THE UNIVERSITY OF TOKYO

A Distributed Index Poisoning Algorithm for Effective Control of Peer-to-Peer Network Applications

Pratama PUTRA and Akihiro NAKAO

The University of Tokyo

Outline

PART I: Introduction	1. Background 2. Objective
PART II: Proposed Solution	1. Best Property: Network Structure 2. Influential Peer Selection Algorithm
PART III: Evaluation	3. Evaluation Settings 4. Evaluation Result

Outline

PART I:
Introduction

1. Background
2. Objective

PART II:
Proposed Solution

1. Best Property: Network Structure
2. Influential Peer Selection Algorithm

PART III:
Evaluation

3. Evaluation Settings
4. Evaluation Result

Generic P2P Problem

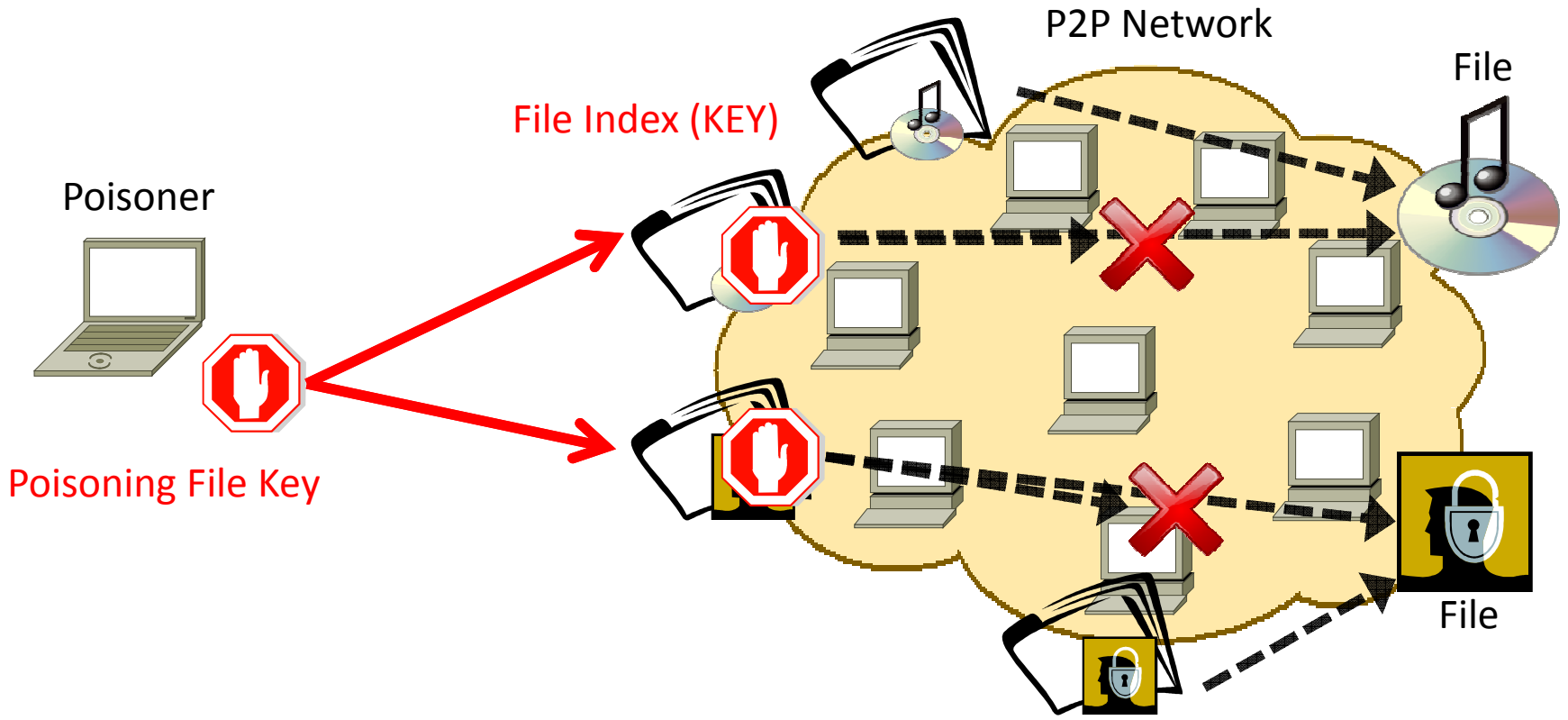
Harmful Content Distributed Using P2P and the Damage They Cause:

- Copyrighted: Causes economic loss to many industries
  - E.g., Music, Videos, Software
- Privacy Data: Can be used for bad purposes
  - E.g., Bank Account, Tax Information, Personal Photos
- Confidential Information: Compromises national security
  - E.g., Military Document
- Illegal Content: Harms the society
  - E.g., Child Pornography

These Content (Files) are Harmful and the Distribution Need to be Controlled

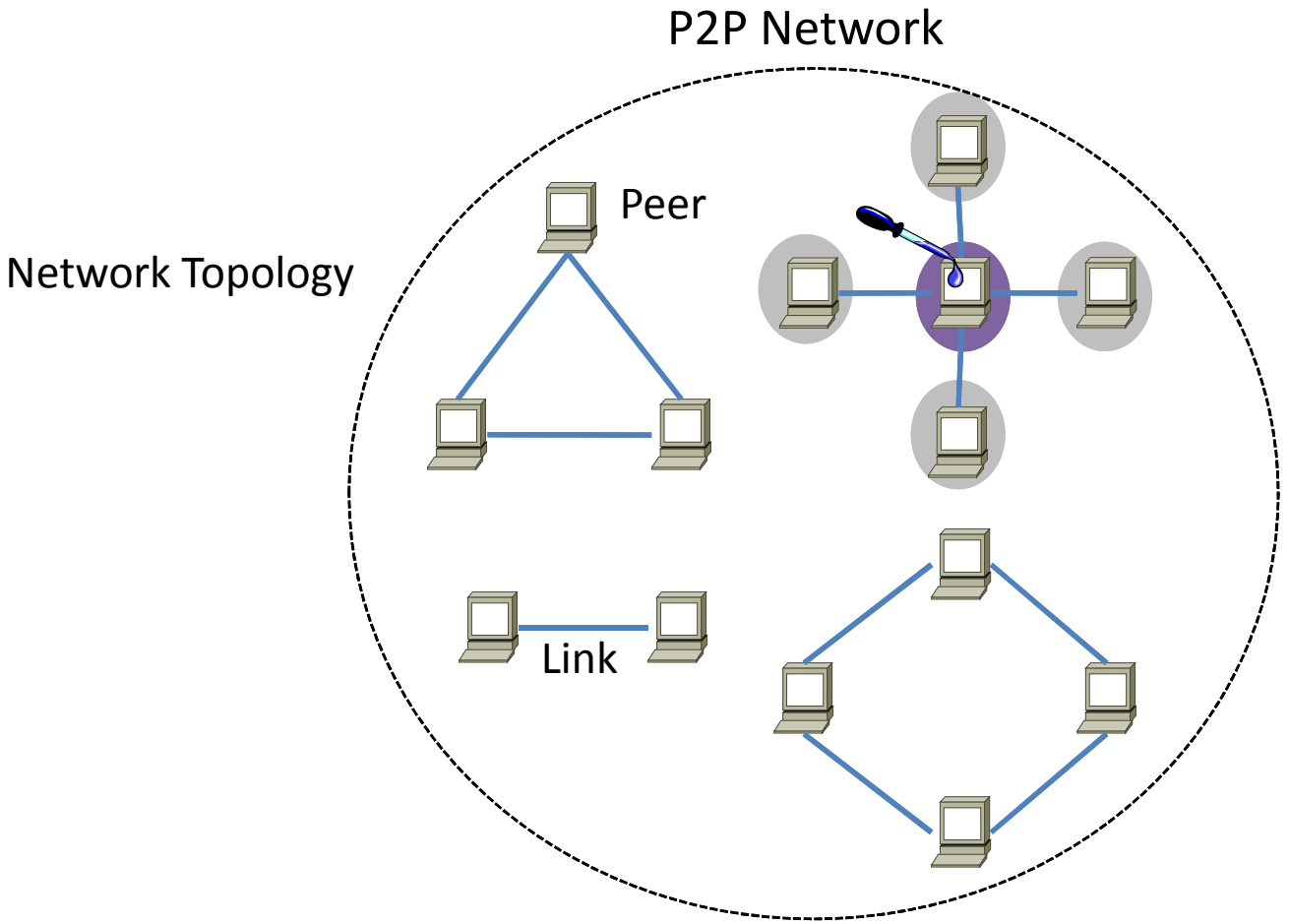
File Distribution Control: Index Poisoning

