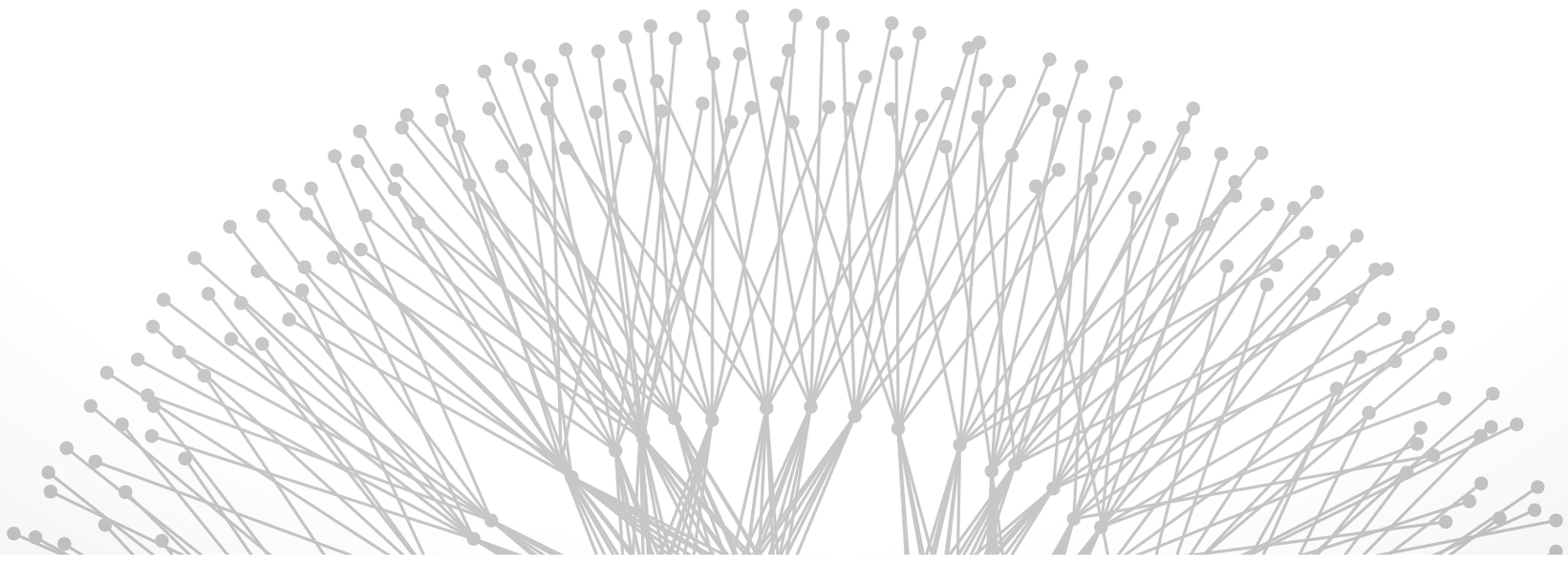
# Network Measurement

Brighten Godfrey
CS 538 April 18 2018



# Science of network measurement

Measurement goes back to the inception of the Internet

By the mid-1990s: Internet and its protocols were big, wild, organic

- Complex system: hard to predict global effects of interacting components
- Distributed multi-party system: can't see everything that's happening

Network measurement moved from "just monitoring" to a science

# Challenge #1: Emergent behavior

Example: Model packet arrivals over time at a link

Simplest common model: Poisson process

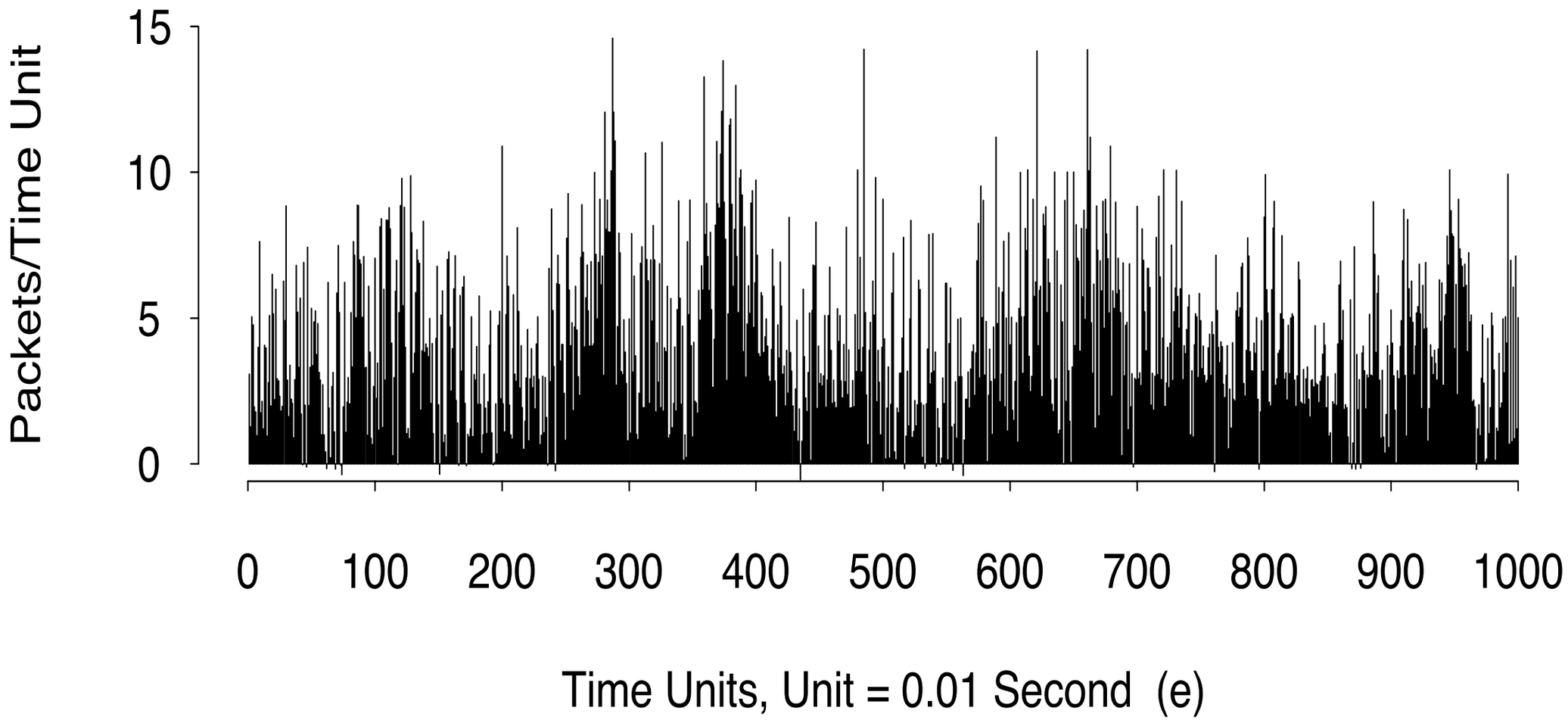
- Parameter: rate $\lambda$ (mean arrivals per unit time)
- Pr[ time till next arrival > t ] = $e^{-\lambda t}$ (exponential dist.)

