Spatial-Temporal Characteristics of Internet Malicious Sources

Zesheng Chen

Florida International University

Chuanyi Ji

Georgia Institute of Technology

Paul Barford

University of Wisconsin-Madison

The 27th Conference on Computer Communications (INFOCOM)
April 14, 2008, Phoenix, AZ, USA

Malicious Sources

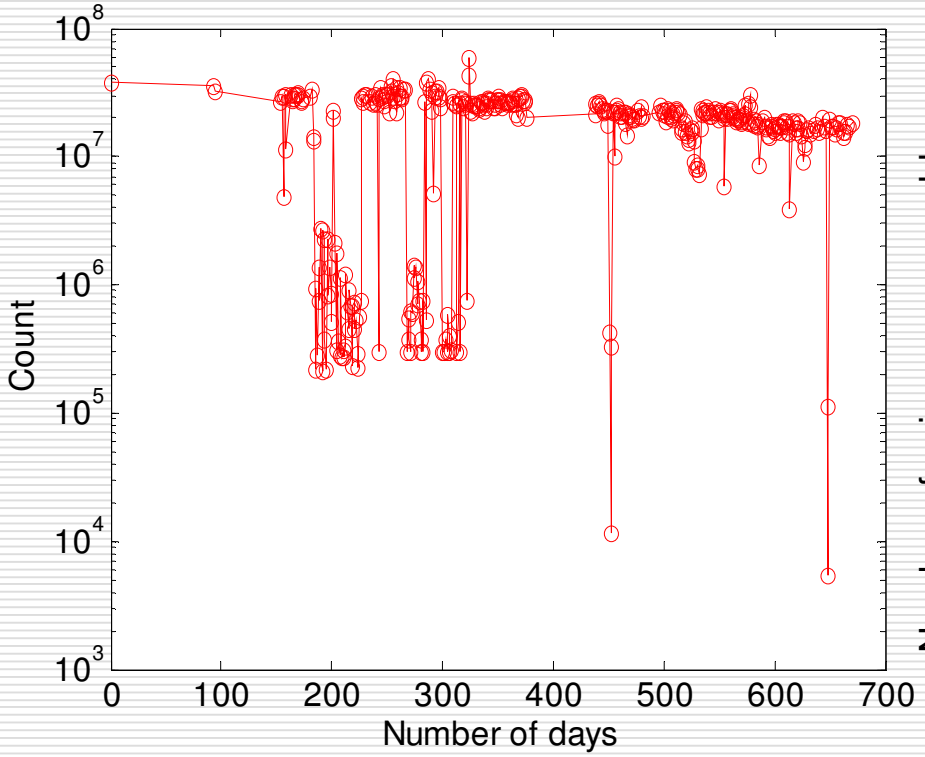
- IRC bots
  - Dos attacks
  - Spam emails
- Worm/virus victims
- Other intruders

Goal of this work: To show the spatial-temporal characteristics of these malicious sources

