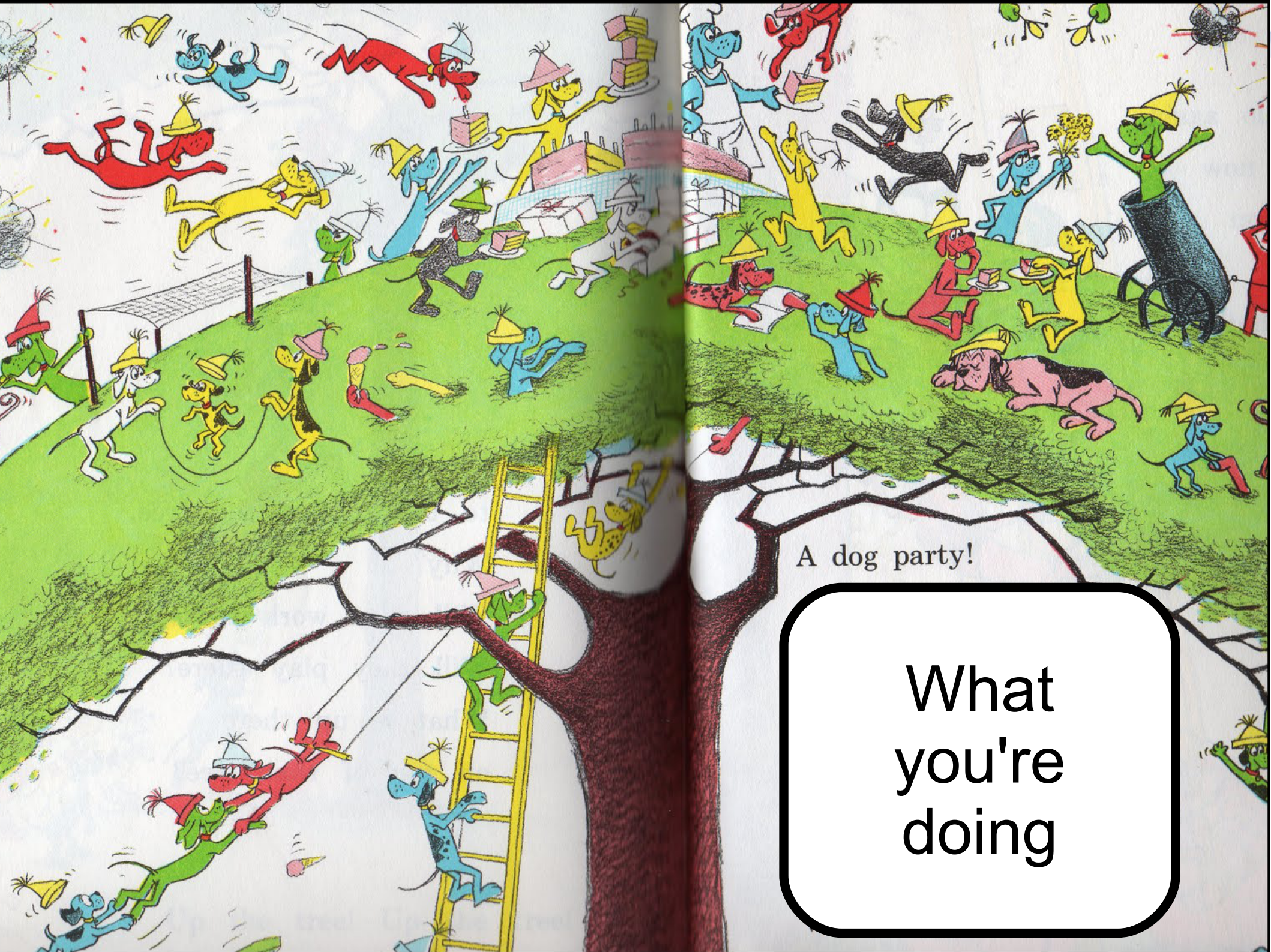
Chad Brubaker, Amir Houmansadr, Vitaly Shmatikov
– UT Austin

Rob Jansen, Aaron Johnson
– US Naval Research Lab


“Nobody knows you're a dog?”


Dogs of the World
Terrier Group
Hound Group
Sporting Group
Toy Group
Non-Sporting Group
Herding Group
Actually, they know
exactly what kind you are

Who your friends are...


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

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

Only 19 left in stock - order soon.

More Buying Choices

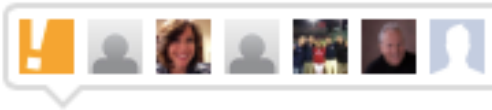
$5.65 new (19 offers)

(55)

Pet Supplies: See all 25,595 items

... based on Rover's Dogbook likes

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



+ 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.

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)

You may need anonymity...

...because they're **actually** out to get you

- LGBTQs in Rednecksville
- Protestors in Repressistan

You may need anonymity...

...or just because most people wear several hats (and don't want them linked)

Family Hat

Hobby Hat

Professional Hat

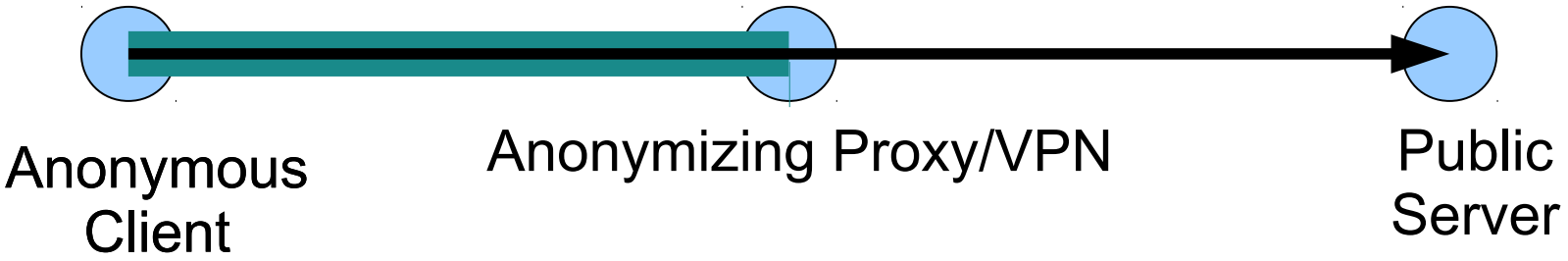
Party Hat

The Complete You

Commercial VPN services

Popular for circumventing the Great Firewall

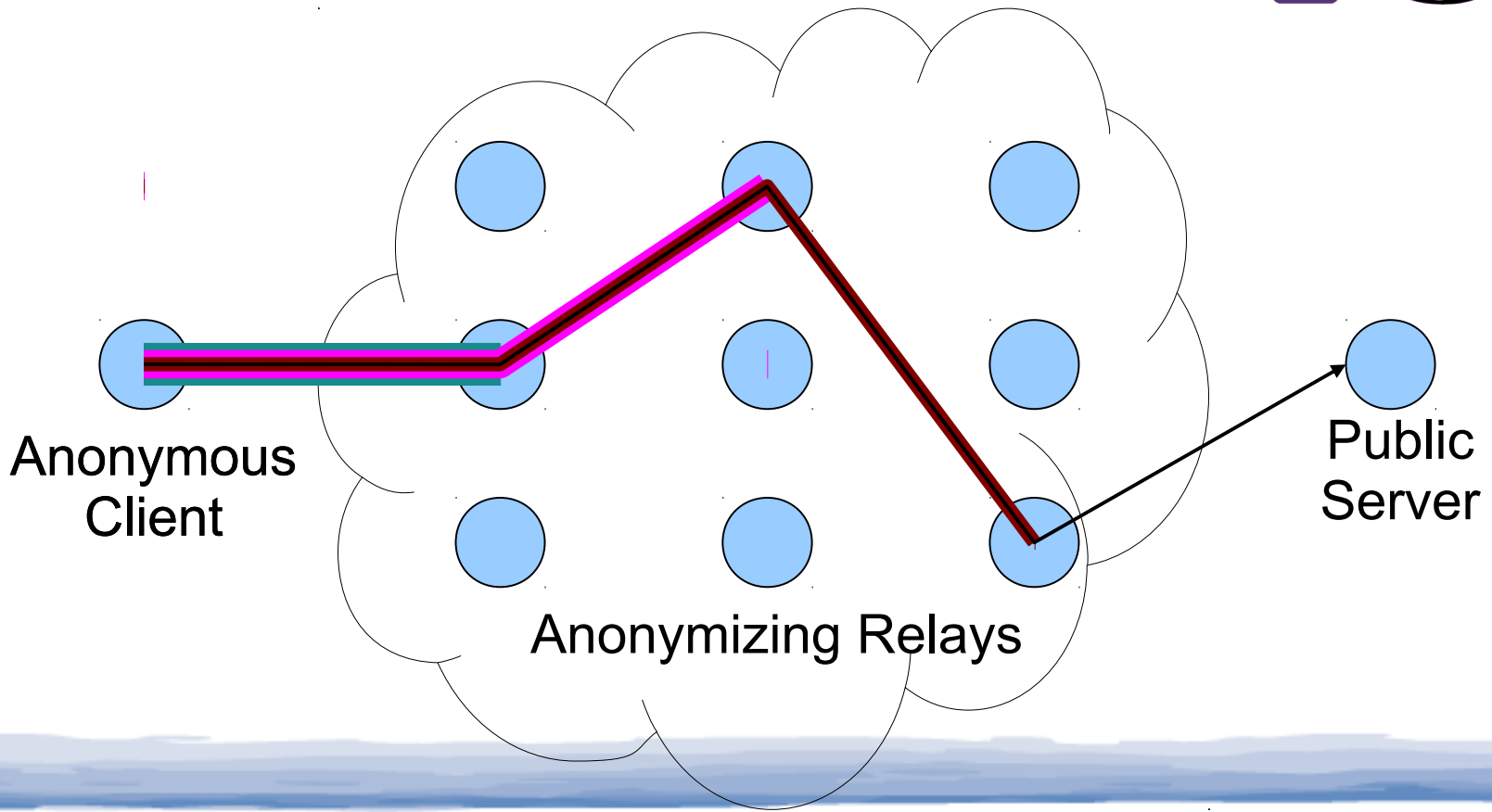
- You build encrypted tunnel with VPN server
- VPN server forwards traffic to destination
- Looks like it's coming from VPN server
- Hope the server operator protects your privacy



The current state-of-the-art

Onion routing tools such as **Tor**

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



Sampled Traffic

Internet-Exchange-Le

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?

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

Low-Resource Routing Attacks Against Tor

Limits of Anonymity in Open Environments

STATISTICAL DISCLOSURE ATTACKS

Traffic Confirmation in Open Environments

Browser-Based Attacks on Tor

Timothy G

- vulnerable to **five**
  - Global traffic a
  - Active attack
  - Denial-of-se
  - Intersection
  - Software exploits
- Question is *when & h*

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)

Dissent: a Clean-Slate Design for Provable, Measurable Anonymity

Builds on fundamentally different primitives

- Verifiable Shuffles, Dining Cryptographers
- Offering provable security properties
- Measurable via formal anonymity metrics

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

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

A New Wave of Anonymity Research?

Other recent alternatives to mixes/onion routing:

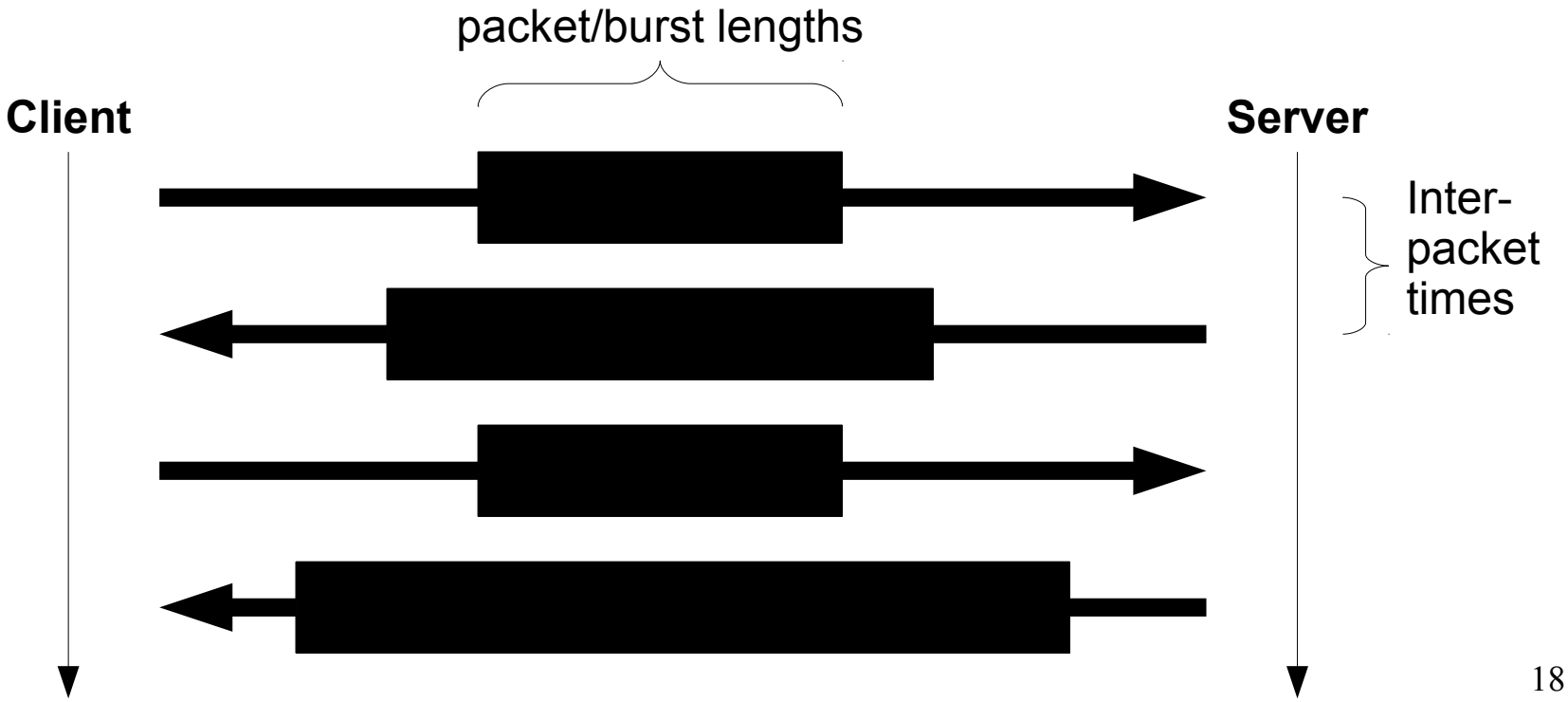
- **Aqua** – Le Blond et al, SIGCOMM 2013
- **CoinShuffle** – Ruffing et al, ESORICS 2014
- **Riposte** – Corrigan-Gibbs et al, Oakland 2015
- **Baffle** – Zamani et al, ICDCS 2015
- **Herd** – Le Blond et al, SIGCOMM 2015
- **Vuvuzela** – van den Hoof, preprint 2015

