What is the Impact of P2P Traffic on Anomaly Detection?

Irfan Ul Haq, Sardar Ali, Hassan Khan, Syed Ali Khayam

School of Electrical Engineering and Computer Science (SEECS),
National University of Sciences and Technology (NUST),
Islamabad, Pakistan

Pre-Stuff ☺

We recently received following Interesting reviews from a very prestigious **Computer and Communiction Security ☺** conference.

"While I am not familiar with these anomaly-detection methods, one would like to first see whether these methods can actually work before being convinced that a distributed correlation scheme based on them can improve the results. This is especially the case since anomaly-based IDS has been cast significant doubt as far back as ten years ago."

After a decade of P2P?

There are things we **know** that we **know**

There are things we **know** that we **don't know**

There are things we **don't know** that we **don't know**

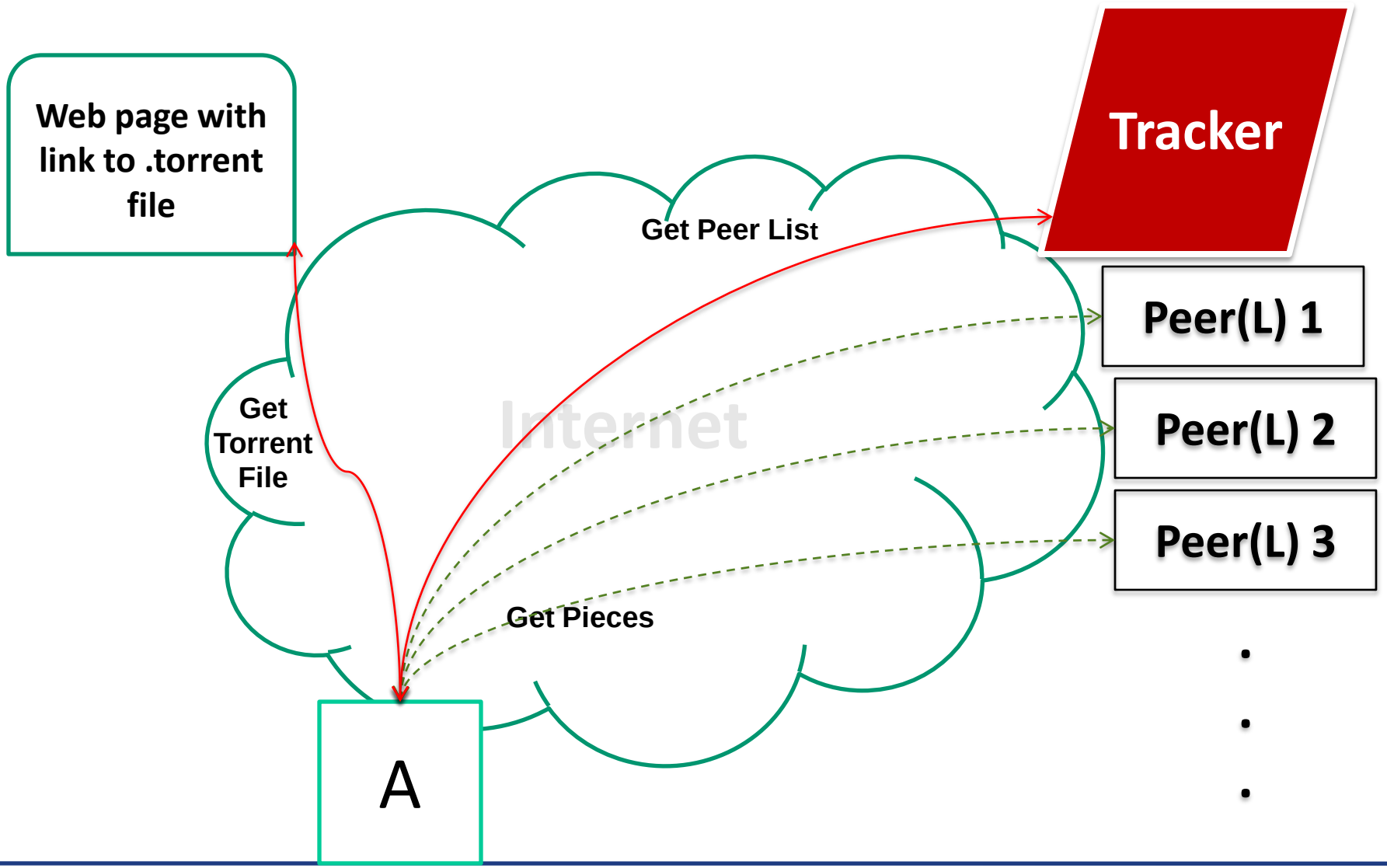
(Donald Rumsfeld)