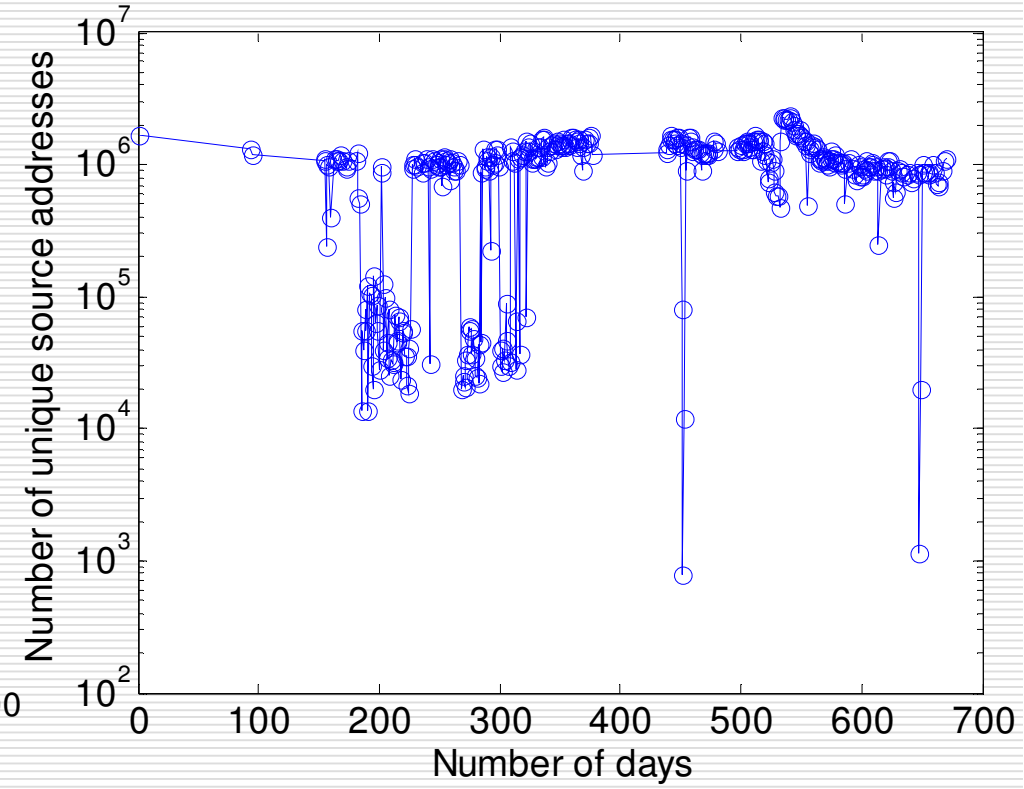
Data Set

- DShield
  - http://www.dshield.org **(part of SANS Institute)**
  - Collects firewall and intrusion detection system (IDS) logs
- Approximately 2,000 organizations
- From Nov. 10, 2004 to Sept. 10, 2006
- ~62Gbytes (after compression)
- 7,535,357,813 records
- 160,590,790 (~2^{27}) distinct source addresses
  - 1 out of 27 host IP addresses is potentially a malicious source