"Index Poisoning" = Breaking pointers (file keys) to content



Polluting (Injecting Poison) the Network with Bogus File Keys Information

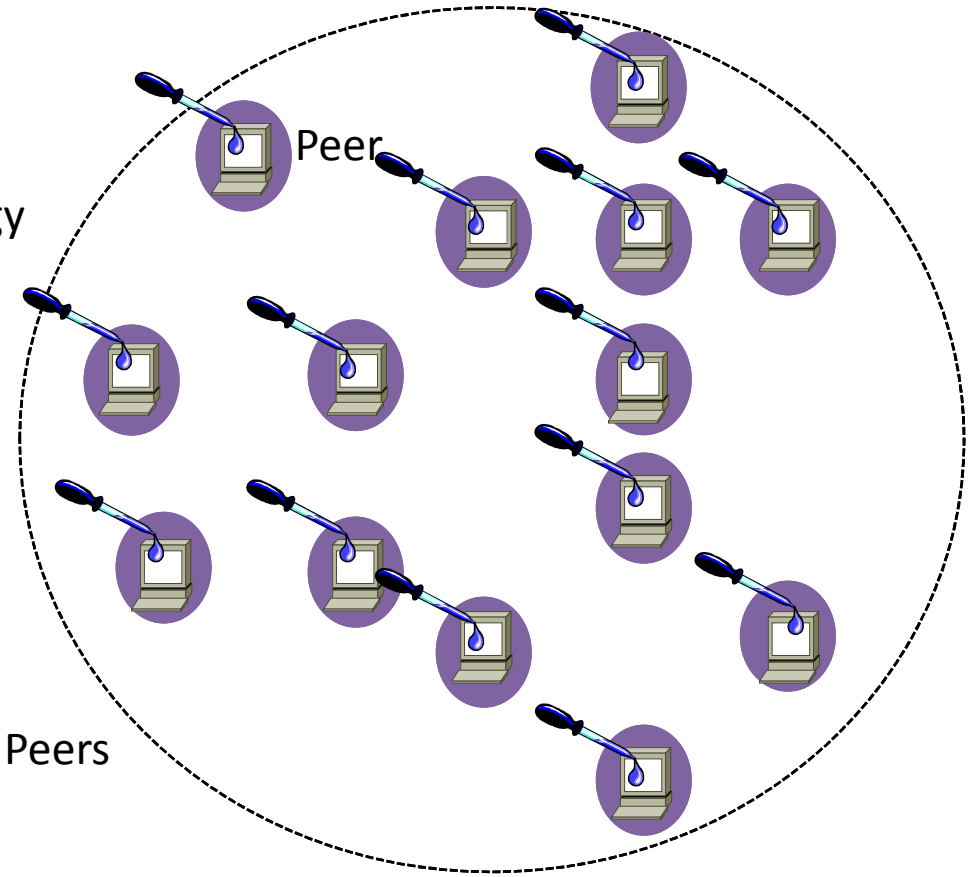
Direct and Indirect Poisoning



Direct Poisoning : Directly Inject Poison to the Target

Indirect Poisoning : Poisoning Effect through Infection from Poisoned Peer

Existing Poisoning (Without Link Info)



Network Topology

Link is not “Visible”

Generated Traffic

“Direct” Poisoning All Peers

Not Effective!!

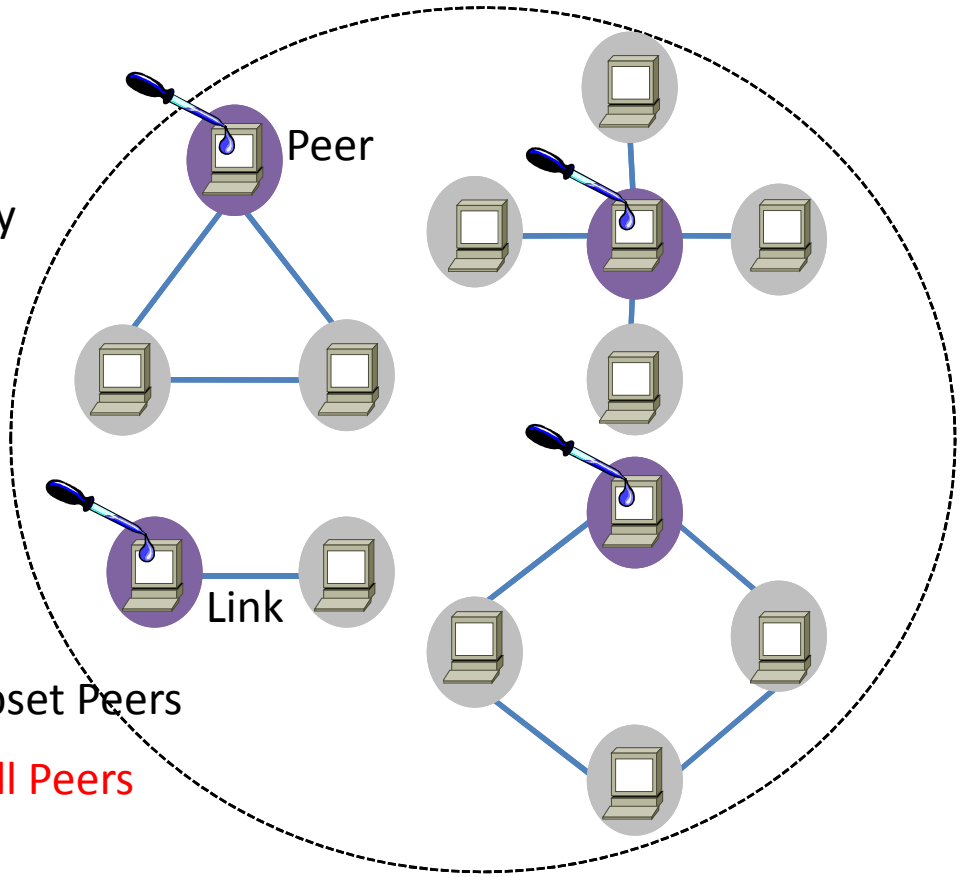
Existing Work: High Coverage Poisoning, but Generate High Traffic