Research Contribution

Research Contribution

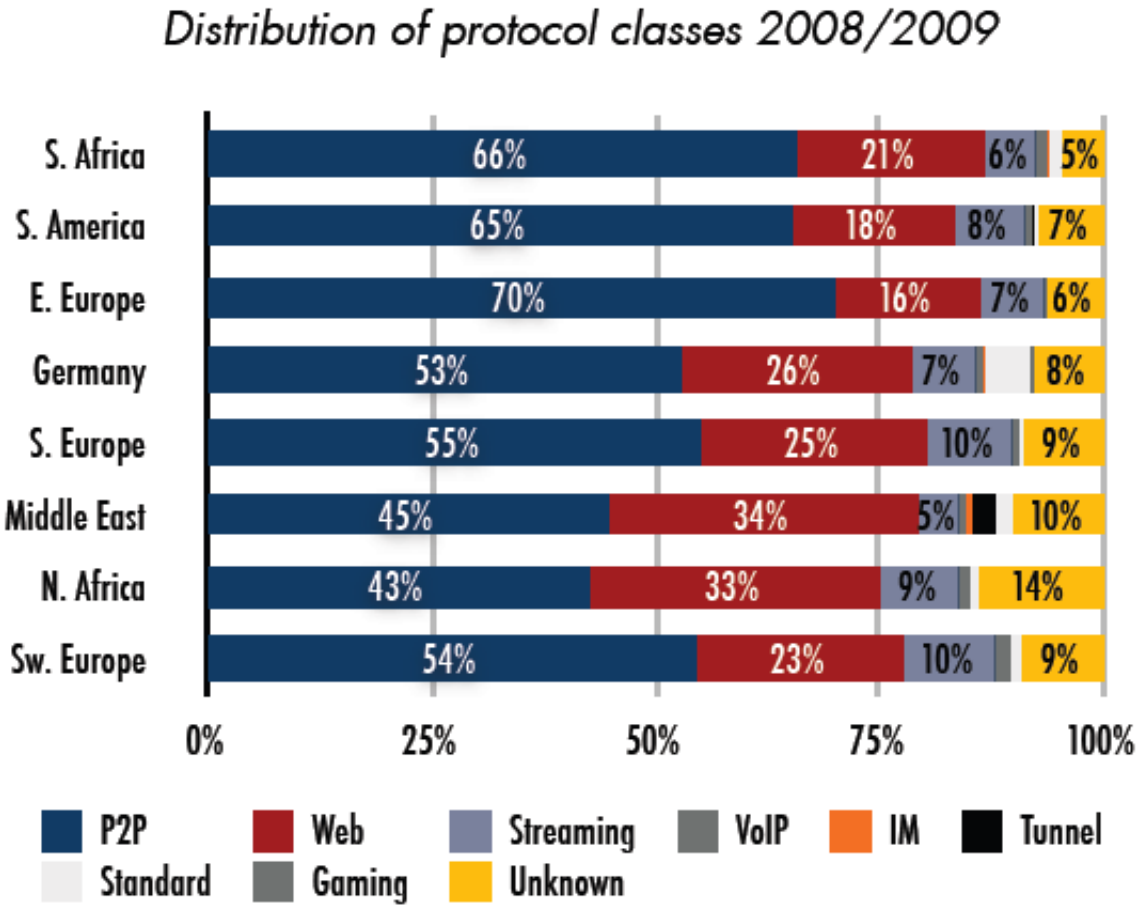
- To provide a first and base-line study on P2P and anomaly detection;
- To compare and evaluate existing anomaly detection techniques and principles under P2P traffic;
- To persuade research community to solve this challenging and worth solving problem;

Unpredictable and Unwanted



Unpredictable and Unwanted cont…

- **40% to 70%** of Internet traffic consists of p2p content*.

