IU-ATC India-UK Advanced Technology Centre of Excellence in Next Generation Networks, Systems and Services

Online Malware Detection in the Cloud

MSN 2014, Coseners House

Angelos K. Marnerides *(ULancs)*,
Michael Watson (ULancs),
Andreas Mauthe (ULancs),
Fadi El-Moussa (BT)

{a.marnerides,m.watson,a.mauthe}@lancs.ac.uk

School of Computing and Communications

Problem description

- Malware is in the majority of cases the triggering point for large-scale attacks.
- Cloud operators & service providers heavily rely on traditional IDS signature-based schemes.
- Essence for online VM sanitization under a statistical approach that overcomes the dependence on pre-known signatures.

Approach

- Measurement-based novelty detector next to the hypervisor.
- Training, prediction & evaluation of system (VMI) and network-related features as gathered by VMs (hypervisor level).
- Scenario: static real-time VM, service-specific sanitization.
- Malware samples: Kelihos , Zeus (+ variants)

One-class SVM

- Supervised novelty detection scheme.
- Special case of traditional 2-class SVMs.
- Accuracy performance depends heavily on the *nu* and *gamma* parameters.
- We use the Gaussian RBF kernel.

Method: Confusion Matrix & Detection Performance Metrics

Confusion Matrix

- Labels: 1 = normal , -1=anomalous
- The detector could be right or wrong
- 4 possible outcomes

		Detector Output	
		1	-1
Expected Output	1	TN	FP
	-1	FN	TP

Metrics

$$Accuracy = \frac{TP + TN}{TP + TN + FP + FN}$$

$$\Pr ecision = \frac{TP}{TP + FP}$$

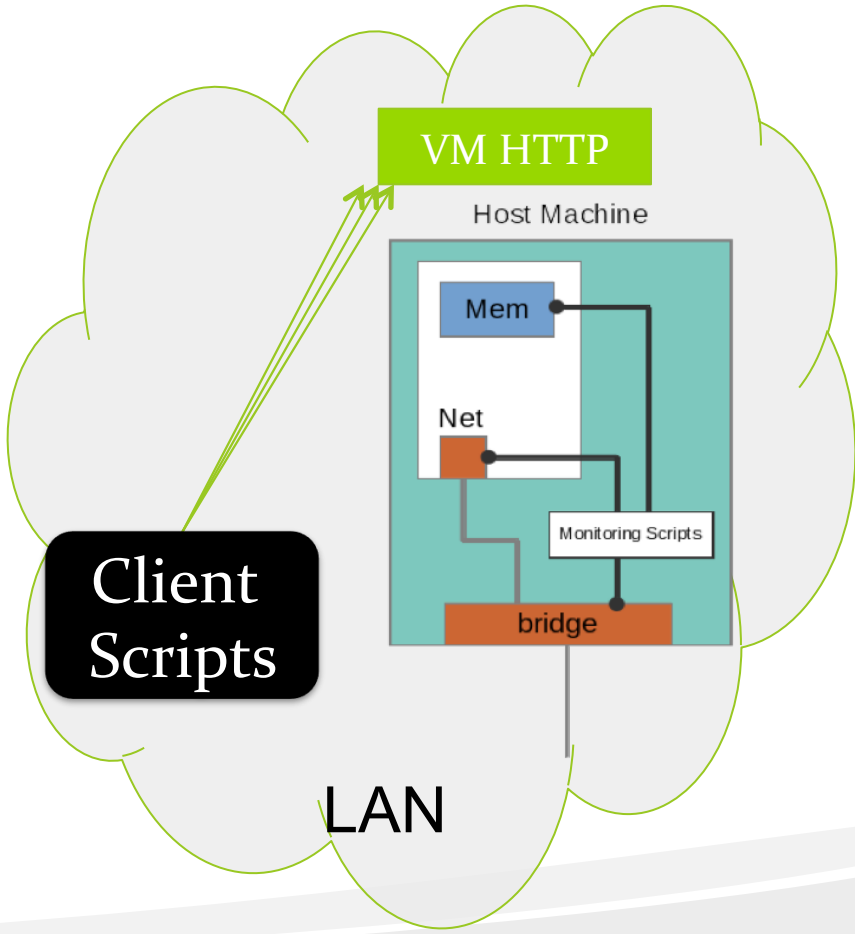
$$\mathrm{Re}call = \frac{TP}{TP + FN}$$