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
- Status and Ongoing Work

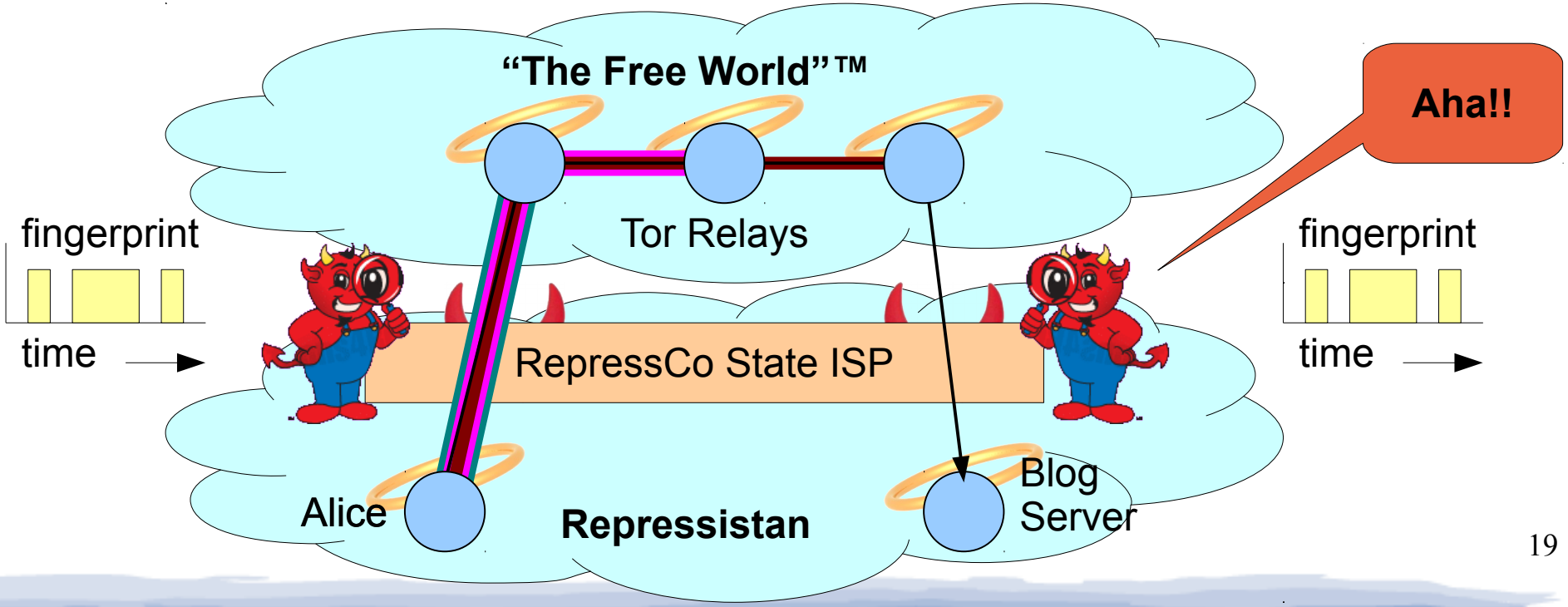
The Traffic Analysis Problem

- 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

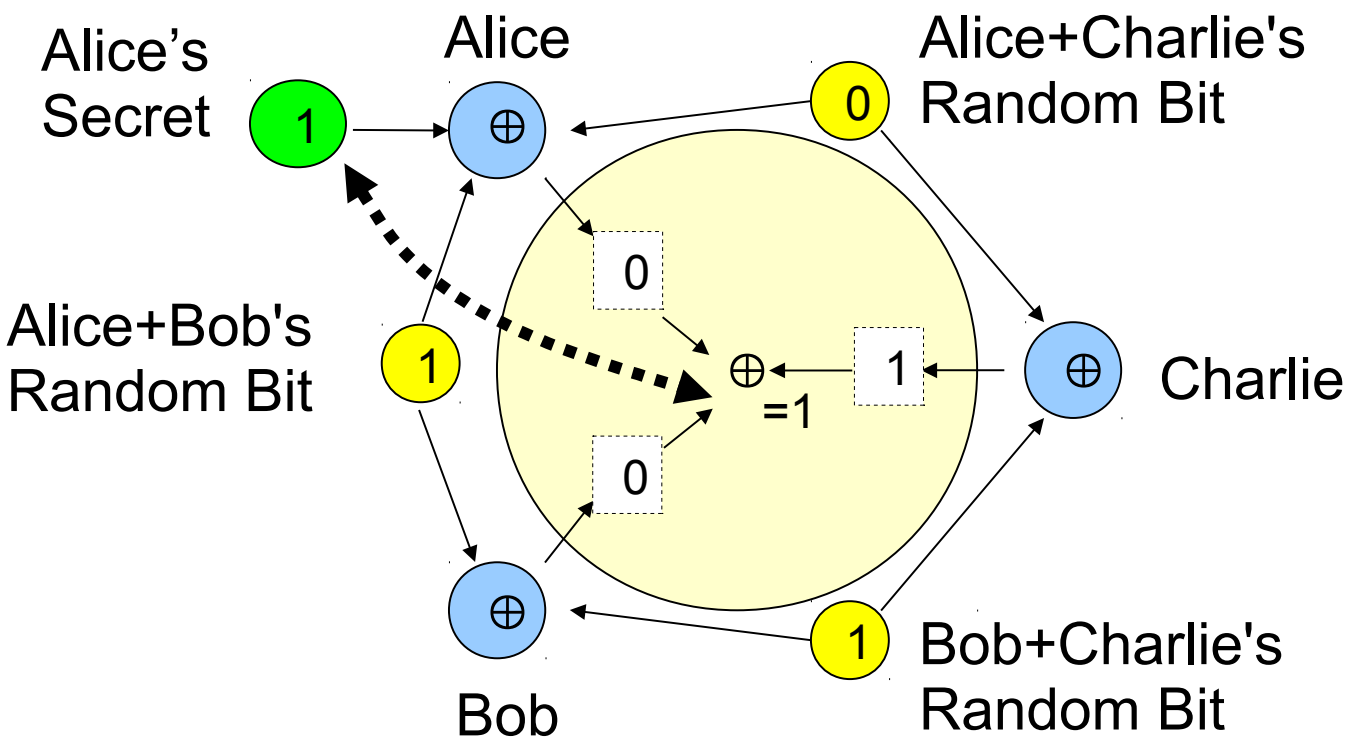
- Alice in Repressistan uses Tor to post on blog server hosted in Repressistan
- State ISP controls *both* entry and exit hops
- Fingerprint & correlate traffic to **deanonymize**



Can We Resist Traffic Analysis?

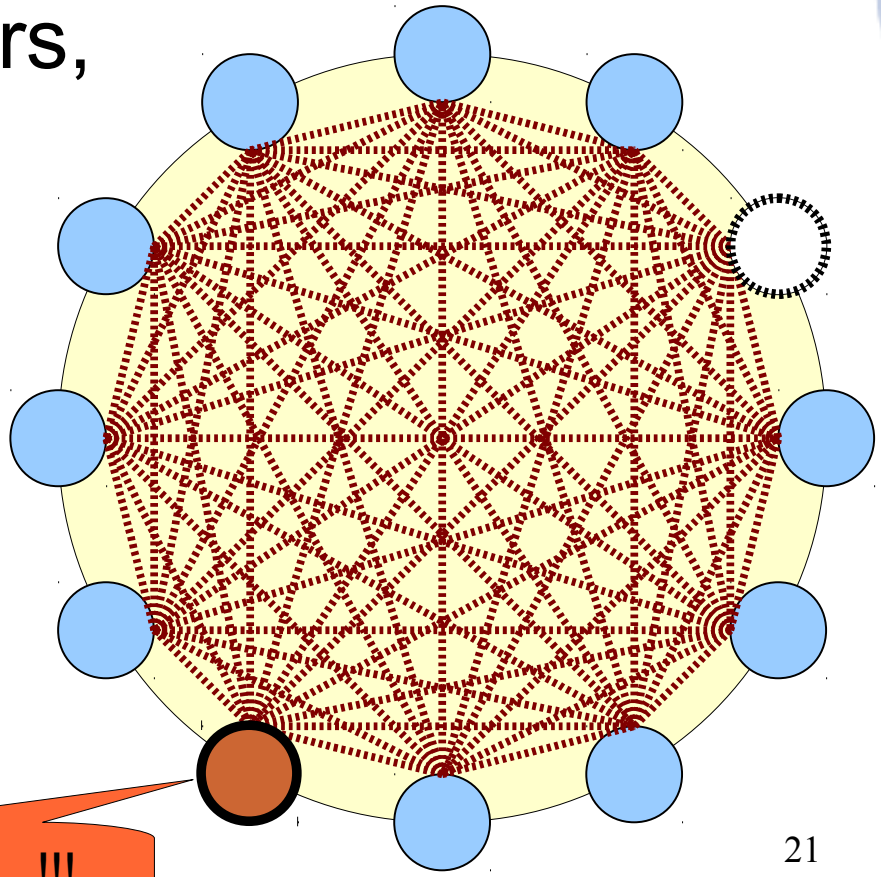
Dining Cryptographers or DC-nets [Chaum '88]

- Key property: provable anonymity within a group



Why DC-nets Doesn't Scale

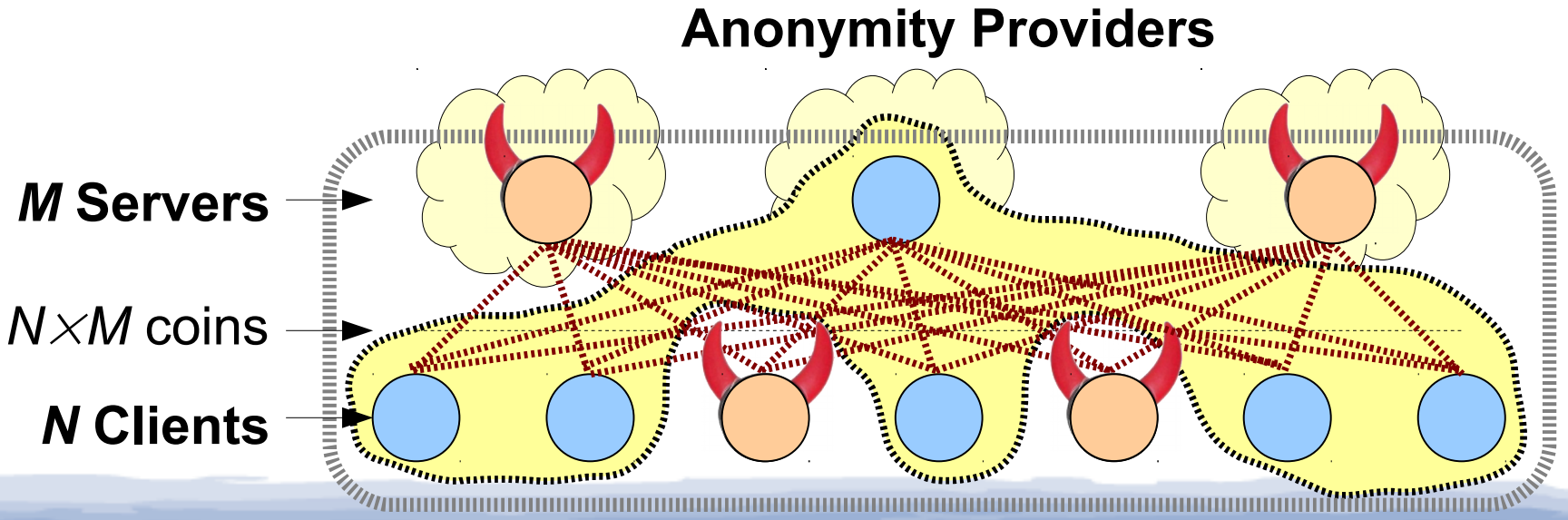
- **Computation cost:** $N \times N$ shared coin matrix
- **Network churn:**
  if *any* participant disappears,
  all nodes must start over
- **Disruption:**
  any single “bad apple”
  can jam communication