Real Case of Index Poisoning*)

- Generates traffic for poisoning single file = 92 Kbps
- More than 3 million copyrighted files
- Traffic required to control = 276 Gbps!!
  - not negligible compared to the legitimate traffic in P2P networks and in the Internet

*) M. Yoshida, S. Ohzahata, A. Nakao , and K. Kawashima, "Controlling File Distribution in Winny Network through Index Poisoning, "*Proceedings of the 23rd International Conference on Information Networking*, pp. 210–214, 2009.

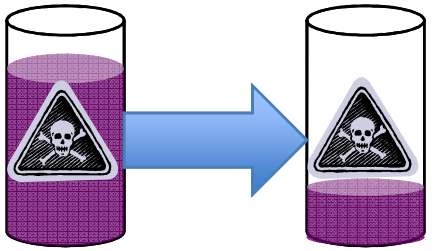
Effective Poisoning (Knowing Link Info)



Network Topology

Link is “Visible”

Generated Traffic



“Direct” Poisoning Subset Peers

“Indirect” Poisoning All Peers

Not Effective!!

Achieve Same Effectiveness with smaller Generated Traffic

= Effective Poisoning → **Our Focus!!**

Outline

PART I:
Introduction

1. Background
2. Objective

PART II:
Proposed Solution

1. Best Property: Network Structure
2. Influential Peer Selection Algorithm

PART III:
Evaluation

3. Evaluation Settings
4. Evaluation Result

Objective

Develop Effective Index Poisoning System to:

- Maximize "influence" of traffic control
- Minimize "generated traffic" for the control

By Limiting Poisoning Scope to Small Number of Influential Peers

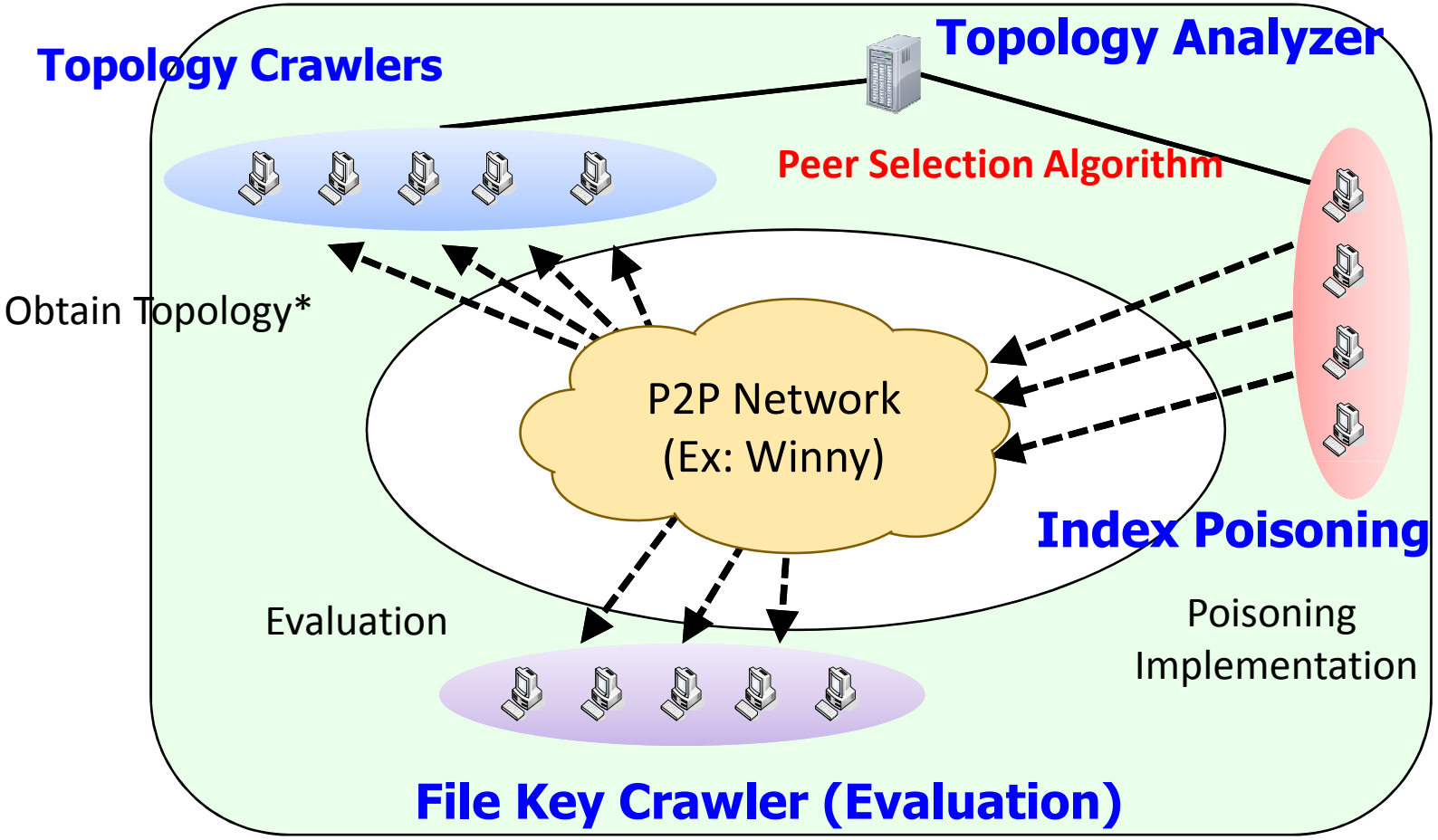
Using ***Winny*** Network as an example

Requirements:

P2P Network Topology Information*

* P. Putra and A. Nakao, "Measuring Peer-to-Peer Network Topology through Geo-Location-Aware Distributed Crawlers, *IEICE Tech. Rep.*, vol 109, no. 228, NS2009-96, pp. 109-114, Oct 2009

Distributed Poisoning System Overview



* P. Putra and A. Nakao, "Measuring Peer-to-Peer Network Topology through Geo-Location-Aware Distributed Crawlers, *IEICE Tech. Rep.*, vol 109, no. 228, NS2009-96, pp. 109-114, Oct 2009