Records and Unique Sources



Records

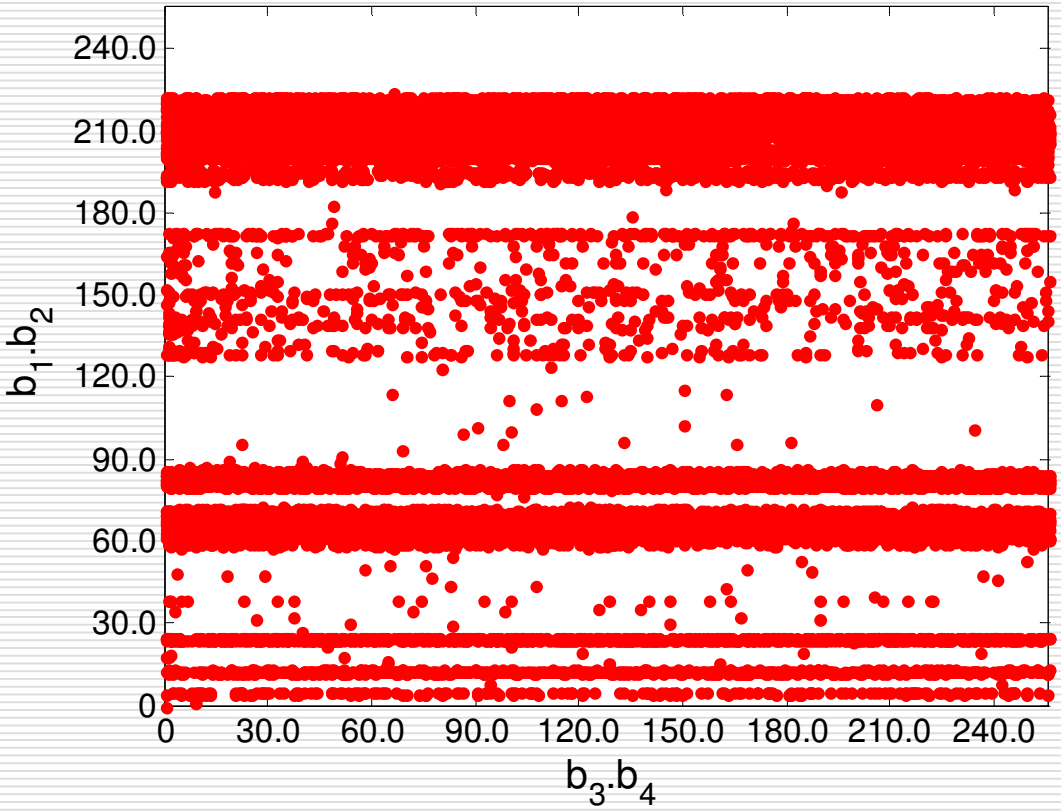


Unique source addresses

Data Dimensions

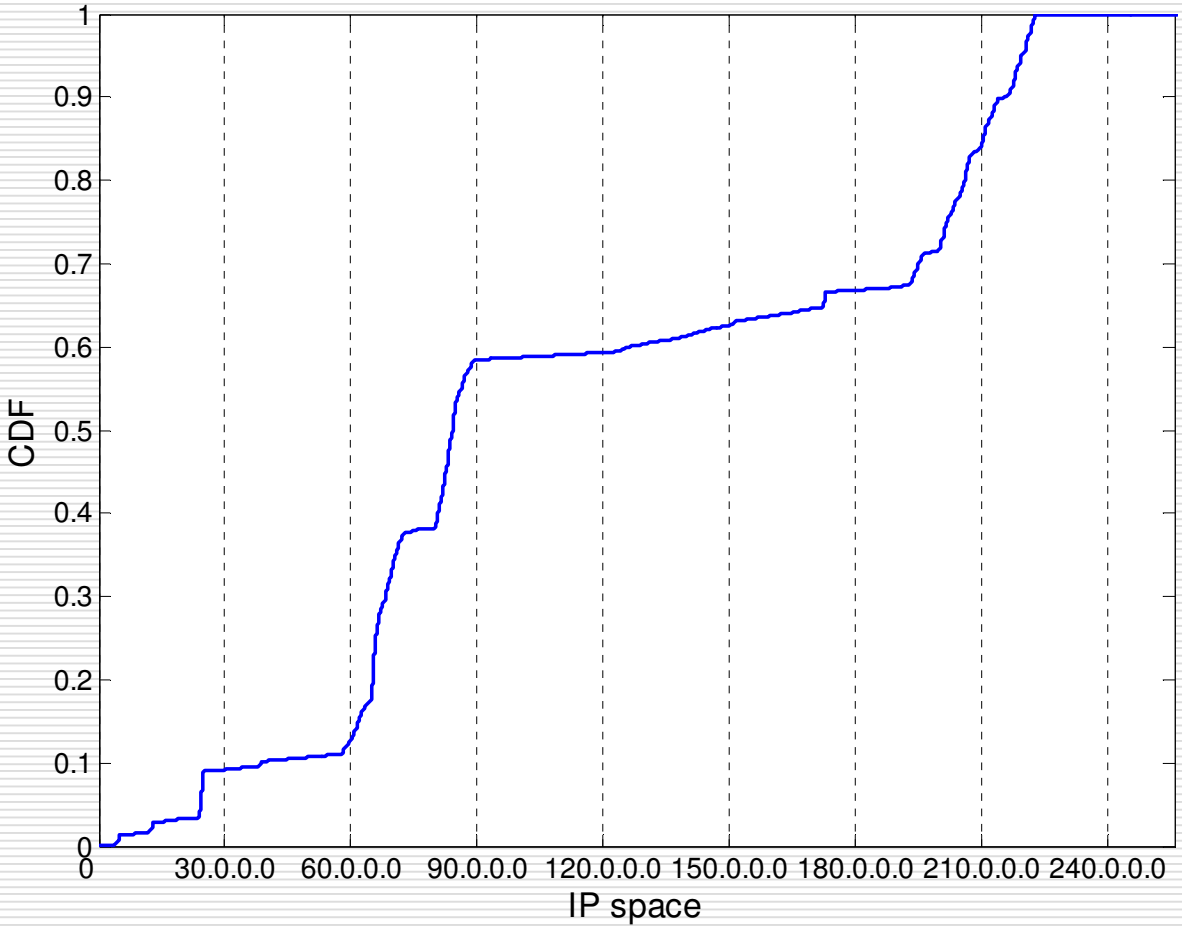
- Spatial behavior
- Temporal behavior
- Exploited protocols

