A Rendezvous-based Paradigm for Analysis of Solicited and Unsolicited Traffic

DUST 2012
May 15, 2012

THE UNIVERSITY of WISCONSIN MADISON

David Plonka
&
Paul Barford
{plonka,pb}@cs.wisc.edu

Outline

- Rendezvous-based Traffic Analysis
  - What is it? Why use it?
  - a DNS rendezvous case study involving office and residential “solicited” traffic
- Darkspace Rendezvous Mechanisms
  - unsolicited and passively solicited traffic
- TreeTop
  - a DNS rendezvous-based analysis tool

    [Plonka & Barford, IMC 2009, SATIN 2011, work in progress]
  - flow export with rendezvous annotations
  - IPv6 performance by service names

Rendezvous-based Traffic Analysis?

- Traffic classification and analysis has focussed on target traffic features (IP headers, DPI, etc.)
- However, Internet hosts learn IP addresses by some *rendezvous* mechanism, e.g.:
  - By static configuration (IP addrs in config files)
  - The Doman Name System (DNS)
  - Application-specific mechanisms (URLs, p2p)
- Inform traffic analysis by considering,
  "How does this host know this IP address?"
  rather than simply,
  "With what IP address did this host interact?"

Why Focus on Rendezvous?

rendezvous, meaning hosts and services "present themselves"

- For standard protocols, rendezvous information is not private and is of low-volume
  - Separate and separable from private payloads
  - Can be monitored in situations where target traffic is high-volume, sampled, or encrypted
- Rendezvous info can indicate when other analysis or classification techniques are effective and not
  - e.g., port-based classification

    [Kim, et al., 2008] [Plonka & Barford, 2011]

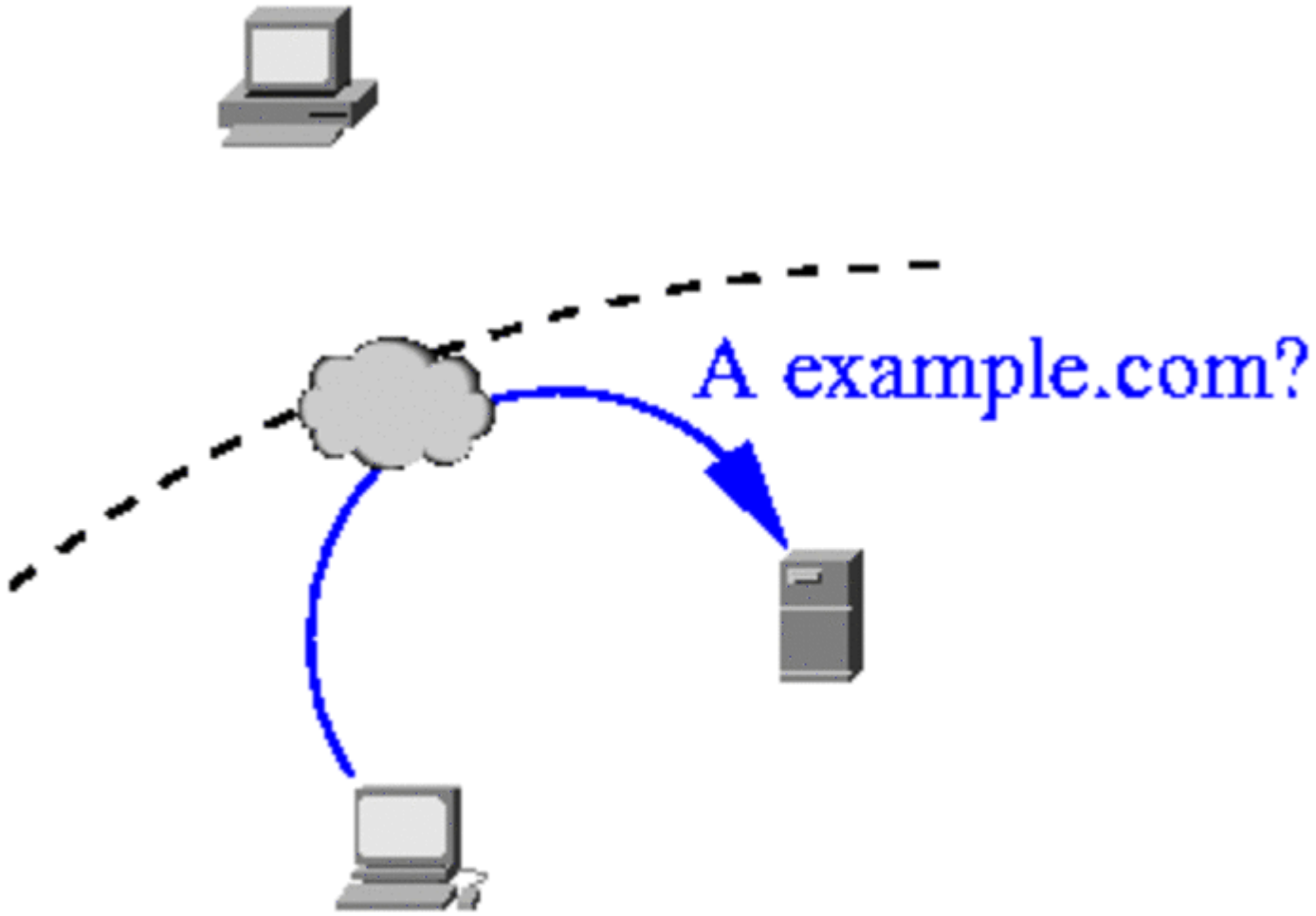
Rendezvous-based Traffic Classification

rendezvous, meaning “present yourselves”

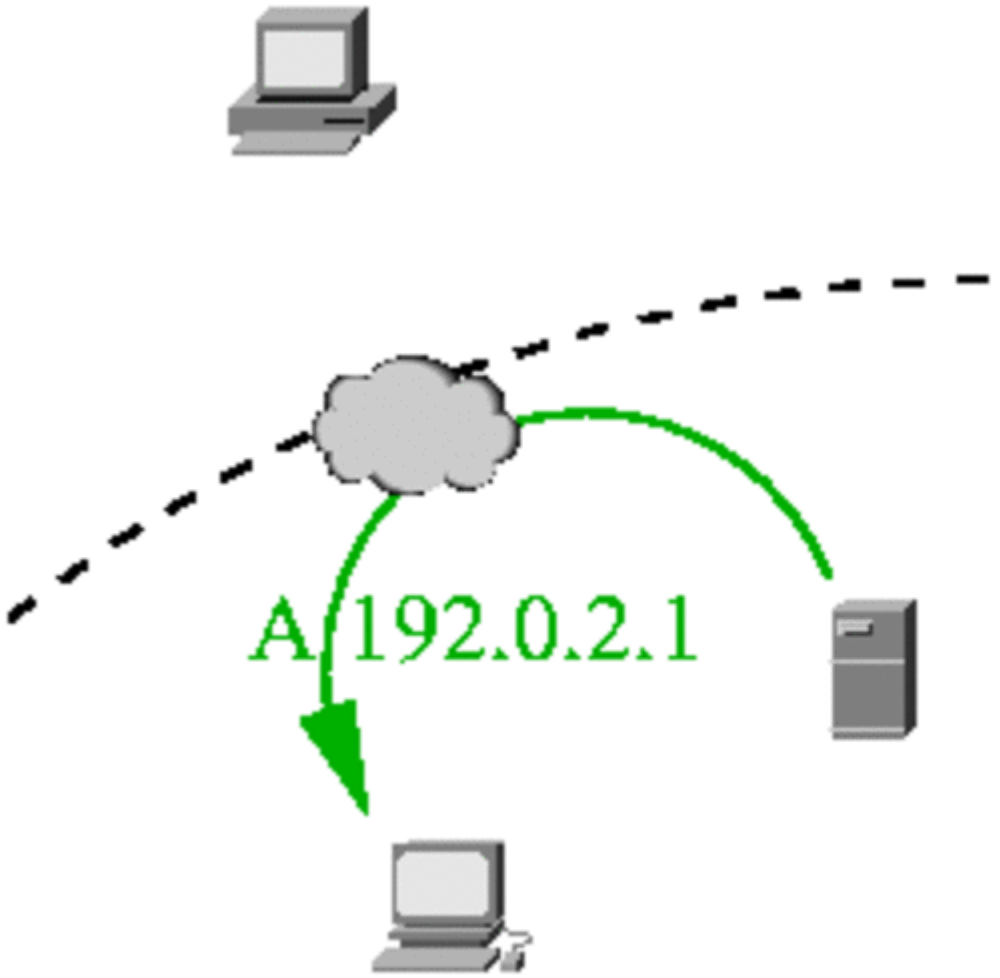
- ***Hypothesis*:** We can inform and improve traffic classification by considering,
  “How does this host know that peer IP address?”

- ***DNS***: Internet hosts regularly use the DNS to find remote IP addresses of the hosts with which they might interact.
  - It is an ***easily separable*** standard, “clear text” protocol.