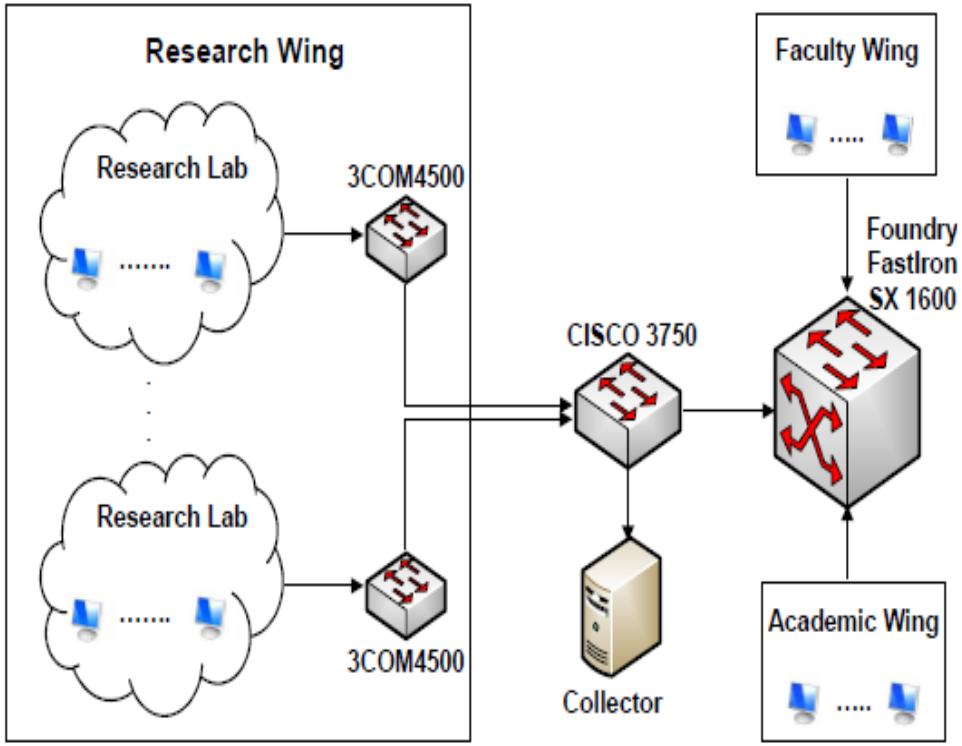


*Ipoque Internet Study Report, http://www.ipoque.com/resources/internet-studies/internet-study-2008_2009

Dataset

- Dataset

Client Name & Version	Session Establish	Traffic Volume
Vuze 4.0	20	685 MB
Flashget 1.9.6	62	60.7 MB
Utorrent 1.8.1	30	1.08 GB
Bit Torrent 6.1.2	134	1.59 GB
Deluge 1.0.7	30	171 MB
Bit Comet 1.0.7	20	57.4 MB
Halite 0.3.1	9	413 MB
eMule 0.49b	203	2.67 GB



Data set is available at http:\\wisnet.seecs.nust.edu.pk\

Evaluation Breakdown

- Accuracy Analysis
  - How much degradation does p2p traffic induce in anomaly detection accuracy (detection and false positive rates)?
  - Which anomaly detection metrics/principles are more sensitive to p2p traffic and why?
  - Does the aggressive nature of p2p traffic dominate some/all attack classes and high-low-rate attacks?

