CSAIL

MIT COMPUTER SCIENCE AND ARTIFICIAL INTELLIGENCE LABORATORY

The Internet's Not a Big Truck:

Toward Quantifying Network Neutrality

Robert Beverly, Steven Bauer, Arthur Berger

{rbeverly,bauer,awberger}@csail.mit.edu

PAM 2007

It's not a big truck.
It's a series of tubes!

- Network Neutrality:
  - Actively debated/discussed by politicians, regulators and researchers
  - But…many definitions!
  - And…no measurements!
  - We focus on one important, well-defined dimension: “**port blocking**”

Internet Port Blocking

- Background
- Methodology
- Initial Results

Internet Port Blocking

- This work: Active/Passive hybrid measurement approach
- Main contribution:
  - Novel leverage of P2P overlay for large-scale Internet measurements
- Promising results:
  - First measurements of “Network Neutrality”

Port Blocking for Policy

- Port blocking: policy control that relies on coupling between applications and port
- IANA Well-known port assignments
- We focus on TCP port blocking, examples:
  - Comcast blocks outgoing port 25 (SMTP, prevent botnet spamming)
  - Michigan blocks incoming ports 135, 137, 139 (Microsoft file sharing)
  - UCI blocks port 1433 (MS-SQL)

Blocking and Neutrality

- ISPs may block for altruistic reasons:
  - MS-SQL worm
  - NetBIOS, etc.
- ISPs may block competing services:
  - Force use of SMTP gateway
  - Madison River Ruling [United States FCC]
- We seek to *inform the debate*
- We *do not* argue legitimacy or justifiability

Measuring Port Blocking

- Design Criteria:
  - *Generality*: Test arbitrary ports
  - *Range*: Test a wide range of networks
  - *Quantity*: Large number of tests
  - *Minimal participation*: Assume no active cooperation from remote hosts
- Approach: Referral Super Peer (RSP)

Internet Port Blocking

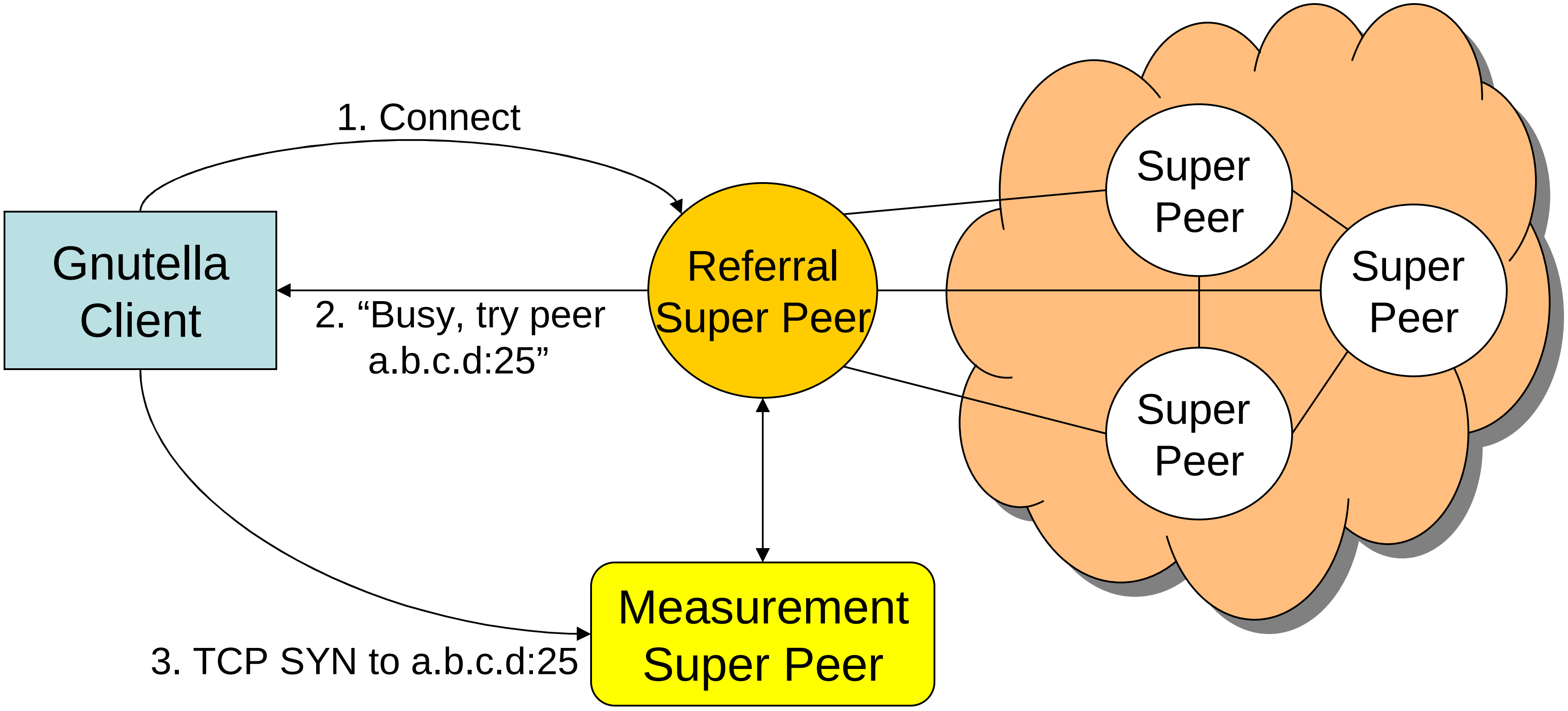
- Background
- Methodology
- Initial Results