“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*

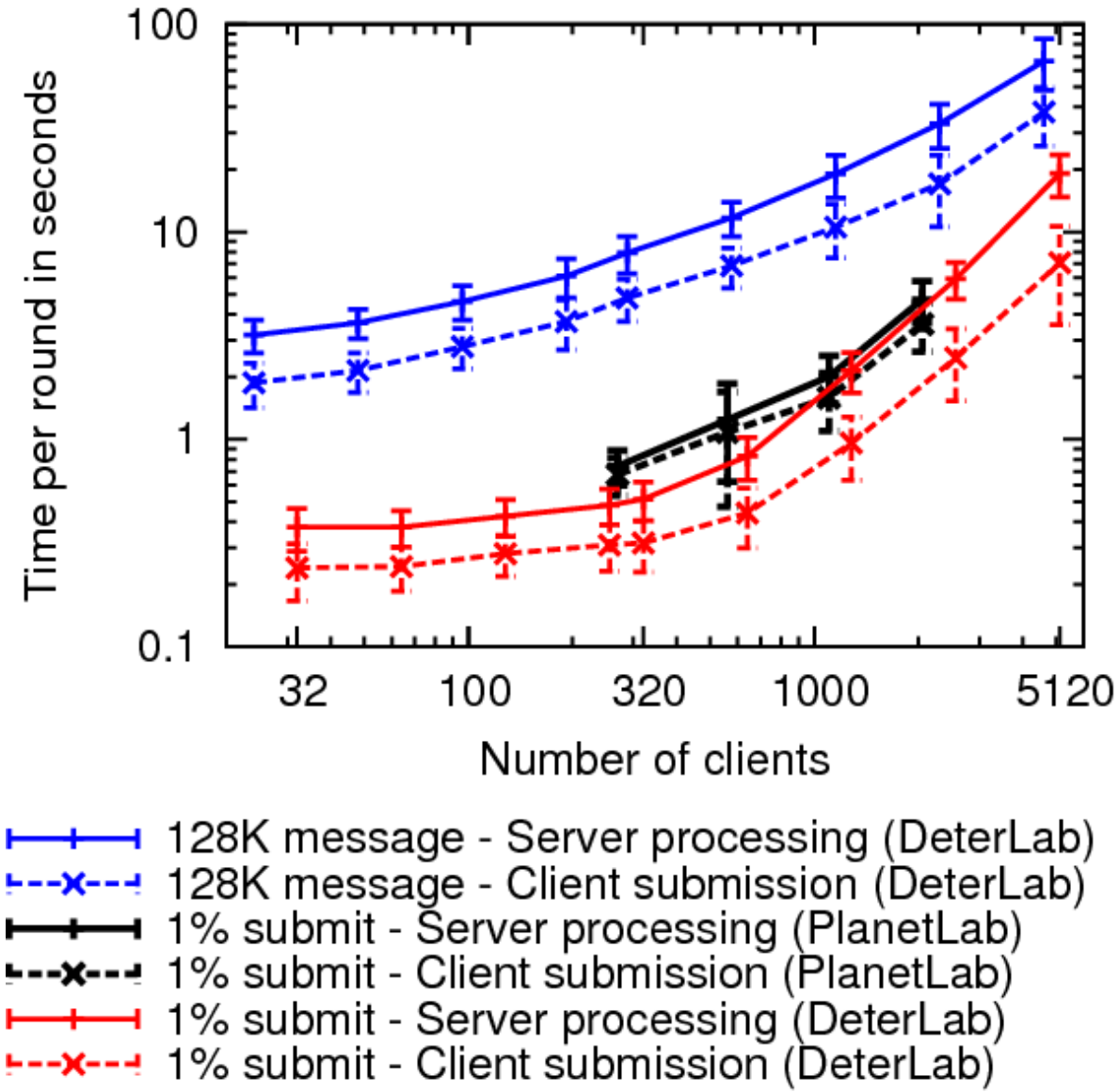


Scaling to Thousands of Clients

100× larger anonymity sets

- (Herbivore, Dissent v1: ~40 clients)

<1 sec latency w/ 1000 clients



Major Limitations

Still scales to “only” thousands of users

- Want to support *millions* of users…
- e.g., by automatically dividing users into groups (as in Herbivore [Sirer], quorums [Zamani], ...)

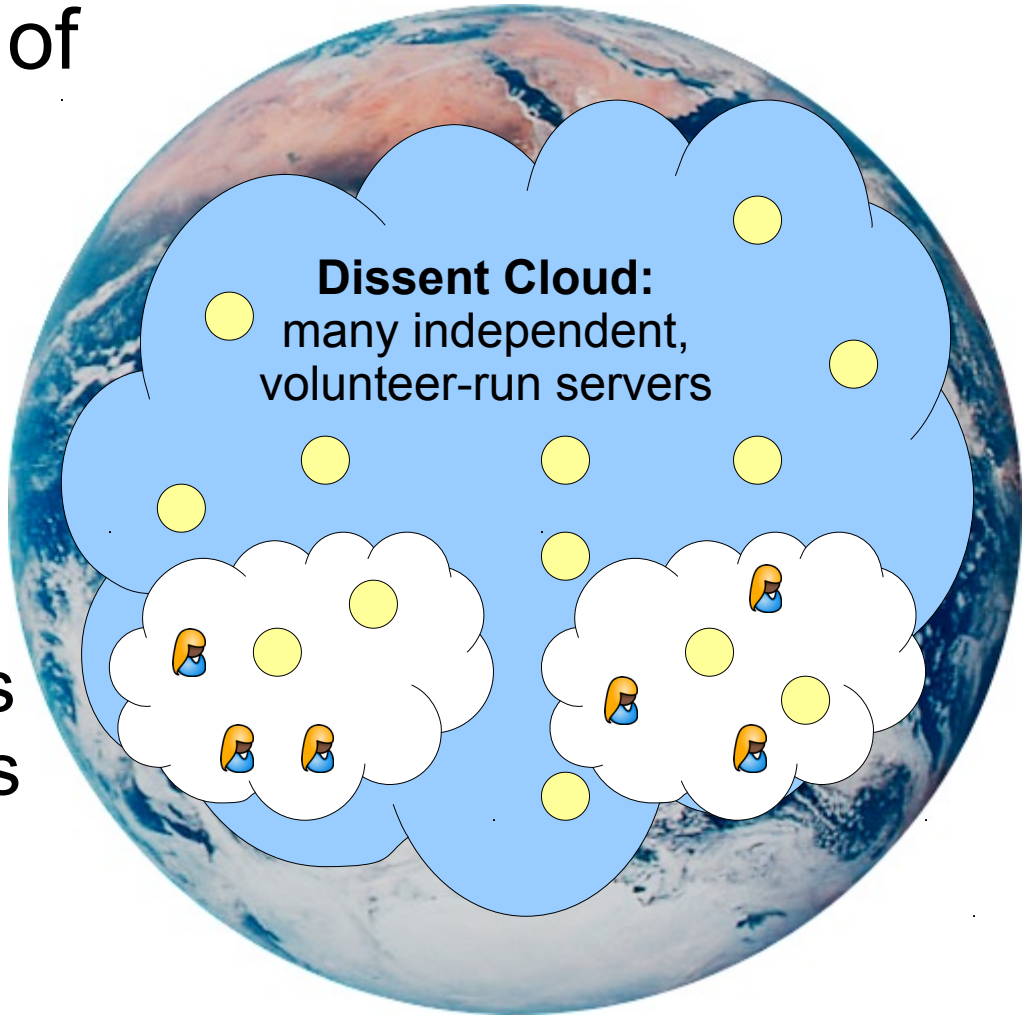
Depends on “carefully chosen” set of servers

- Needs be *automatically chosen* from server list
- But then server directory and random choice becomes security-critical attack target

Ongoing: Dissent at Large Scales

Decentralized directory of Dissent servers

- User-controlled Group formation
  - Trustworthy random server selection
  - Tunable anonymity vs performance tradeoffs



Key building block:
Strongest-Link Cothorities (ongoing work)

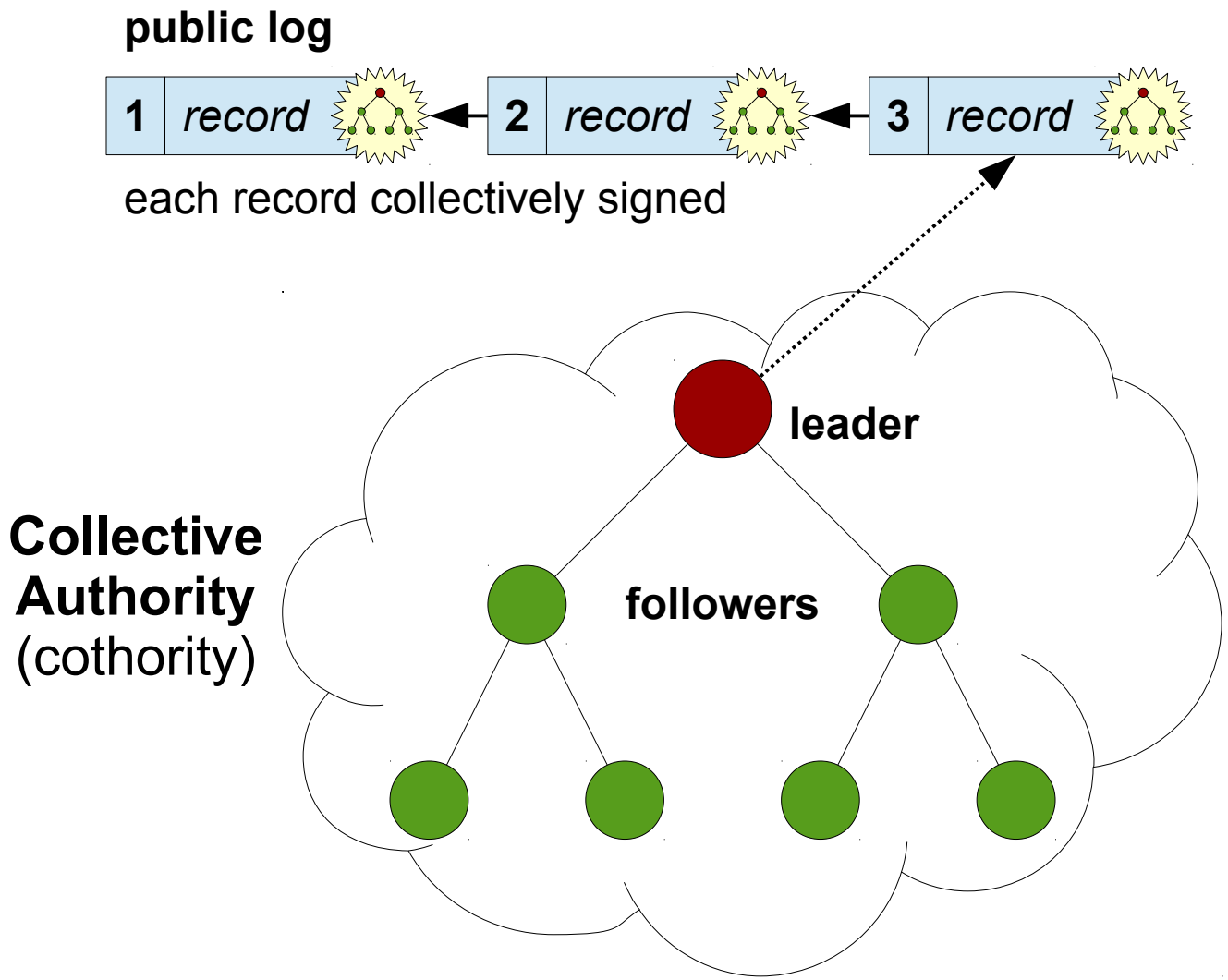
Cothorities: Collective Authorities

Thousands of servers form *single* replicated state machine, Byzantine consensus group