Properties

- Memoryless: Even knowing entire history gives no clue as to next arrival time
- Number of arrivals in a given time interval concentrates around expected value

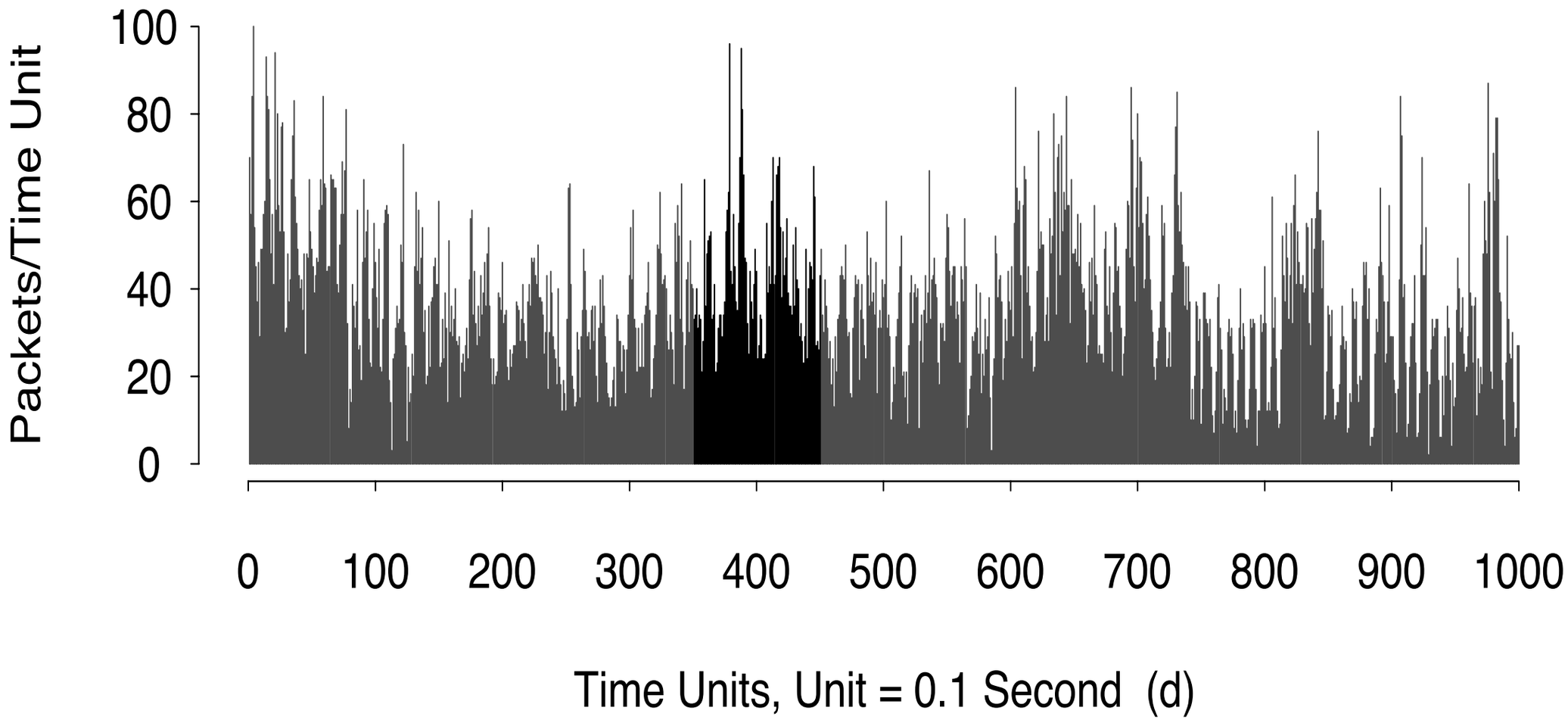
# Temporal patterns of traffic

"On the Self-Similar Nature of Ethernet Traffic"
Leland, Taqqu, Willinger, Wilson, SIGCOMM 1993



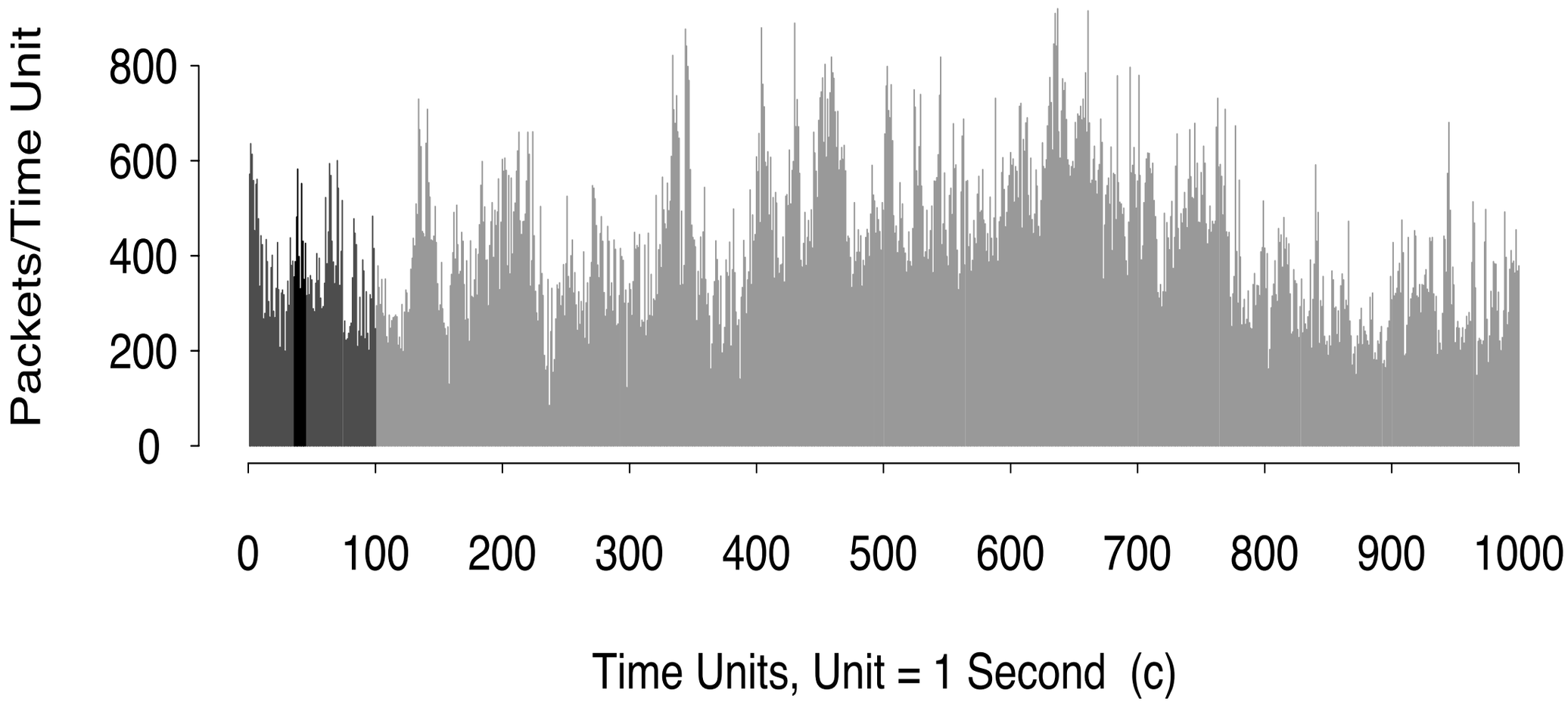
# Temporal patterns of traffic

"On the Self-Similar Nature of Ethernet Traffic"
Leland, Taqqu, Willinger, Wilson, SIGCOMM 1993