Referral Super Peer

- RSP is a “normal” Gnutella Super Peer
- Abides by Gnutella protocol
- Bootstraps into Super Peer Mesh with standard GWebCache mechanisms

Induces clients which connect to the RSP to probe for port blocking as part of their natural overlay formation process

Infrastructure High-Level



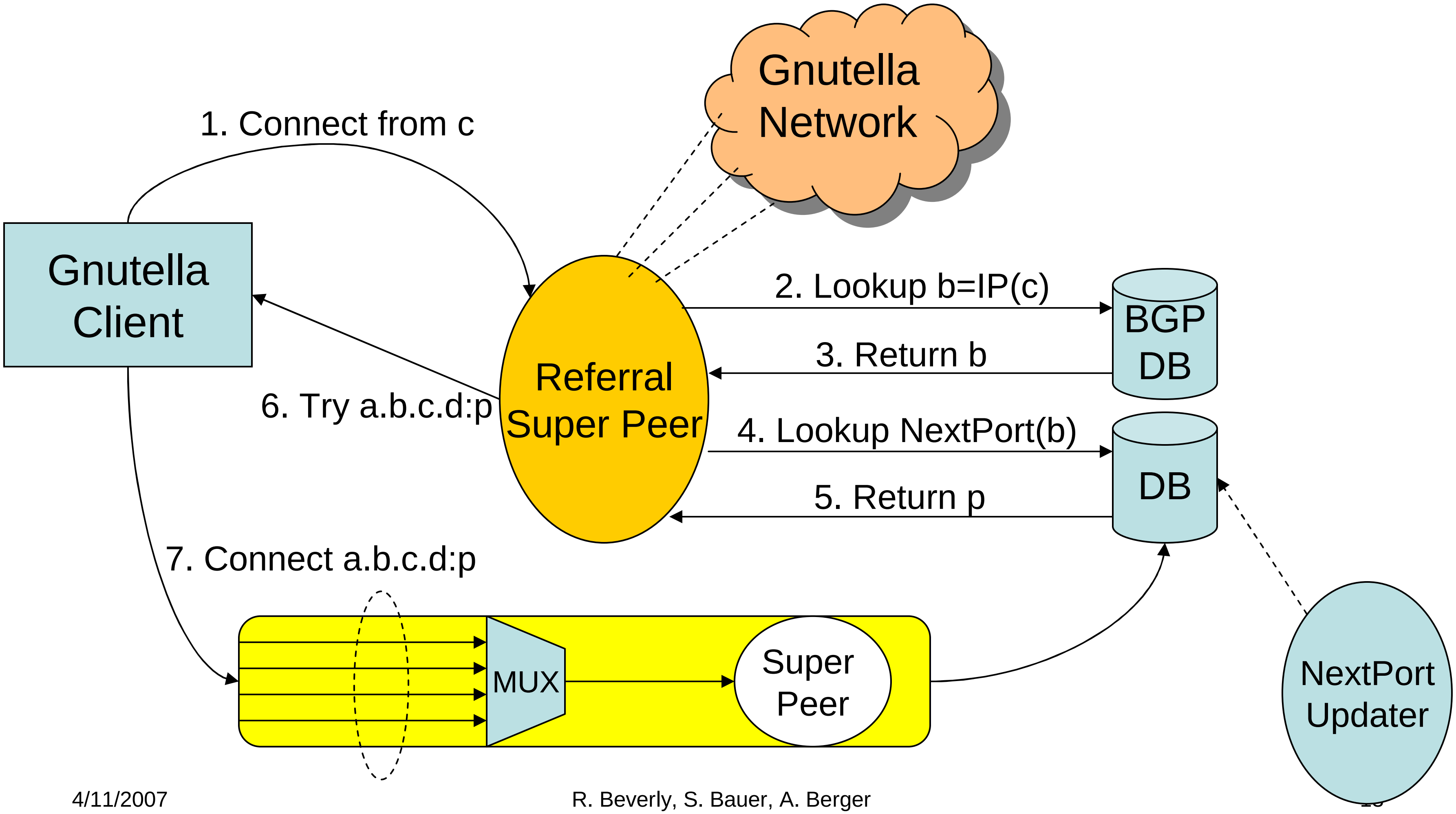
RSP is Innocuous!

- Does not disrupt or degrade overlay
- RSP and Measurement SP do not serve any content (no legality question)
- RSP only redirects clients (not harmful)
- Measurement SP is a real SP, once connected, clients receive service
- In fact, long-lived, high-bandwidth Super Peers help Gnutella network

Infrastructure High-Level

- Want to measure at a BGP prefix granularity:
  - Tie system into BGP database
- Maintain per-IP per-CIDR state:
  - Tie system to a SQL database
- Bias initial search toward contentious ports: P2P, SMTP, VPN, VoIP, etc.

Full Methodology



A Map of Internet Port Blocking

- Devil in the details…
- Consider a busy referral for port *p* to client *c* residing in CIDR b
- Observe TCP SYN from *c* for *p*:
  - *p* is not blocked on path from *b*
  - *b* is neutral to applications using *p*
- No TCP SYN from *c* for *p* implies either:
  - *p* is blocked on path from *b*
  - *c* ignored referral

Probabilistic Inference

- Empirical prior probability
- For 99.5% probability that *i* non-responsive referrals indicates *b* blocks *p*:
  - $P(n(p,b)=0|H(p,b,i)=0) = 0.995$
- Solution (see paper for formal derivation):
  - $i=\log_{0.9}(0.005) \approx 50$

Must send and not observe responses for ~50 referrals to clients in *b* for port *p* to conclude that *p* is blocked on the path from *b*

Why Gnutella?