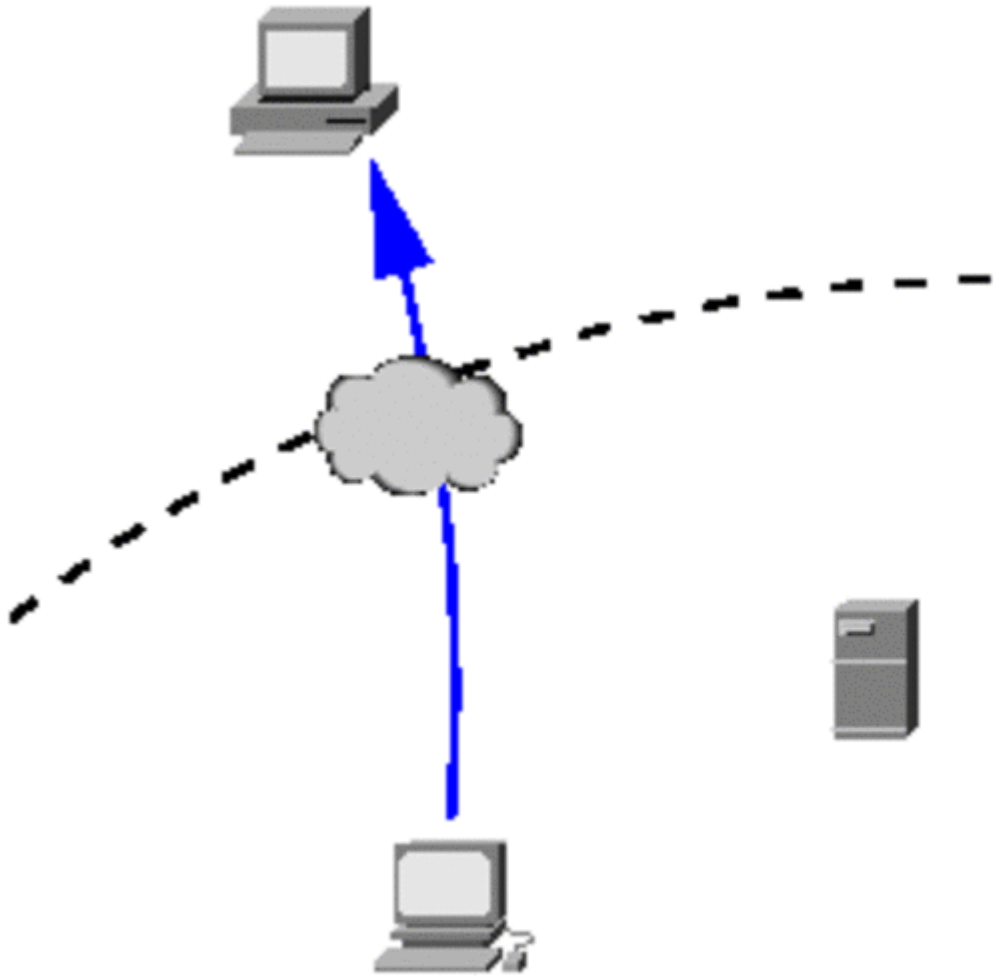
DNS Rendezvous: (1) Query



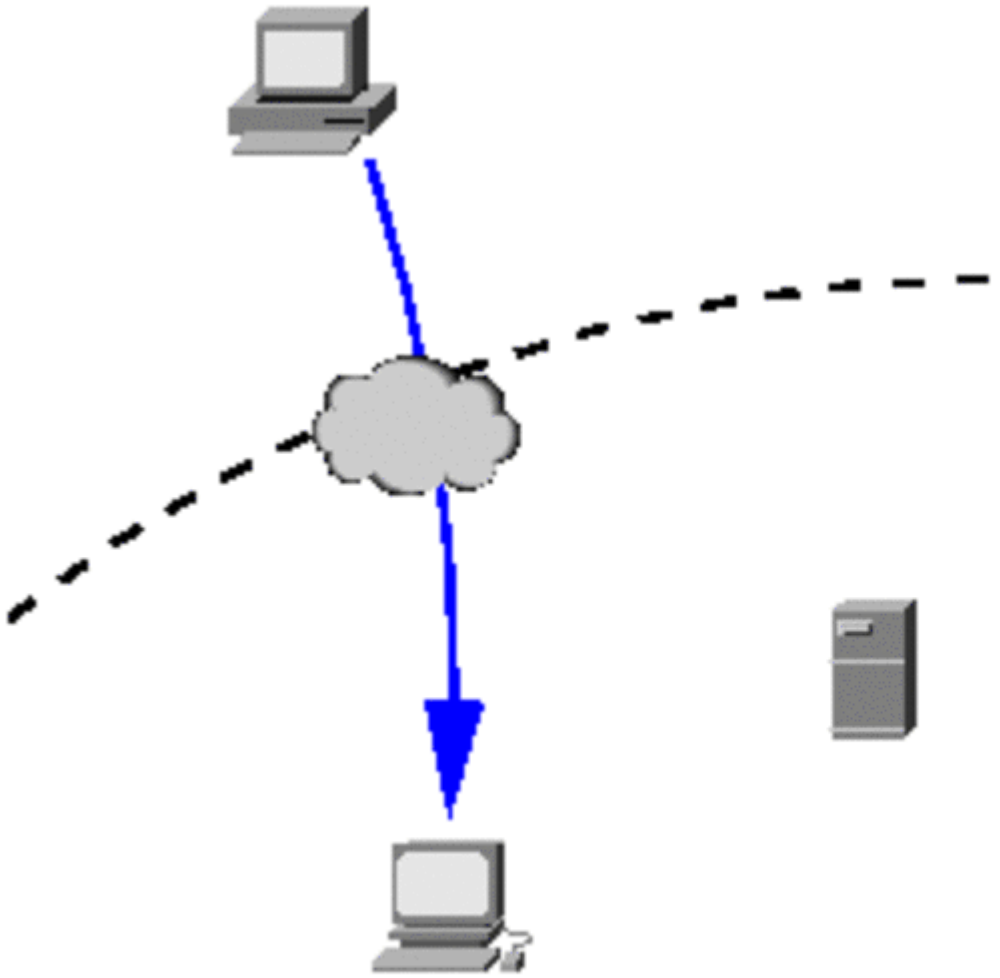
DNS Rendezvous: (2) Response



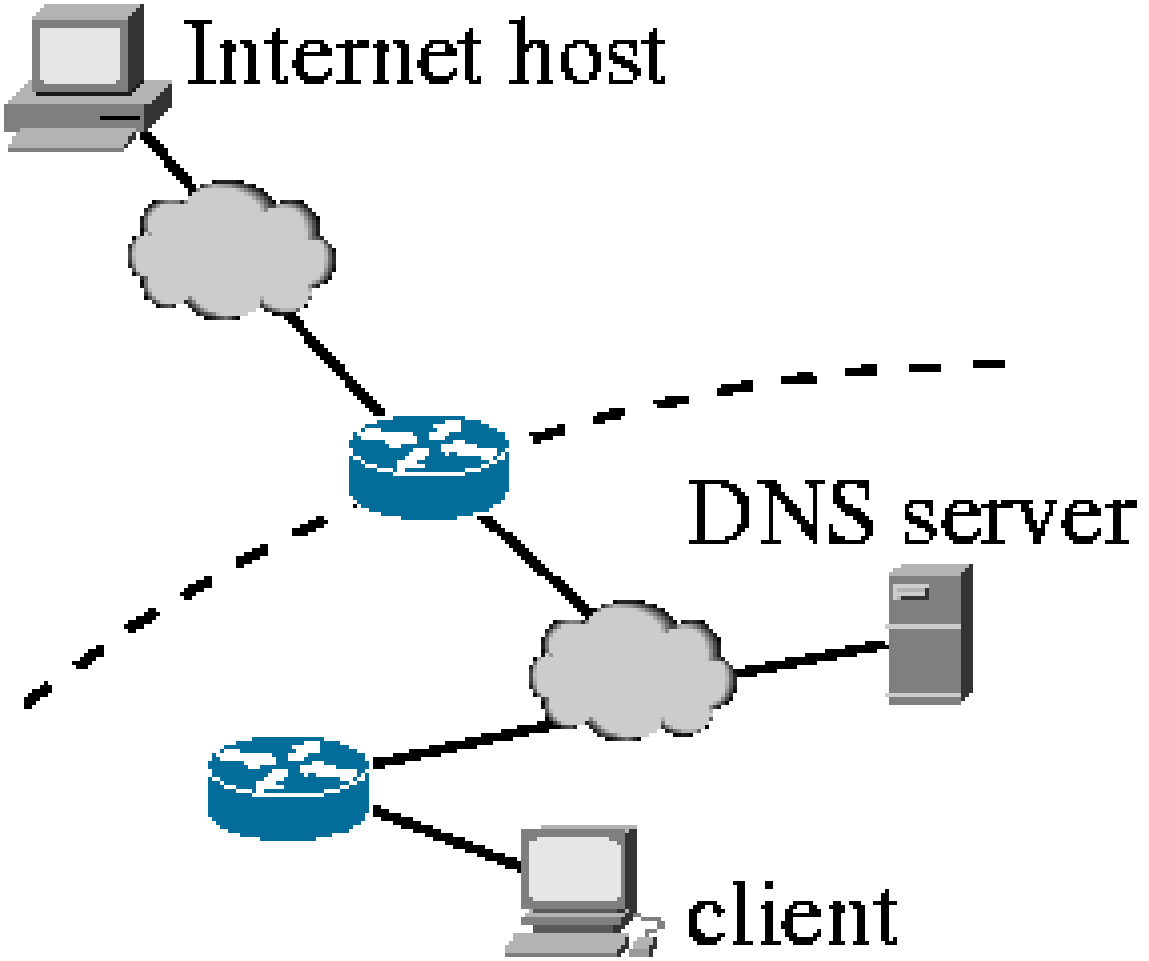
DNS Rendezvous: (3) Outbound



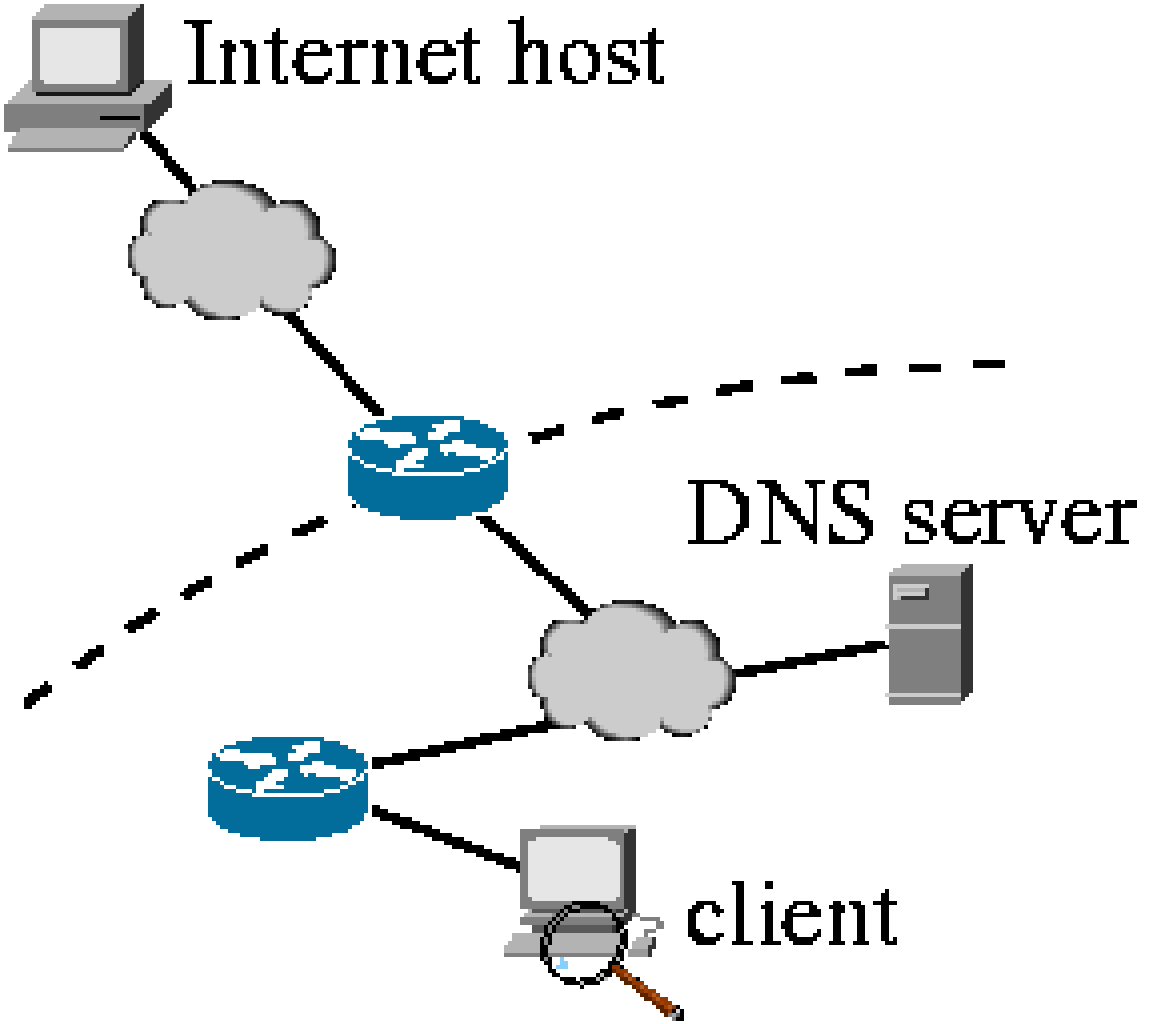
DNS Rendezvous: (4) Inbound



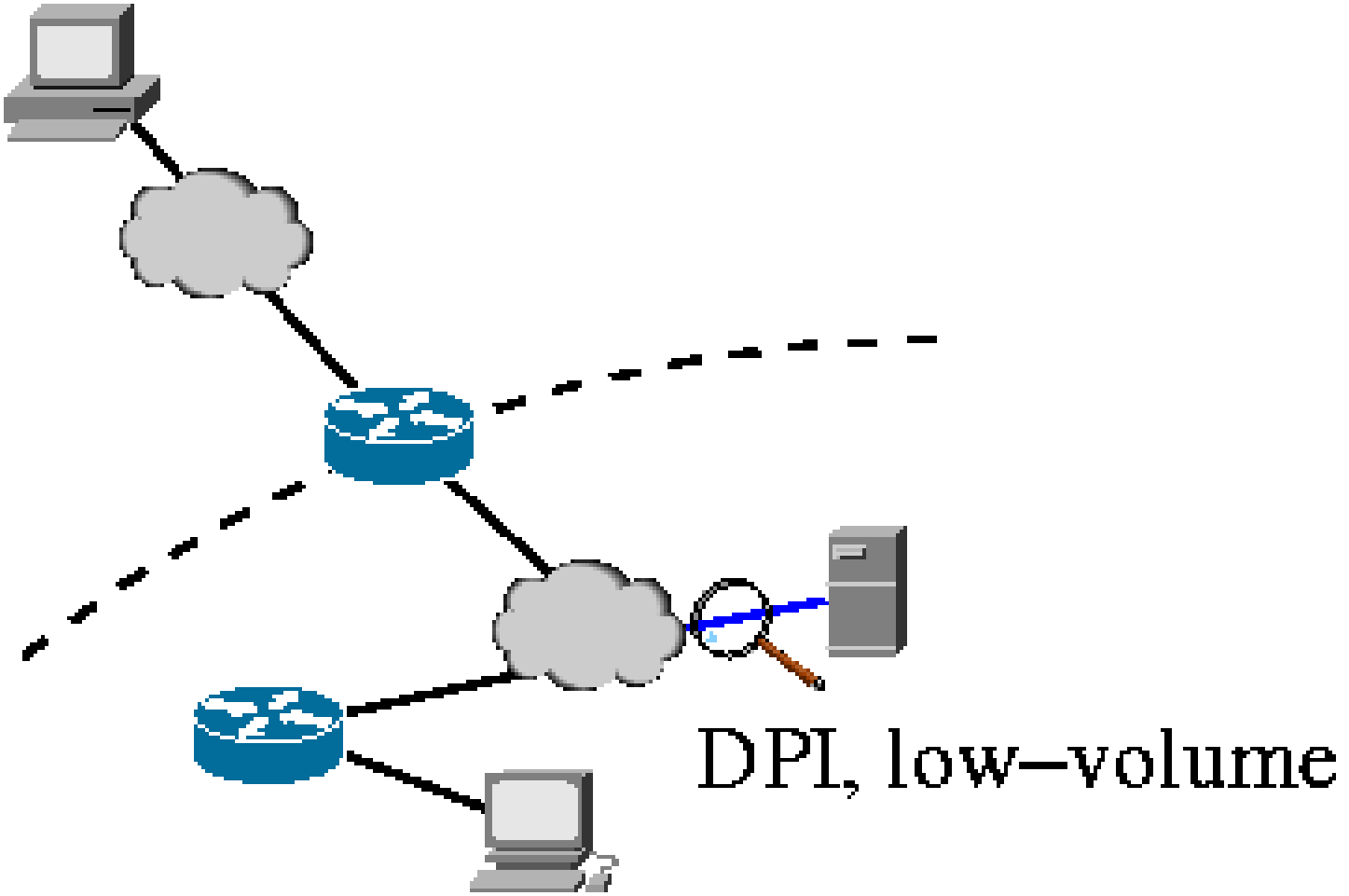