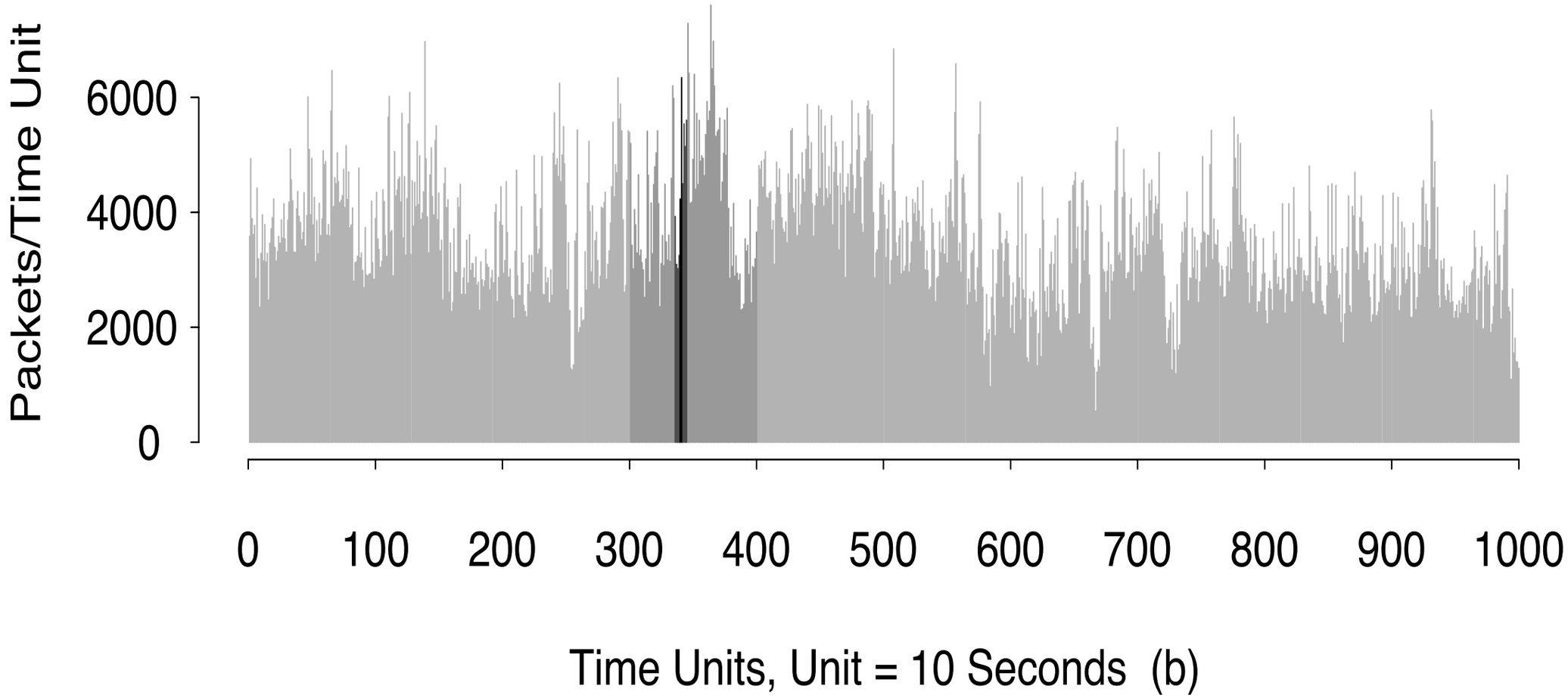
# Temporal patterns of traffic

"On the Self-Similar Nature of Ethernet Traffic"
Leland, Taqqu, Willinger, Wilson, SIGCOMM 1993



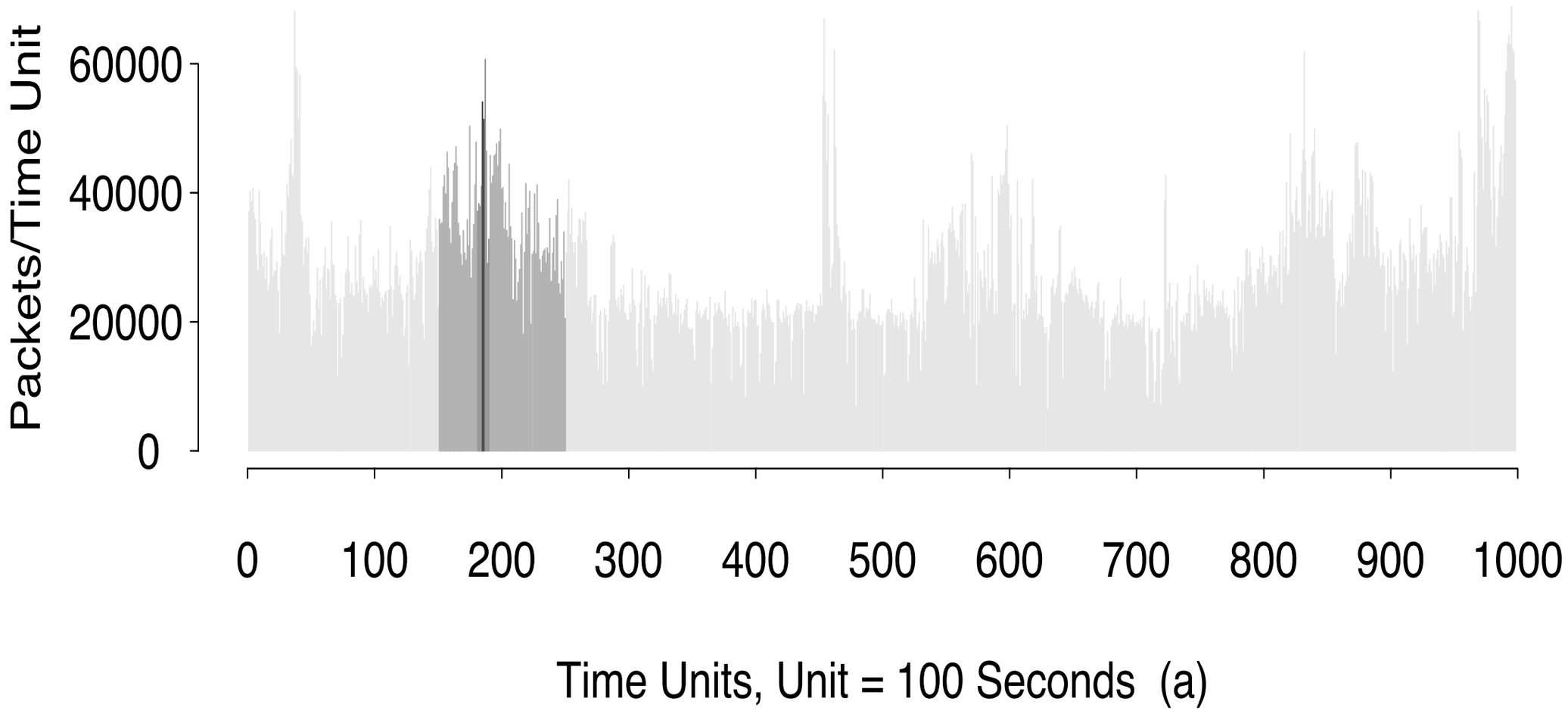
# Temporal patterns of traffic

"On the Self-Similar Nature of Ethernet Traffic"
Leland, Taqqu, Willinger, Wilson, SIGCOMM 1993