- Exploit the Gnutella P2P overlay to easily:
  - Globally advertise a service
  - Draw (lots of) incoming connections toward us
  - Gnutella is estimated at ~3.5M users
- Test large portions of the Internet topology
- Method is general; any service which allows arbitrary `IP:port` redirection suffices
- Current work using same ideas with HTTP

Internet Port Blocking

- Background
- Methodology
- Initial Results

Measurement Bias

- Unbiased measurements from non-trivial portion of Internet (~31k prefixes ≈15% of Internet)
- Cannot measure networks that disallow Gnutella content filtering
- RSP listens on non-default port to avoid Gnutella port blocking
- Networks we don’t measure could block more, fewer or different ports than we find

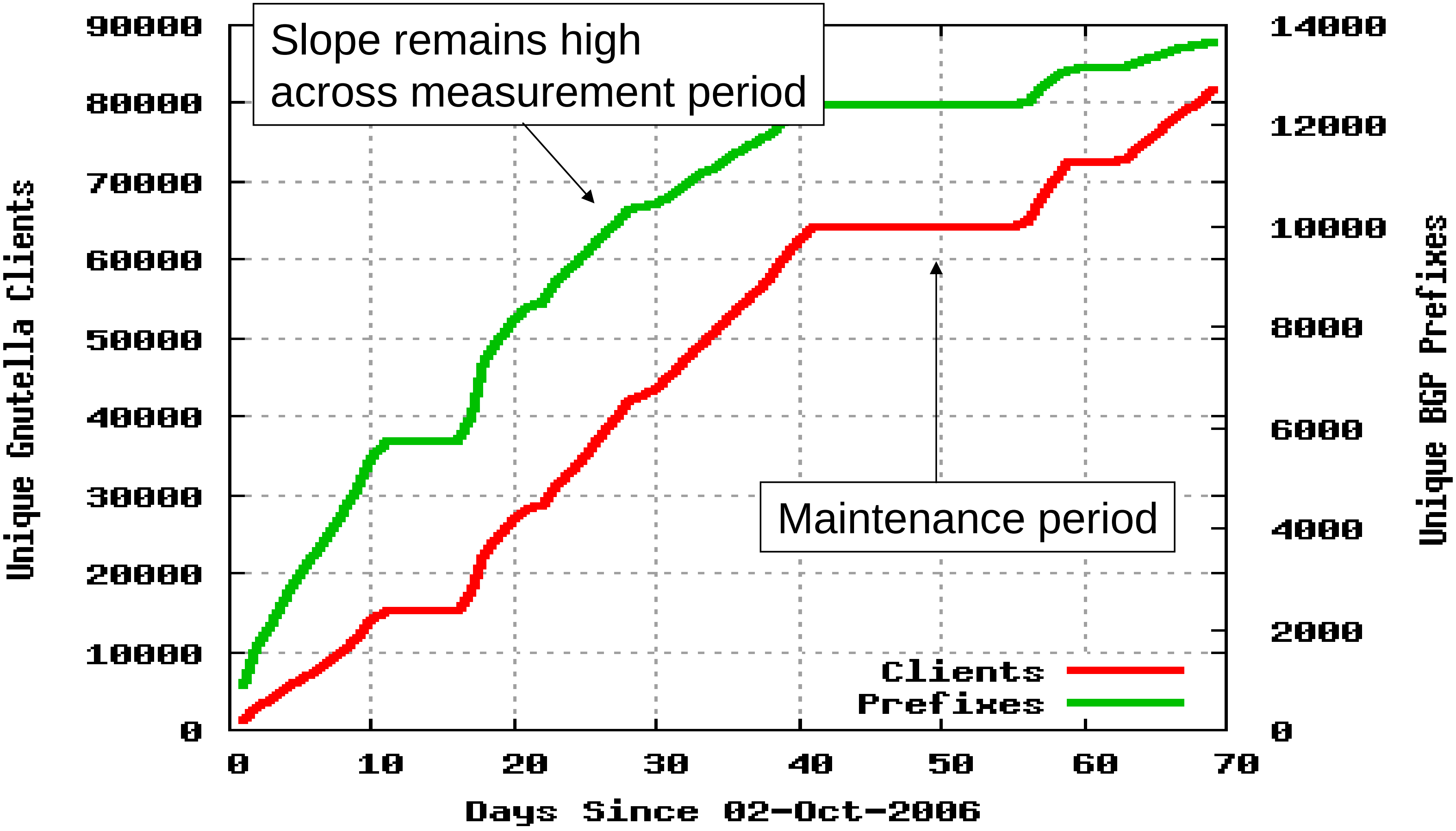
Efficacy of Methodology

- Collected data for two months: October to December 2006
  - ~31k unique BGP prefixes
  - ~1M TCP connections
  - ~72k unique Gnutella clients
  - ~150k referrals sent

Size of Network

- First question: what is the rate of new unique clients and BGP prefixes?