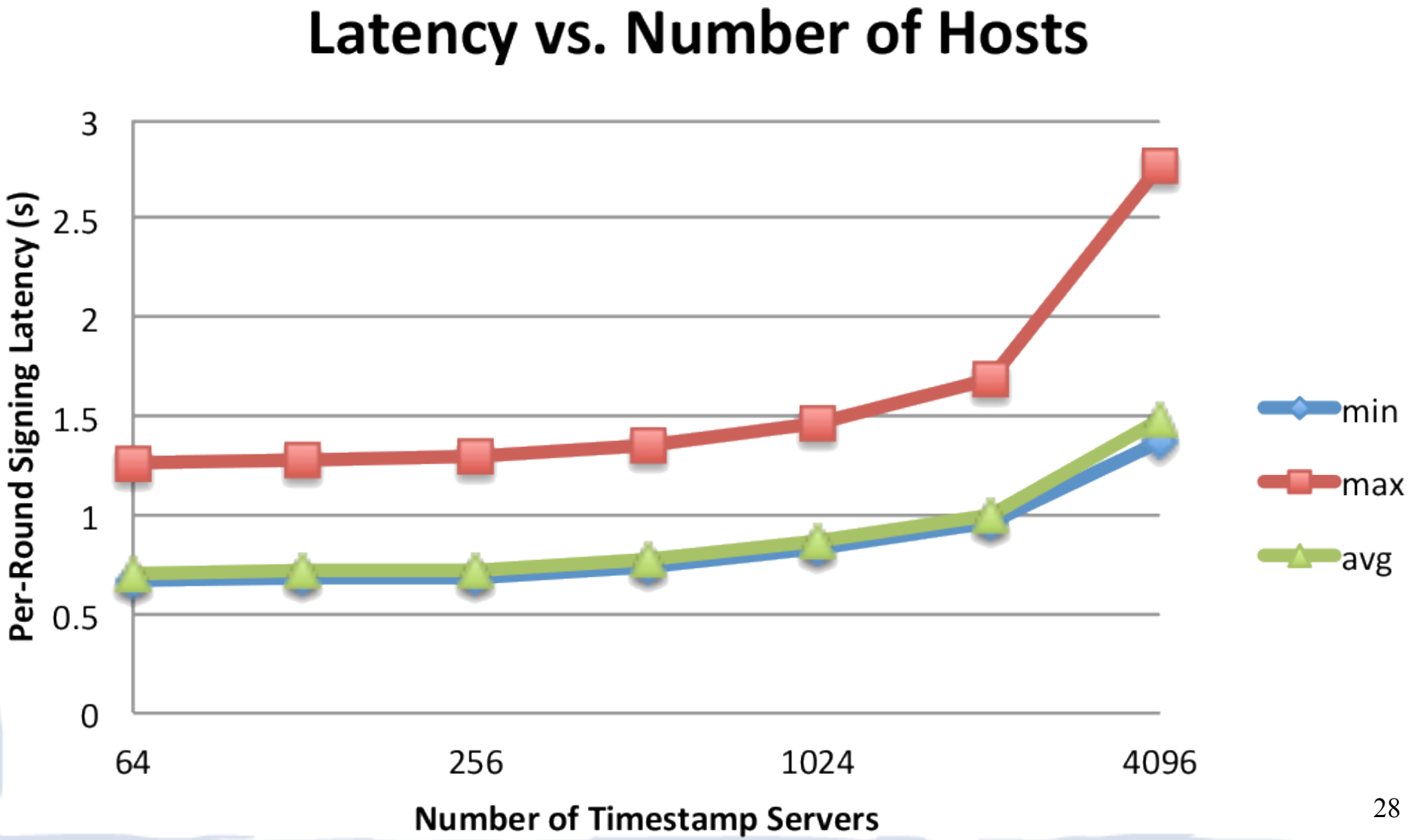
- Collectively agree on directory of servers
  - No need to trust 8 “special” servers as in Tor
- Collectively toss unknown, unbiasable coins
  - Even if colluding nodes go offline strategically
- Collectively sign and witness log entries
  - Clients/users can verify via *single* signature check

Details (preprint): http://arxiv.org/abs/1503.08768

Cothorities: Collective Authorities



Cothorities Scaling Results



Experimentation Lessons

For both Dissent and Cothorities, need to answer the question “how big can this protocol scale?”

- We *always* needed many more testbed nodes than were easily/cheaply available
- Therefore used virtualization, oversubscription (e.g., 16 Dissent processes per physical node)
- But then when the protocol stops scaling, is that the *protocol* or the *oversubscription*?

Experimental Testbed Wishlist

More systematic experiment scaling support

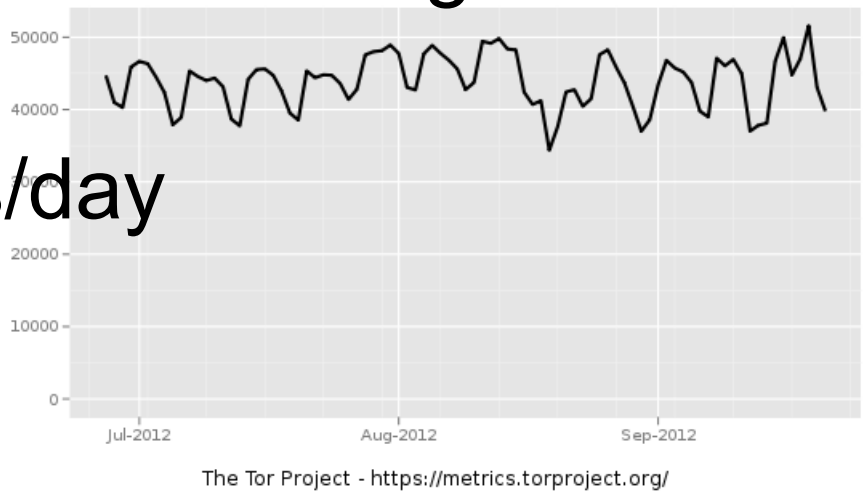
- More testbed nodes (of course, always)
- More, better, easier-to-deploy virtual nodes
  - Knob: machines, VMs, containers, processes
- Large-scale, queue-able “batch” jobs
  - Support for both “long” and “wide” allocations
- Tools to validate oversubscribed experiments
  - Same topology, different # vnodes per machine
  - Validation-based auto-tuning, incremental growth?

Talk Outline

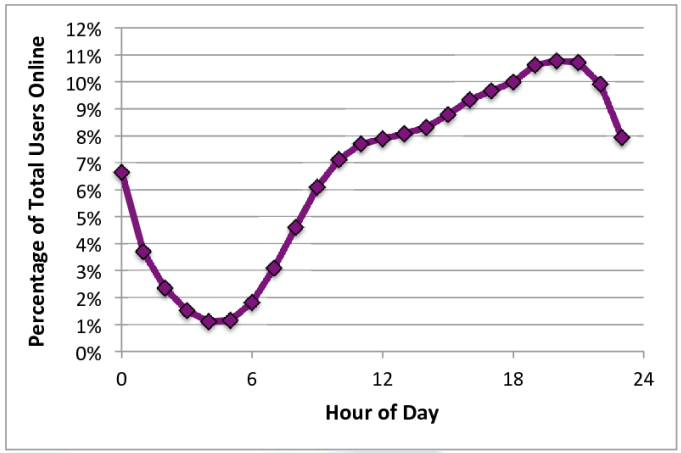
- ✔ Why Anonymity?
- ✔ Current State of the Art
- **Grand Challenges in Anonymity**
  - ✔ Global traffic analysis
  - ✔ Active interference attacks
  - **Intersection attacks**
  - De-anonymizing exploits
  - Accountability provisions
- Status and Ongoing Work

How anonymous are you *really?*

- Bob in Dictatopia posts via Tor to blog hosted in “The Free World”™
- Tor Metrics: 50,000 users/day connect from Dictatopia
  - Good anonymity, right?



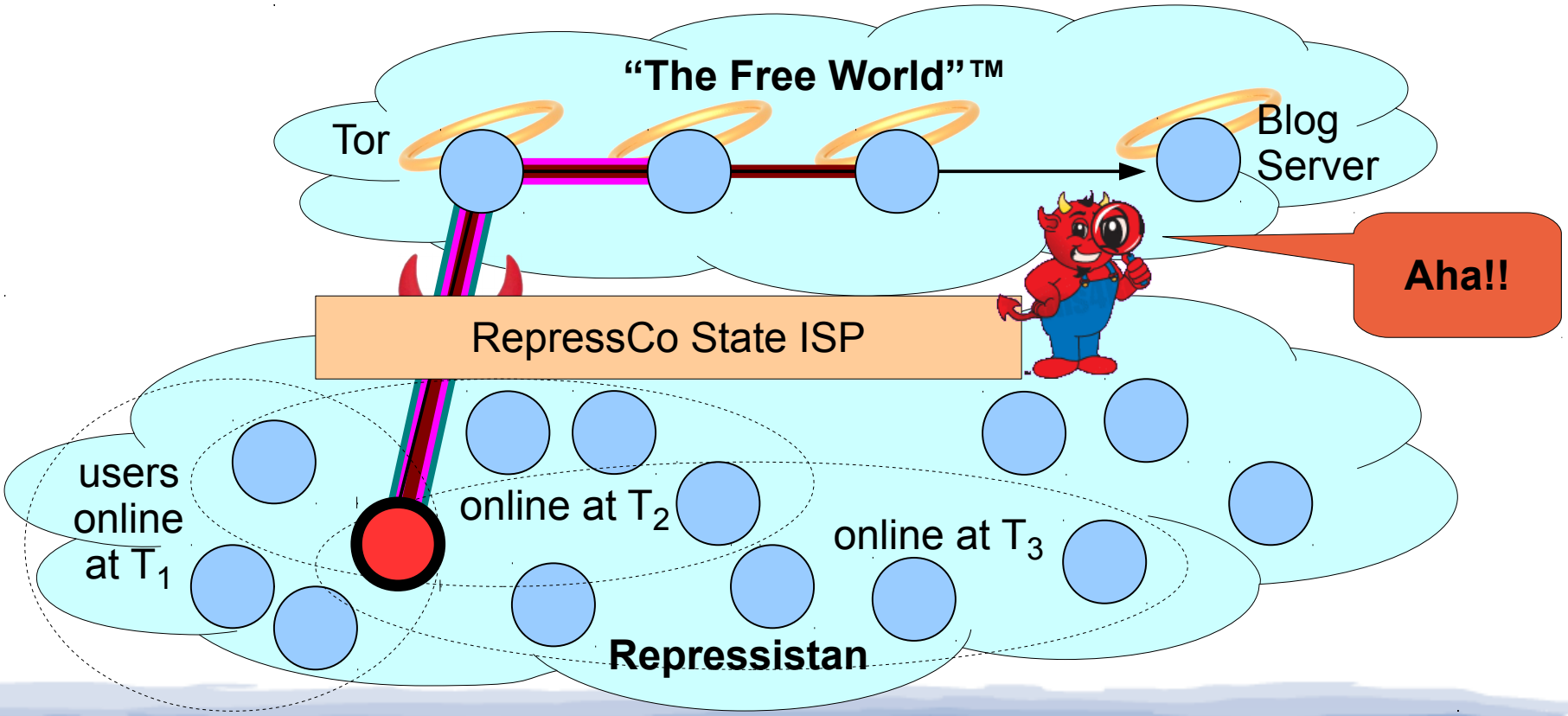
- But ISP logs tell police when users are online; blog post has timestamp
  - How many users are online ***at same time Bob posts?***
    - ~5,000 at 7PM?
      ~500 at 5AM?



The Intersection Attack Problem

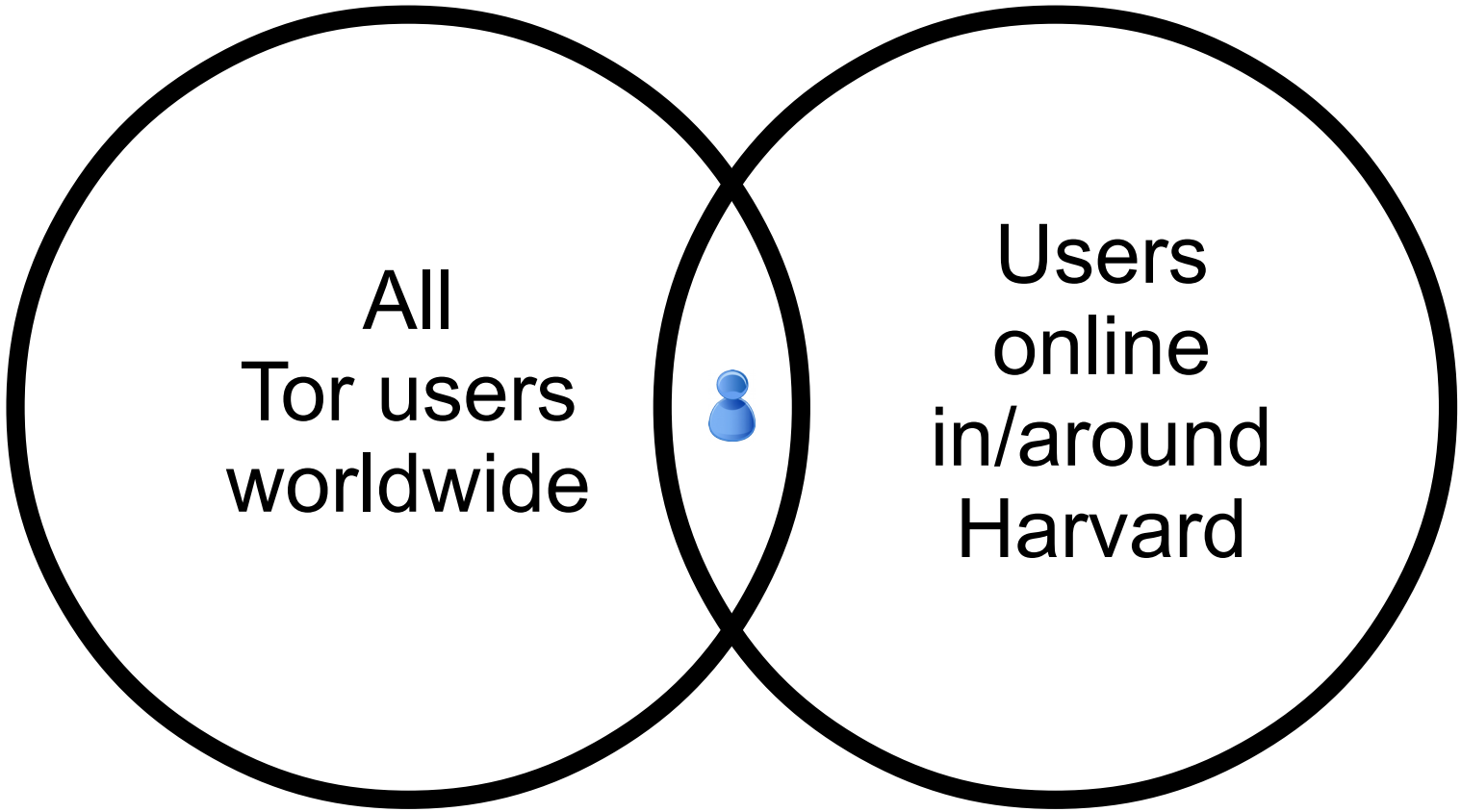
Kate signs posts with pseudonym “Bob”