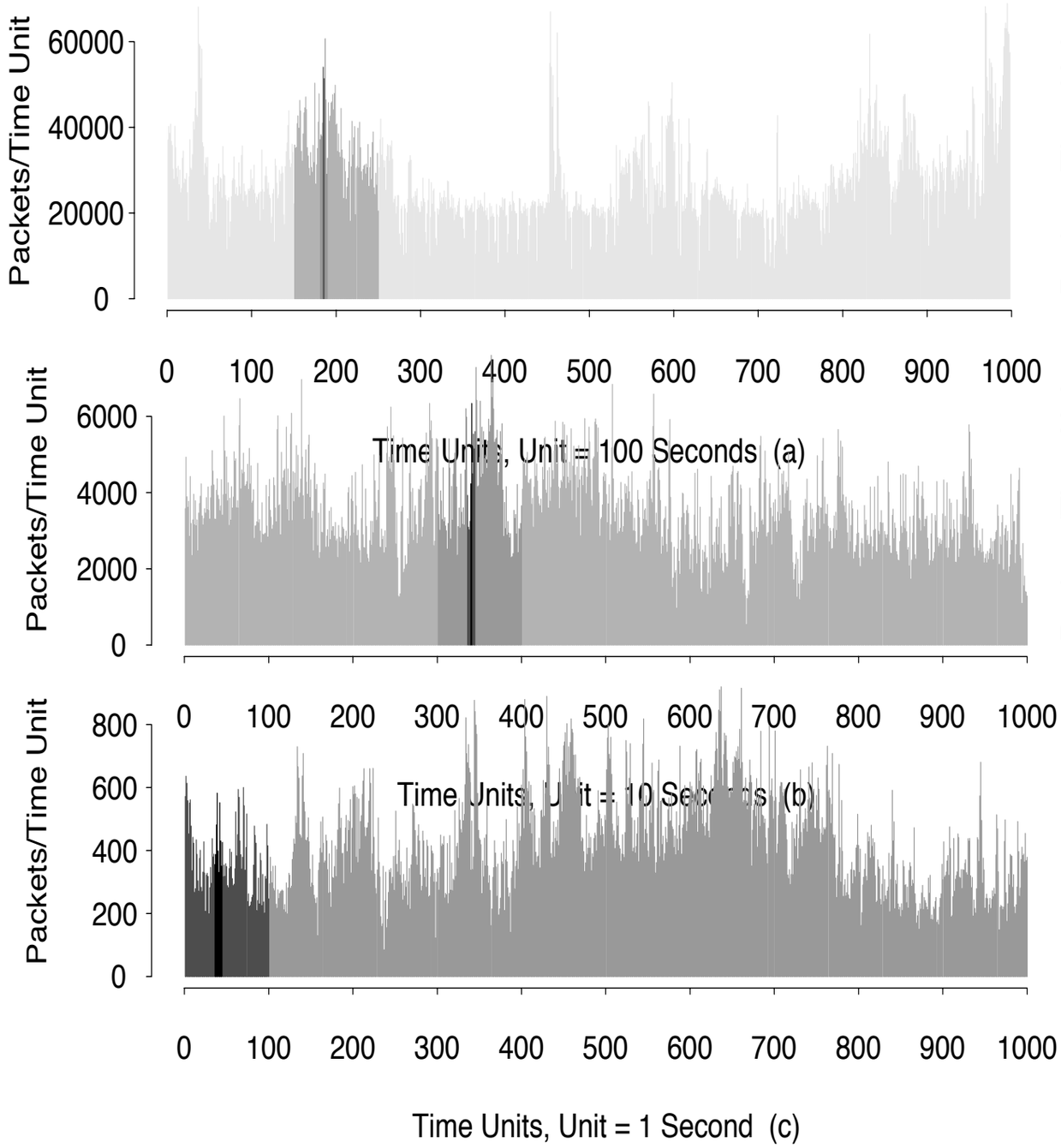
# Temporal patterns of traffic

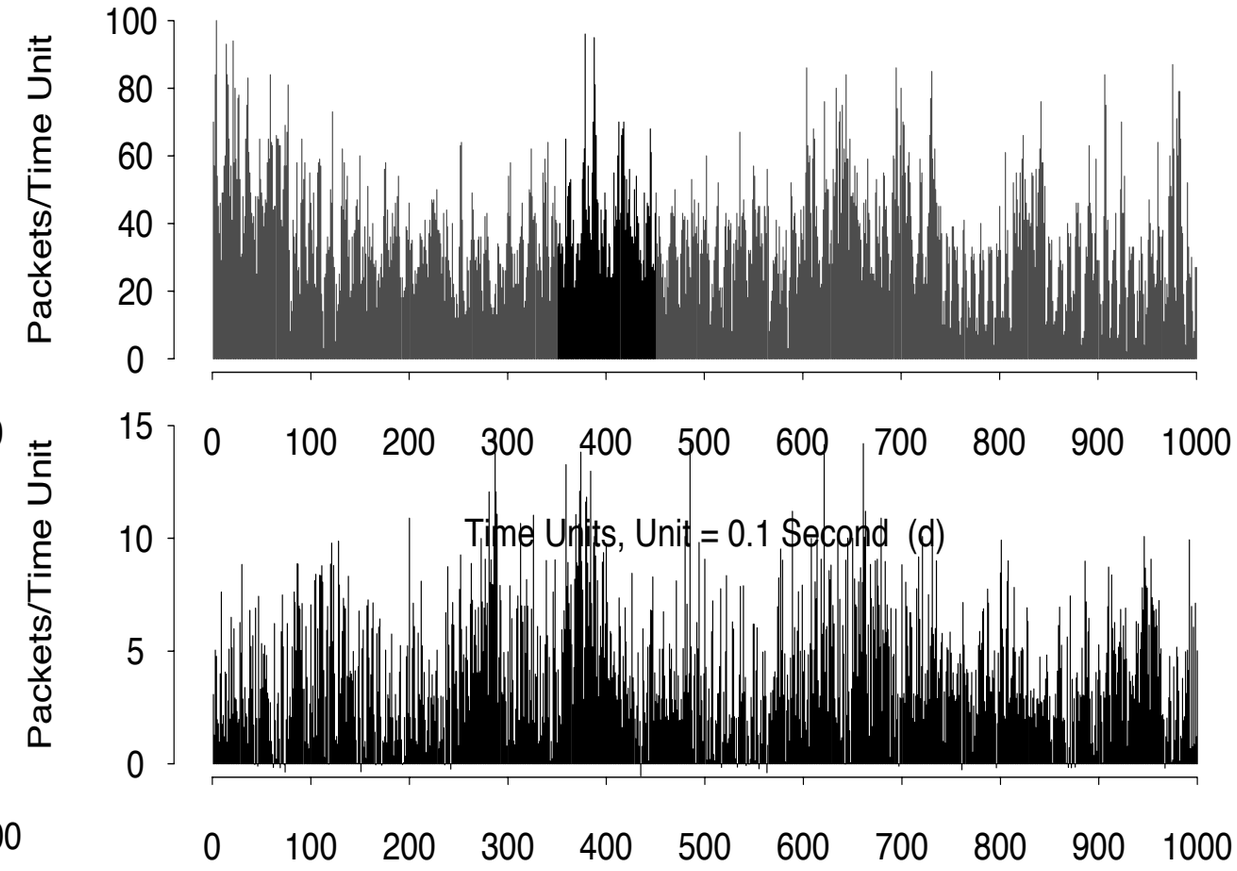
"On the Self-Similar Nature of Ethernet Traffic"
Leland, Taqqu, Willinger, Wilson, SIGCOMM 1993



# Temporal patterns of traffic

"On the Self-Similar Nature of Ethernet Traffic"
Leland, Taqqu, Willinger, Wilson, SIGCOMM 1993




