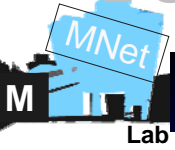


Detecting Peer-to-Peer Activity by Signaling Packet Counting



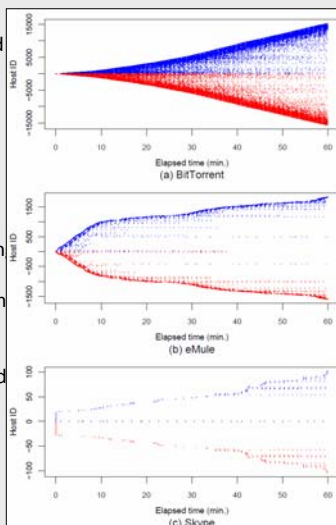
Chen-Chi Wu¹, Kuan-Ta Chen², Yu-Chun Chang¹, and Chin-Laung Lei¹
¹National Taiwan University ²Academia Sinica

Background

- Peer-to-peer traffic now constitutes a substantial volume of Internet traffic
- Identify P2P applications is essential for network traffic management
 - Service differentiation
 - Capacity planning
 - QoS provisioning
- Traditional approaches
 - Port numbers [1]
 - Specific signatures in packet payloads [2]
 - Transport layer information [3]

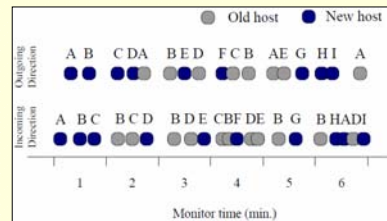
Motivation

- Problems in traditional approaches
 - Port numbers
 - Dynamic port numbers
 - Application layer signatures
 - Encrypted packet payloads
 - Proprietary protocols
 - Transport layer approach
 - These works do not pay attention on recognizing particular P2P applications associated with the monitored traffic
- Our approach recognizes P2P applications running on monitored hosts based on the **signaling behavior**
 - The signaling behavior between peers are fundamental and unlikely to be changed
 - Each application possesses distinctive features of signaling behavior
- Visual demonstration of signaling activity patterns
 - Host IDs: hosts that ever contacted with the monitored host
 - Each dot indicates a signaling packet



Methodology

- Keep track of the signaling packets sent from and received by the monitored host for a period of monitor time
- Characterize the signaling behavior of the monitored host on two levels
 - Host level: based on the number of new or old hosts
 - Message level: based on the number of new or old packets



- Derive features from signaling packet stream

| Host level |
|---|
| Ratio of new / old hosts (mean, sd)* |
| Growth rate of new / old hosts (mean, sd) |
| Correlation coefficient between the number of new and old hosts |

| Message level |
|---|
| Ratio of new / old packets (mean, sd) |
| Growth rate of new / old packets (mean, sd) |
| Correlation coefficient between the number of new and old packets |
| Alternate rate of new and old packets (mean, sd) |

*Standard deviation

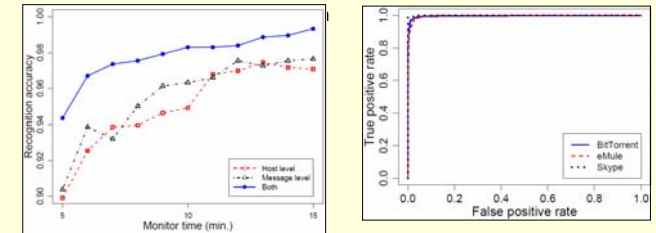
- Classifier design
 - Apply support vector machine (SVM)
- Training phase
 - Derive features for each training signaling stream
 - Label each training stream with the category of P2P applications
- Identification phase
 - Derive features for monitored signaling stream
 - Determine the P2P applications

Performance

- Use traces captured on hosts running BitTorrent, eMule or Skype
- Perform 10-fold cross validation on the recognition accuracy
 - The effect of different feature sets
 - The effect of the length of the monitor time

Evaluation Results

- Correctly recognize 97% of samples with only the message-level features within a monitor time of 15 minutes
- Achieve 99% accuracy with both feature sets
- The ROC curves evidence the high true positive rate and low false



Conclusion

- Recognize P2P applications running on monitored hosts based on the signaling behavior of applications
- Our approach is useful for three reasons
 - It does not access packet payload
 - Only the signaling traffic is required
 - It is able to identify particular P2P applications

Future Work

- Consider more P2P applications
- Identify multiple P2P applications simultaneously running on a host

- S. Sen, and J. Wang. "Analyzing peer-to-peer traffic across large networks," *IEEE/ACM Transactions on Networking*, 12(2):219-232, 2004
- S. Sen, O. Spatscheck, and D. Wang. "Accurate, scalable in-network identification of p2p traffic using application signatures," in *Proceedings of WWW'04*
- T. Karagiannis, A. Broido, M. Faloutsos, and K. Claffy. "Transport layer identification of p2p traffic," in *Proceedings of IMC'04*



Department of Electrical Engineering
National Taiwan University

Institute of Information Science
Academia Sinica

