Strengthening Forensic Investigations of Child Pornography on P2P Networks

Marc Liberatore[1] Brian Neil Levine[1] Clay Shields[2]

[1]University of Massachusetts Amherst
[2]Georgetown University

Conference on emerging Networking EXperiments and Technologies (CoNEXT 2010)

UMASS AMHERST

Outline

- Measurement, forensics, and investigations
- Measurements of P2P distribution of child pornography (CP)
- Tagging, a technique for improving the value of evidence

Measurement vs. Forensics

Network measurement is a sampling of relevant information about a network. Network measurement aims to meet a scientific standard.

Forensic measurement is a set of measurements used to establish identity, intent, and actions. Forensic measurement aims to meet a legal standard.

RoundUp — A Tool for P2P Investigations

We built and deployed RoundUp, a tool for forensic measurement of Gnutella.

RoundUp is in use by over 600 investigators in ICAC, as well as at the FBI.

RoundUp measures Gnutella traffic, and can forensically measure specific traffic.

(Liberatore, Erdely, Kerle, Levine, and Shields in [DFRWS2010])

Finding Candidates

Goal

Find evidence of a crime through observations on the Internet.

Evidence:

- may be direct or hearsay
- includes files of interest, hash values, filenames
- is ultimately associated with a user (IP address? GUID?)

Use the p2p system to find candidates for further investigation.

This process is measurement!

Evidence

A candidate is chosen for further investigation, by jurisdiction, type/quantity of files, observed history.

The investigator directly connects to:

- determine all files shared by a peer
- find other corroborating evidence (IP, GUID, vendor id)
- perform a single-source download

This process should be forensic measurement!

Subpoena and Search Warrant

Network investigation done; shoe-leather work remains:

- Subpoena ISP for DHCP records / billing information
- Search warrant for premises — written broadly

Once on site:

- Examine media and seize if appropriate
- Validate that evidence on media corresponds to network observations

Identifying Offenders

Investigators use observed IPs to obtain search warrants.

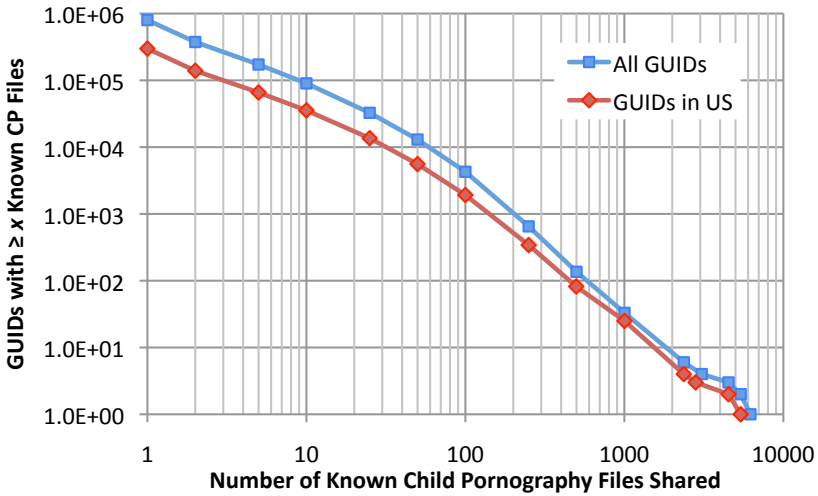
Investigators use network (IP) and application (GUID, PeerIds) identifiers to identify offenders, link observations, discern intent.