Outline

PART I:
Introduction

1. Background
2. Objective

PART II:
Proposed Solution

1. Best Property: Network Structure
2. Influential Peer Selection Algorithm

PART III:
Evaluation

3. Evaluation Settings
4. Evaluation Result

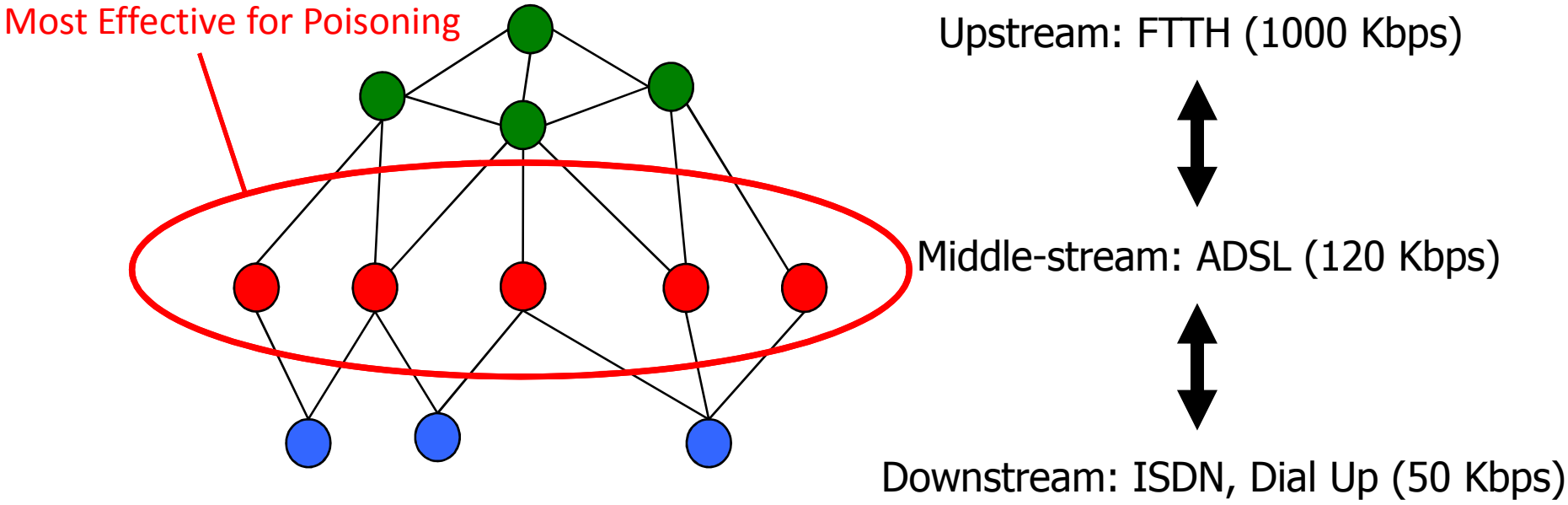
How to Use Topology?

- From Topology (Peer and Link), We Can Obtain Network Properties*
- Network Property Example
  - Node Degree
  - Network Cluster
  - Network Structure
  - Etc.
- **Network Structure** is the Best for Index Poisoning (In Winny Case)
- We Need to Infer Network Structure!!

* P. Putra and A. Nakao, ”Measuring P2P network topology through geo-location-aware distributed crawlers,” *8th Asia-Pacific Symposium on Information and Telecommunication Technologies (APSITT),* pp.1-6, 15-18 June 2010

Best Property*: Network Structure

- Structure of Winny**



- Structure determined by Peer's Bandwidth Declaration & File Key Carried by **Search Query** distributed more easily from down/middle to upstream
- We confirm with experiment that middle-stream (120 Kbps) is the best for Index Poisoning

* P. Putra and A. Nakao, "Measuring Peer-to-Peer Network Topology through Geo-Location-Aware Distributed Crawlers, *IEICE Tech. Rep.*, vol 109, no. 228, NS2009-96, pp. 109-114, Oct 2009