A Snapshot on Spatial Distribution



August 6, 2005

Spatial Distribution for Entire Trace



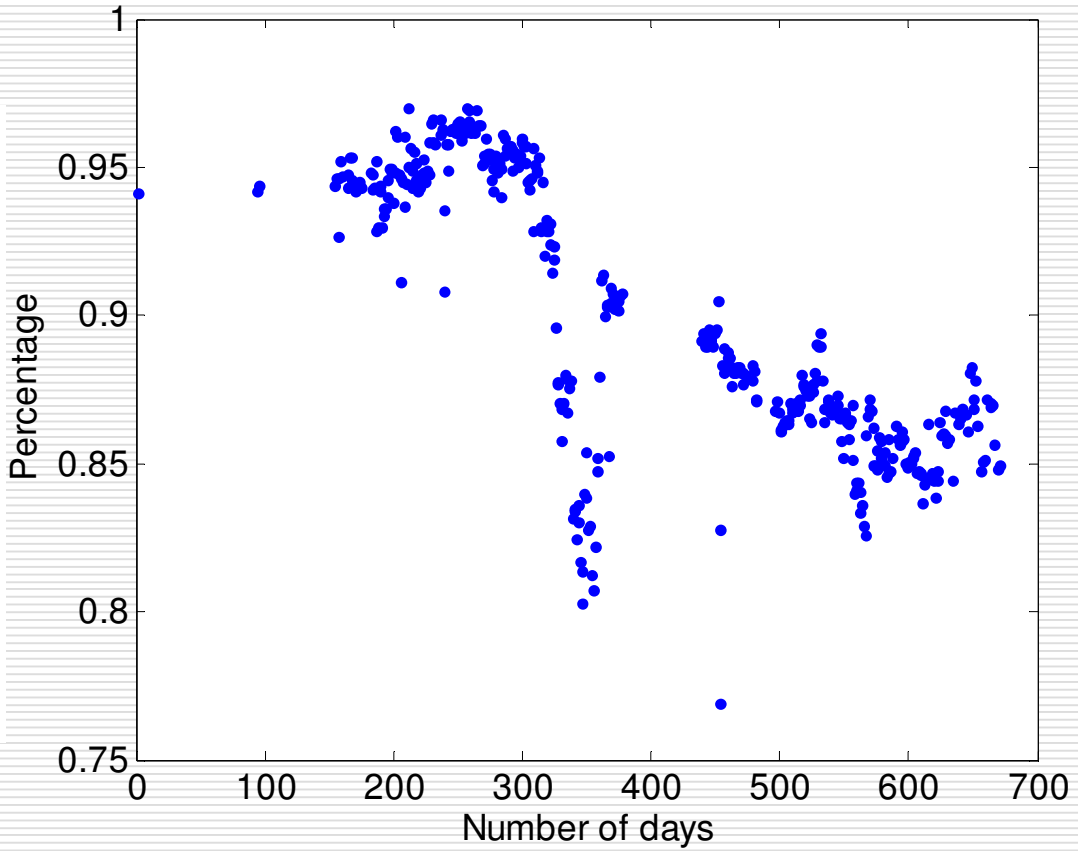
Malicious sources are clustered!