$$F - score = 2 \times \left(\frac{\Pr ecision \times \mathrm{Re}call}{\Pr ecision + \mathrm{Re}call} \right)$$

$$G - mean = \sqrt{\Pr ecision \times \mathrm{Re}call}$$

Experimental Setup

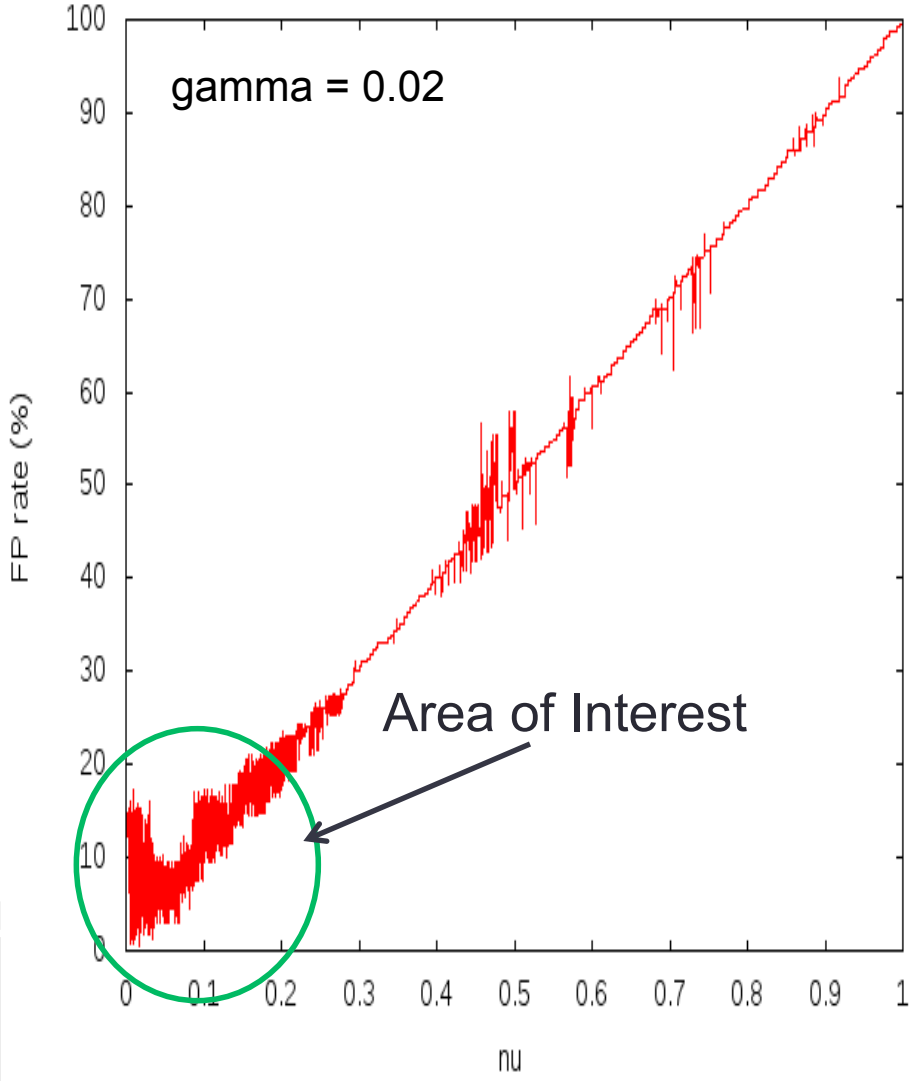
- Host : Xen 4.1 with the XAPI toolstack, Dom0 OS: Ubuntu 12.10, VM OS : Windows XP (SP3) running an HTTP server
- Custom clients sending randomly-sized requests to the server & user activity on the VMs.
- Kelihos & Zeus variants injection at some point in time
- 21 features on per-process and per-unidirectional flow (3 second snapshots)