Rate of New Clients



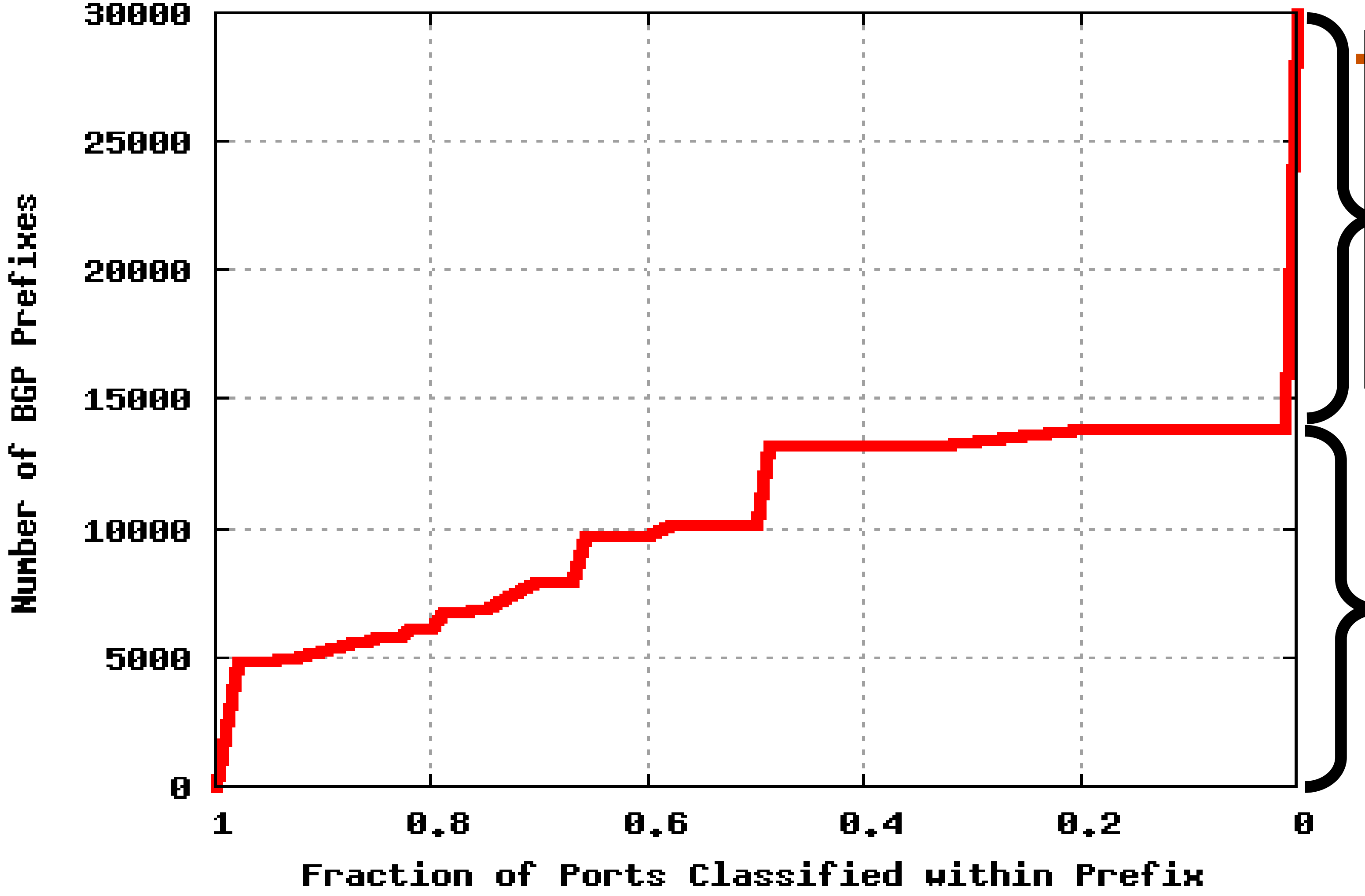
Slope remains high across measurement period

Maintenance period

System Performance

- Second question: how well does the system allow us to make inferences?

Fraction of Ports Classified Within Each Prefix



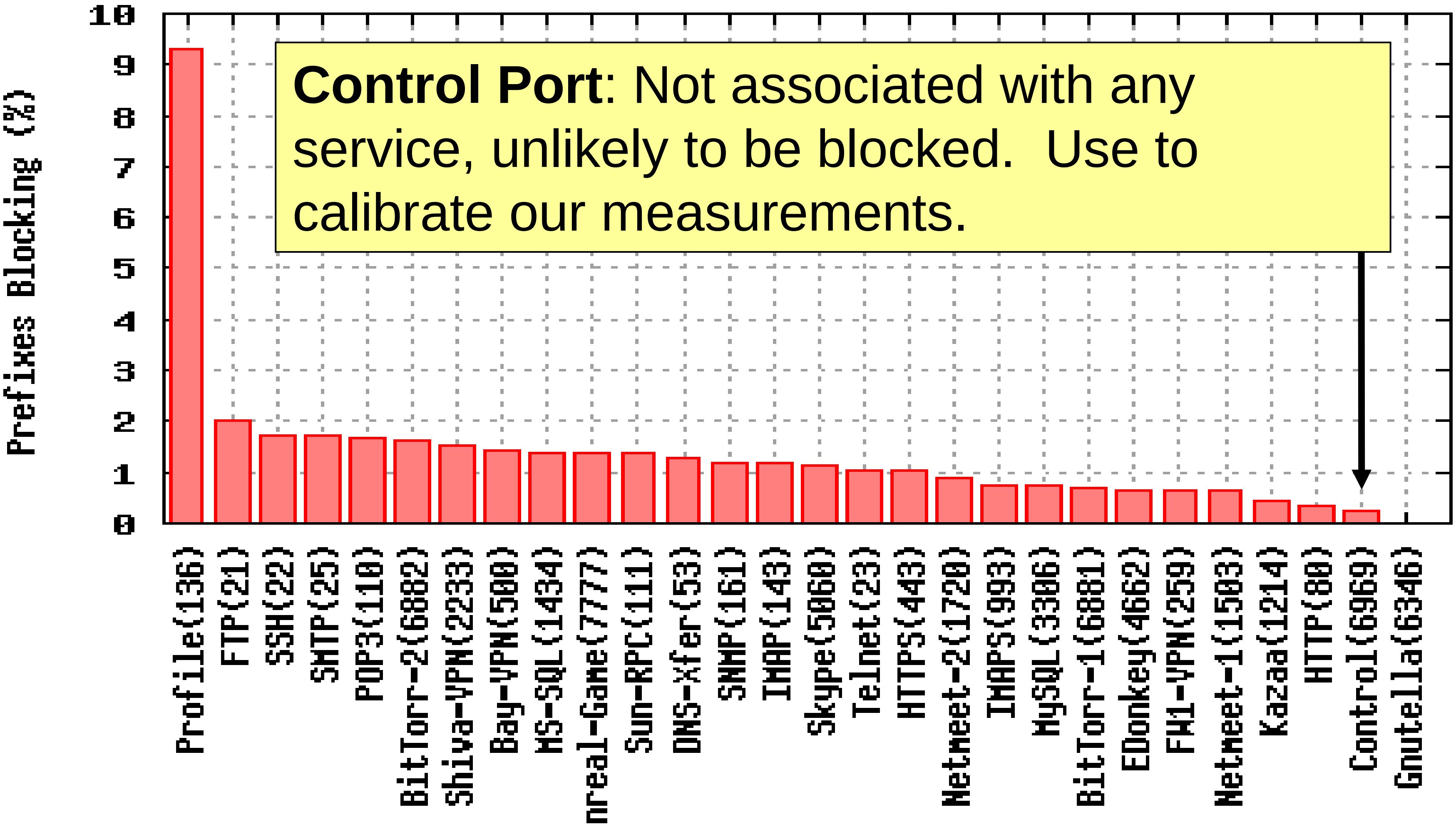
Can’t make any inferences over ~%55 of all prefixes