Traffic Observation Points



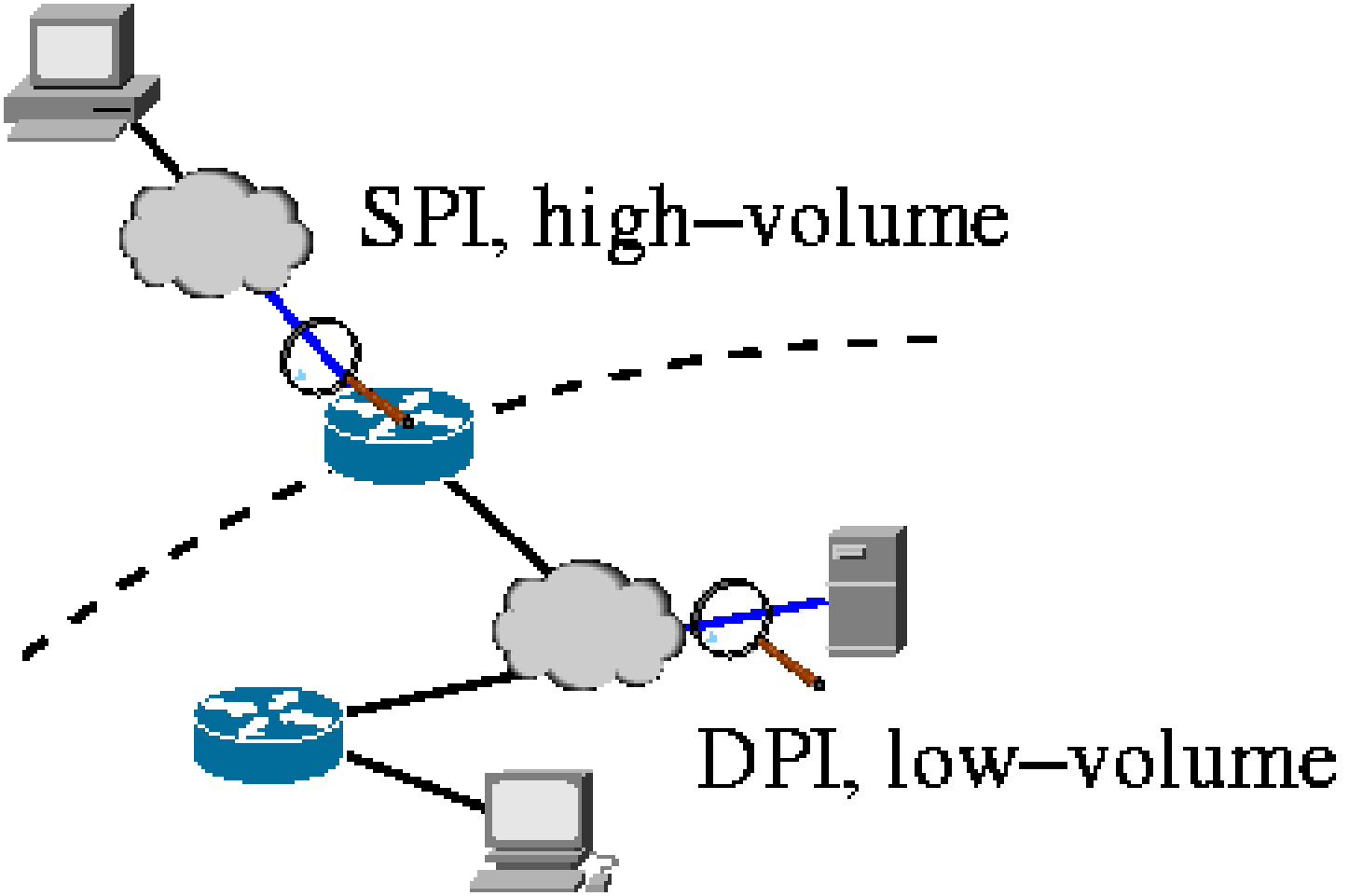
Traffic Observation Points



Traffic Observation Points

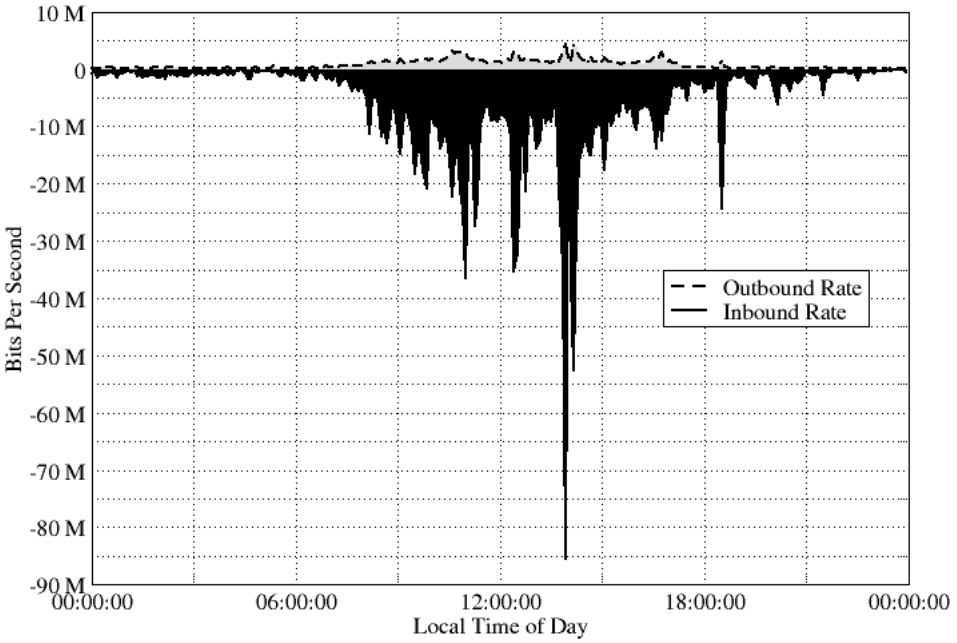


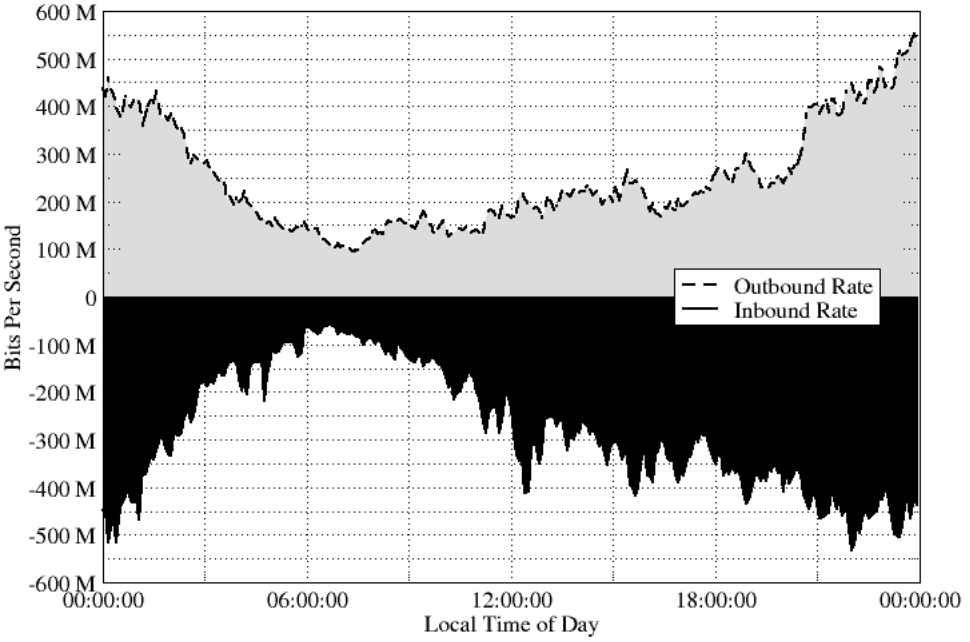
Traffic Observation Points



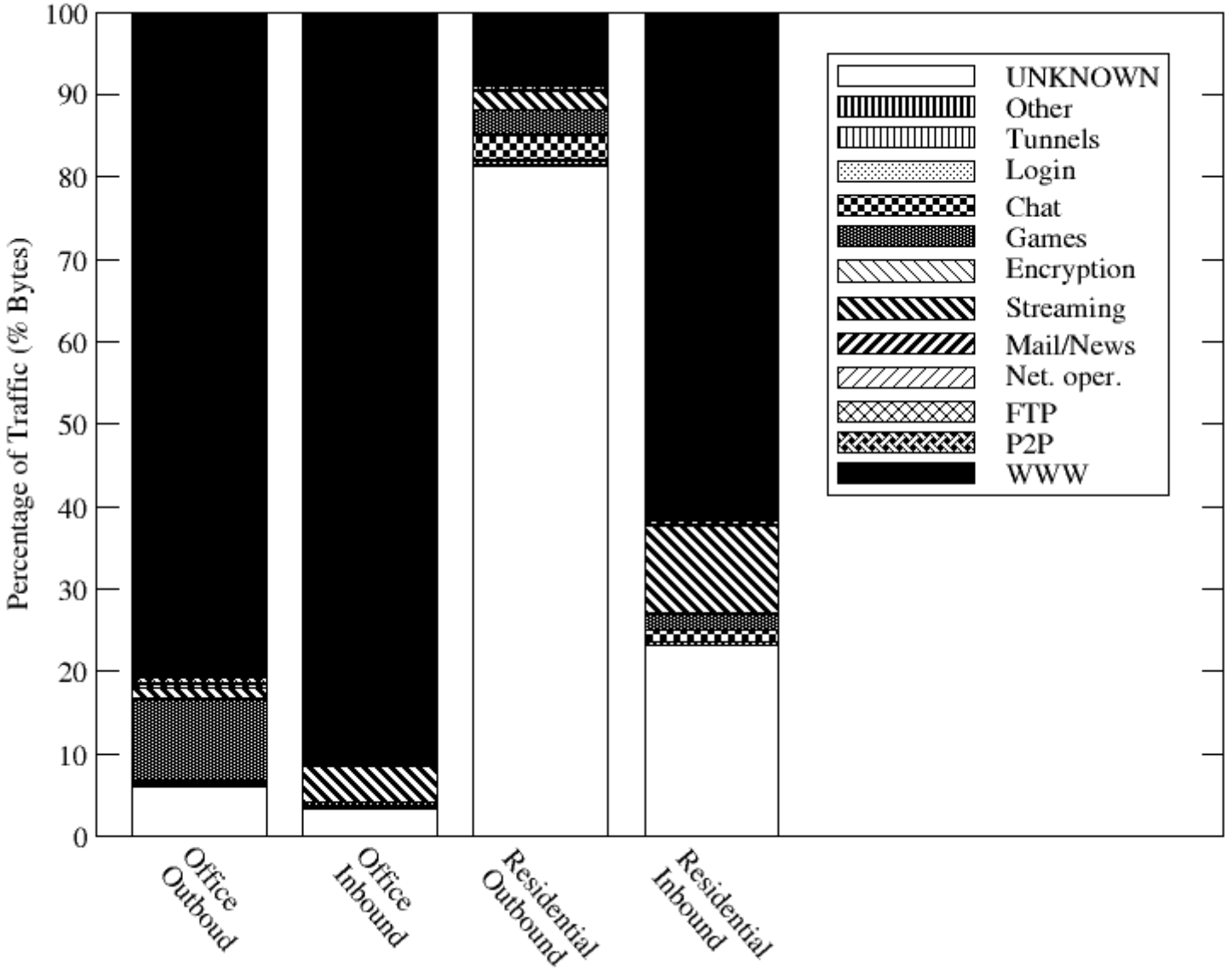
Characteristics of Data Sets

Data Set	Date	Day	Duration	Clients	Unique NOERROR FQDNs	DNS Reply Pkts	Average DNS Reply Utilization	Average Wide-Area Outbound / Inbound Utilization
Office	2009-04-17	Fri	24h	614	19.4 K	560 K	12.2 Kbps	753 Kbps / 5.66 Mbps
Residential	2009-04-17	Fri	24h	9,819 (5,344)	(143 K)	15.7 M	360 Kbps	244 Mbps / 276 Mbps