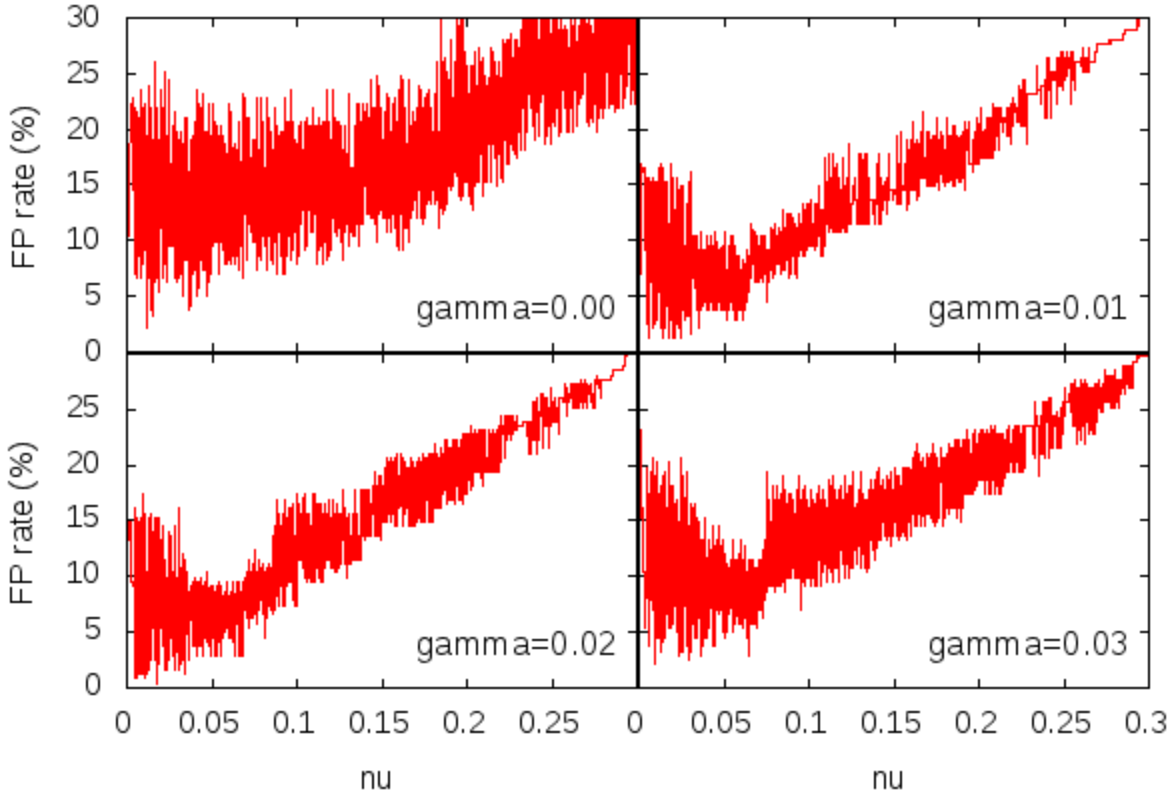
SVM Parameter selection vs. FPR

- Iterative creation of classifiers with different parameters using the same training dataset.