Classified >50% of tested ports for ~13k prefixes

Initial Results

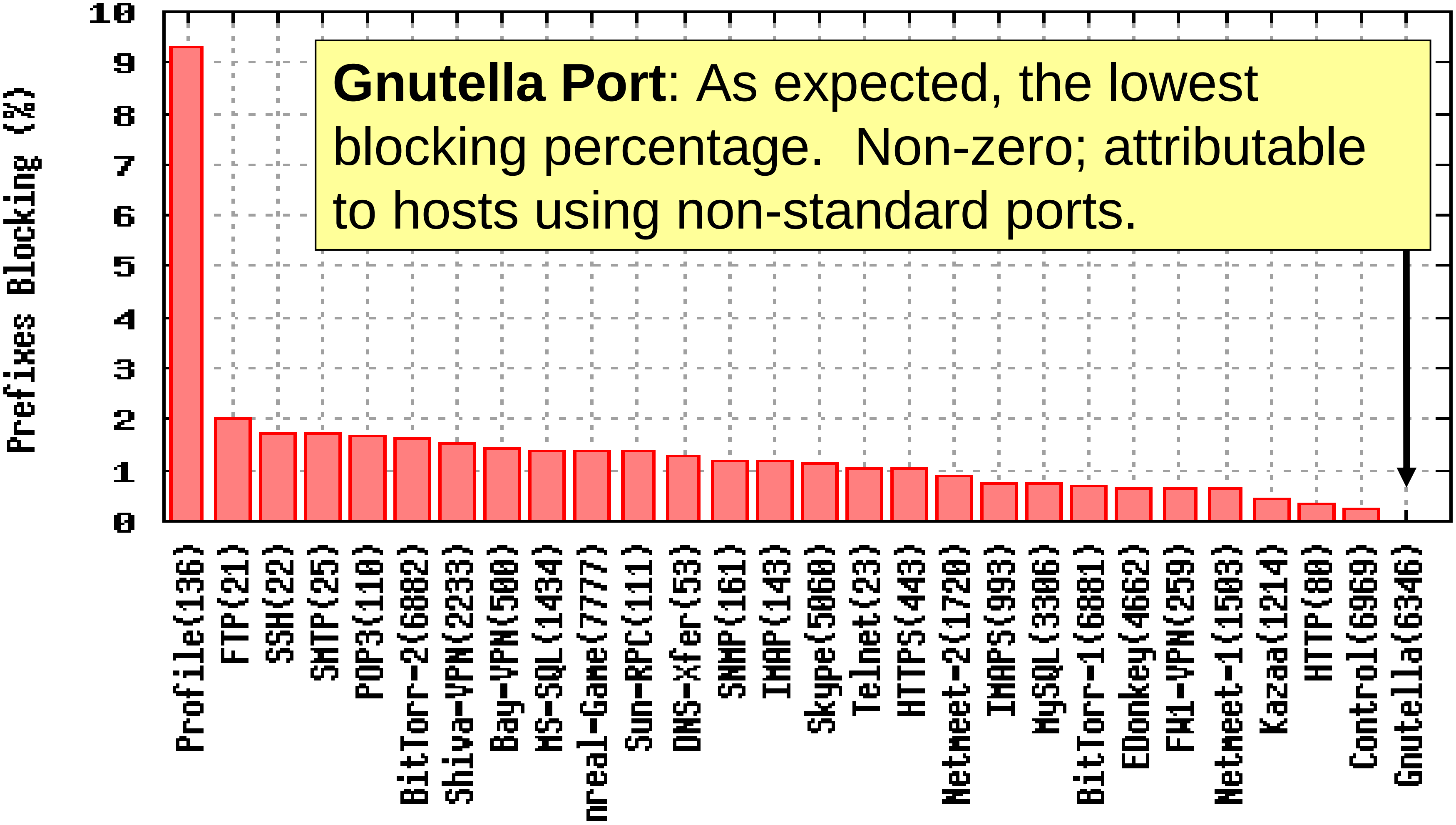
- Given our observations, which ports are more likely to be blocked relative to others?