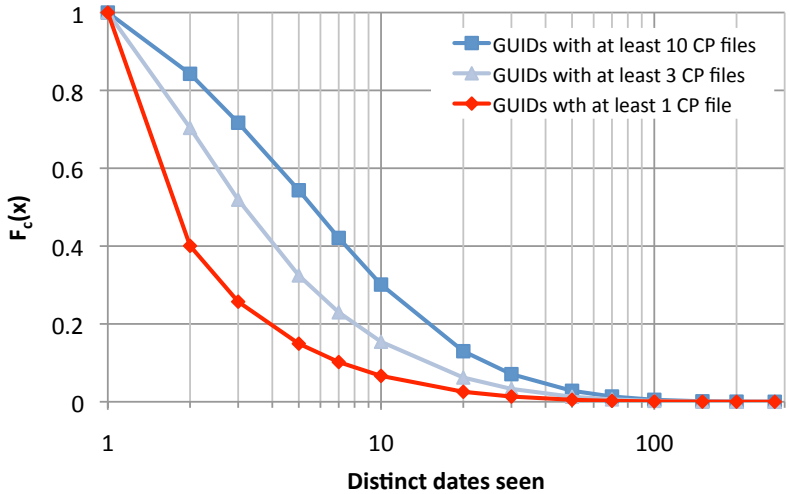
What did investigators observe?
How good (reliable, consistent, etc.) are IPs and GUIDs?

Measurement Summary

From 2009-10-05 through 2010-03-02:

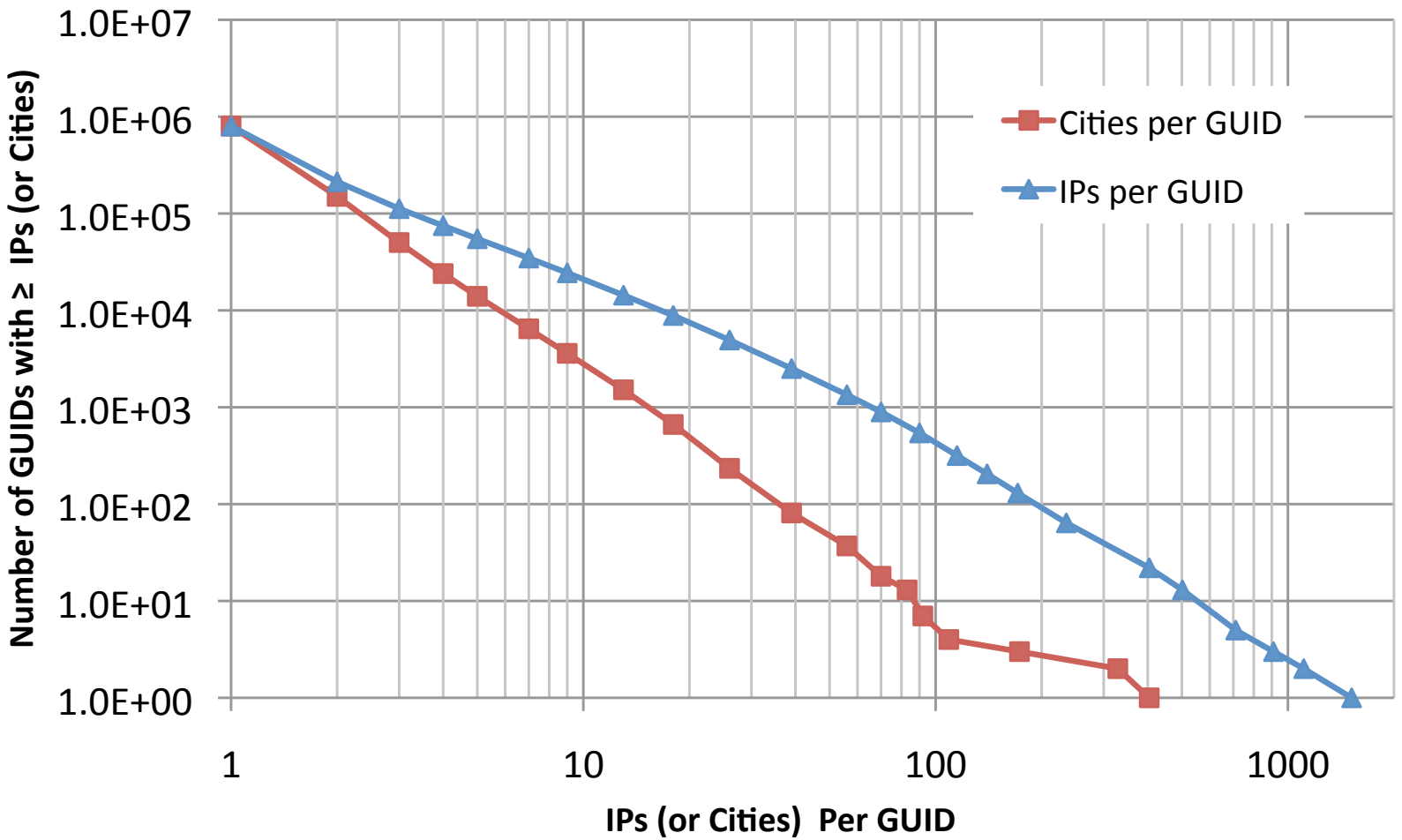
- 3.07 million IP addresses
- 799,556 GUIDs
- 19,000 distinct items of CP (by hash)