**Bursty at all resolutions; Not captured by simple Poisson traffic model!**

# Challenge #2: Lack of visibility

Only a fraction of the system is visible

- For what we can observe, the cause is not obvious

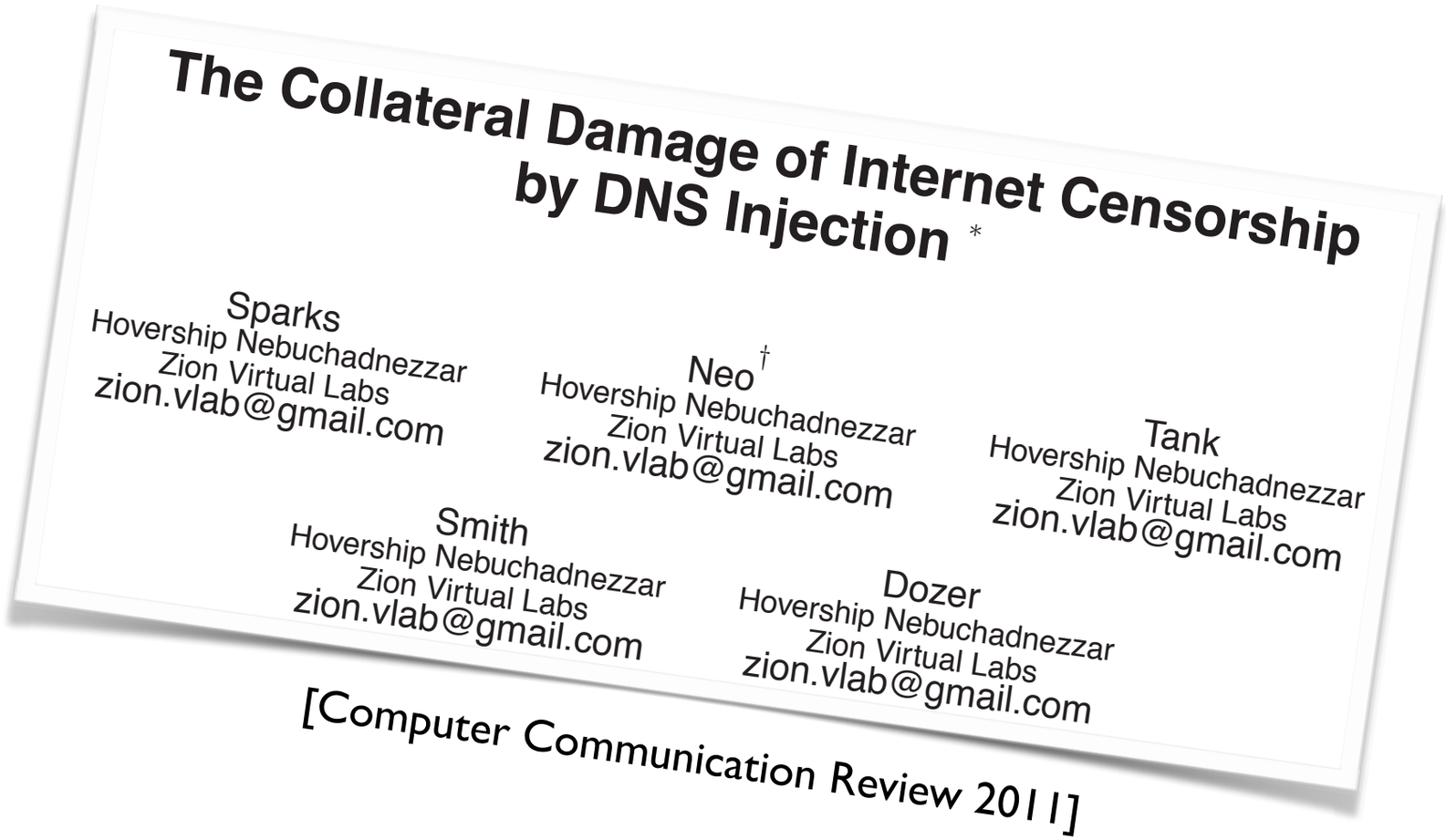
Foundational work by Vern Paxson in the mid 1990s

- "End-to-End Routing Behavior in the Internet", SIGCOMM 1996
- Loops, asymmetry, instability
- Established Internet measurement methodology: "looking inside the black box" via end-to-end measurements

| Name | Description |
|---|---|
| adv | Advanced Network & Services, Armonk, NY |
| austr | University of Melbourne, Australia |
| austr2 | University of Newcastle, Australia |
| batman | National Center for Atmospheric Research, Boulder, CO |
| bnl | Brookhaven National Lab, NY |
| bsdi | Berkeley Software Design, Colorado Springs, CO |
| connix | Caravela Software, Middlefield, CT |
| harv | Harvard University, Cambridge, MA |
| inria | INRIA, Sophia, France |
| korea | Pohang Institute of Science and Technology, South Korea |
| lbl<br>lbli | Lawrence Berkeley Lab, CA<br>LBL computer connected via ISDN, CA |
| mid | MIDnet, Lincoln, NE |
| mit | Massachusetts Institute of Technology, Cambridge, MA |
| ncar | National Center for Atmospheric Research, Boulder, CO |
| near | NEARnet, Cambridge, Massachusetts |
| nrao | National Radio Astronomy Observatory, Charlottesville, VA |
| oce | Oce-van der Grinten, Venlo, The Netherlands |
| panix | Public Access Networks Corporation, New York, NY |
| pubnix | Pix Technologies Corp., Fairfax, VA |
| rain | RAINet, Portland, Oregon |
| sandia | Sandia National Lab, Livermore, CA |
| sdsc | San Diego Supercomputer Center, CA |
| sintef1<br>sintef2 | University of Trondheim, Norway<br>University of Trondheim, Norway |
| sri | SRI International, Menlo Park, CA |
| ucl | University College, London, U.K. |
| ucla | University of California, Los Angeles |
| ucol | University of Colorado, Boulder |
| ukc | University of Kent, Canterbury, U.K. |
| umann | University of Mannheim, Germany |
| umont | University of Montreal, Canada |
| unij | University of Nijmegen, The Netherlands |
| usc | University of Southern California, Los Angeles |
| ustutt | University of Stuttgart, Germany |
| wustl | Washington University, St. Louis, MO |
| xor | XOR Network Engineering, East Boulder, CO |

[Paxson's vantage points]

# Collateral Damage of Censorship



**The Collateral Damage of Internet Censorship by DNS Injection** *


Sparks
Hovership Nebuchadnezzar
Zion Virtual Labs
zion.vlab@gmail.com

Neo †
Hovership Nebuchadnezzar
Zion Virtual Labs
zion.vlab@gmail.com

Tank
Hovership Nebuchadnezzar
Zion Virtual Labs
zion.vlab@gmail.com

Smith
Hovership Nebuchadnezzar
Zion Virtual Labs
zion.vlab@gmail.com

Dozer
Hovership Nebuchadnezzar
Zion Virtual Labs
zion.vlab@gmail.com


[Computer Communication Review 2011]

# Collateral Damage

DNS injection censorship causes collateral damage, censoring outside its jurisdiction