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



The Bomb Hoax Attack

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



Buddies [CCS '13]

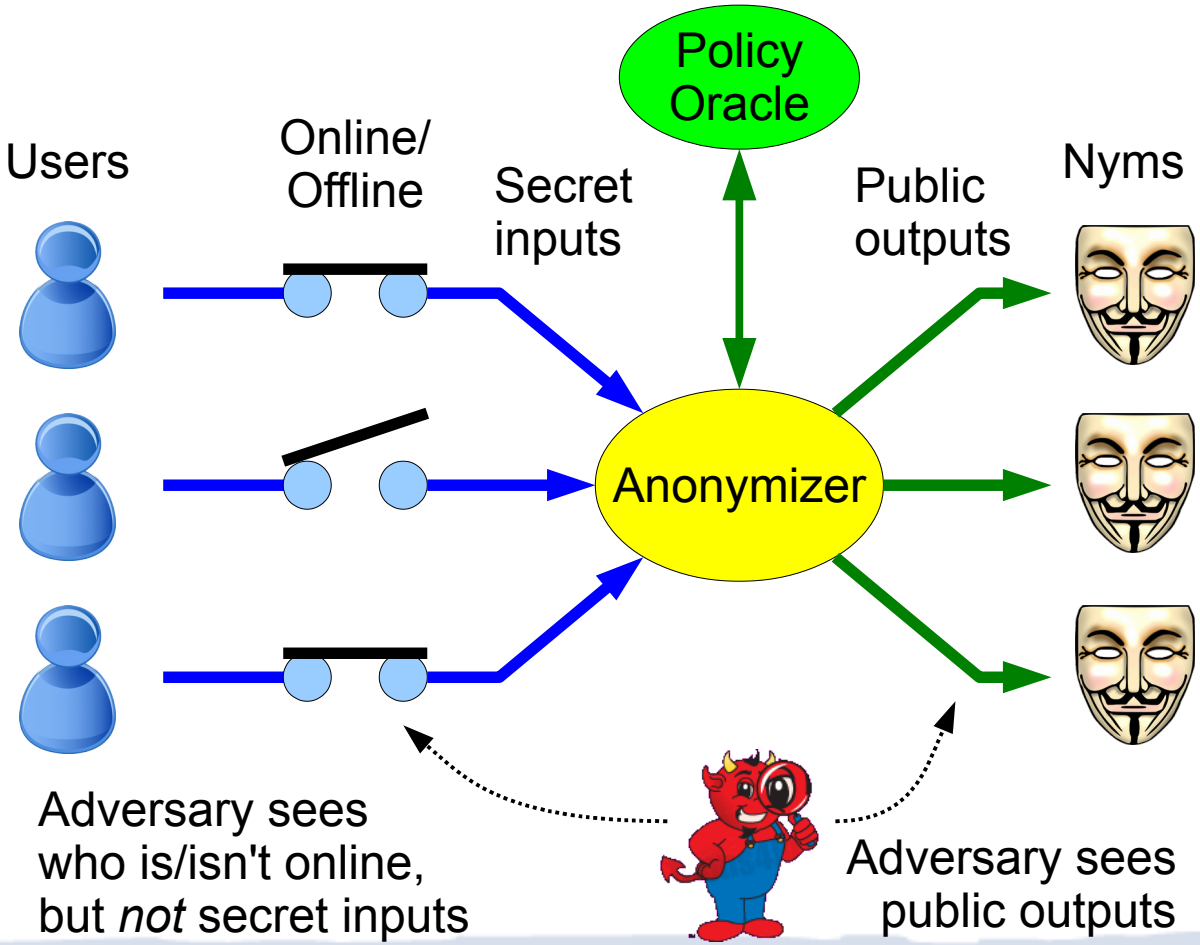
First attempt at building intersection attack resistance into a practical anonymity system

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*



Computing Anonymity Metrics

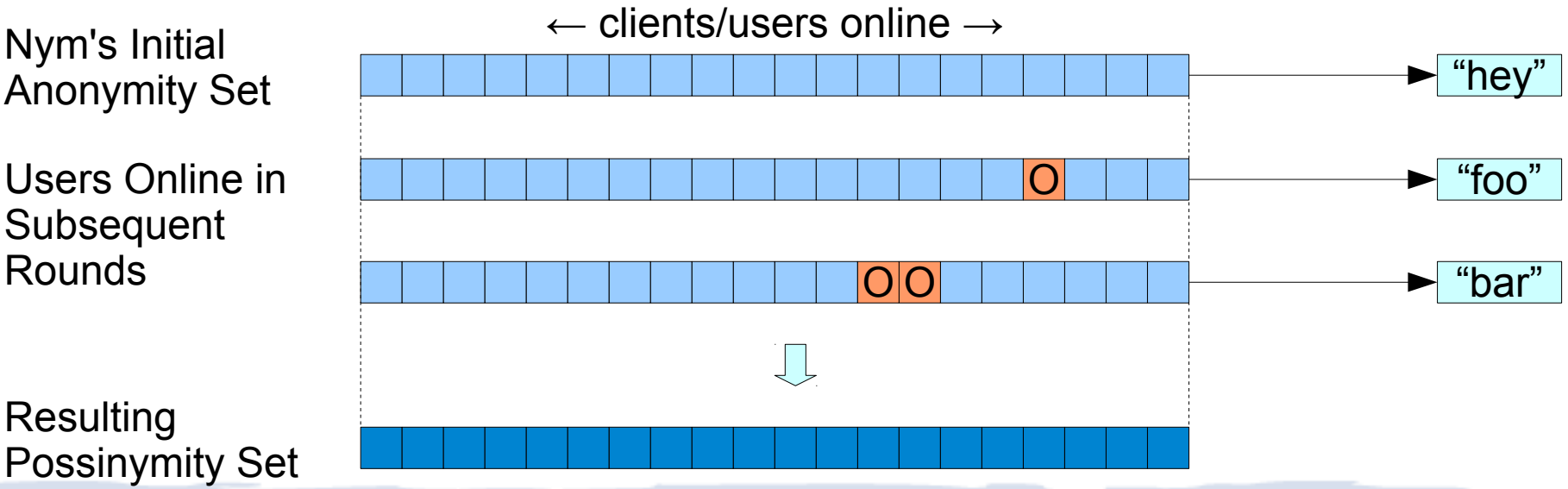
Policy Oracle *simulates an adversary's view*

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

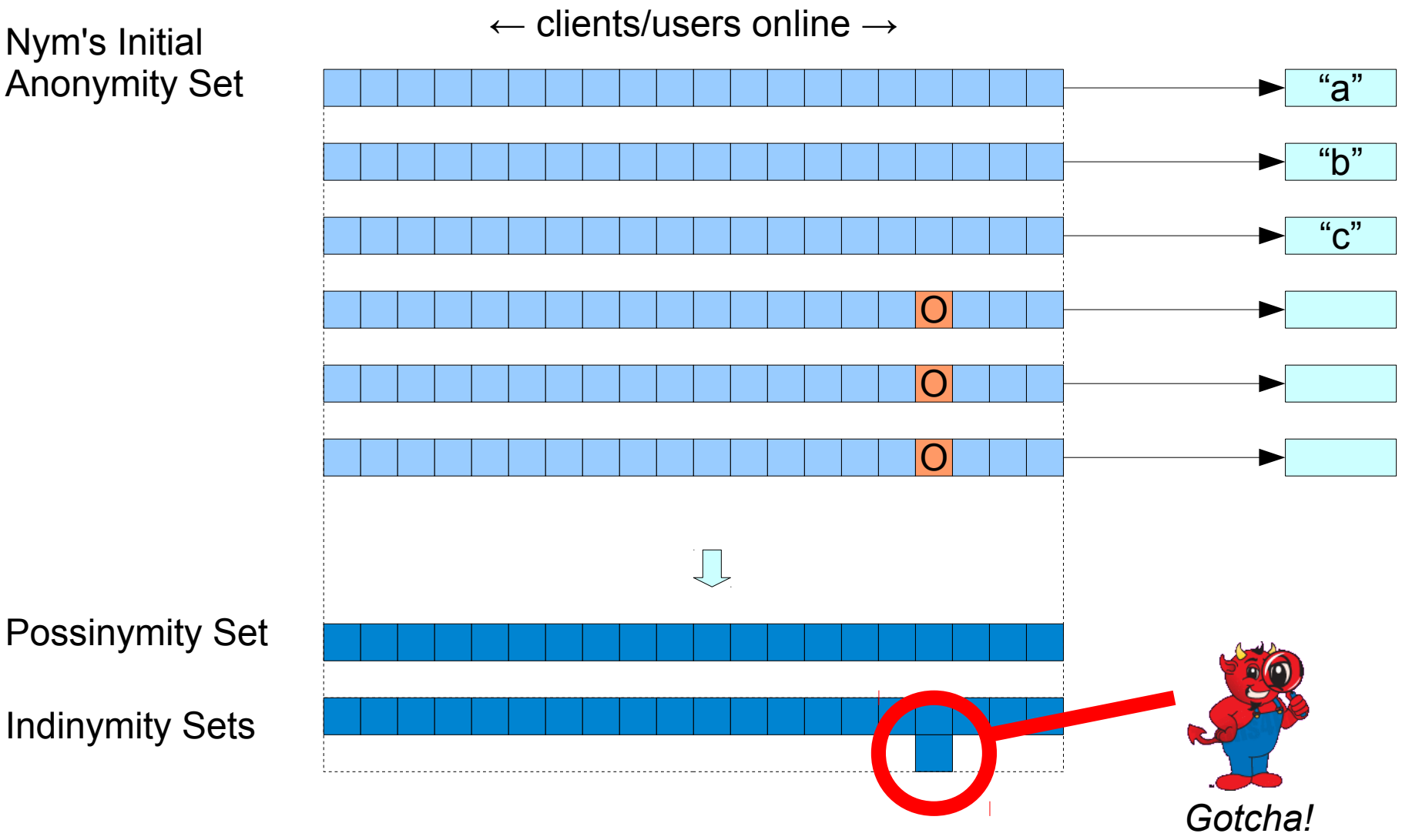
Possinymity: Possibilistic Deniability

Set of users who *could conceivably* own Nym

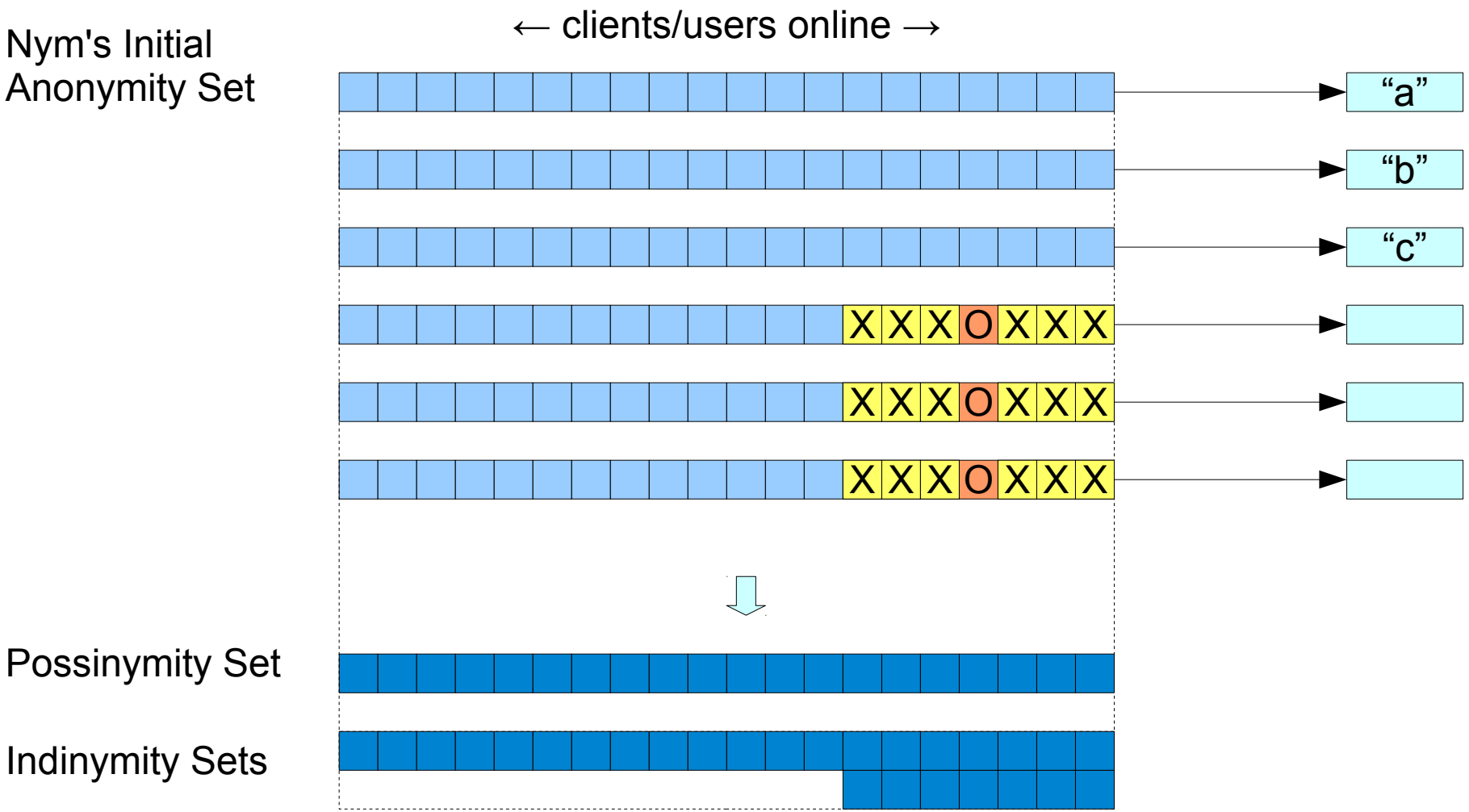
- Intersection of sets of all users *online and unfiltered* in rounds where *a message appears*
- Simplistic, but may build “reasonable doubt”