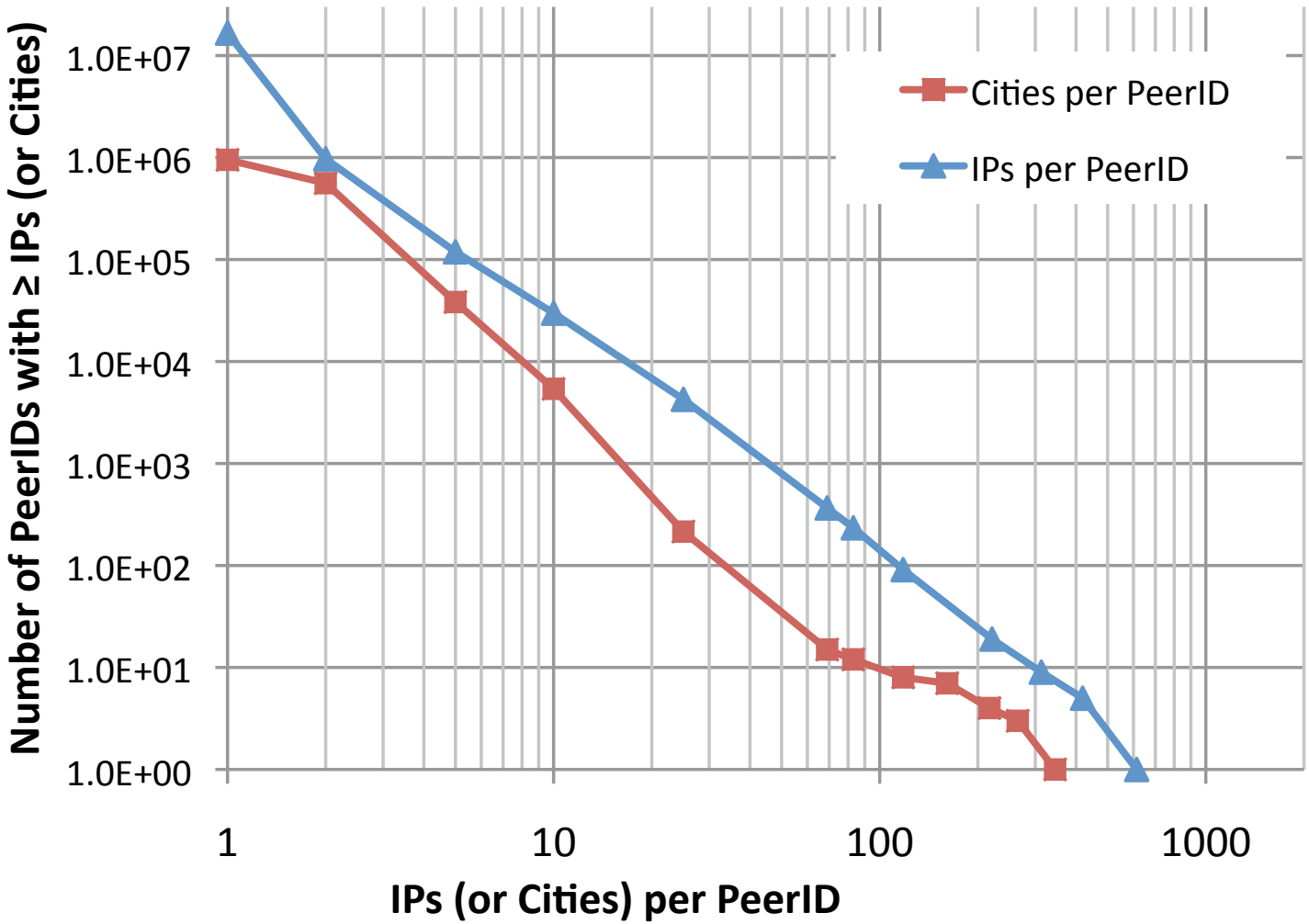
Gnutella IDs

Many GUIDs are mapped 1:1 to IPs – but not all!