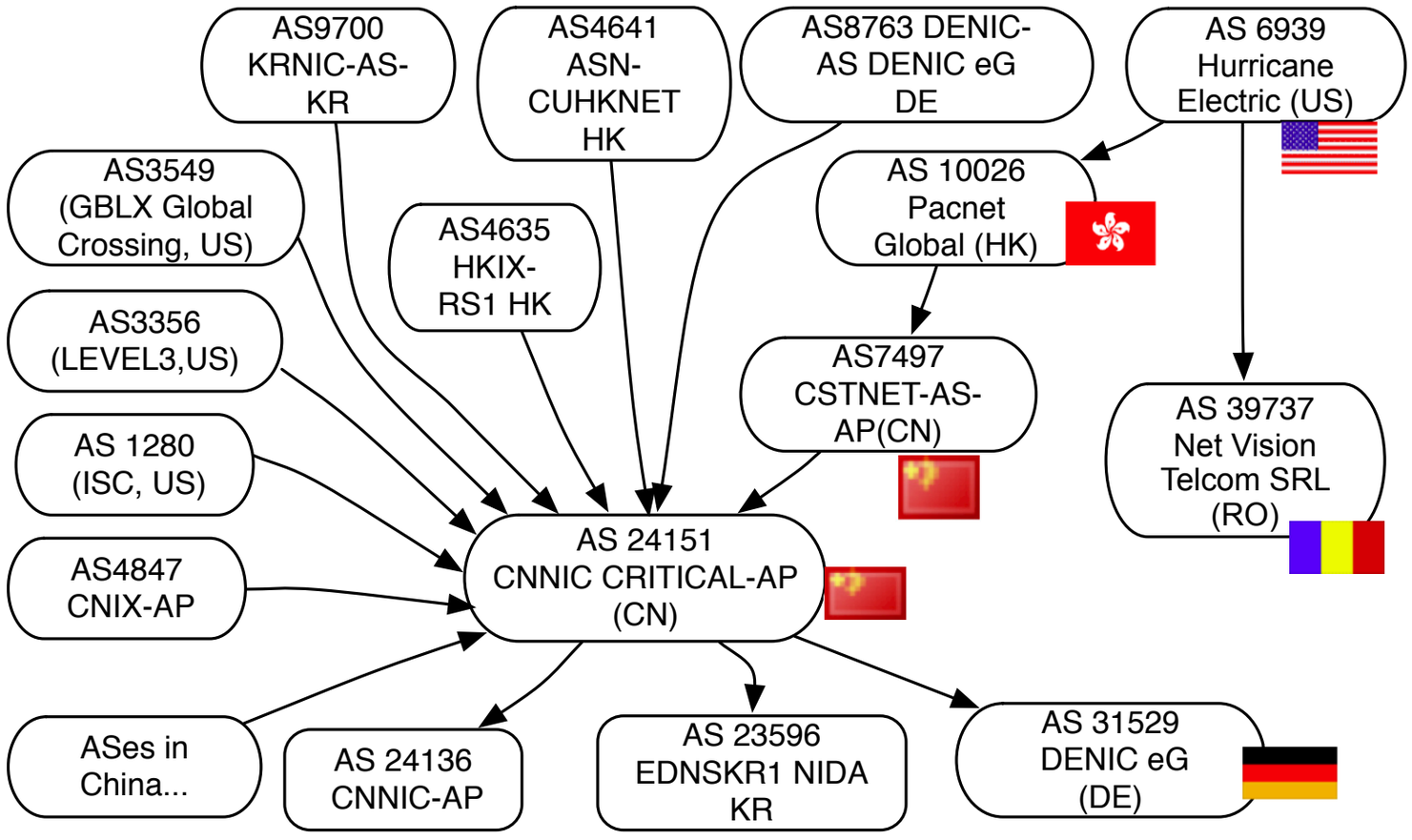



**Figure 5: Topology of ASes neighboring CNNIC**

[Figure from Anon, CCR'11]

# Collateral Damage

DNS injection censorship causes collateral damage, censoring outside its jurisdiction

## Causes

- DNS lookup involves contacting multiple servers iteratively
- Each step may be anycasted to many potential servers
- Any intermediate server or transit path could cause injected censorship

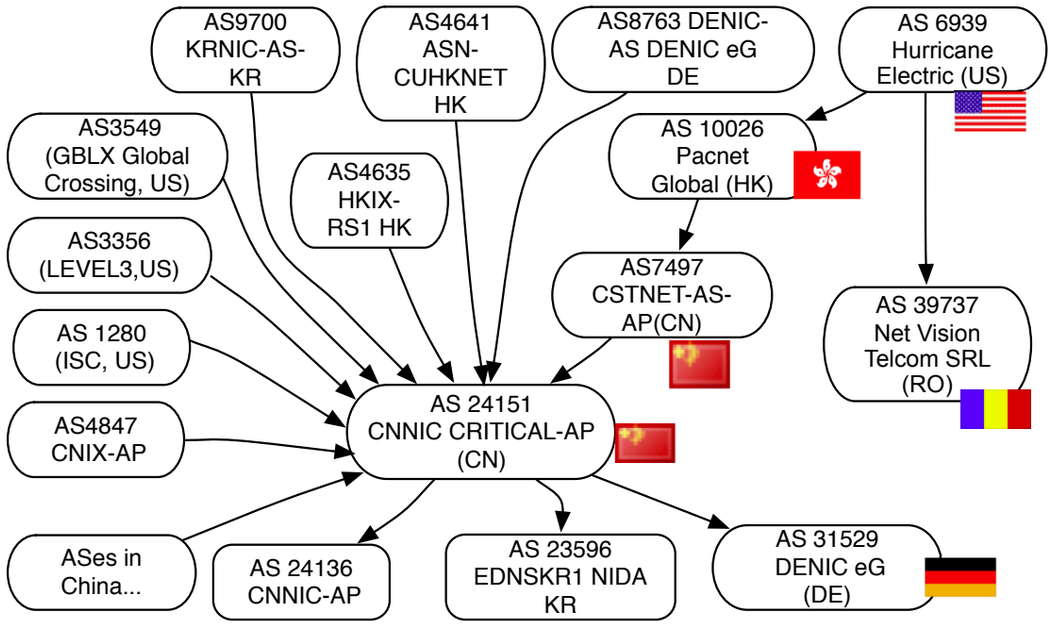


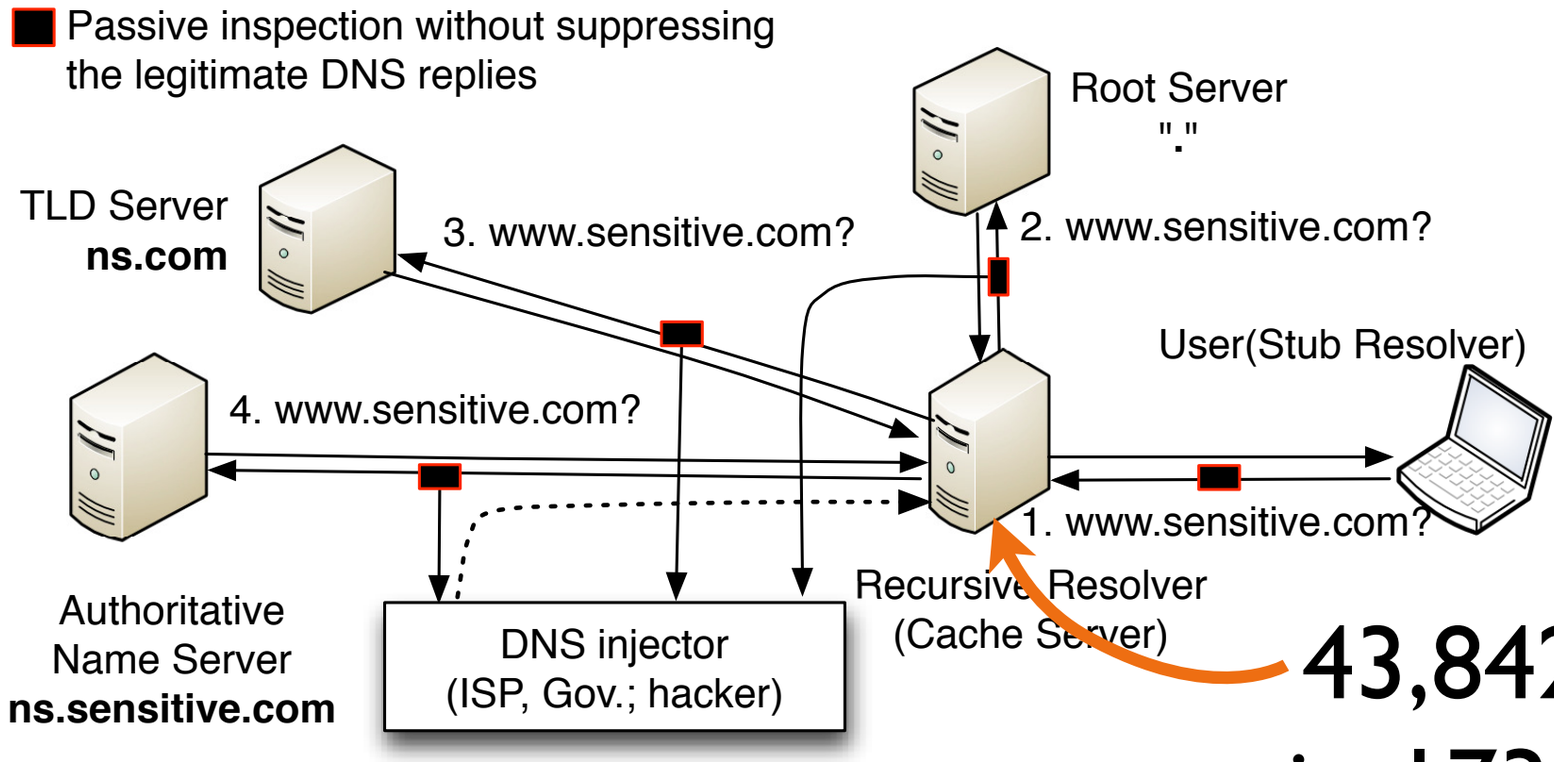

**Figure 5: Topology of ASes neighboring CNNIC**

[Figure from Anon, CCR'11]

# Vantage points

Need many vantage points to create global picture of collateral damage!

- How could we do this?