** Isamu Kaneko, "The Technology of Winny," ASCII, 2005.

Outline

PART I:
Introduction

1. Background
2. Objective

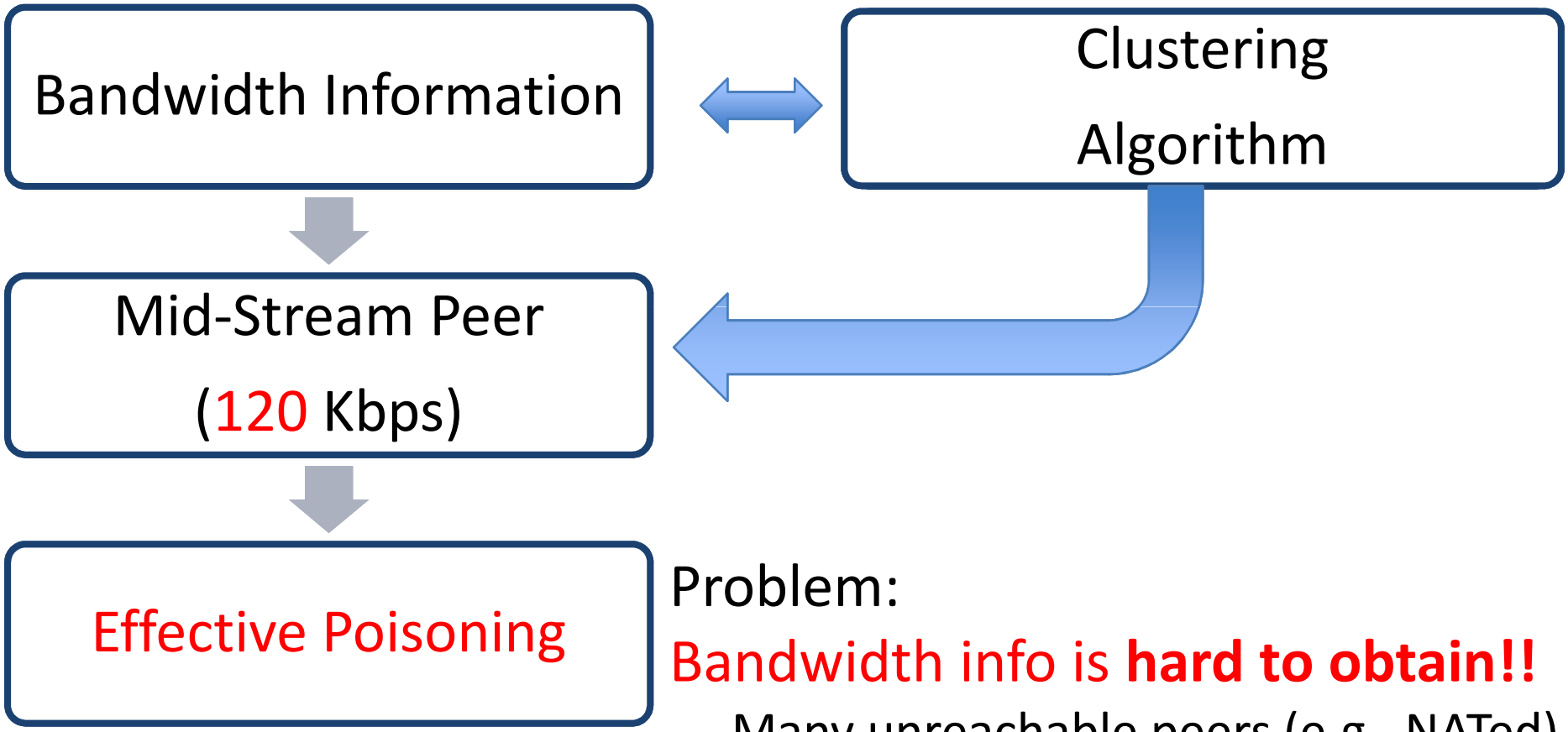
PART II:
Proposed Solution

1. Best Property: Network Structure
2. Influential Peer Selection Algorithm

PART III:
Evaluation

3. Evaluation Settings
4. Evaluation Result

Problem Overview



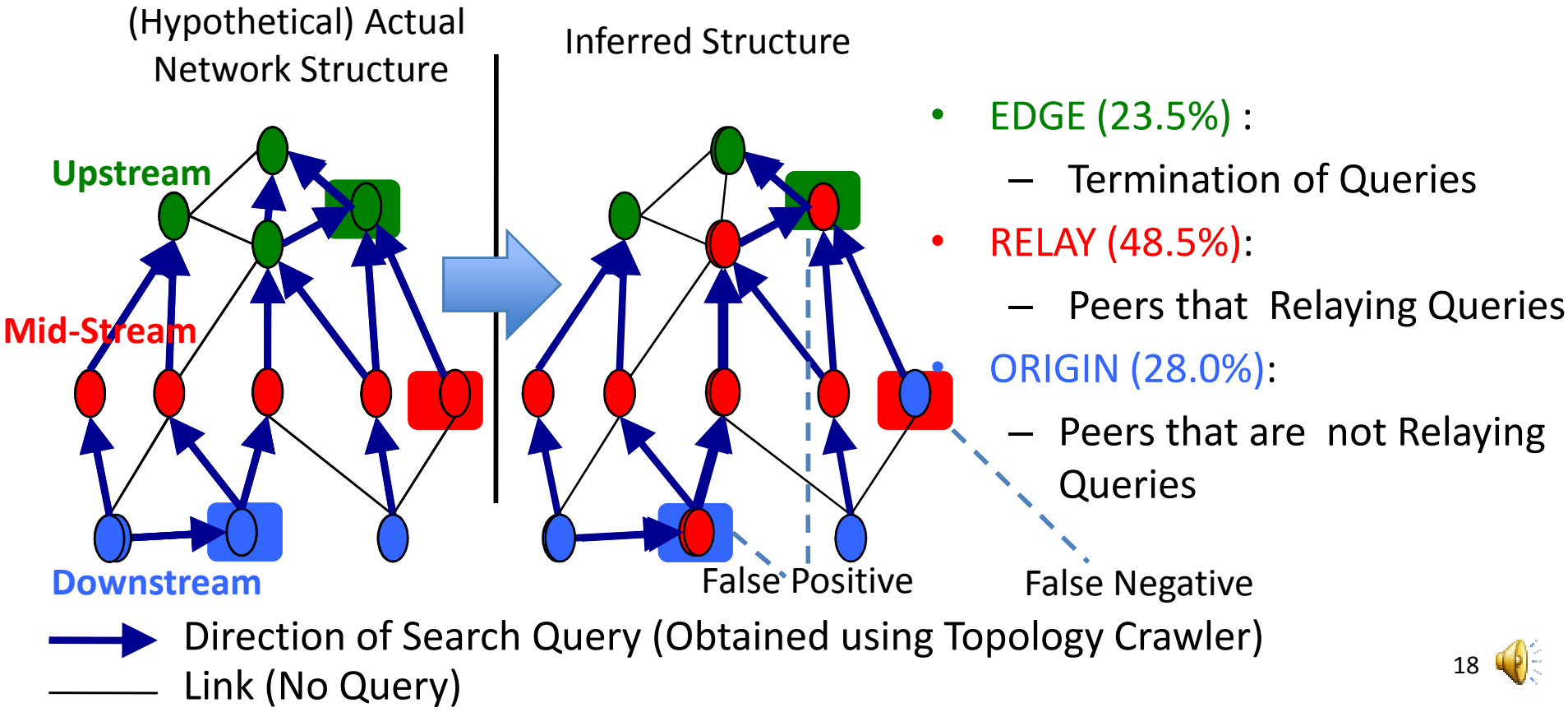
Problem:

Bandwidth info is **hard to obtain!!**

- Many unreachable peers (e.g., NATed)