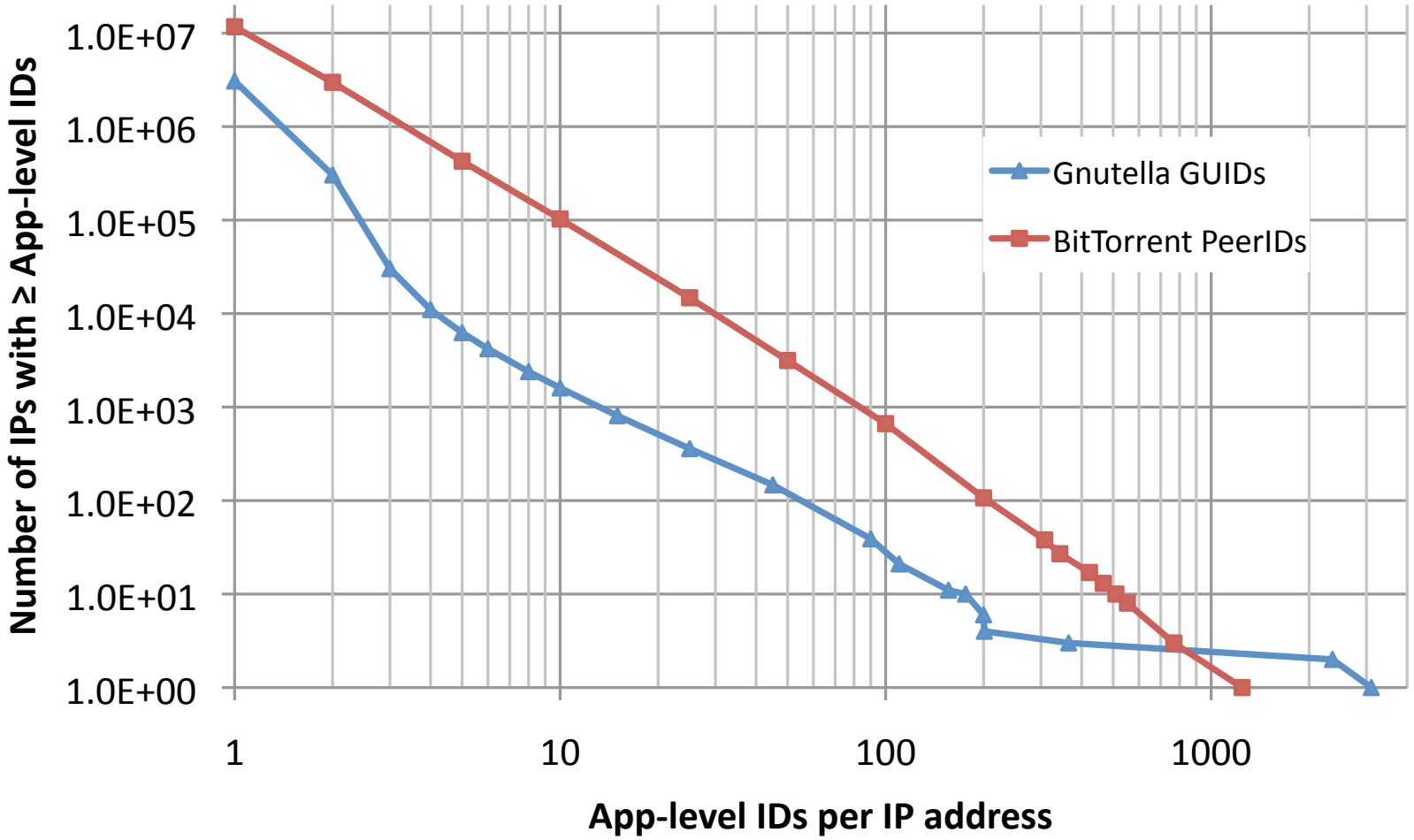
BitTorrent IDs

Same trends are present in BitTorrent:



IPs to GUIDs Also Unreliable

Again, many are 1:1, but not all.



What's Going On Here?

Many anomolies can be explained:

- One GUID observed in 329 cities, using 398 IP addresses — actually a botnet
- Many GUIDs stay in the same geographic area — mobile users
- IPs with several GUIDs may be NAT
- Some clients generate new IDs per download
- Tor

But we know this list isn't exhaustive.
And we can't always map anomolies to explanations.

What to Do?

Recall goals: to identify offenders, link observations, discern intent.

Proposed Solution

Tag network traffic such that recoverable markings are left on suspect's machine.

An Analogy

Pay drug dealers with marked bills, recover bills on arrest.

The Tagging Process

Deliver data to remote clients with tagged bits; recover bits on arrest. Key concerns:

- Finding appropriate **vectors** for tag delivery.
- Ensuring tags are **covert**.
- Quantifying the **false positive** rate.

Vectors and Covert Tags

An ideal vector allows arbitrary input, persists indefinitely, and is detrimental to disable.