- False Positive Rate (FPR) computation and search for the optimal values for *nu* and *gamma.*

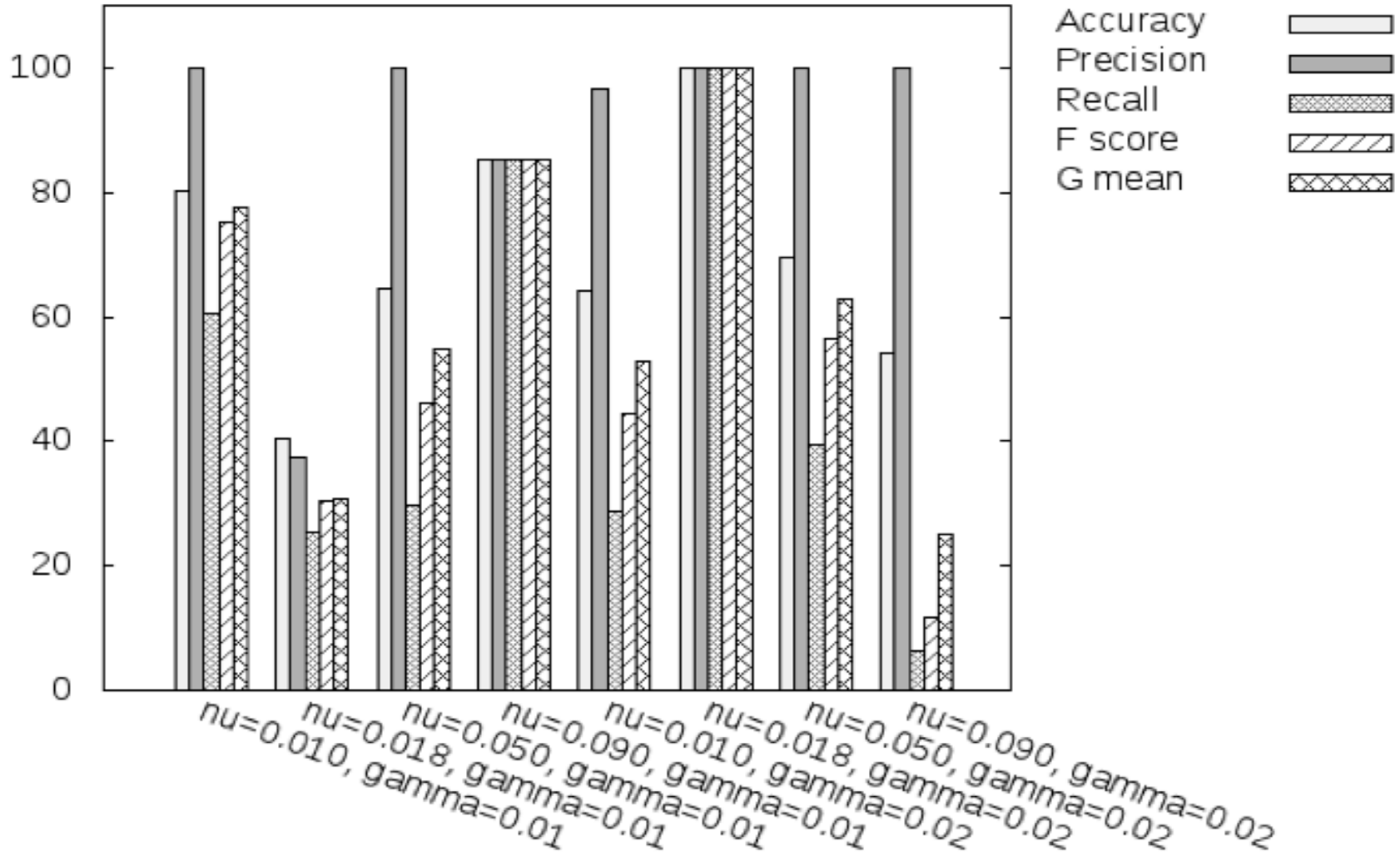


SVM Parameter search & selection (cont..)