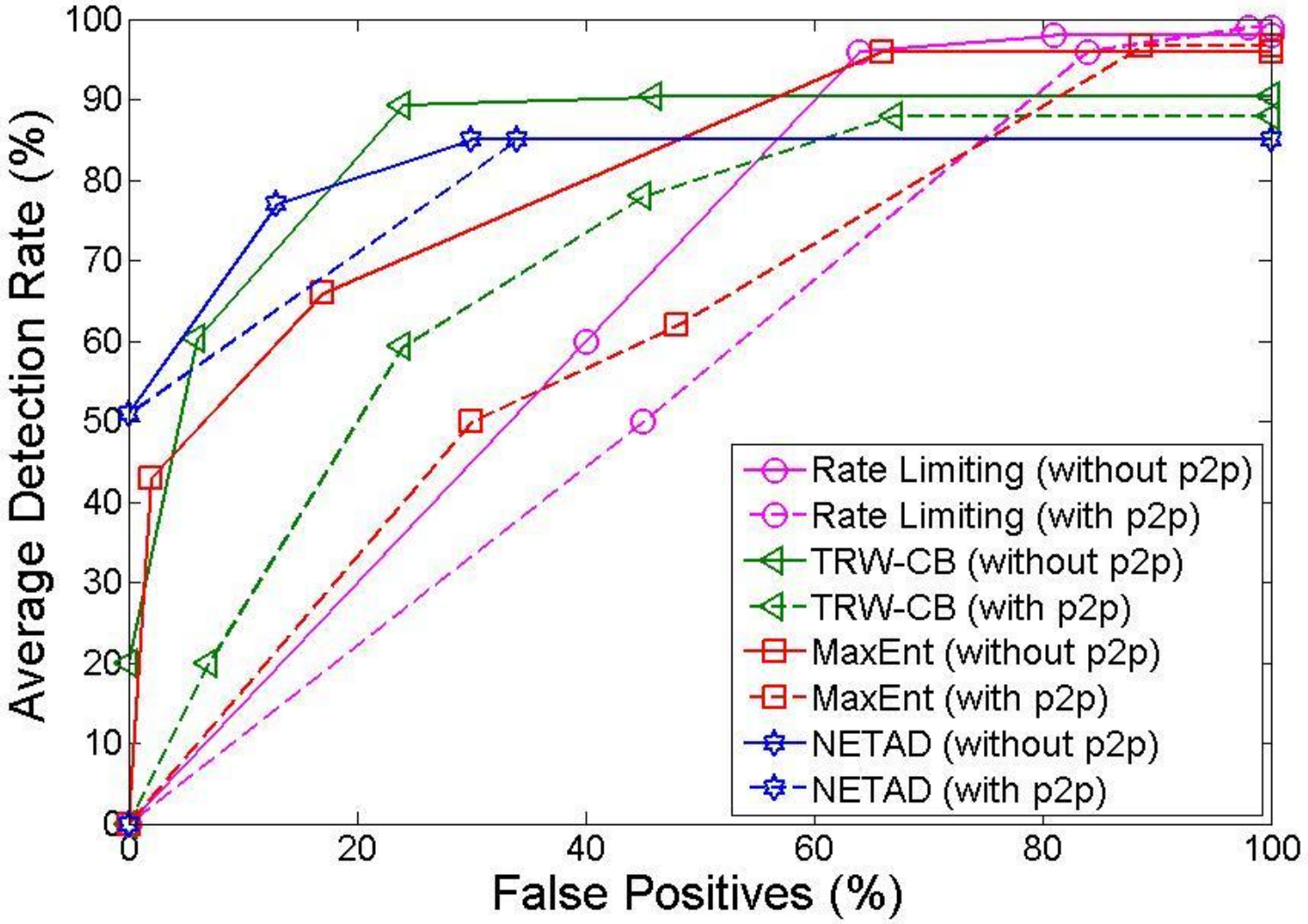
Evaluation Breakdown cont…

- Training Analysis
  - Can an anomaly detector handle p2p traffic if it is trained on a dataset containing p2p traffic?
- Mitigation Strategy
  - Can a pragmatic solution be designed to make an anomaly detector insensitive to the p2p traffic?
  - Can existing public p2p traffic filtering solutions mitigate the torrent effect?

Accuracy Analysis

- How much degradation does p2p traffic induce in anomaly detection accuracy (detection and false positive rates)?
- Which anomaly detection metrics/principles are more sensitive to p2p traffic and why?

How much degradation?