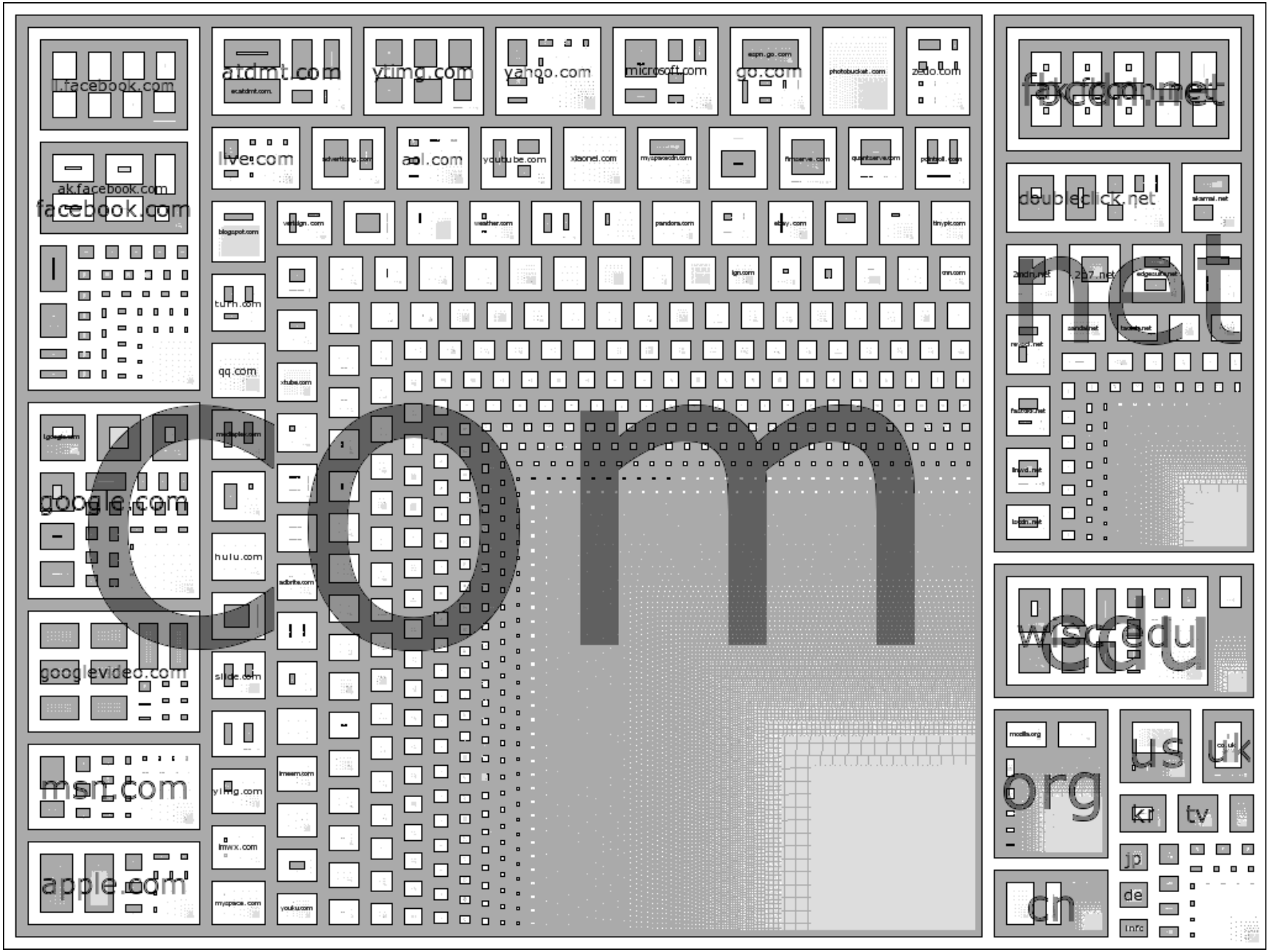




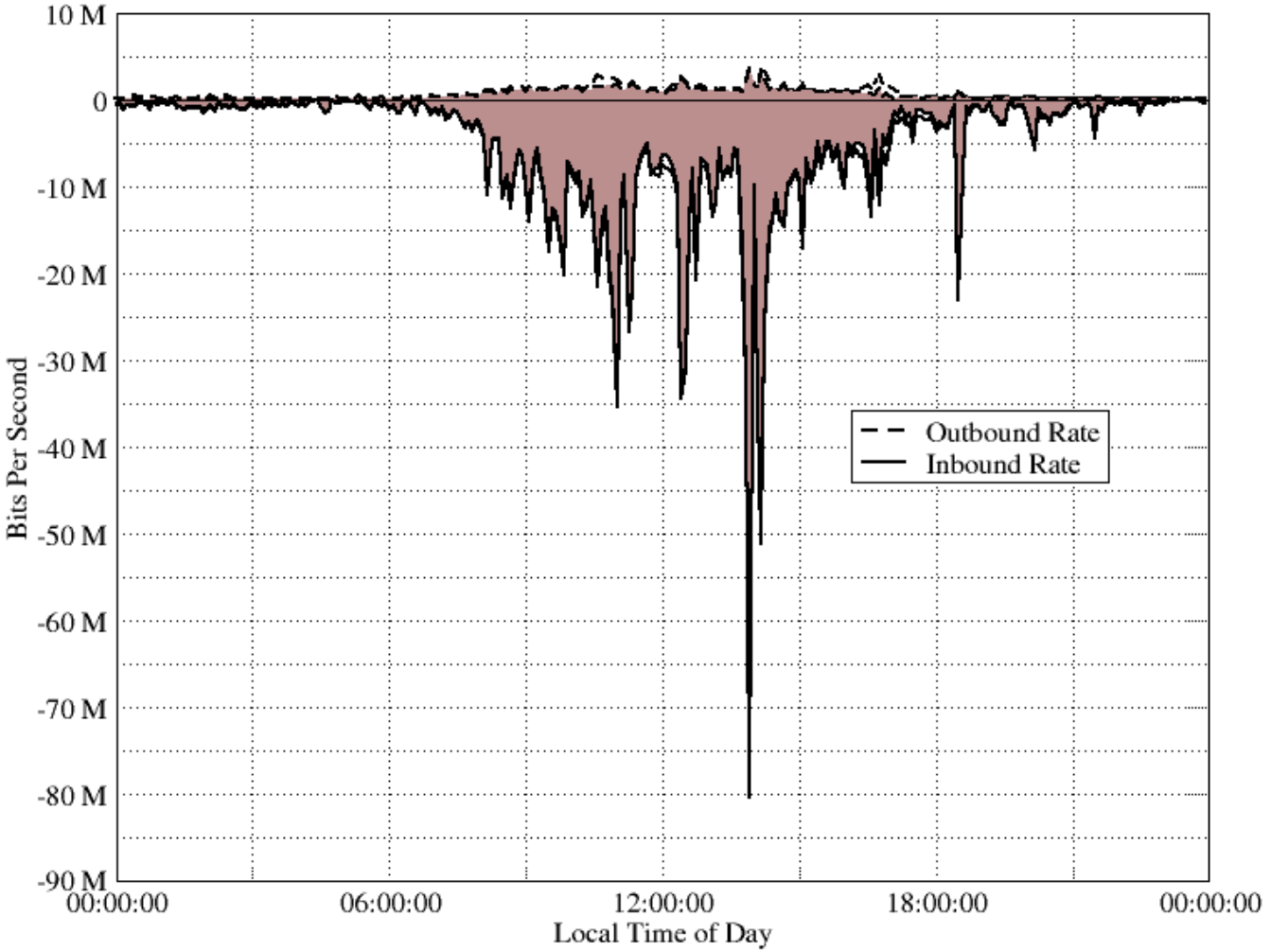
Target Traffic Classification: Port-based method