We'll take what we can get, for example:

- BitTorrent peer IP caches
- DNS cache entries
- p2p payload data
- log files

Ideally we'd find them by automated (static?) analysis.

We'll tag with bit strings that have no overt meaning.

Example Tags

BitTorrent peer caches store IPs:

```
{'ip': '83.253.52.14',
 'port': 6886,
 'prot': 1,
 'src': 'Tracker'}, ... },
{'ip': '87.7.101.196',
      'port': 54650,
      'prot': 1,
      'src': 'PeerExchange'},
...
```

Values can be added to a peer's cache through peer exchange.
Investigators can use these IPs (which may be spoofed) as a tag.

More Tags

Vuze log files record all unknown PeerIDs:

```
- [2009] Log File Opened for Vuze 4.2.0.2
- [0406 09:16:22] unknown_client [LTEP]:
"Unknown KG/2.2.2.0" / "KGet/2.2.2"
[4B4765742F322E322E32],
Peer ID: 2D4B47323232302D4947755337614 94E45425245
- [0406 09:22:14] mismatch_id [LTEP]:
"BitTorrent SDK 2.0.0.0" / "BitTorrent SDK 2.0"
[426974546F7272656E742053444B20322E30],
Peer ID: 2D4245323030302D2759514731415 95027646262
```

PeerIDs are arbitrary, 20-byte values.

`sha1(‘‘Detective John Doe, case #1234, ...’’)`
would make a great PeerID-based tag.

False Positive Rate

Let tags be of length n.