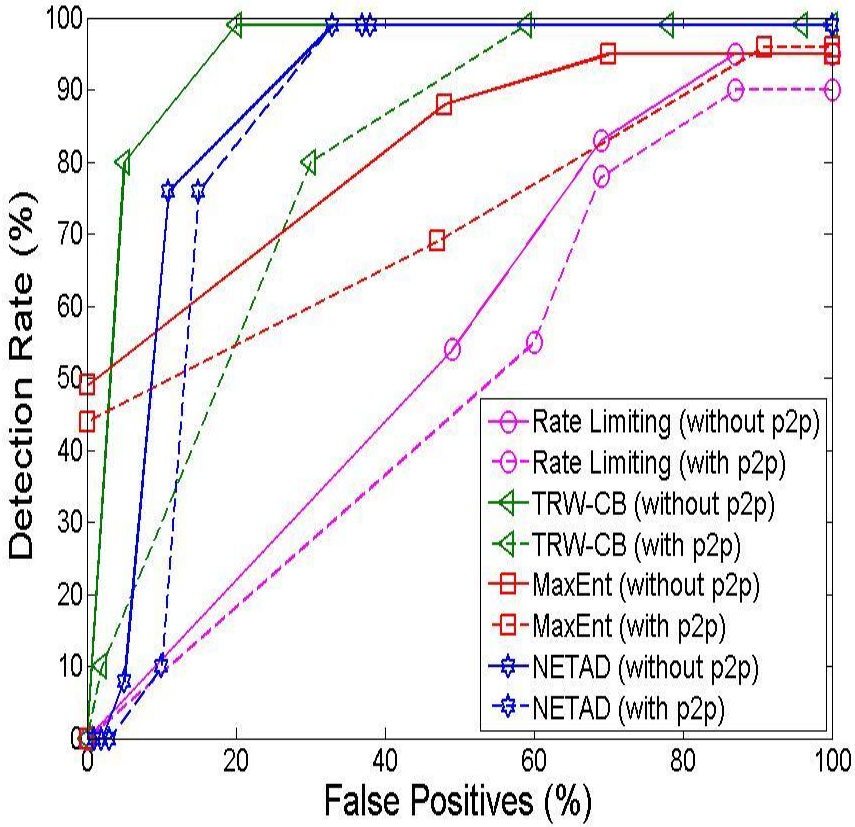


Accuracy Analysis cont…

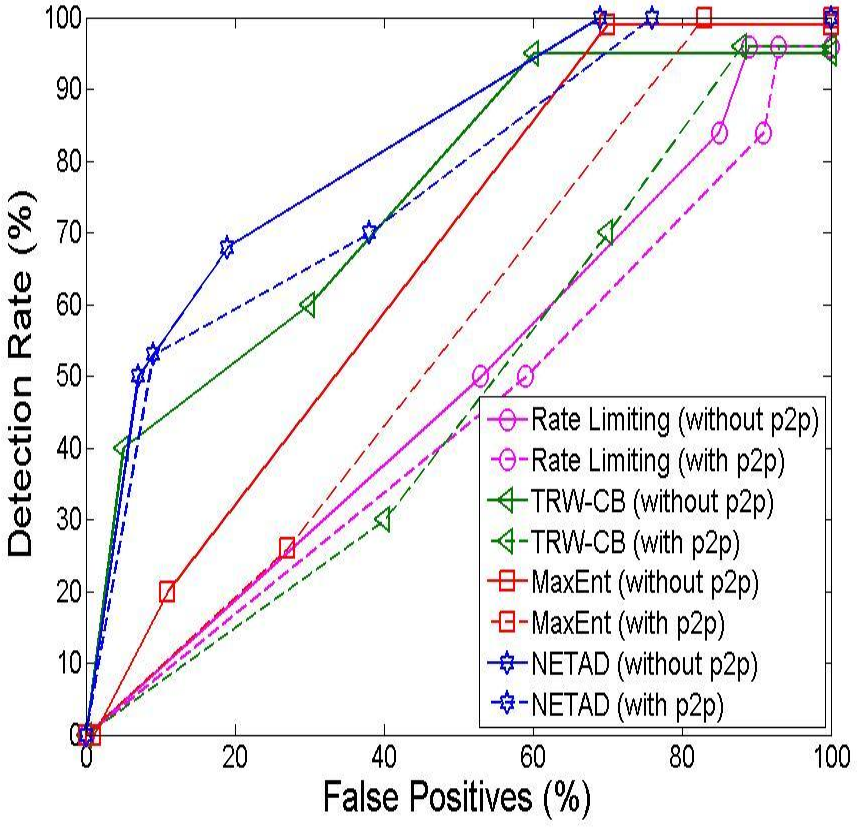
- Does the aggressive nature of p2p traffic dominate some/all attack classes and high-low-rate attacks?

Attack type, high- and low-rate impact?