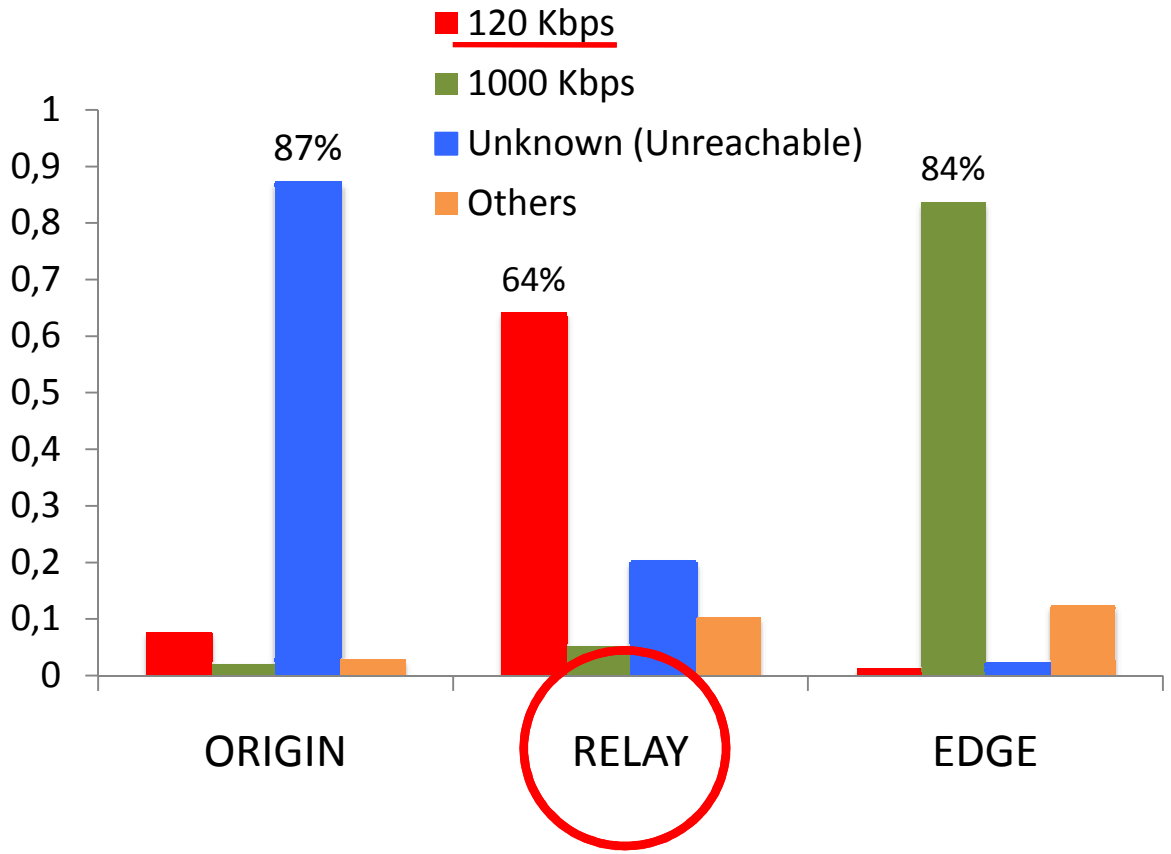
Level-1 Clustering Algorithm

- Infer the Network Structure by Looking at Peers Behavior in Search Queries
- May not Exactly the Same with Actual Structure, but Close Enough



- EDGE (23.5%) :
  - Termination of Queries
- RELAY (48.5%):
  - Peers that Relaying Queries
- ORIGIN (28.0%):
  - Peers that are not Relaying Queries

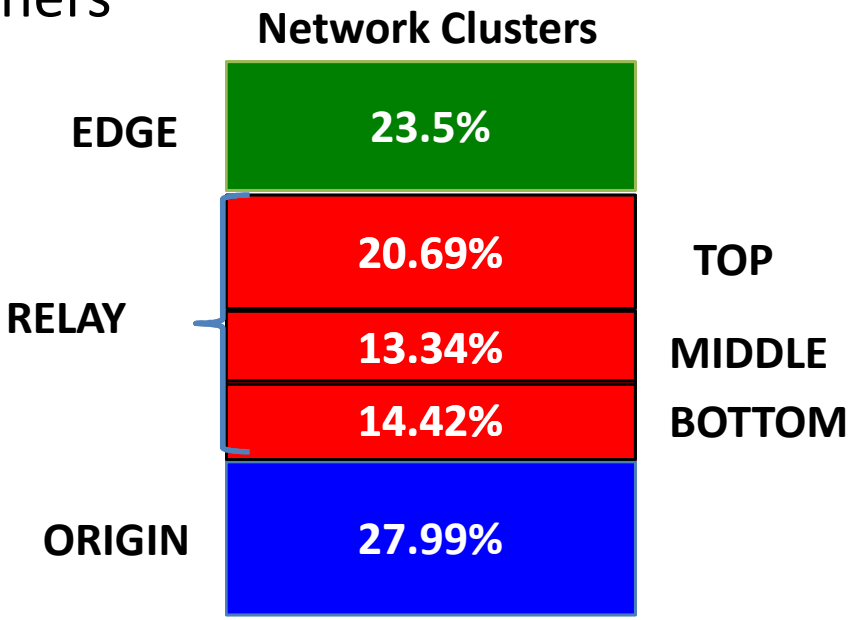
Level-1 Clusters Characteristic (Based on Bandwidth Declaration)



- Each Clusters are Dominated with Peers with Specific Bandwidth Declaration
- RELAY Seems to be Most Effective, Since Contains many 120 Kbps Peers
- RELAY (48%) need to be divided into smaller clusters to reduce more traffic

Level-2 Clustering Algorithm

- Divides RELAY by Looking at Peers Proximity with EDGE and ORIGIN
  - TOP : Peers Adjacent to EDGE
  - BOTTOM : Peers Adjacent to ORIGIN
  - MIDDLE : Others



- MIDDLE and BOTTOM are Likely the Most Effective Clusters (contains many 120 kbps, occupy small portion, High and Fast Key-spread*)

* Confirmed with experiment

Outline

PART I:
Introduction

1. Background
2. Objective

PART II:
Proposed Solution

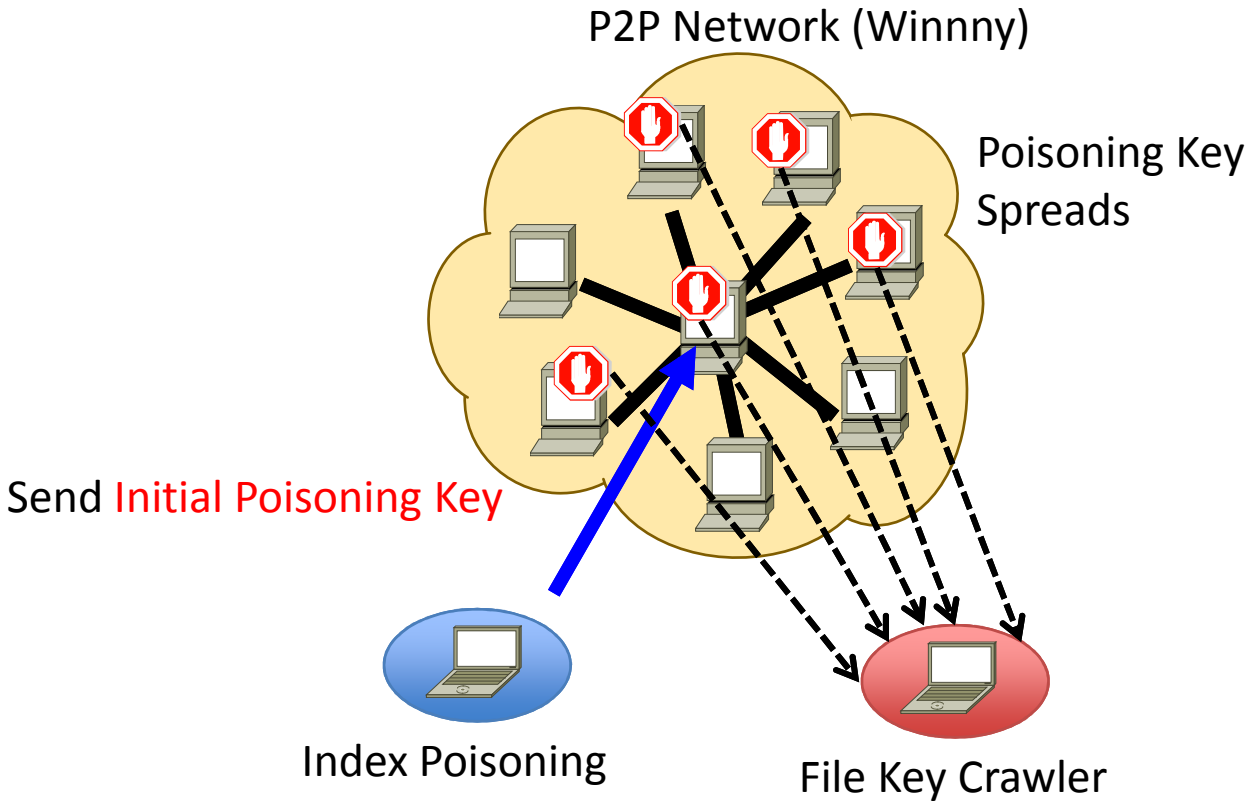
1. Best Property: Network Structure
2. Influential Peer Selection Algorithm