Control Port



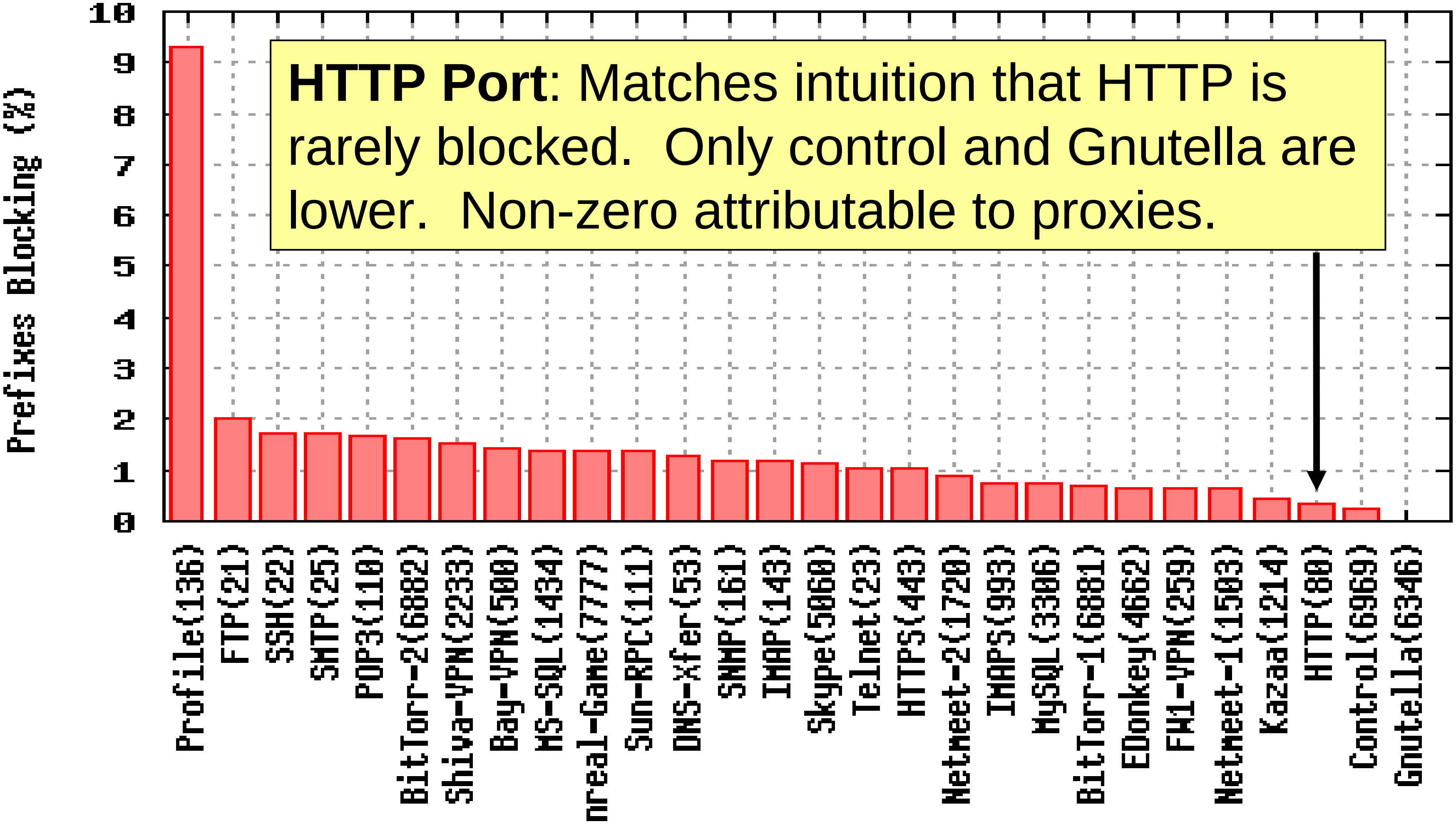
Control Port: Not associated with any service, unlikely to be blocked. Use to calibrate our measurements.