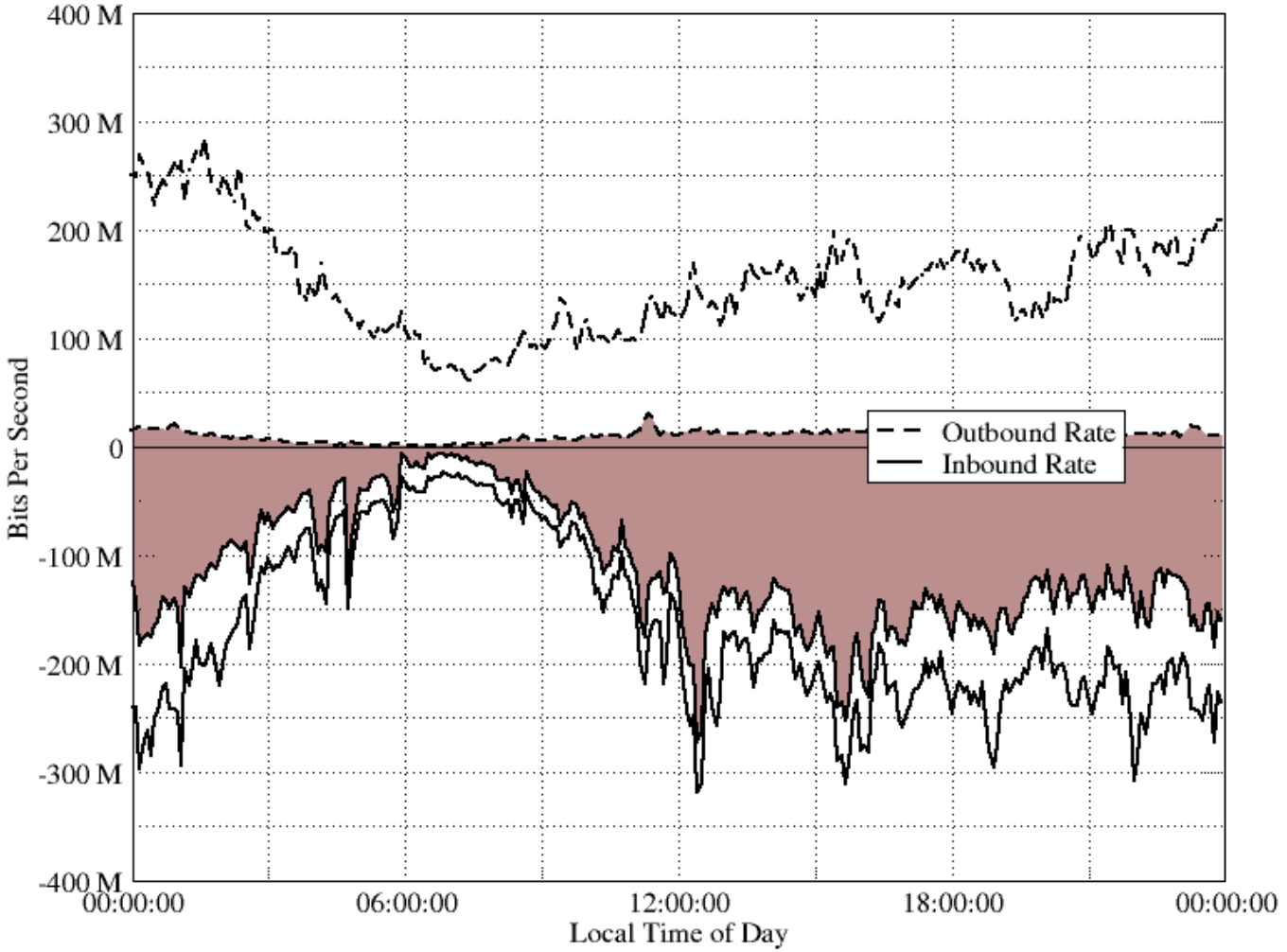
Residential: Domain Popularity



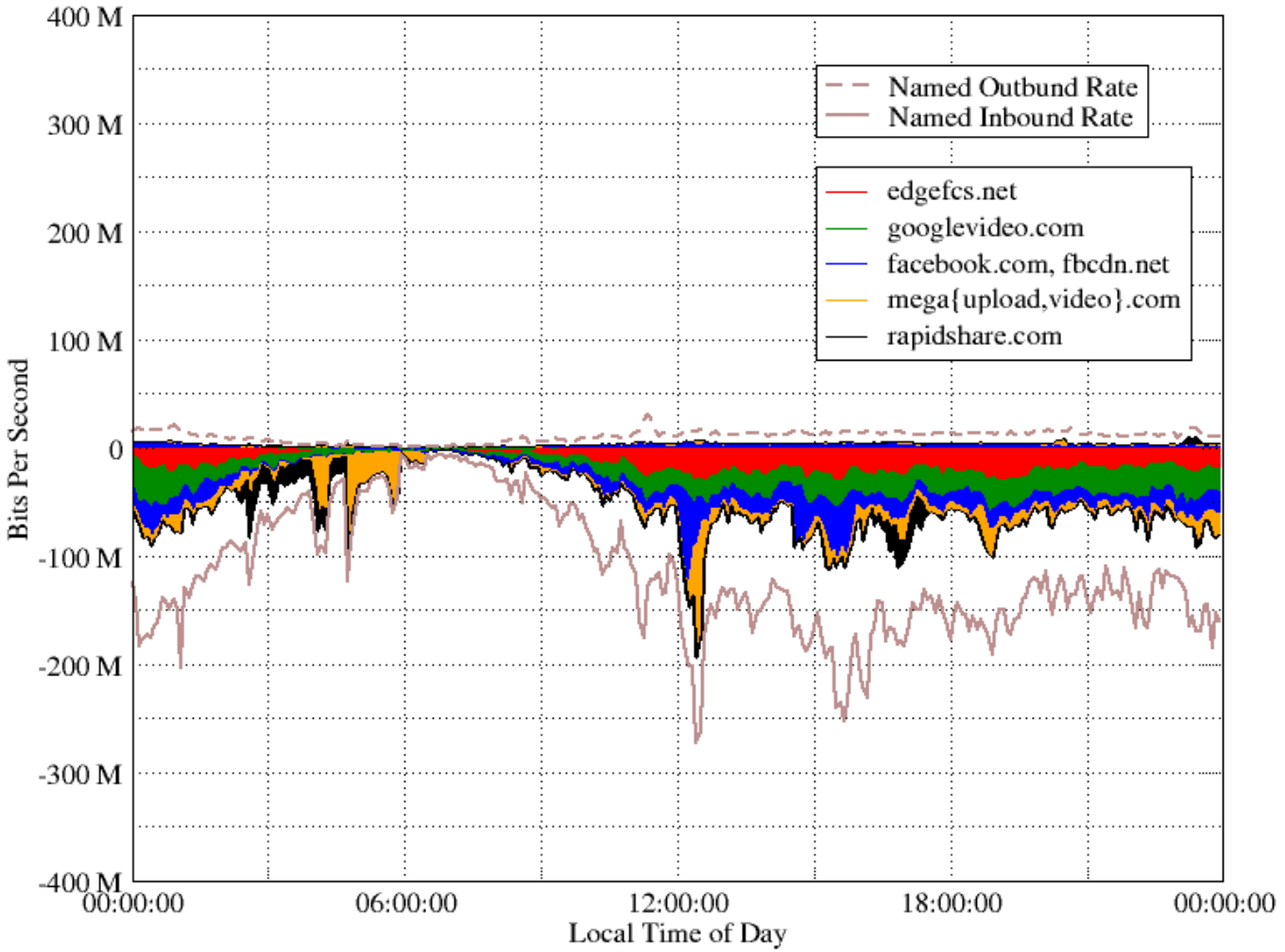
Office Target Traffic Classification: “named” and “unnamed”



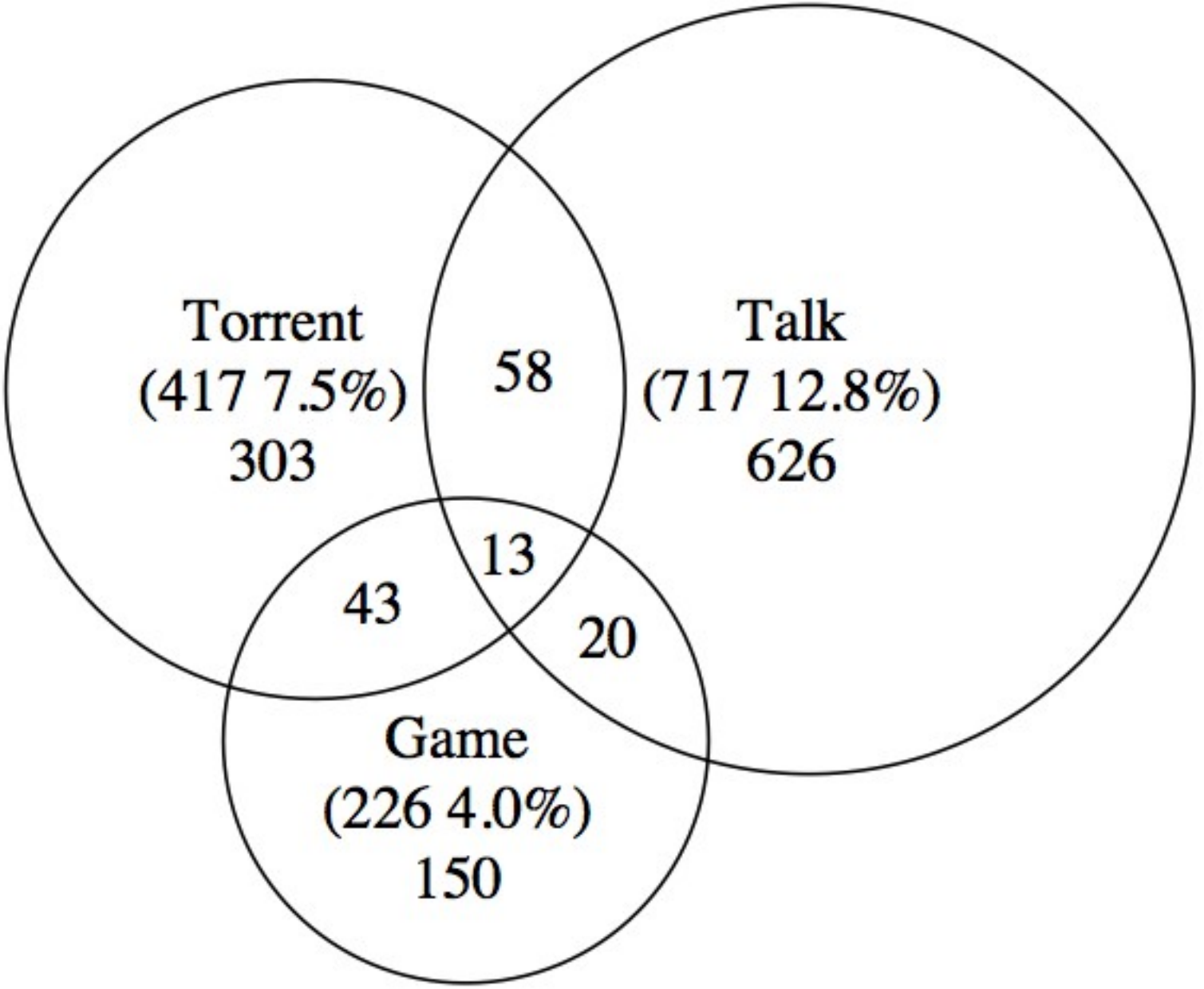
Residential Target Traffic Classification: “named” and “unnamed”