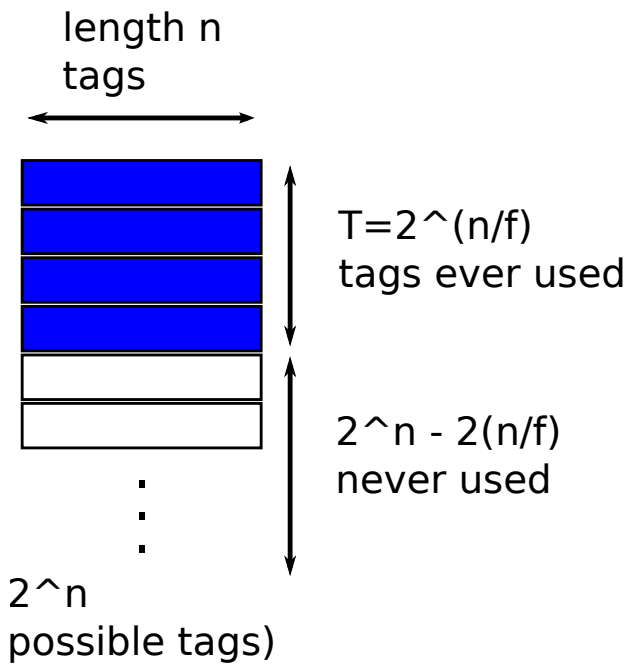
Assume a priori a number of taggable events $T = 2^{n/f}$, where $f > 1$.

If an investigator recovers L candidate tags from a machine:

$$\begin{aligned} & Pr\{\text{False positive}\} \\ = 1 - & Pr\{\text{no candidate matches}\} \\ & = 1 - (1 - \frac{2^{n/f}}{2^n})^L \end{aligned}$$

But often vectors have small n: If $L = 2000$ and $n \leq 32$, the chances of a false positive is greater than 3%. Too high?

Tagging table:



Alternate Tagging Techniques: Ordered Subsets

Solution: Break each tag into k subtags that fit constraints.

Subtags can be stored in a preserved order (e.g., a log file):

$$\begin{aligned} Pr\{\text{False positive}\} &= 1 - Pr\{\text{no full tag matches}\} \\ &\leq 1 - \left(1 - \binom{L}{k}\frac{1}{2^n}\right)^{2^{\frac{n}{f}}} \end{aligned}$$

Without ordering, there are several other approaches.