80-20 Rule (Pareto Principle)

TABLE I
9 RANGES OF /16 SUBNETS OR /16 CLUSTERS.

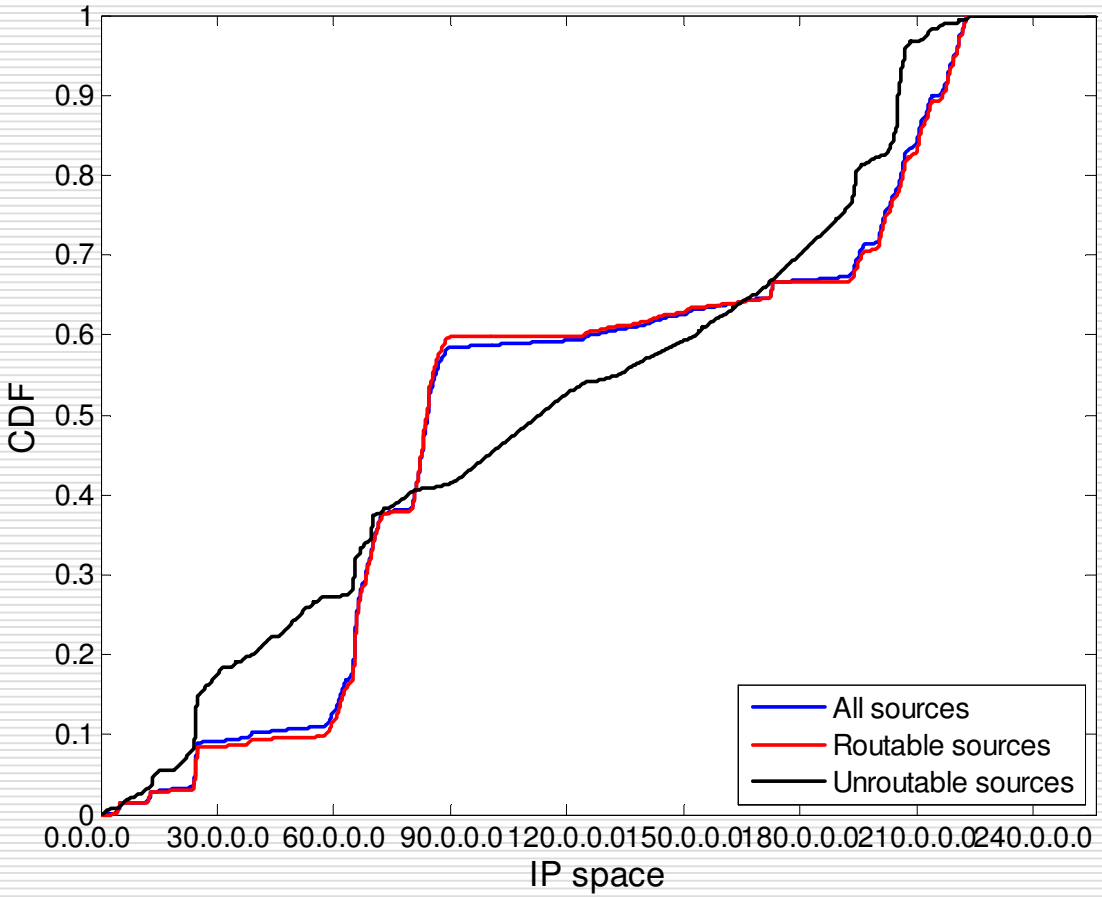
59.0.0.0/16 ~ 72.57.0.0/16
80.0.0.0/16 ~ 86.141.0.0/16
192.38.0.0/16 ~ 196.204.0.0/16
198.53.0.0/16 ~ 213.255.0.0/16
216.6.0.0/16 ~ 222.253.0.0/16
4.0.0.0/8
12.0.0.0/8
24.0.0.0/8
172.0.0.0/8