The “Statistical Disclosure” Problem

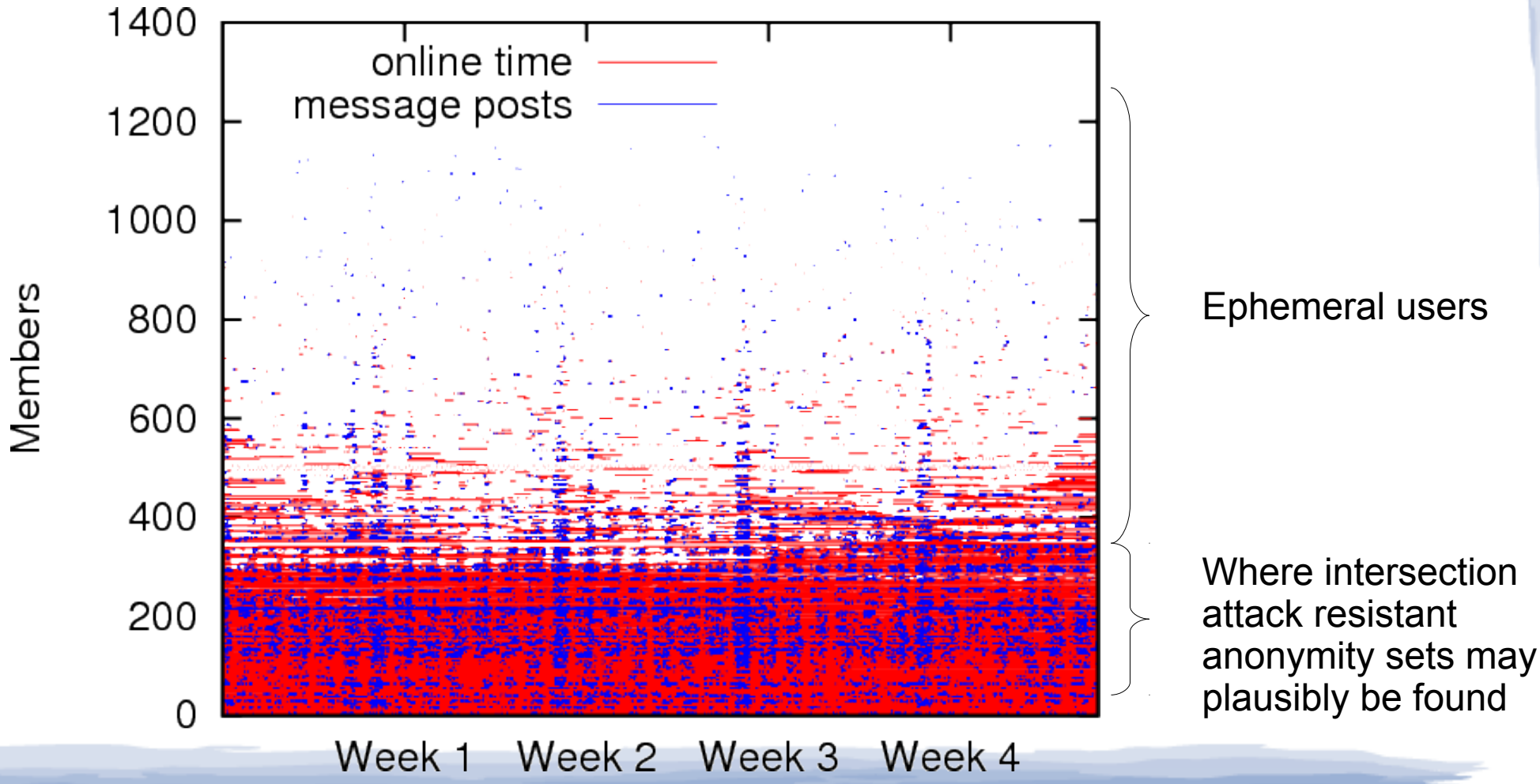


How Dissent Preserves Indinymity

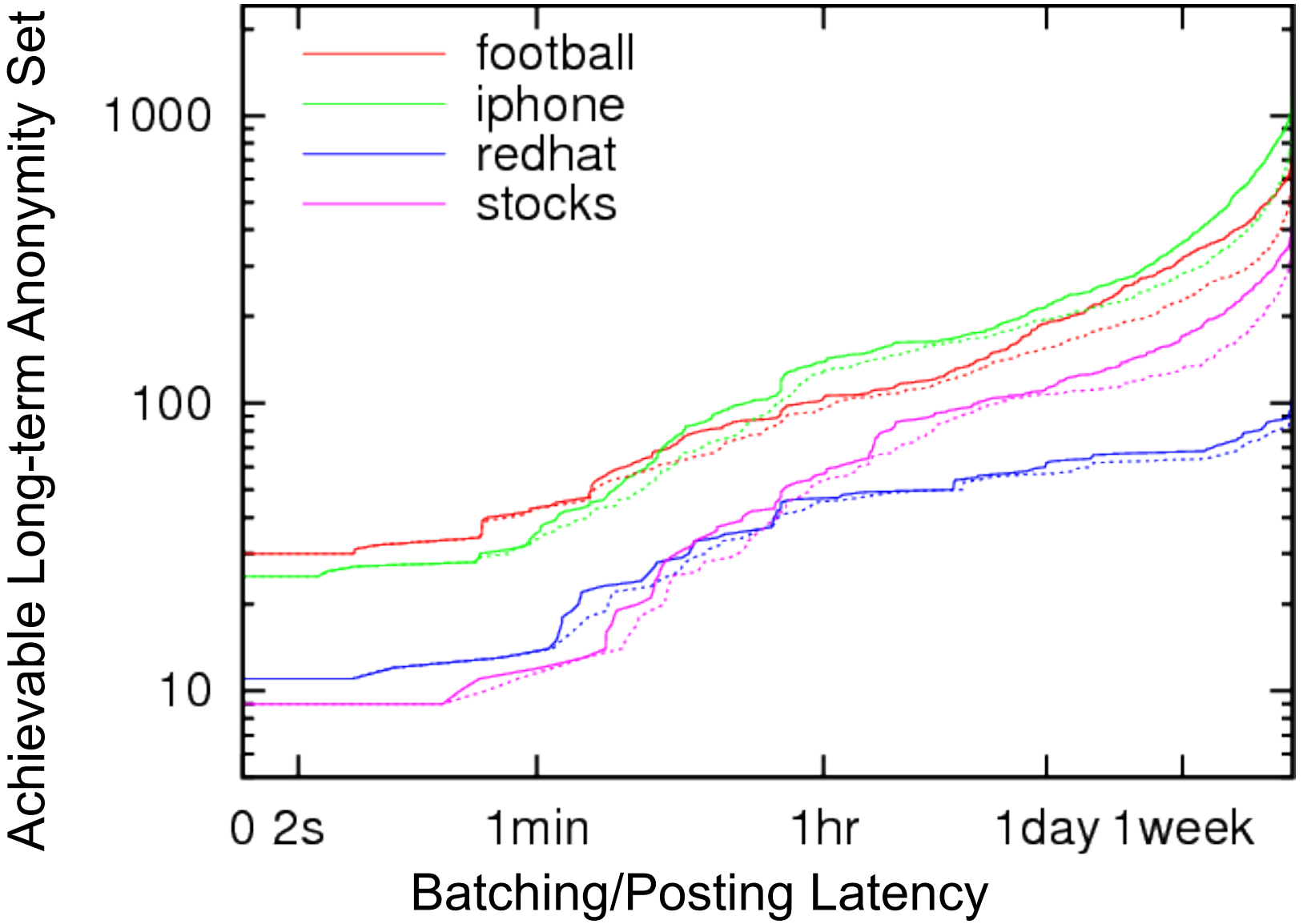


How effective? Depends on users...

Analysis based on IRC online status traces



Achievable anonymity fundamentally depends on *latency tolerance*



Major Limitations

To get good answers from simulation study, we needed “realistic” network data traces:

- “Realistic” P2P *network topology* data
- “Realistic” network *dynamics/churn* data: when clients come and go, get disconnected
- “Realistic” *user behavior* data: when users load/unload the app, etc.

...all for a prototype with no “real users” yet

Experimentation Lessons

Data-driven experimentation has become critical

- Need to be able to find relevant datasets, incorporate them readily into experiments

The “right” dataset to use may not be clear

- IRC was messaging-oriented, included user online/offline times needed for Buddies
- But online/offline times from, e.g., BitTorrent trace may be more *behaviorally* suitable
  - BitTorrent users are “asked” to remain online

Experimental Testbed Wishlist

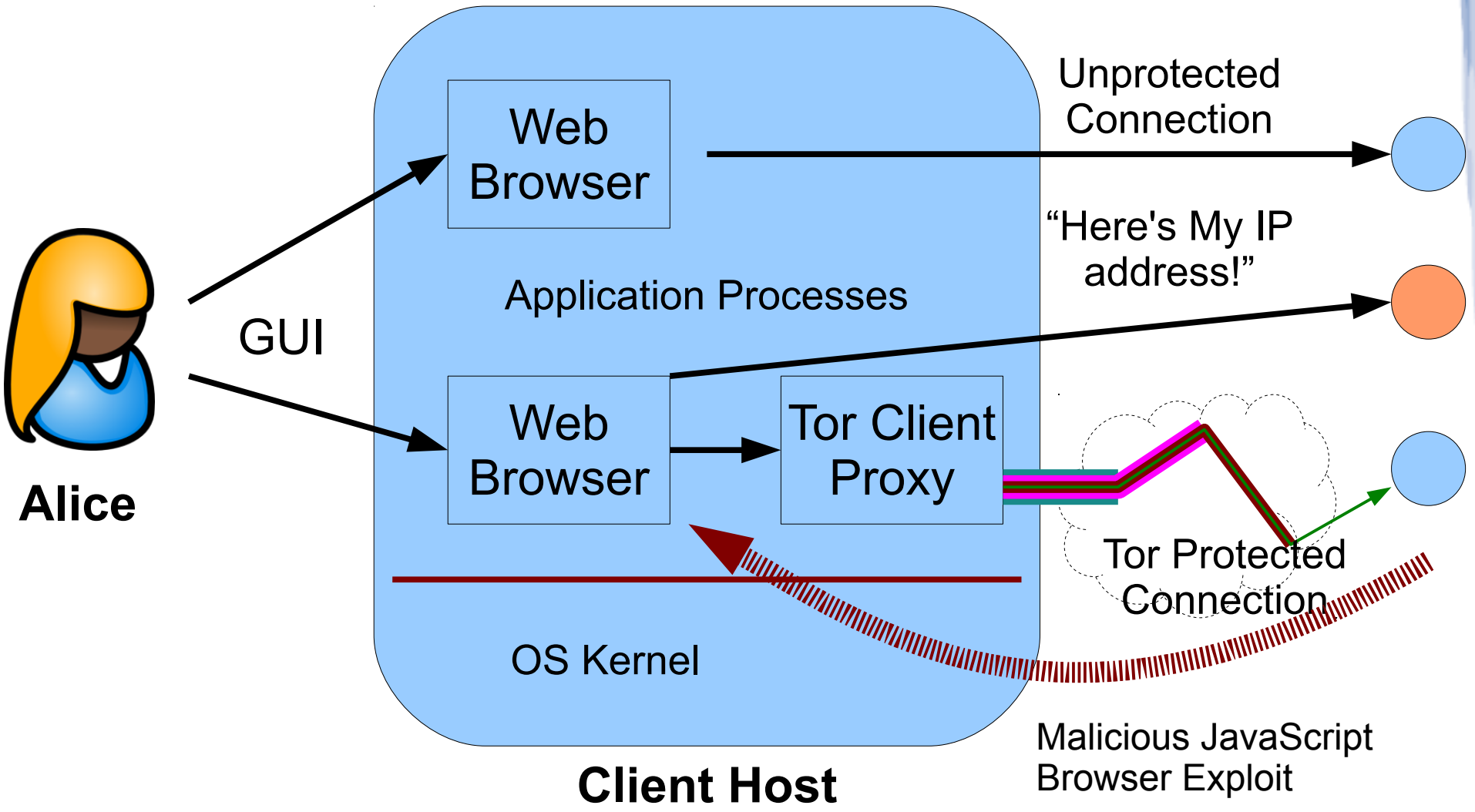
Integrated data/trace-driven experimentation