Attack Type : Portscan



High-rate



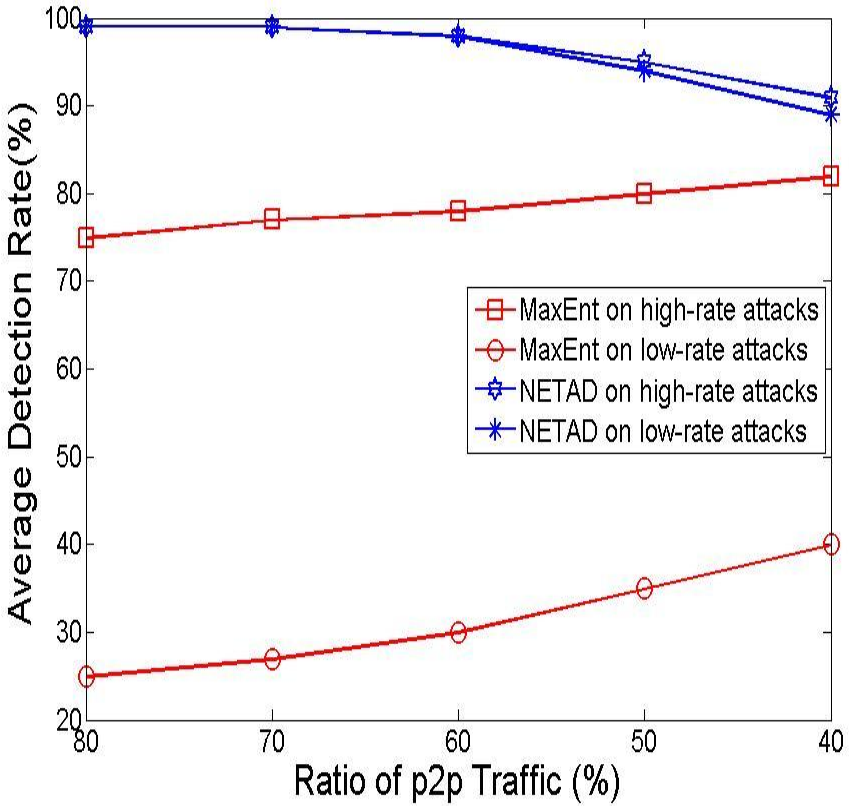
Low-rate

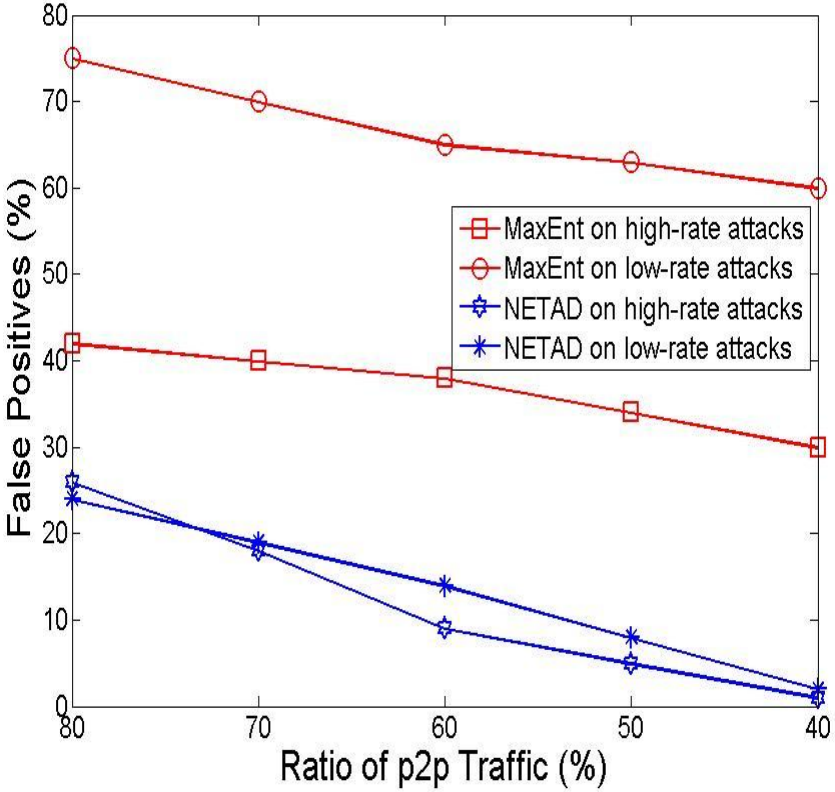
Analysis of Training Impact

- Can an anomaly detector handle p2p traffic if it is trained on a dataset containing p2p traffic?