PART III:
Evaluation

3. Evaluation Settings
4. Evaluation Result

Evaluation Experiment

- Evaluation Metric:
  - Coverage $= \frac{\text{Poisoning Key Holders}}{\text{Total Peers}}$
  - Traffic Measured using Packet Capture



Count the Number of Poisoning Key Holders

Comparison of Poisoning Method

- MIDDLE Cluster Poisoning
  - Directly Poison MIDDLE peers (13%)
- MIDDLE + BOTTOM Cluster Poisoning
  - Directly Poison MIDDLE (13%) and BOTTOM (14%)
  - Two Key Lifetime Options: 20 and 40 minutes

Our Methods

- All Peers Poisoning (Existing Method)
  - Directly Poison 100% Peers

Our Methods Poisons Smaller Target, Achieve the Same Effectiveness

Outline

PART I:
Introduction

1. Background
2. Objective

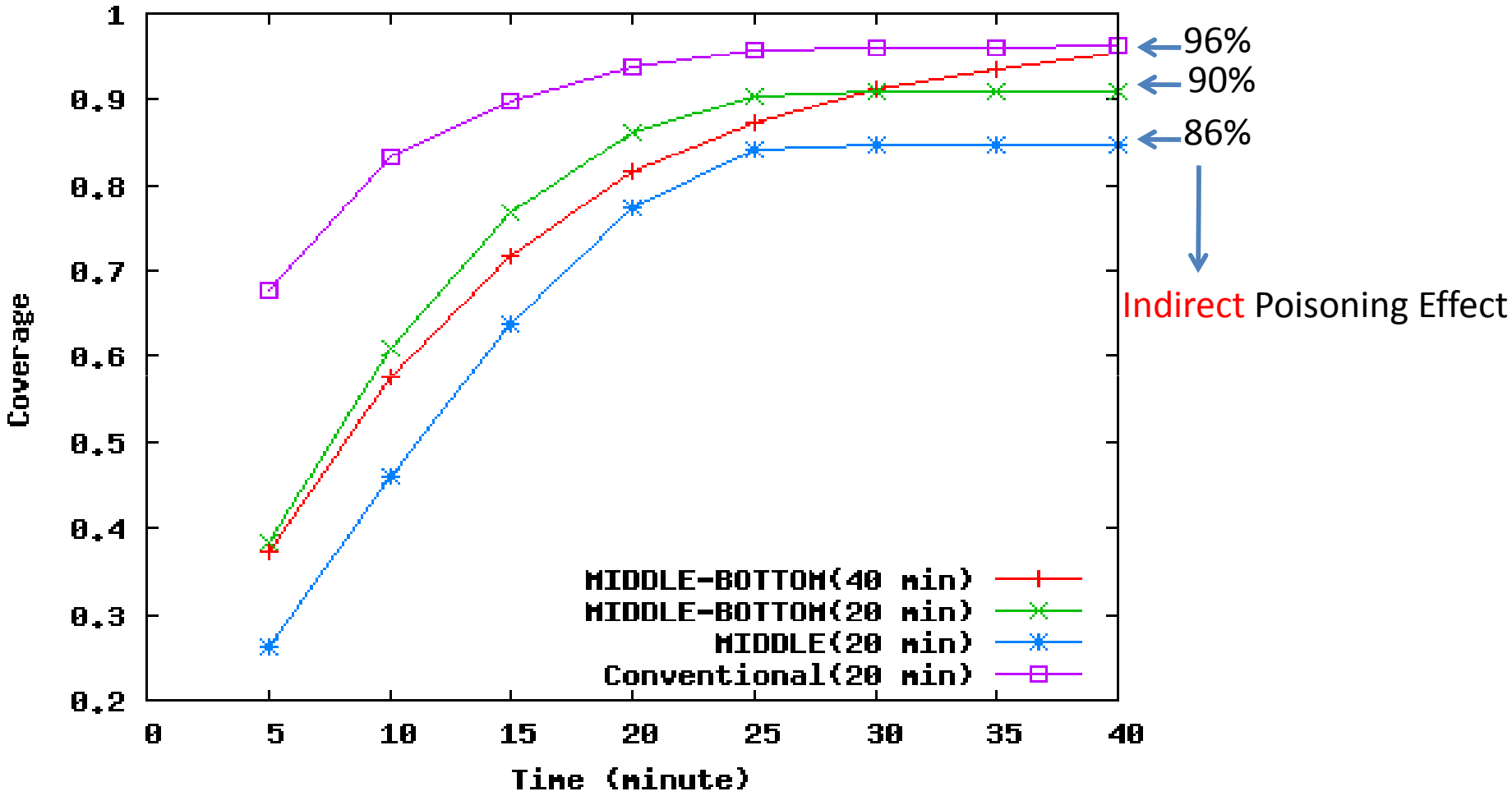
PART II:
Proposed Solution

1. Best Property: Network Structure
2. Influential Peer Selection Algorithm

PART III:
Evaluation

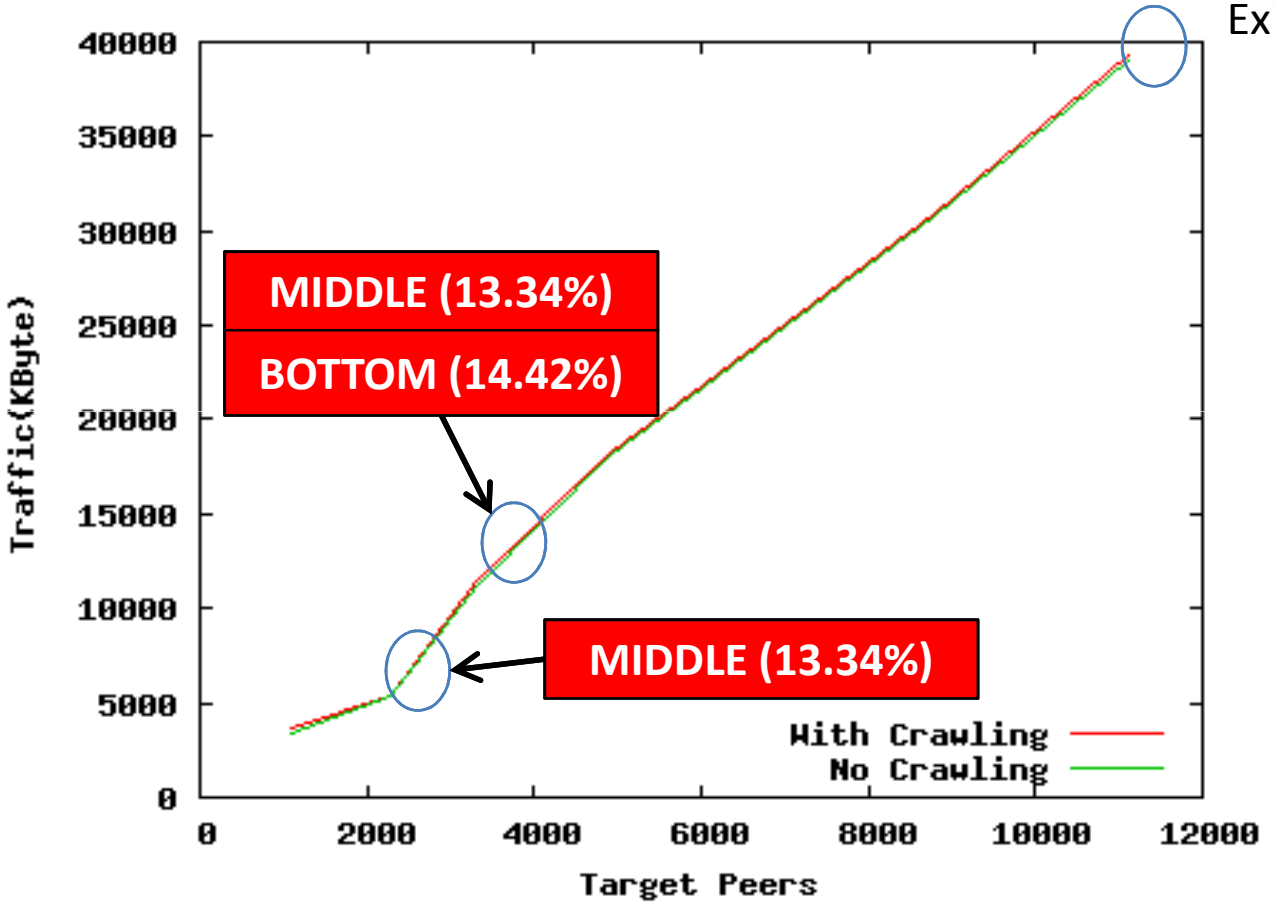
3. Evaluation Settings
4. Evaluation Result

Coverage (Indirect Poisoning Effect)



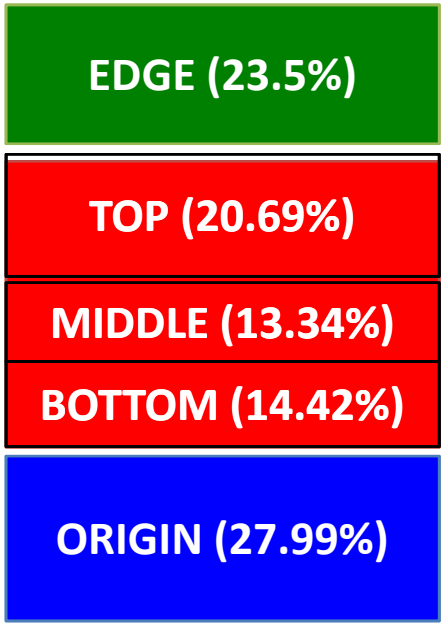
- Increasing Key Lifetime Gives more Time for Indirect Poisoning Effect to take place in the Network, so the MIDDLE + BOTTOM Poisoning can Achieve High Coverage as the Existing Work (96%)

Generated Traffic
(for Controlling Single File Distribution)



Existing Work

Network Clusters

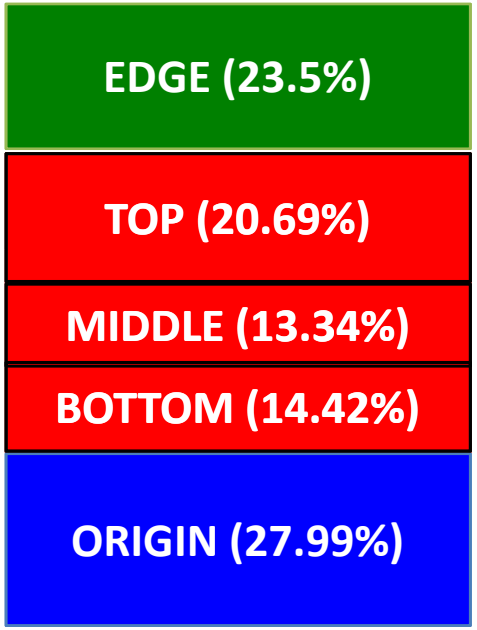


- With the same performance, Proposed Method Reduces Traffic to 13% (MIDDLE) and 27%(MIDDLE + BOTTOM)

Evaluation Summary

Method	Coverage	Control Traffic
All Peers (Existing Work)	96%	40 MB/File
MIDDLE	86%	7 MB/File
MIDDLE + BOTTOM	96%	15 MB/File

Network Clusters



Conclusion

- We propose and verify efficient poisoning method:
  - minimal control traffic (13—27%)
  - the same effectiveness as the existing work (96% Coverage)