- Currently testbeds, topology/trace repos are separate things in separate places
- Build library of “standard” virtual topology datasets easy to instantiate on testbed?
  - And how to rescale “realistically” to any size (see Internet topology rescaling work)
- Library of “standard” network dynamics traces easy to apply *dynamically* on testbeds
  - e.g., simulating “realistic” churn on P2P nets

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
- Status and Ongoing Work

Typical Anonymity 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*

Attacking Tor: how the NSA targets users' online anonymity

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

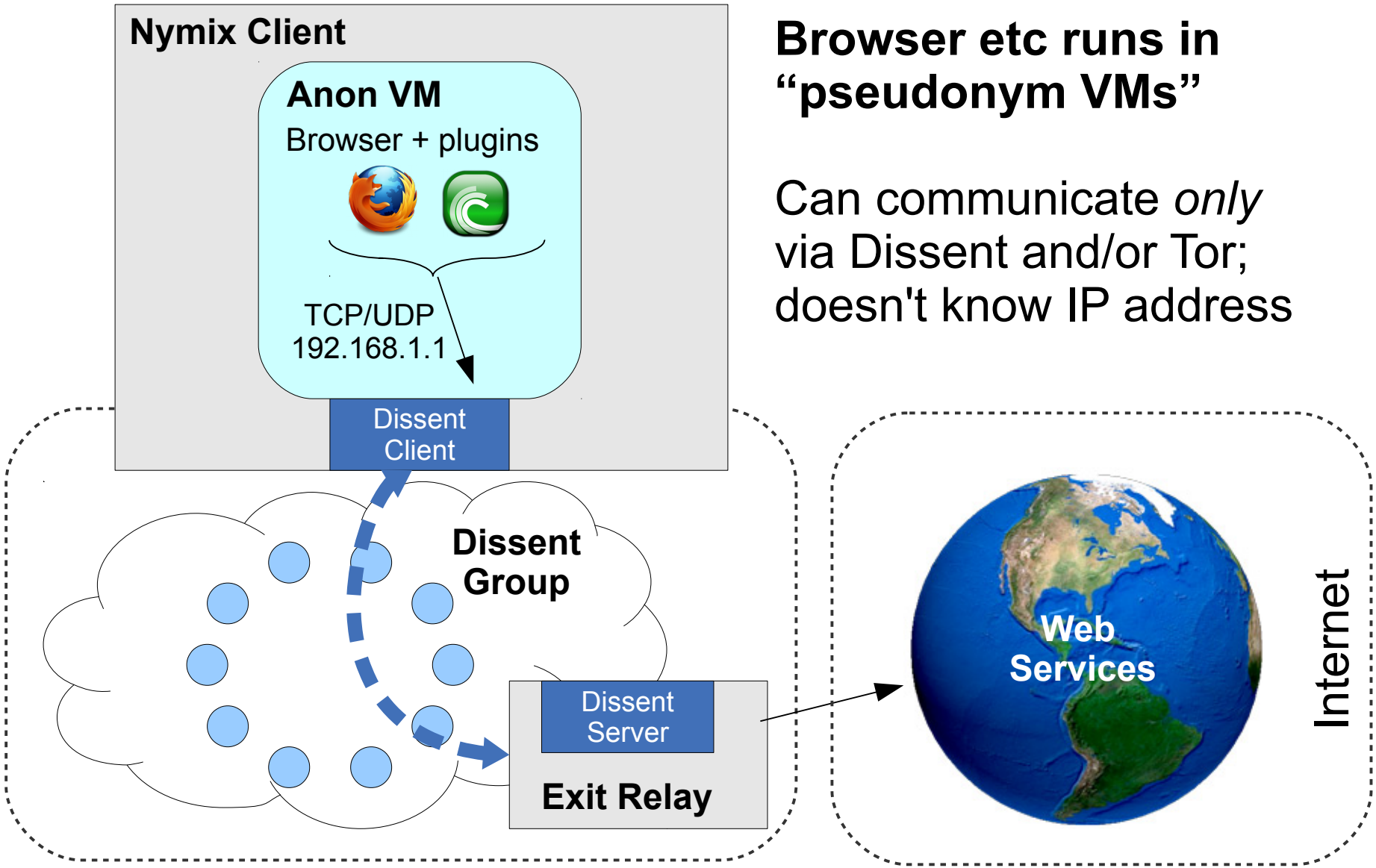
Op MULLENIZE and beyond - Staining machines

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.

Nymix [TRIOS '14]: VM-hardened Anonymous Clients



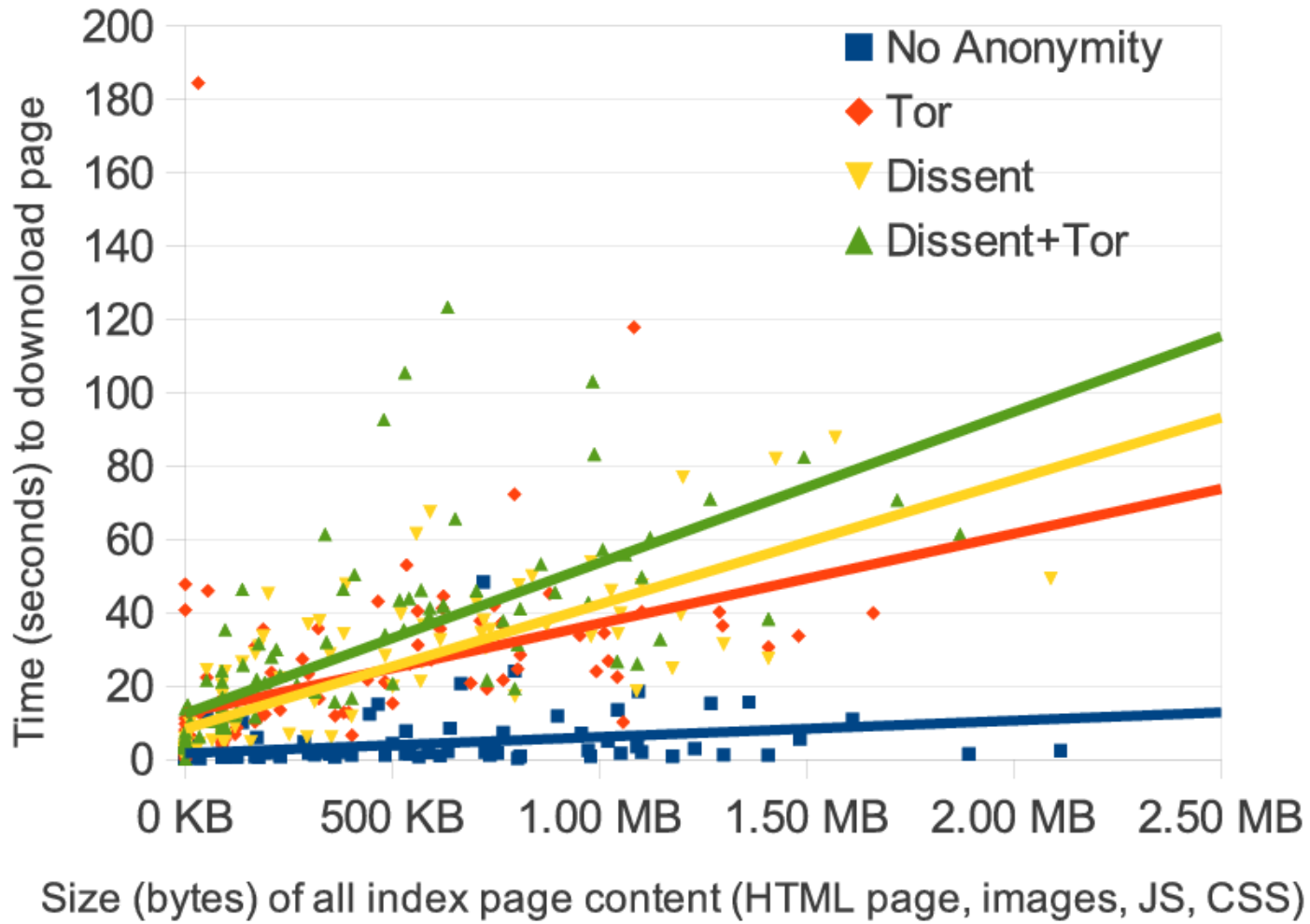
Browser etc runs in "pseudonym VMs"

Can communicate *only* via Dissent and/or Tor; doesn't know IP address

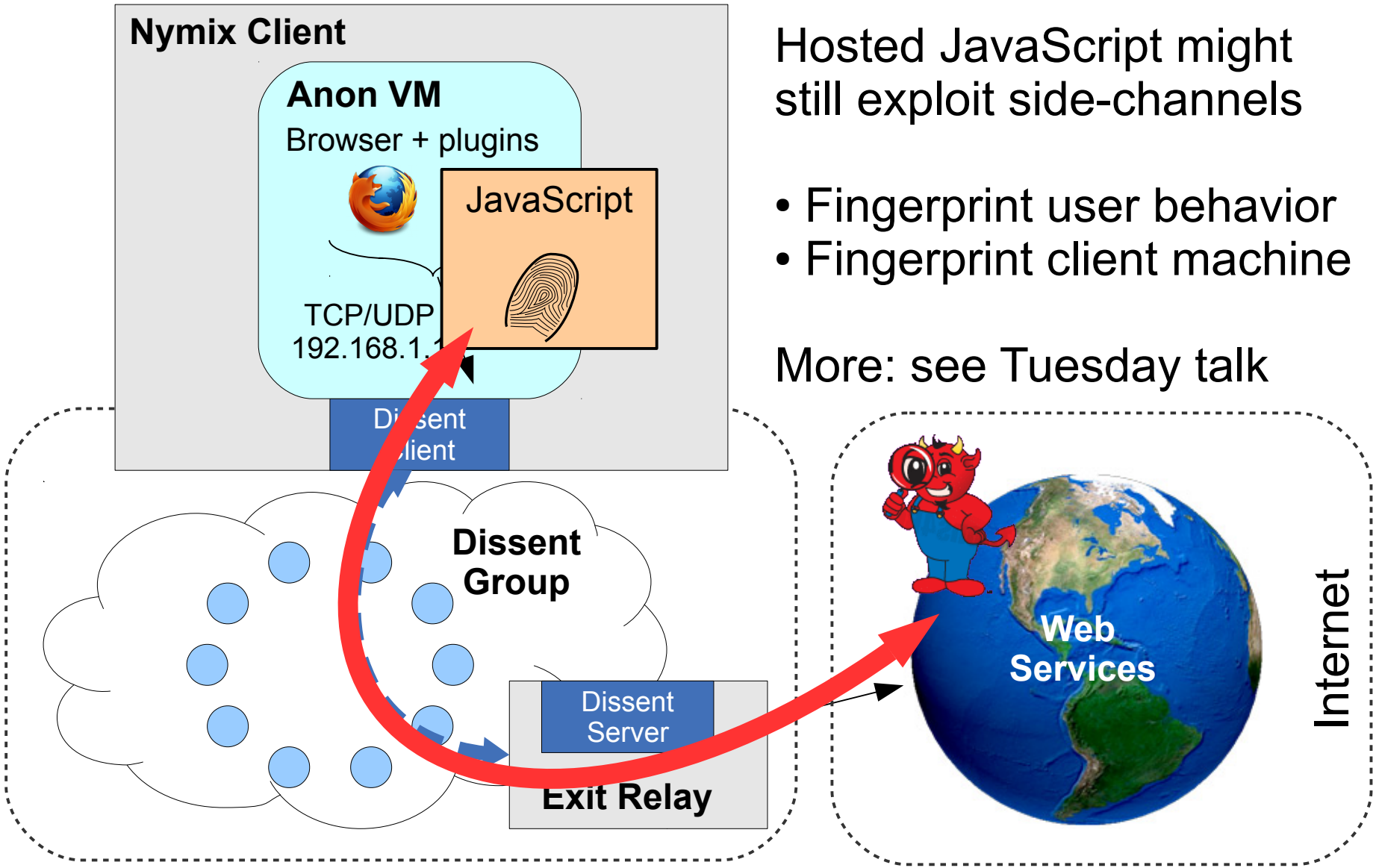
WiNon Browsing Latency

5 servers,
24 clients,
WiFi LAN
→ usability
comparable
to Tor

Illustrative only –
"apples-to-
oranges"