Impact of training?

Testing : With Torrent

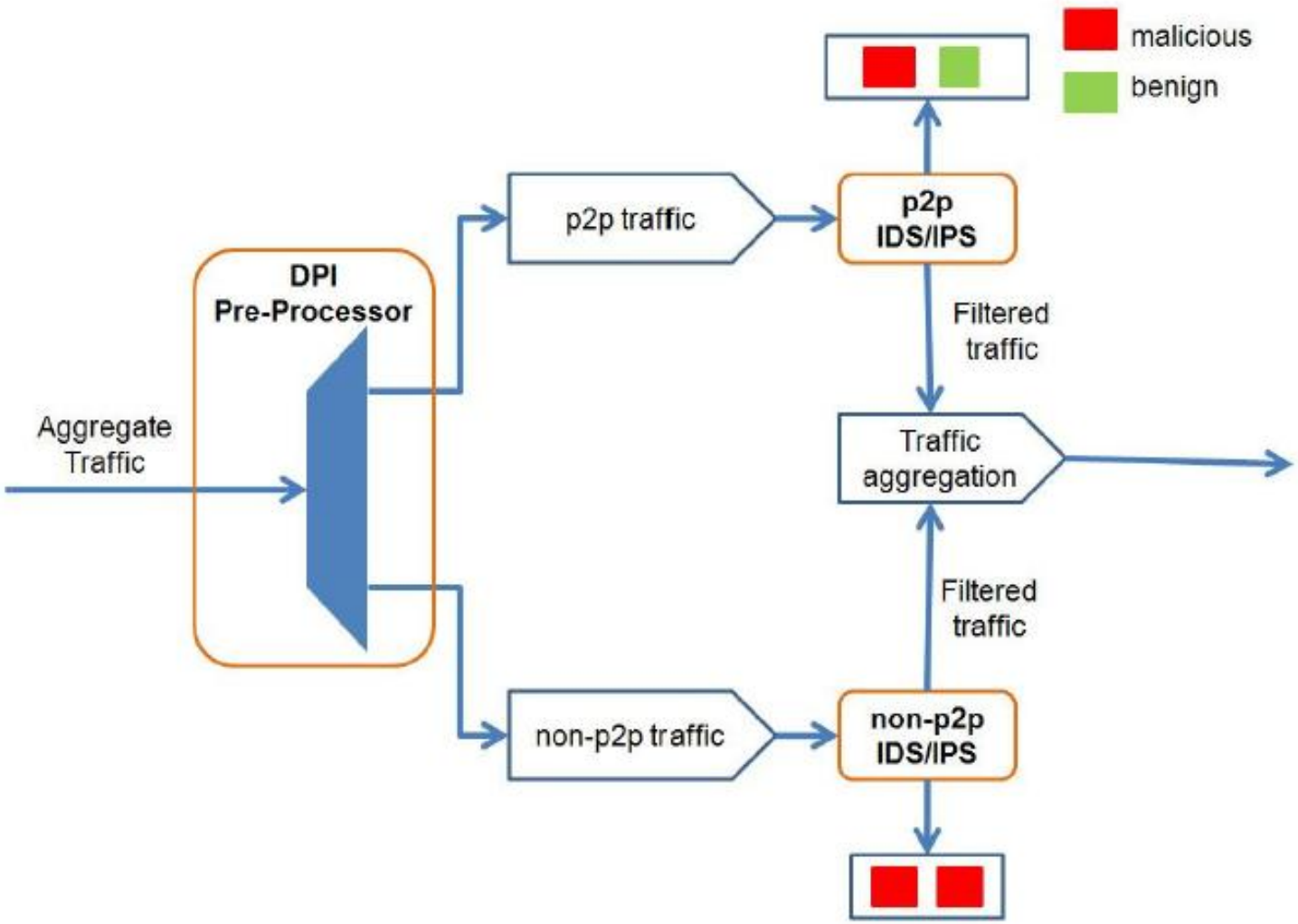




Mitigation Strategy

- Can a pragmatic solution be designed to make an anomaly detector insensitive to the p2p traffic?
- Can existing public p2p traffic filtering solutions mitigate the torrent effect?