- Reducing Control Traffic lifts the limitation in the number of controllable files in existing index poisoning
- Future Work:
  - Reduce more traffic to increase controllable file number
  - Apply proposed method for other control method
  - Apply proposed method for other P2P network

References

J. Liang, N Naoumov and K. W. Ross, “The Index Poisoning Attack in P2P File-Sharing Systems,” *Proc. IEEE Infocom 2006,* pp. 1—12, 2006.

X. Lou and K. Hwang, “Proactive Content Poisoning To Prevent Collusive Piracy in P2P File Sharing,” *IEEE Transactions on Computers TC 2007,* 2008.

P. Dhungel, X. Hei, K. W. Ross, and N. Saxena, “The Pollution Attack in P2P Live Video Streaming: Measurement Results and Defenses,” *Proc. P2P-TV 2007,* pp 323—328, 2007.

J. Kong, W. Cai and L. Wang, “The Evaluation of Index Poisoning in BItTorrent,” *Proc. IEEE ICCSN*, pp. 382—386, 2010.

M. Yoshida, S. Ohzahata, A. Nakao , and K. Kawashima, “Controlling File Distribution in Winny Networkthrough Index Poisoning, ”*Proceedings of the 23rd International Confer-ence on Information Networking*, pp. 210–214, 2009.

D. Stutzbach and R. Rejaie, “Capturing Accurate Snapshots of the Gnutella Network,” Global Internet Symposium, pp. 127-132, 2005.

J. Exposto, J. Macedo, A. Pina, A. Alves , and J. Rufino, “,Geographical Partition for Distributed Web Crawling, ” *Proceedings of the 2005 Workshop on Geographic Information Retrival,* pp. 55–60, 2005.

H.Nakazawa et al., “A method of Controlling Search Network in Winny Network, ” IEICE Tech. Rep., vol. 109, no. 362, IN2009-122, pp. 71–76,2010.

S. Brin and L. Page,“ The Anatomy of a Large-Scale Hypertextual Web Search Engine, ”In Proceedings of the Seventh International World Wide Web Conference, pp. 107–117, 1998.

C. Xie, S. Guo, R. Rejaie, and Y. Pan, “Examining graph properties ofunstructured peer-topeer overlay topology, ”Global Internet Symposium(GI ’07), pp.13–18, 2007.