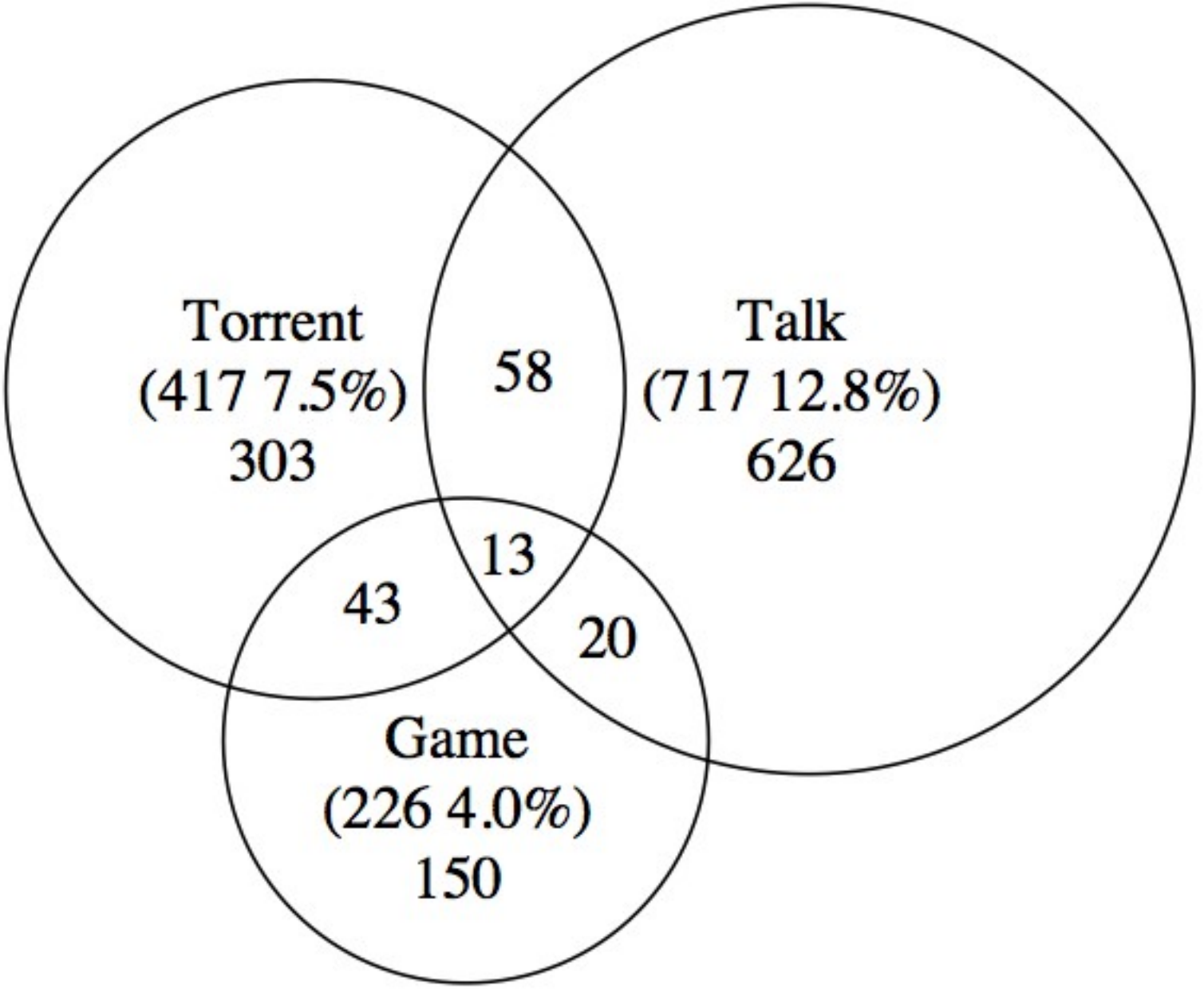
Residential Target Traffic Classification: “named” by popular domains



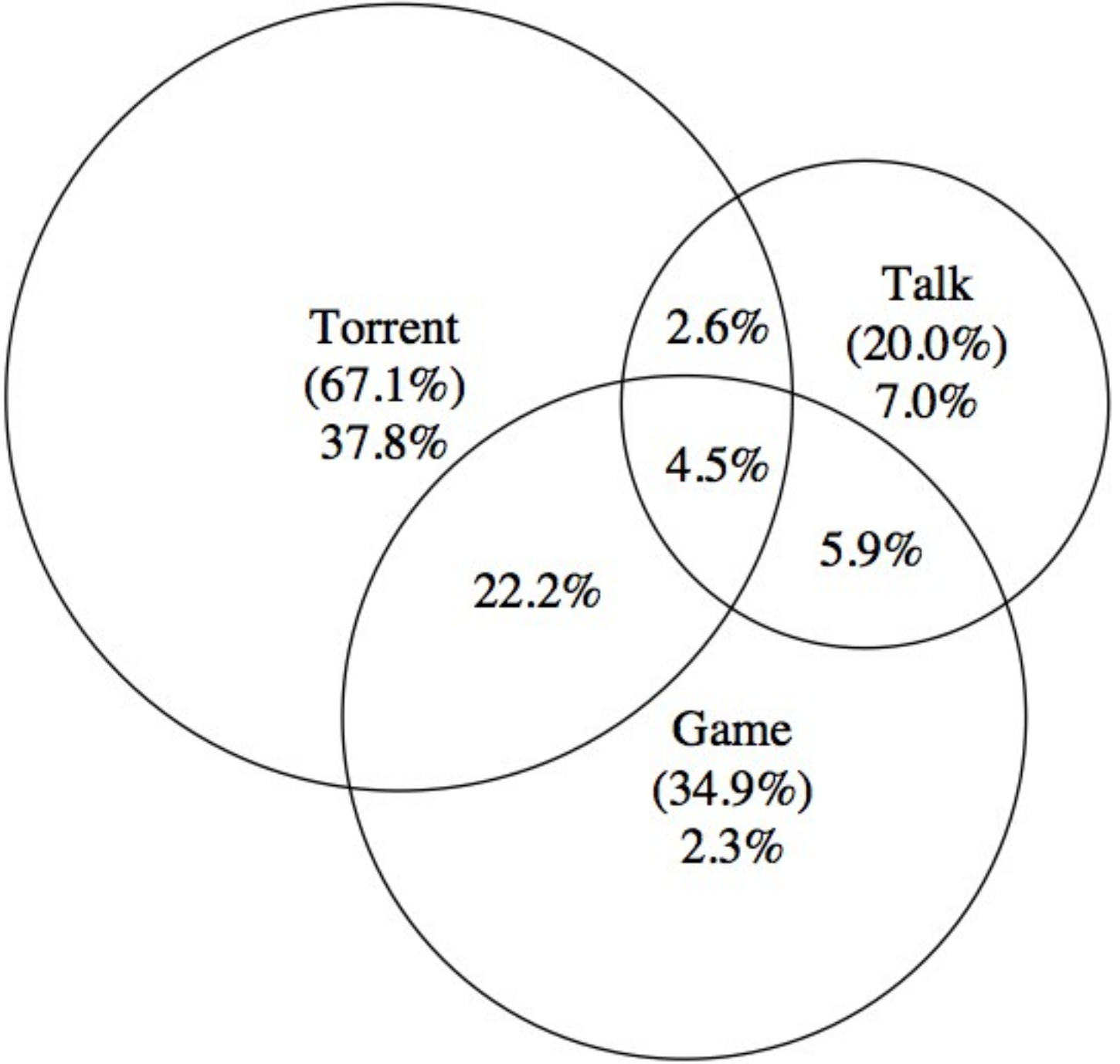
Host Profiling and Reputation based on Rendezvous Information



Residential Hosts Classification by P2P Host Profile (1 day)