Mitigation Strategy



Mitigation Strategy *cont…*

Table 1. Evaluation of OpenDPI and KPC on Encrypted P2P Traffic

	Classified as P2P	Classified as Unknown	Classified as non-p2p
OpenDPI	3.80%	96.20%	0%
KPC	64.70%	25.30%	0%

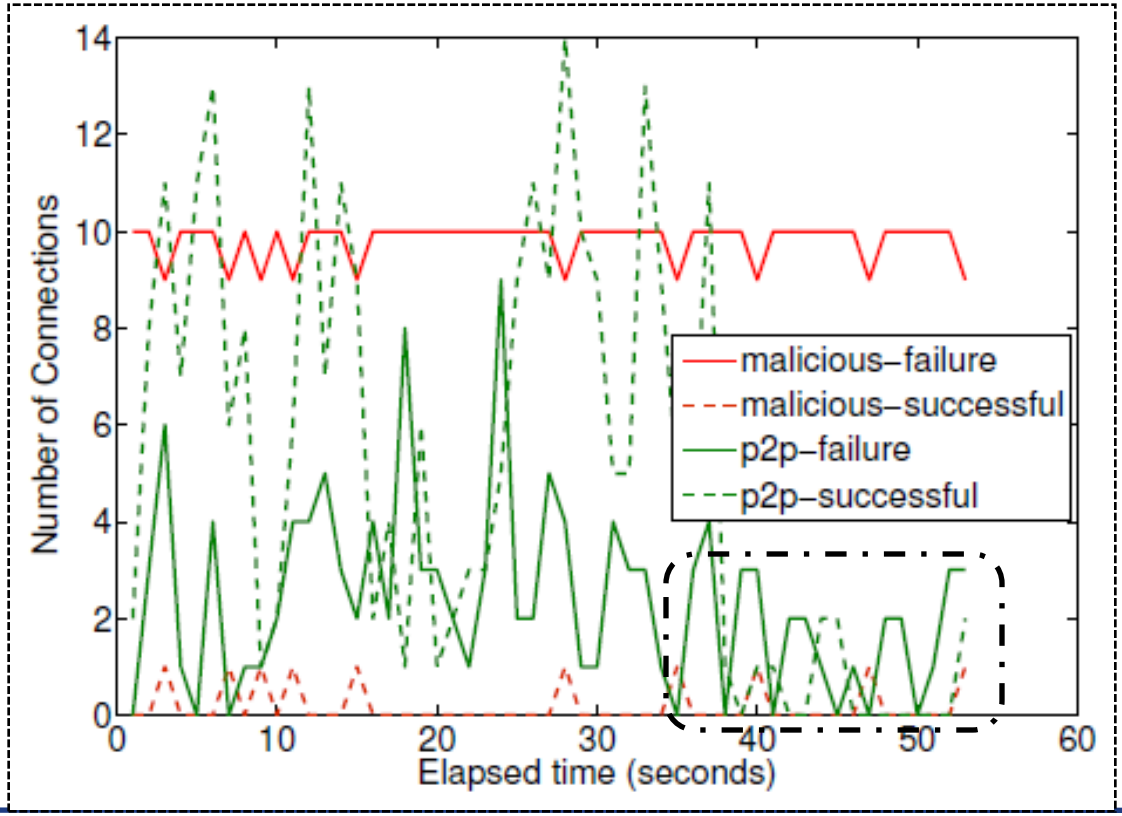
Table2. Mitigating P2P Effect Using P2P Traffic Classifiers Based Traffic Filtering (DR= Detection Rate; FP= False Positive; KPC= Karagiannis' Payload Classifier)

	Rate Limiting		TRW-CB		MaxEnt		NETAD	
	DR%	FP%	DR%	FP%	DR%	FP%	DR%	FP%
No Filtering	50	45	60	22	62	48	65	25
OpenDPI	56	43	64	12	63	32	70	17
KPC	60	40	70	6	66	17	77	13

Future Work

Future Work

- Development of an ADS that works under p2p traffic.
- Impact of traffic splitting based on the application layer protocols on detection.
- Mis-configured P2P traffic and AD?



Questions