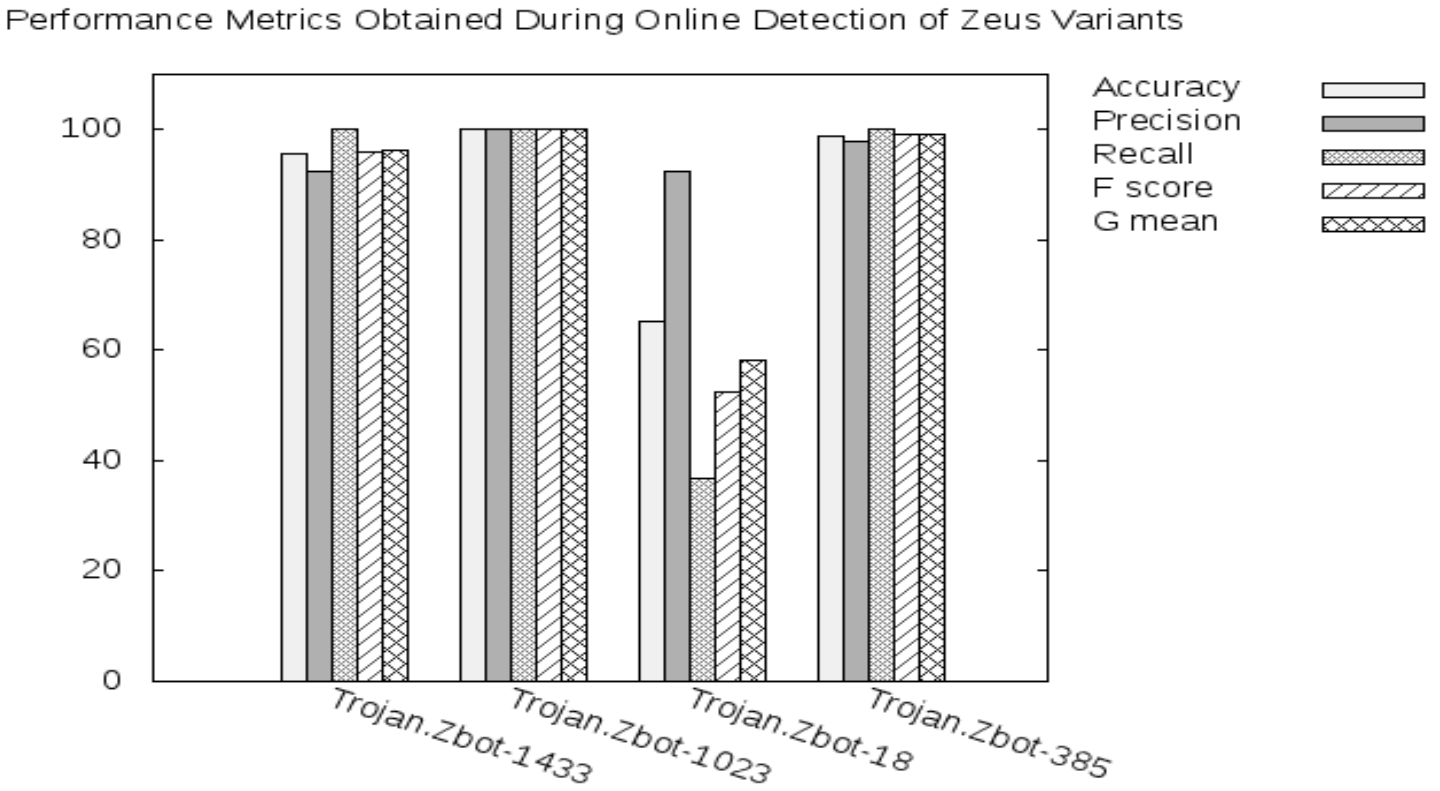


- Each iteration follows a roughly linear relationship between nu and FPR for a given gamma.
- The lowest FPR is at gamma = 0.02, nu = 0.018