Gnutella Blocking



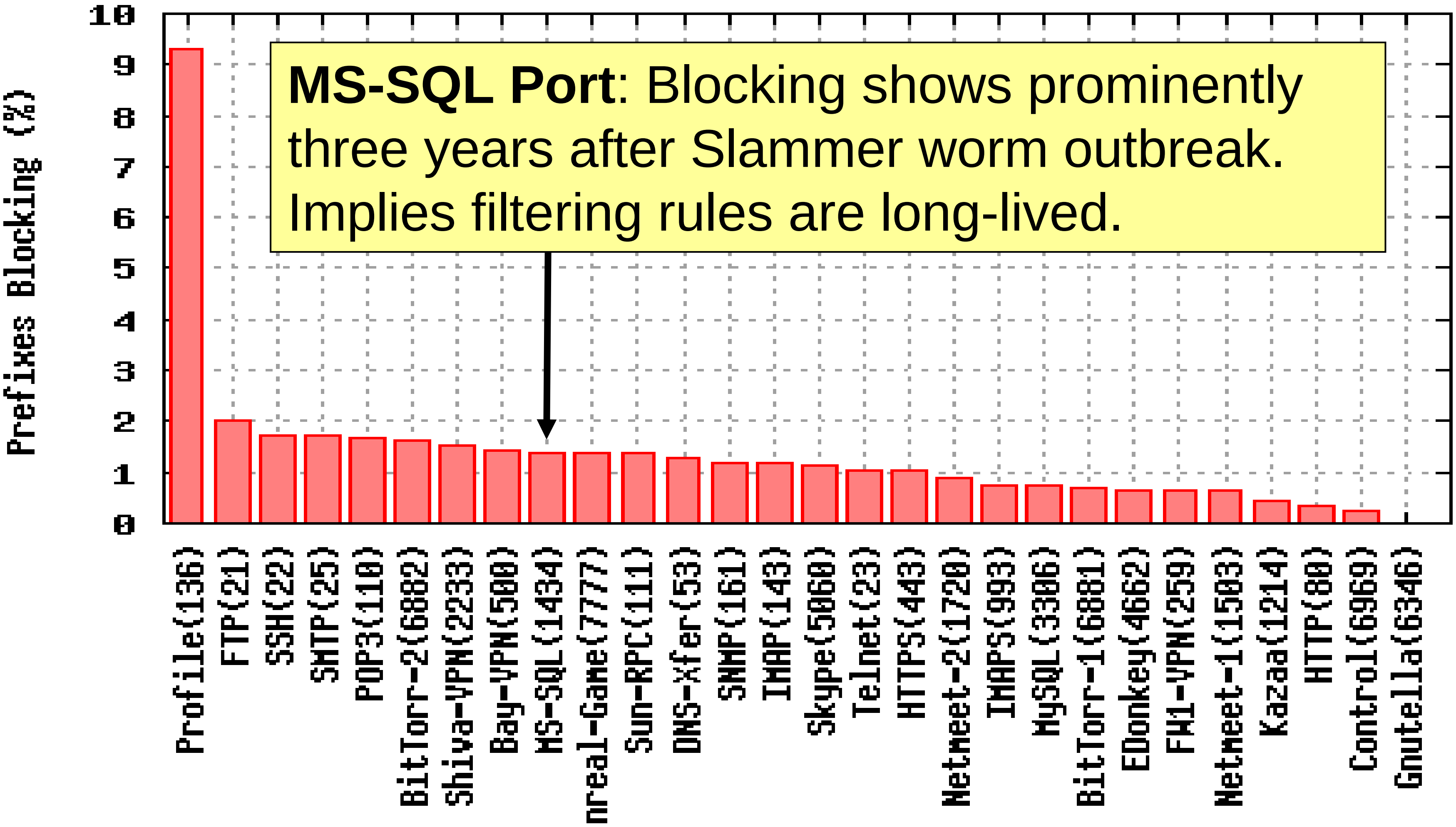
Gnutella Port: As expected, the lowest blocking percentage. Non-zero; attributable to hosts using non-standard ports.

HTTP Blocking



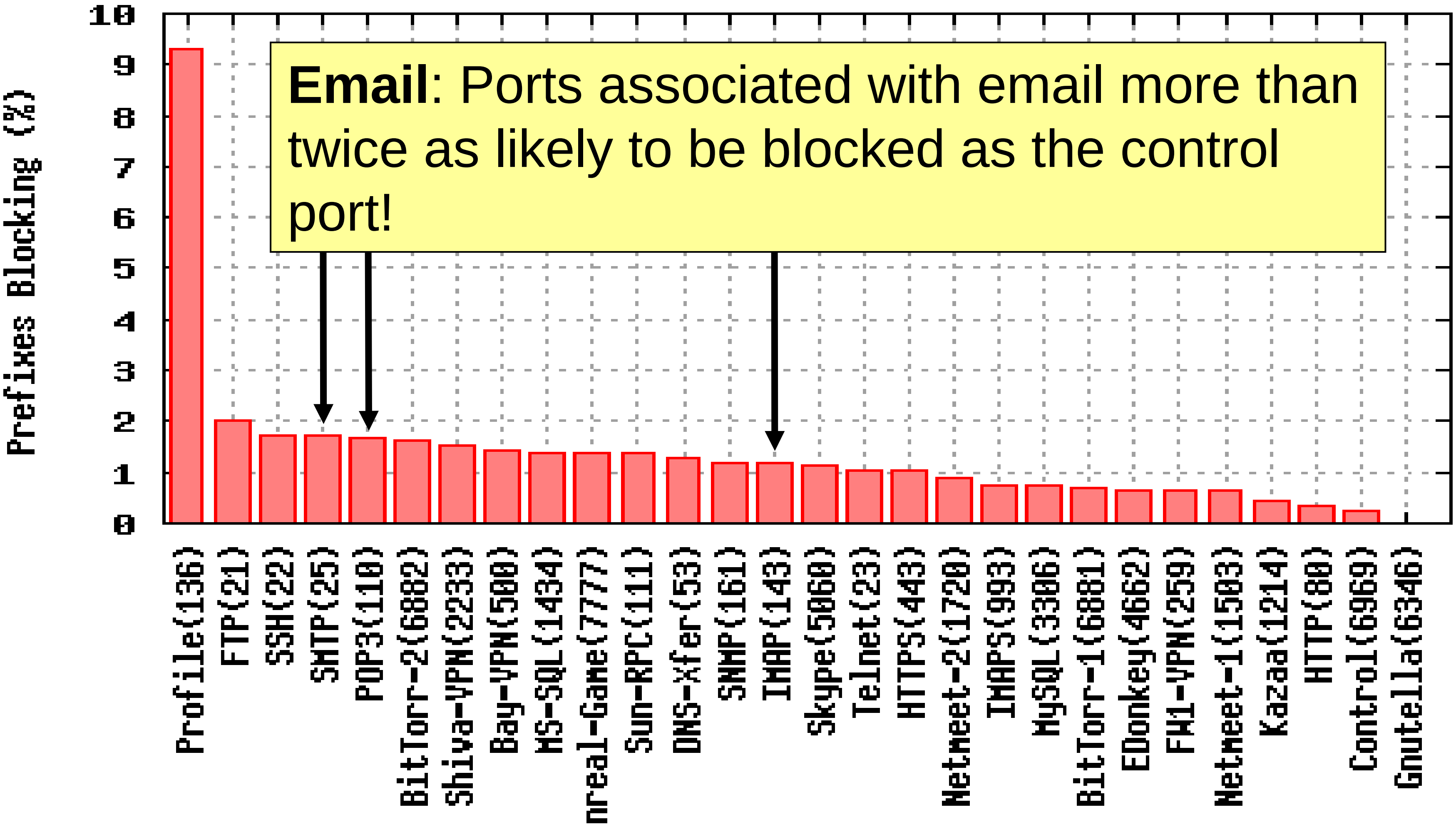
HTTP Port: Matches intuition that HTTP is rarely blocked. Only control and Gnutella are lower. Non-zero attributable to proxies.