43,842 of these in 173 countries

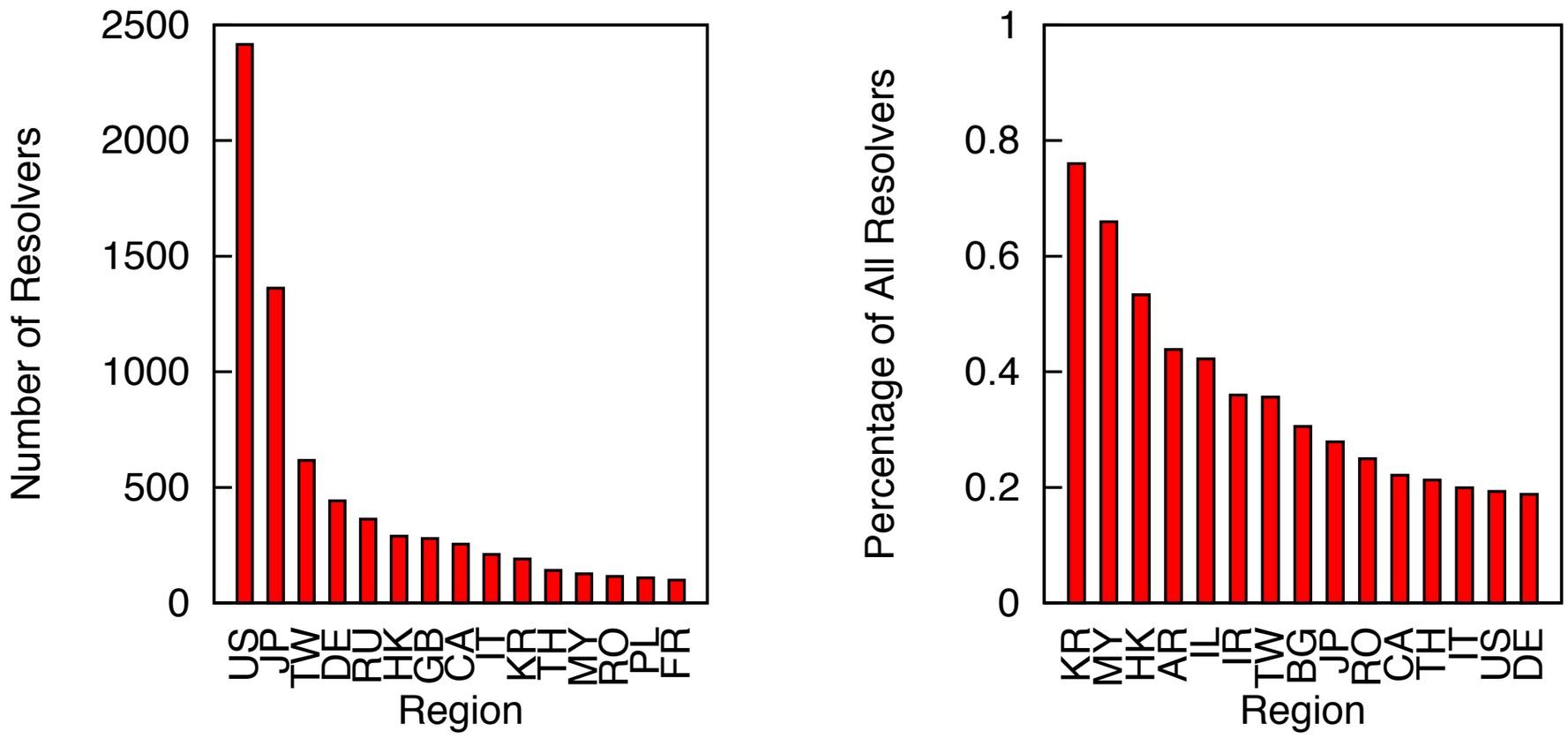
**Figure 1: DNS query process and DNS injection**

[Figure from Anon, CCR'11]

# Results

26% of resolvers tested have at least some pollution!

Most commonly polluted: names in TLD .de:




(a) Number of affected resolvers. (b) Percentage of affected resolvers.

**Figure 3: Distribution of affected resolvers for TLD `.de`.**

[Figure from Anon, CCR'11]

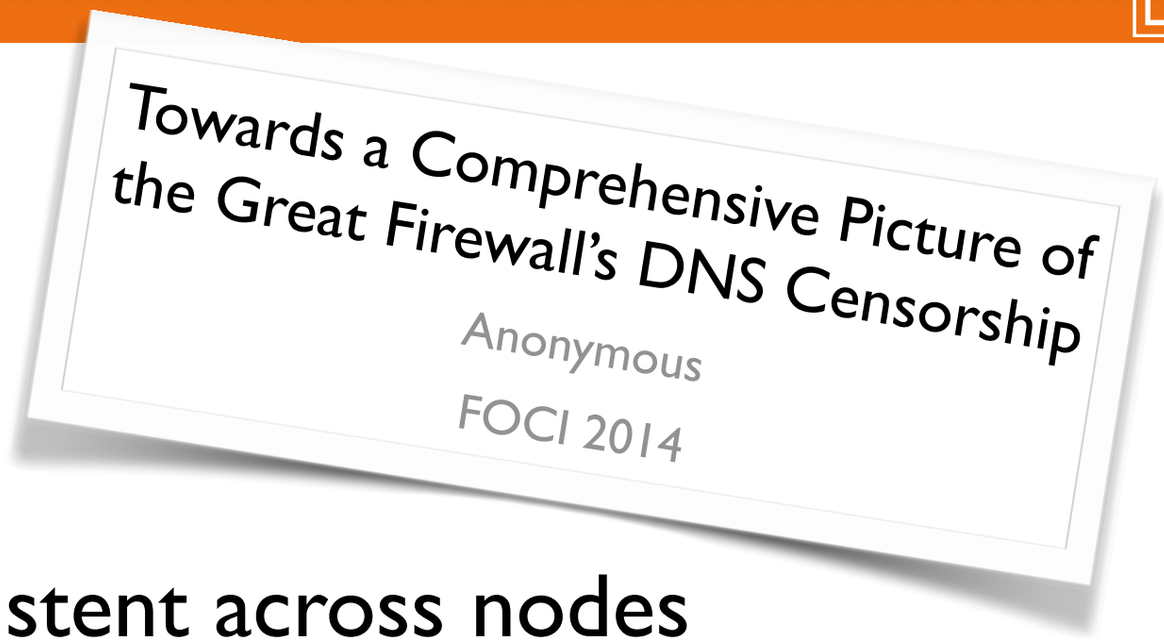
# Discussion

How could you counteract this censorship?

How could service providers offer protection?

How could an individual client protect itself?

# Towards a Comprehensive Picture




## Key points

- Centrally managed, consistent across nodes
- Pervasive (99.9% polluted)
- Deployed at edge of country
- At one node
  - Load balancing based on (src, dst) IP across 360 processes
  - 2800 censored responses per sec

# Discussion

"Our results may overestimate the GFW injector locations due to the problem of false negatives"