Major Open Challenge: Fingerprinting via Side-Channels



Hosted JavaScript might still exploit side-channels

- Fingerprint user behavior
- Fingerprint client machine

More: see Tuesday talk

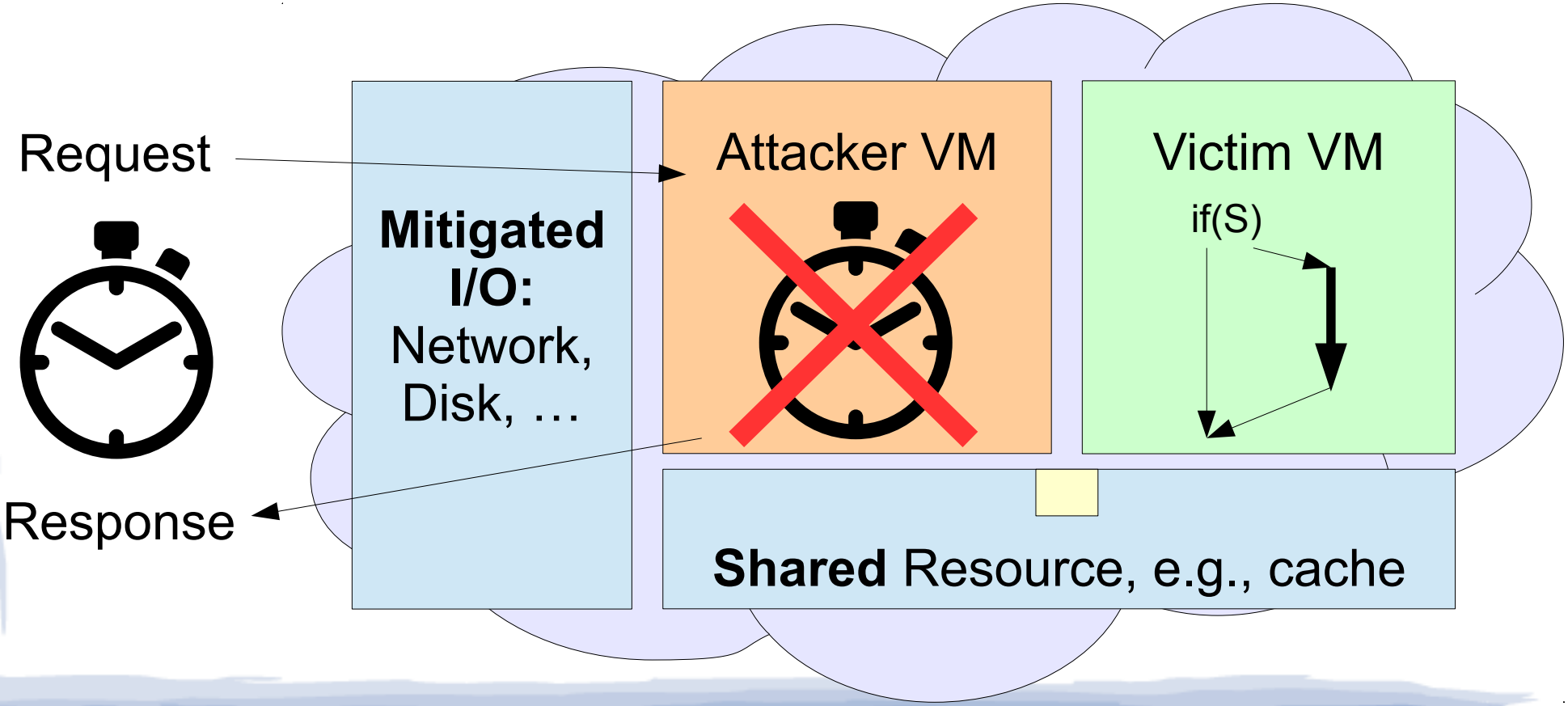
Related: Google Bouncer

Server farm ("testbed"): runs submitted Android apps, attempts to detect malware

- But what if malware knows about, tries to evade the Bouncer?
  - Many ways to fingerprint, differentiate server vs client machines: timing, CPU, etc.
  - Just "play nice" if testbed/honeypot detected
- Key problem: **Bouncer needs "anonymity"!**

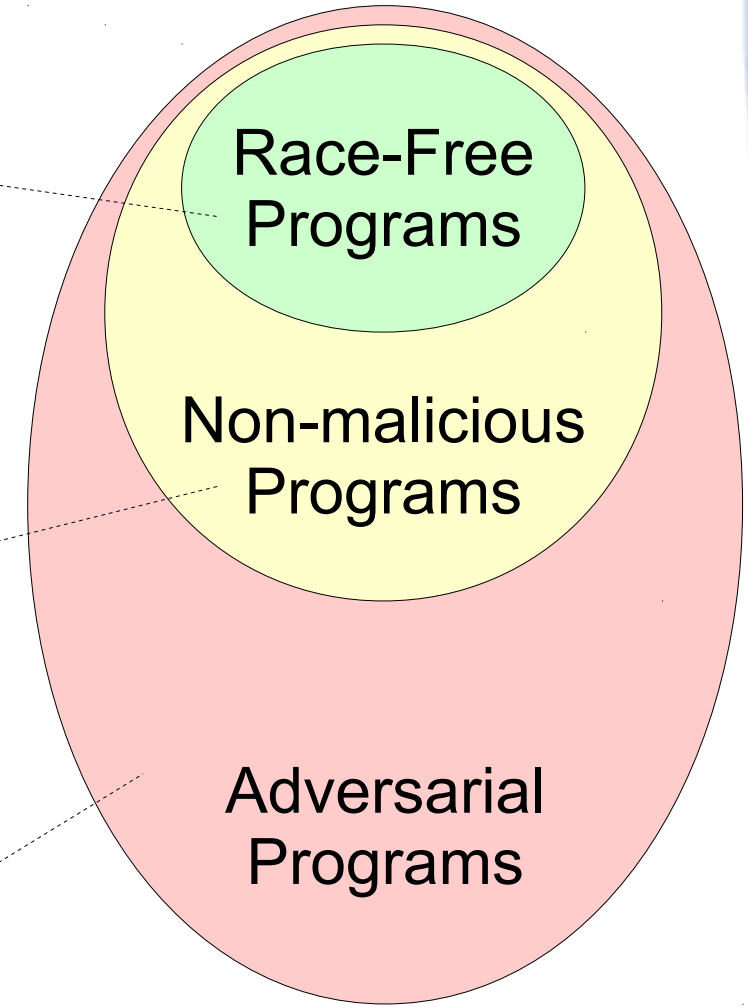
Ongoing: Side-Channel Mitigation

Use **secure, system-enforced determinism** to close or rate-limit leakage via side-channels

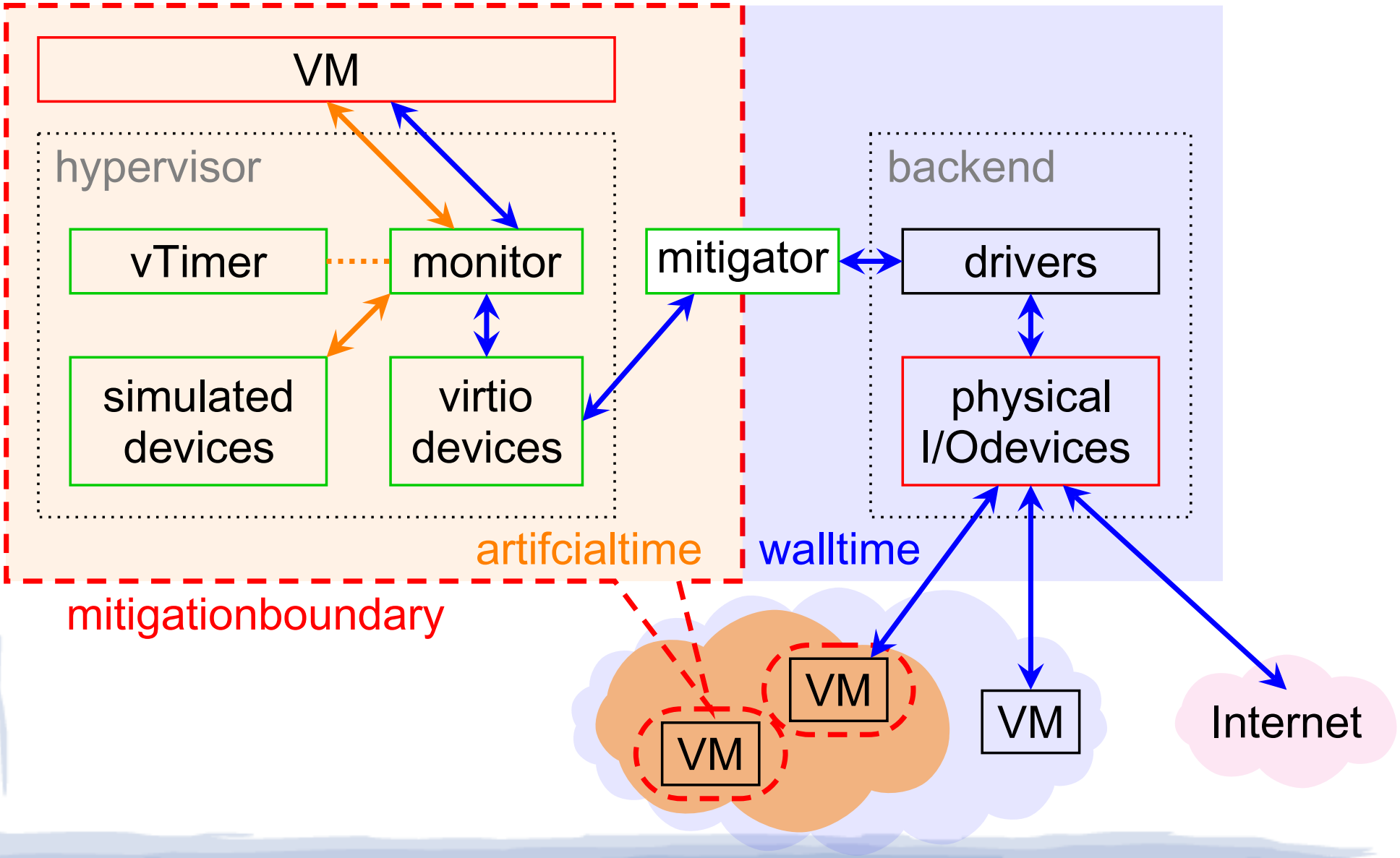


“Strengths” of Determinism

- **Weak Determinism:**
  typically library-implemented,
  works on *race-free* code
  [Grace, Kendo, …]
- **Strong Determinism:**
  typically library-implemented,
  works on *non-malicious* code
  [CoreDet, Dthreads, …]
- **Secure Determinism:**
  system-enforced,
  works on *adversarial* code
  [Determinator, Deterland]



Deterland Hypervisor Architecture



Experimentation Lessons/Wishlist

Testbeds are not just for lab experimentation; increasingly they're used in security-critical roles

- Need stronger indistinguishability from clients
  - Even when executing adversarial code
- Need determinism for multiple purposes:
  - Experiment repeatability, debugging
  - Reproducible research
  - Protecting “anonymity” of testbed nodes used for honeypots, malware analyzers

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
- **Status and Ongoing Work**

Dissent: Status and Ongoing Work

- Proof-of-concept works, code available
  - https://github.com/DeDiS
  - **Preliminary:** not at all feature-rich, user-friendly
  - **Don't** use it [yet] for security-critical activities!
- Next-generation prototype in progress
  - Decentralized anonymity at **large scales**
  - Community-area anonymous WiFi at **low latencies**
  - Anonymity applications such as **Dissent Town Hall**

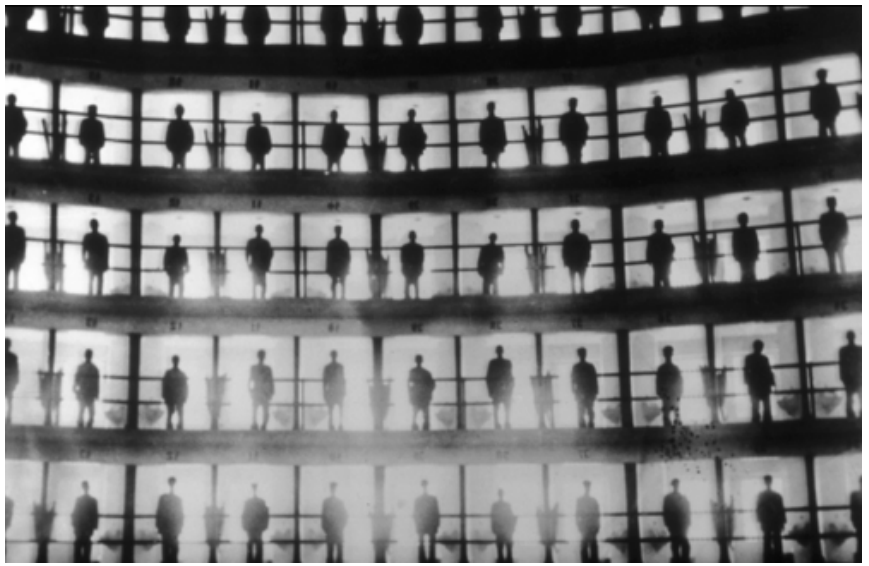
Experimentation Lessons Learned *(probably not for first or last time)*

- Evaluating how protocols scale
  - Never enough nodes, need to oversubscribe
  - Wish: testbed support for (re)scaling, validation
- Finding datasets for trace-driven experiments
  - Best datasets often unclear, often need several
  - Wish: integrate data repositories with testbeds
- Repeatability: not just for convenience anymore
  - Protect “users” and “bouncers” from fingerprinting
  - Wish: secure determinism for clients & testbeds

Conclusion

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

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



Dissent took a few early steps toward solutions (and we learned a lot getting there!)

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