MS-SQL Blocking



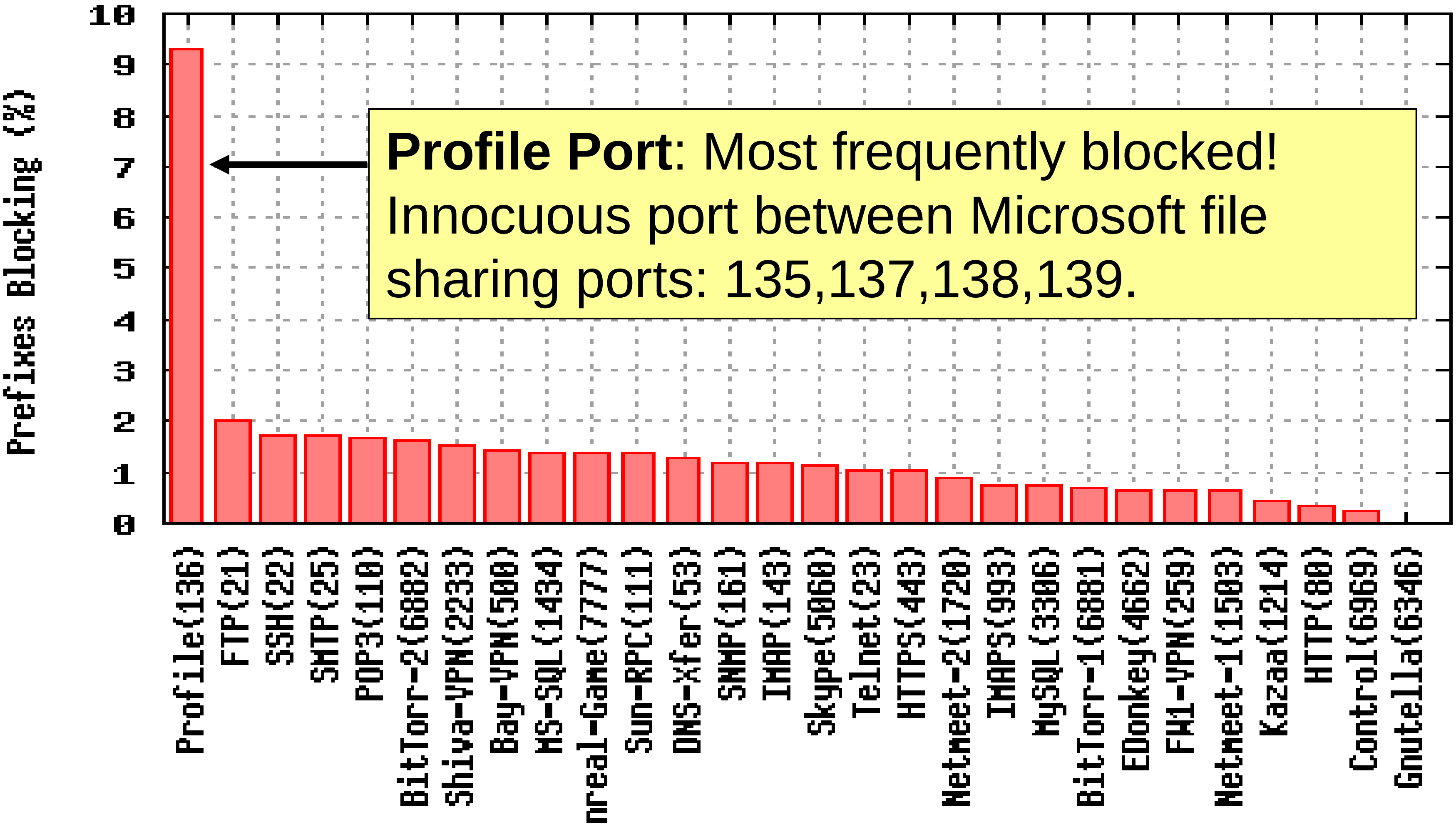
MS-SQL Port: Blocking shows prominently three years after Slammer worm outbreak. Implies filtering rules are long-lived.

Email Blocking



Email: Ports associated with email more than twice as likely to be blocked as the control port!

Collateral Damage



Profile Port: Most frequently blocked! Innocuous port between Microsoft file sharing ports: 135,137,138,139.

Future Analysis

- Determine relationship between blocking and type of prefix (business, .edu, ISP, etc)
- Determine geographical distribution of blocking
- Use AS topology to make inferences on where filtering is employed
- Evolution of blocking over time

Future Work

- Continue to collect measurements, increase our degree of confidence
- TCP Traceroutes:
  - Port-specific traceroutes to determine ingress filtering properties
  - Traceroutes allow us to determine where blocking occurs, filtering asymmetry, etc.
- Second methodology in progress employing HTTP using techniques outlined in this work

Research Summary

- Novel use of P2P overlay for measurement
- First measurements of Internet port blocking
- Initial results suggest promising avenue for systematic large-scale measurement

Thanks! Questions?