“unnamed” Target Traffic by P2P Profile



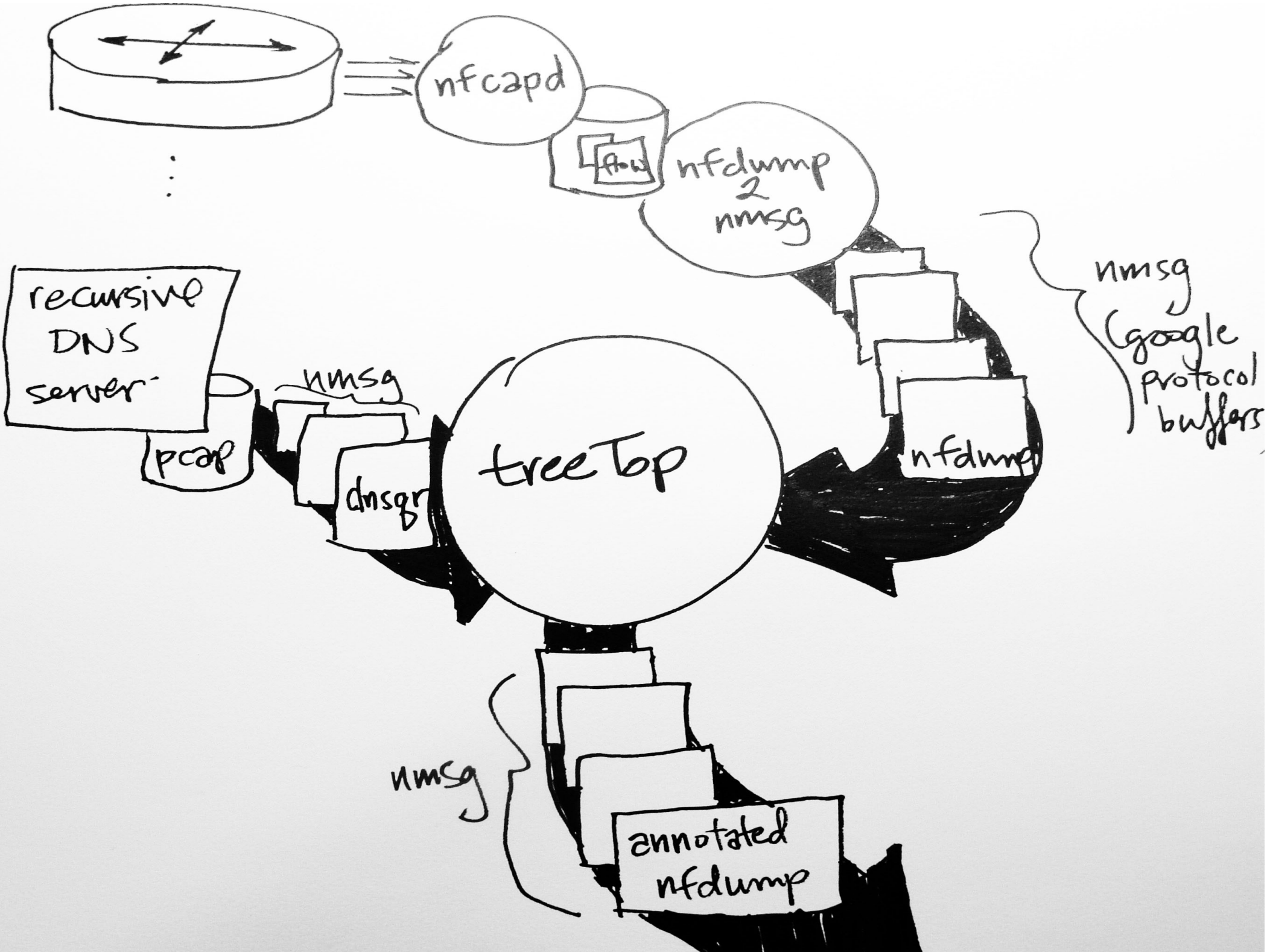
Results Summary: Traffic Classified (% bytes)

Data Set	Port-known	DNS-named and Port-known	DNS-named	DNS-named and DNS-Profiled
Office Out	93.9%	80.5%	81.8%	91.9%
Office In	96.6%	91.8%	93.2%	95.4%
Residential Out	18.6%	6.2%	6.7%	83.5%
Residential In	76.9%	58.3%	67.9%	88.2%

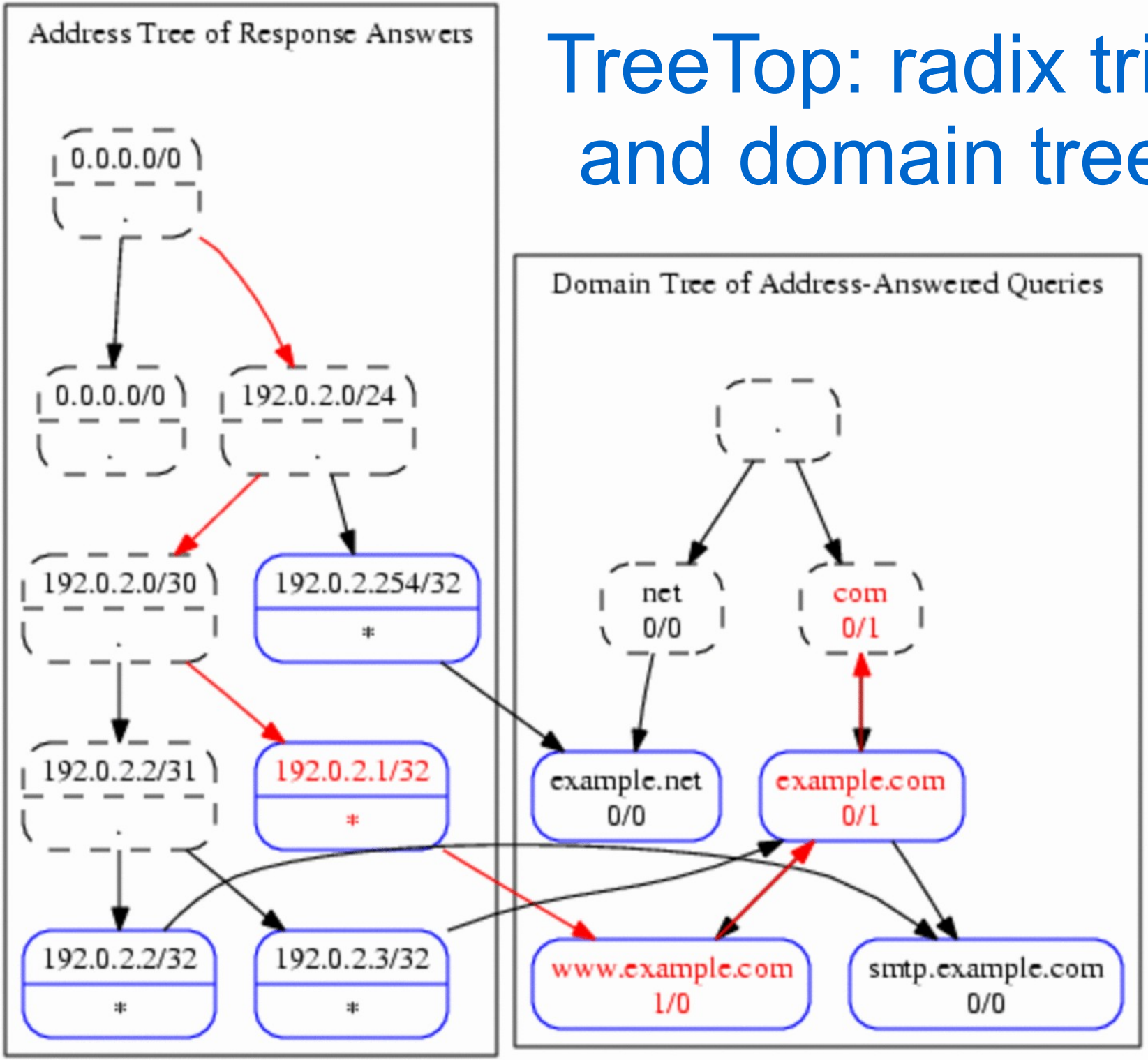
Rendezvous in Darkspace/Grayspace?

- ***Darkspace and Unsolicited***: a host uses *some* technique to choose remote/peer IP addresses
  - **Algorithm**, e.g., scanning a contiguous set of IP addresses in series, choosing IP addresses at random
  - **Bug**, e.g. D-link products connect to 45.52.84.48, the 7-bit string “-4T0”, believed to be a stray value left in an uninitialized 32-bit integer meant to store an SMTP server's IP address [Yegneswaran, Barford, Plonka, 2004]
  - **Misconfiguration** or stale configuration, e.g., SNMP traps to various 45/8 addresses from Interop events
  - IP prefixes become **encumbered** by legacy roles

TreeTop:
Rendezvous-annotated Flow Export


nfcapd
flow
nfdump
2
nmsg
recursive
DNS
server
nmsg
pcap
dnsqr
treeTop
nfdump
nmsg
Google
protocol
buffers
nmsg
annotated
nfdump

TreeTop: radix tries and domain trees



[3 private slides redacted]

Discussion

- In what circumstances can we **trust** rendezvous information for traffic classification or host profiling/reputation?
- Tap rendezvous methods other than the DNS; e.g., application-specific methods (WWW, P2P); are they discoverable, **separable** and clear?
- Should we alter or invent rendezvous protocols to better inform classification and packet treatment?
- Is rendezvous a useful unifying analysis concept?

A Rendezvous-based Paradigm for Analysis of Solicited and Unsolicited Traffic

FIN

THE UNIVERSITY of WISCONSIN MADISON

David Plonka
&
Paul Barford
{plonka,pb}@cs.wisc.edu