20% of IP address space



Unroutable Sources

- BGP routing table information
- Dec. 19, 2006
- 217,025 prefixes
- 7.3% unroutable



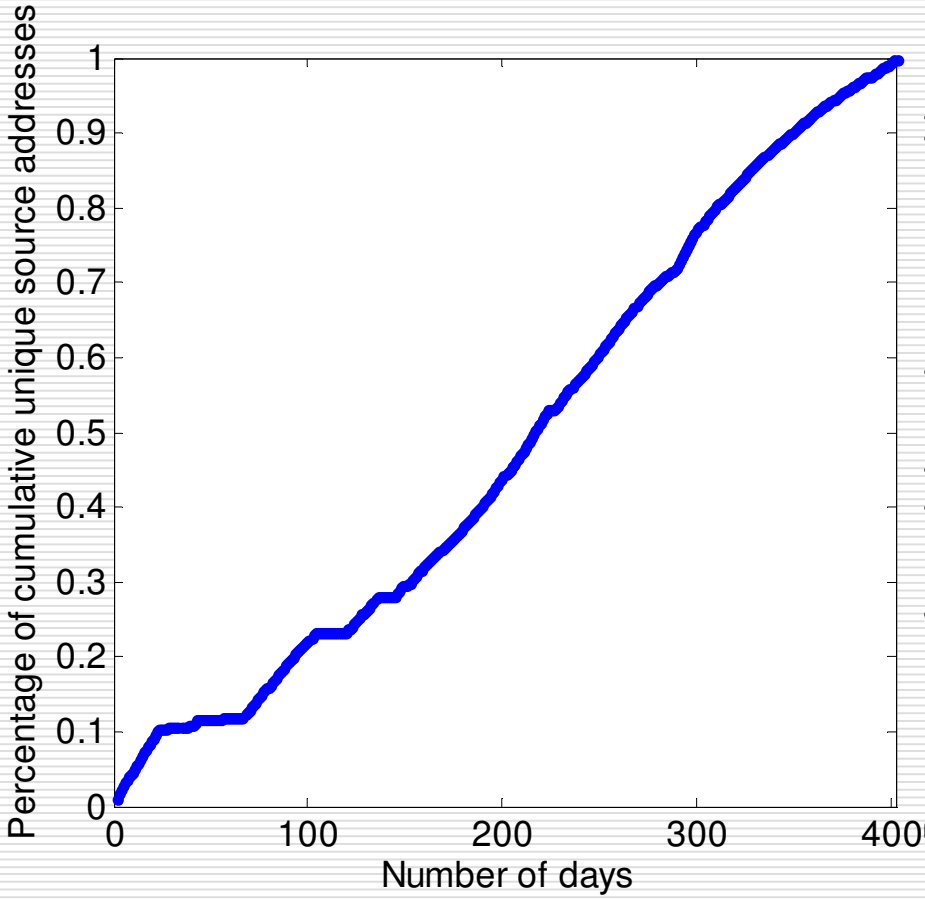
Top 20 Prefixes

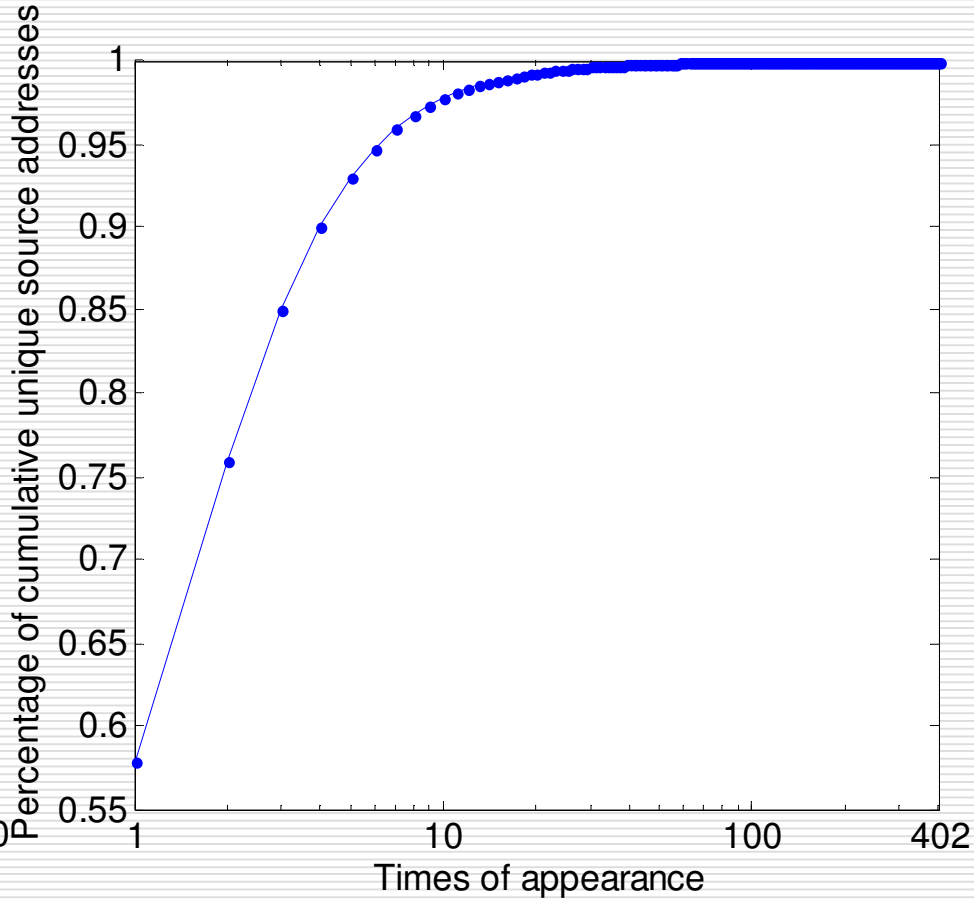
TABLE II
TOP 20 PREFIXES.

IP prefix	AS#	# sources	Country	ISP
210.0.0.0/8	7474	3643064	AU	OPTUSONLINESERVICES-AP
84.128.0.0/10	3320	3115791	DE	DEUTSCHE TELEKOM AG
61.0.0.0/8	4678	2359184	IN	NATIONAL INTERNET BACKBONE
172.128.0.0/10	1668	2045237	US	AMERICA ONLINE
4.0.0.0/8	3356	1961772	US	LEVEL 3 COMMUNICATIONS INC
12.0.0.0/8	7018	1851807	US	ATT LINCROFT ORT
65.128.0.0/11	209	1781094	US	QWEST COMMUNICATIONS CORPORATION
65.192.0.0/11	701	1416188	US	UUNET TECHNOLOGIES INC
38.0.0.0/8	174	1031860	US	PERFORMANCE SYSTEMS INTERNATIONAL INC
80.128.0.0/11	3320	994636	DE	DEUTSCHE TELEKOM AG
83.0.0.0/11	5617	956003	PL	VOIP SERVICES BY POLISH TELECOM
82.224.0.0/11	12322	935487	FR	PROXAD / FREE SAS
65.112.0.0/12	209	881941	US	HARVARD UNIVERSITY
12.0.0.0/9	7018	760014	US	ATT LINCROFT ORT
86.128.0.0/10	2856	742166	US	BT-CENTRAL-PLUS
65.0.0.0/12	6389	737370	US	BELLSOUTH.NET INC
4.0.0.0/9	3356	598696	US	LEVEL 3 COMMUNICATIONS INC
81.128.0.0/11	2856	583225	UK	BT-N3SP
217.224.0.0/11	3320	567963	DE	DEUTSCHE TELEKOM AG
172.192.0.0/12	1668	530767	US	AMERICA ONLINE