Alternate Tagging Techniques: Unordered Subsets

We can subtag k times per observation

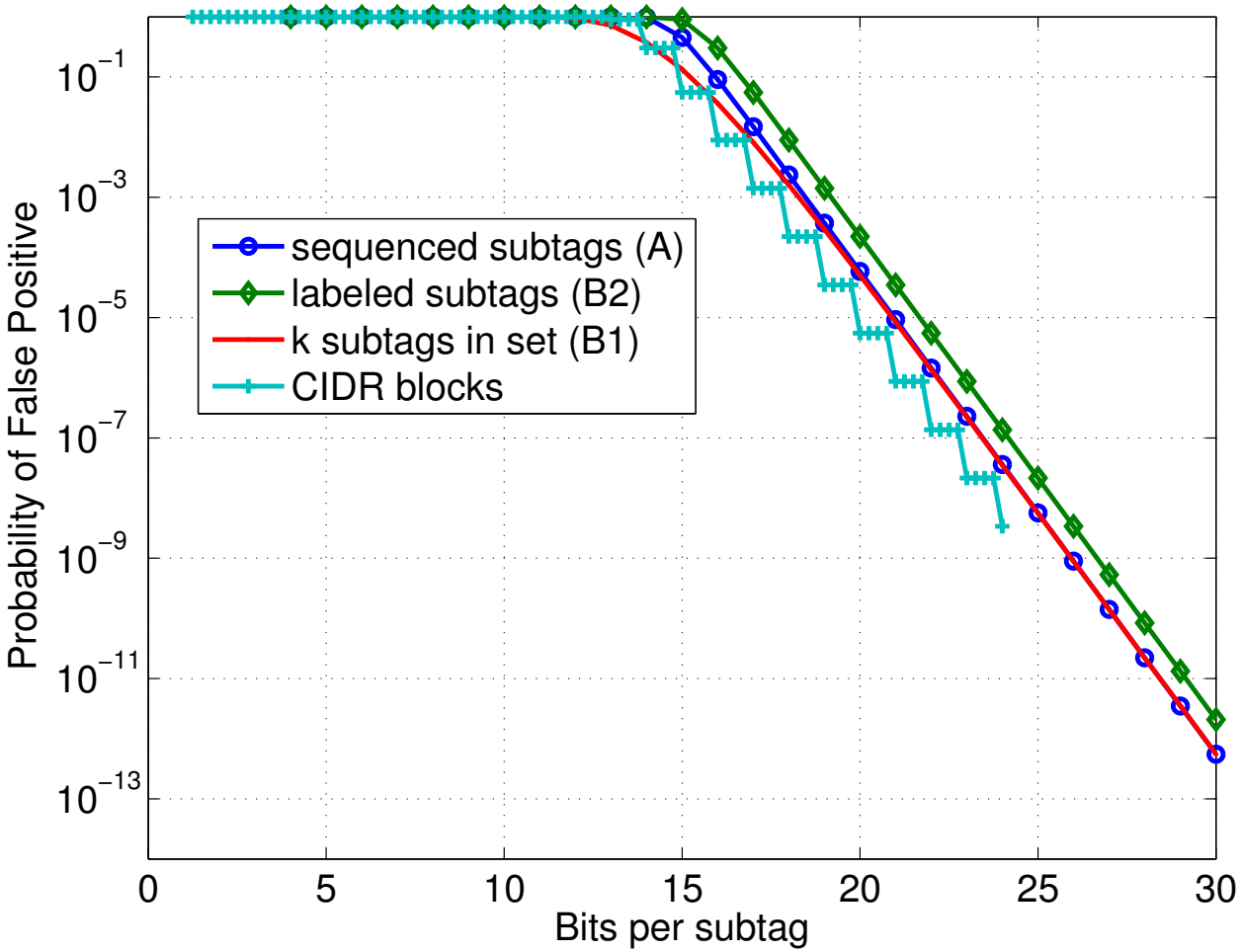
$$\begin{aligned} Pr\{\text{F.P.}\} &= Pr\{k \text{ or more of } L \text{ subtags match}\} \\ &= 1 - \sum_{i=0}^{k-1} \binom{L}{i} (2^{\frac{n}{fk} - \frac{n}{k}})^i (1 - (2^{\frac{n}{fk} - \frac{n}{k}}))^{L-i} \end{aligned} \quad (1)$$

We can reserve bits to impose order:

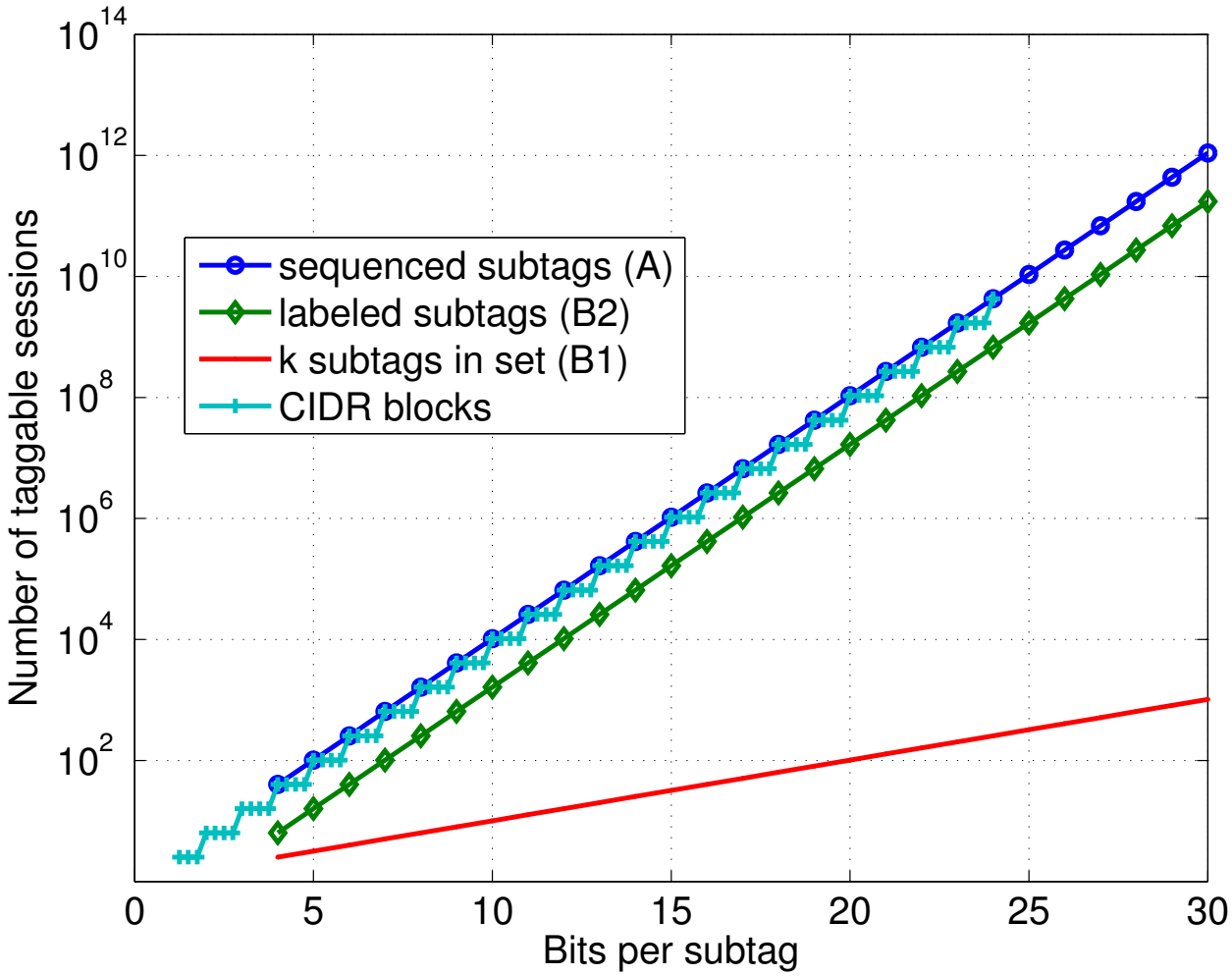
$$\begin{aligned} Pr\{\text{F.P.}\} &= 1 - Pr\{\text{none of } (\tfrac{L}{k})^k \text{ subtags match}\} \\ &= 1 - \left(1 - \frac{2^{rk/f}}{2^{rk}}\right)^{(\frac{L}{k})^k} \end{aligned}$$

Subtags can contain implicit ordering (e.g., fixed CIDR bits): a special case of built-in reserved bits.

More Bits, Lower FP Probability



More Bits, More Taggable Sessions



Conclusions

- Forensic measurements have different standards and goals from typical network measurements.
- Network and application-level identifiers may suffice for probable cause, but are not 100% reliable.
- Tagging allows for the flexible creation of forensically verifiable identifiers.

Acknowledgments

This work was supported in part by National Institute of Justice Award 2008-CE-CX-K005 and in part by the National Science Foundation awards CNS-0905349, CNS-1018615 and DUE-0830876. The opinions, findings, and conclusions or recommendations expressed in this publication are those of the authors and do not necessarily reflect those of their employers, the U.S. Department of Justice, the National Science Foundation, or ICAC.