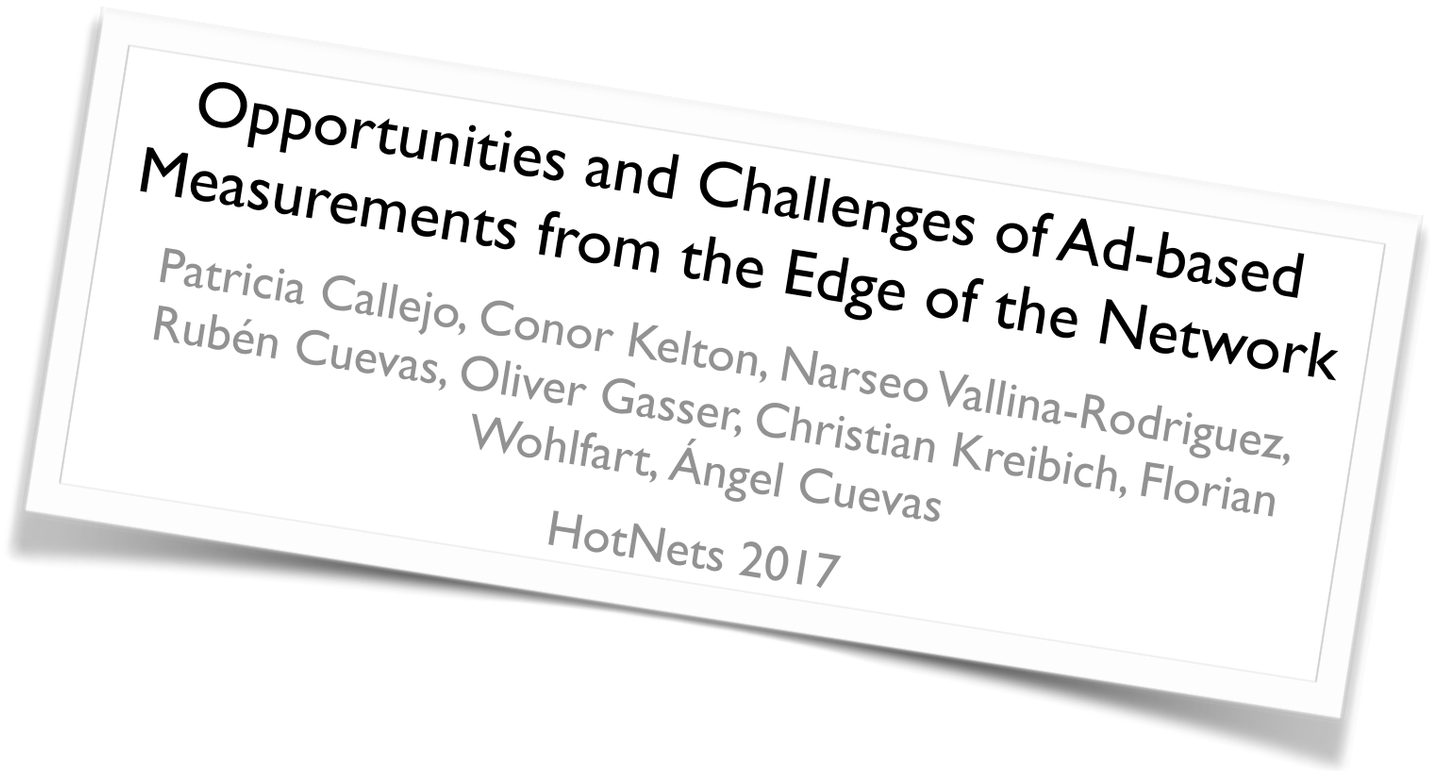
- If packets are dropped, wouldn't that cause us to miss a polluted response and underestimate GFW locations?

You can hack the Internet to infer surprising information!

- Indirect probes via King method
- Traceroutes to pointpoint censor locations
- TTL and IP ID tracking

# Discussion

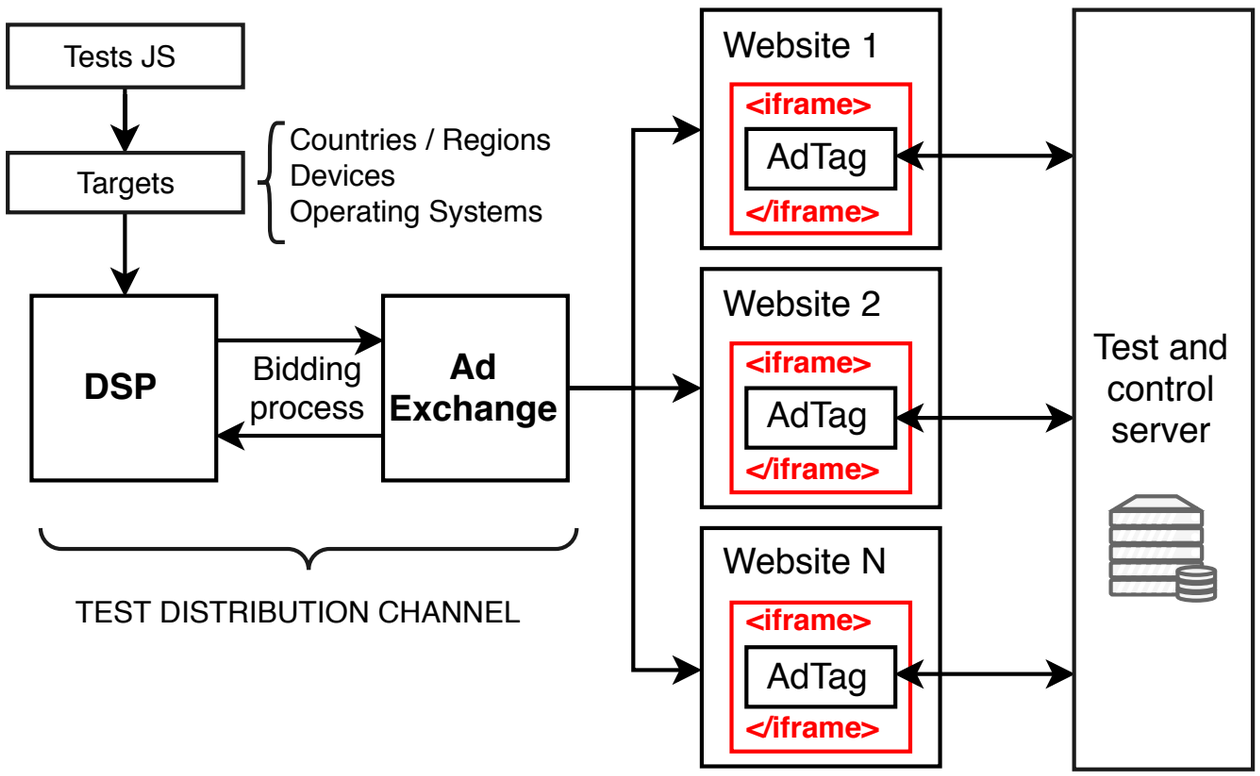
Even more vantage points are possible!



# AdTag [HotNets'17]

Platform: "programmatic advertising"

- Advertisers bid for placement in client's requested pages
- HTML5 iframe isolated from parent page
- Restricted in various ways (JavaScript making certain browser-supported API calls like WebSocket, WebRTC)
- …but allows connections to researcher's chosen server



# AdTag [HotNets'17]

High bang for the buck

- $0.10 starting "CPM" (cost per mille) at this Demand Side Platform (ad broker)

Requires careful attention to ethical concerns

- E.g. may contact illicit sites without client knowing!
- May need to be even more conservative than an IRB

Cost: about $312

| Project | Nodes†/IPs* | ASes | Countries | Time | Deployment strategy |
|---|---|---|---|---|---|
| *AdTag* | 2,500,000* | 20,700 | 185 | 7 days | Targeted ads |
| RIPE Atlas | 9,300† | 3,300 | 181 | 6 years | Testbed / Dedicated node |
| Archipelago | 181† | 146 | 60 | 10 years | Testbed / Dedicated node |
| Netalyzr | 2,200,000* | 14,500 | 196 | 6 years | Crowdsourcing / Mobile app, browser applet |
| Luminati | 1,300,000* | 14,700 | 172 | 5 days | P2P-based VPNs |

**Table 1: Comparison of a global AdTag campaign with previous studies in terms of network coverage, measurement duration, and deployment strategy. (*: number of sessions; †: number of nodes)**

# A word of caution

> *The most important difference between computer science and other scientific fields is that:* We build what we measure. *Hence, we are never quite sure whether the behavior we observe, the bounds we encounter, the principles we teach, are truly principles from which we can build a body of theory, or merely artifacts of our creations. ... this is a difference that should, to use the vernacular, 'scare the bloody hell out of us!'*

– John Day

# Announcements

Next time: Future ISP networks

Assignment 2 due Friday 11am