Based on http://www.ip2location.com/

Temporal Behavior



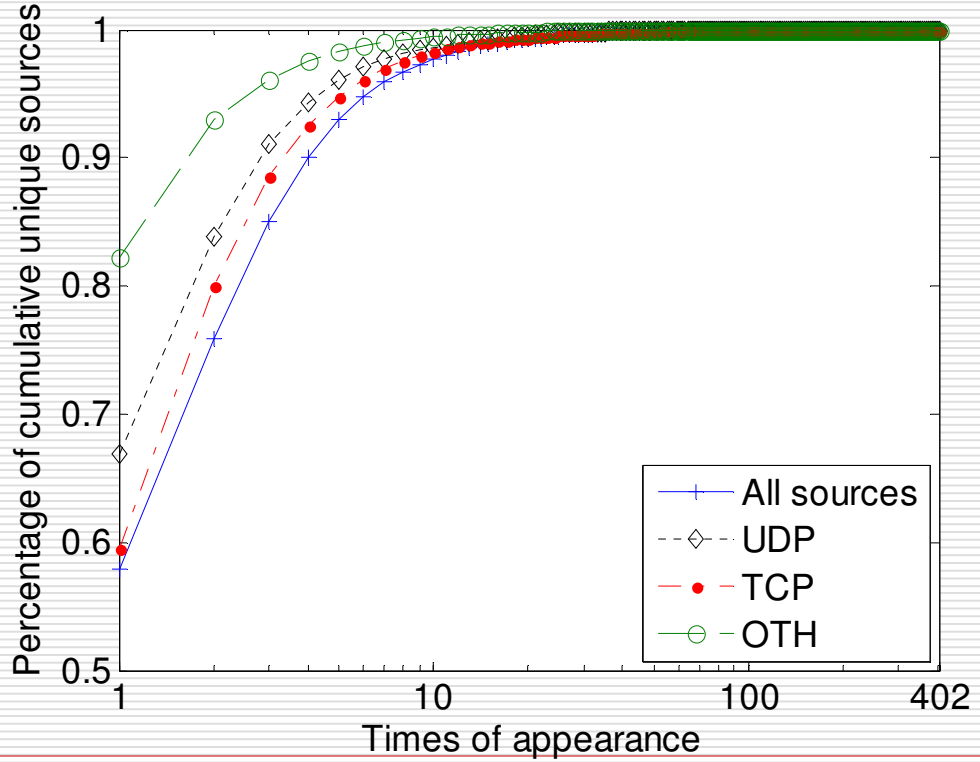


Most of sources have a short lifetime!

Exploited Protocols

TABLE IV
PERCENTAGES OF SOURCES THAT USE DIFFERENT TRANSPORT PROTOCOLS.

Protocols	TCP	UDP	OTH	TCP+UDP	TCP+OTH	UDP+OTH	TCP+UDP+OTH
Percentage	39.28%	76.35%	11.25%	19.06%	5.20%	5.95%	3.33%



Malicious sources prefer UDP!

Implications

- Malicious sources are clustered
  - 80-20 rule
  - 20% IP addresses
  - Reputation systems
- Most of sources have a short lifetime
  - Difficulty in using Blacklist
- Malicious sources prefer UDP
  - Another dimension for detection

Conclusions and Future Work

- Study the spatial-temporal characteristics of malicious sources
- Base on a huge trace over 22 months
- Find that 1 out of 27 host IP addresses is potentially a malicious source
- Discover that more than 80% of sources are clustered in the same 20% of the IPv4 address space over time
- Incorporate spatial-temporal features of malicious sources into the defense systems

Thanks for your attention