Kelihos detection under various SVM parameters

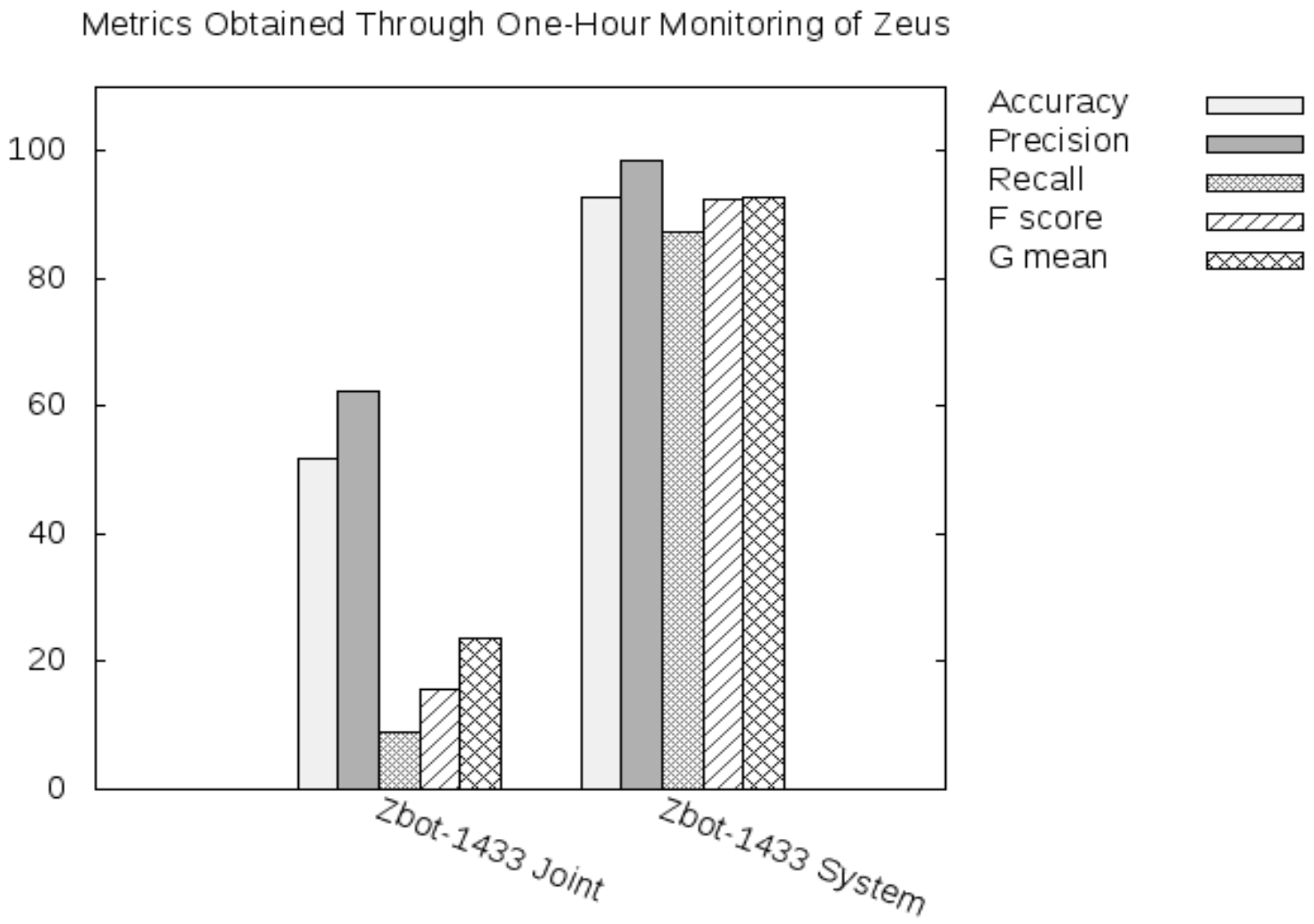


Online detection of Zeus variants



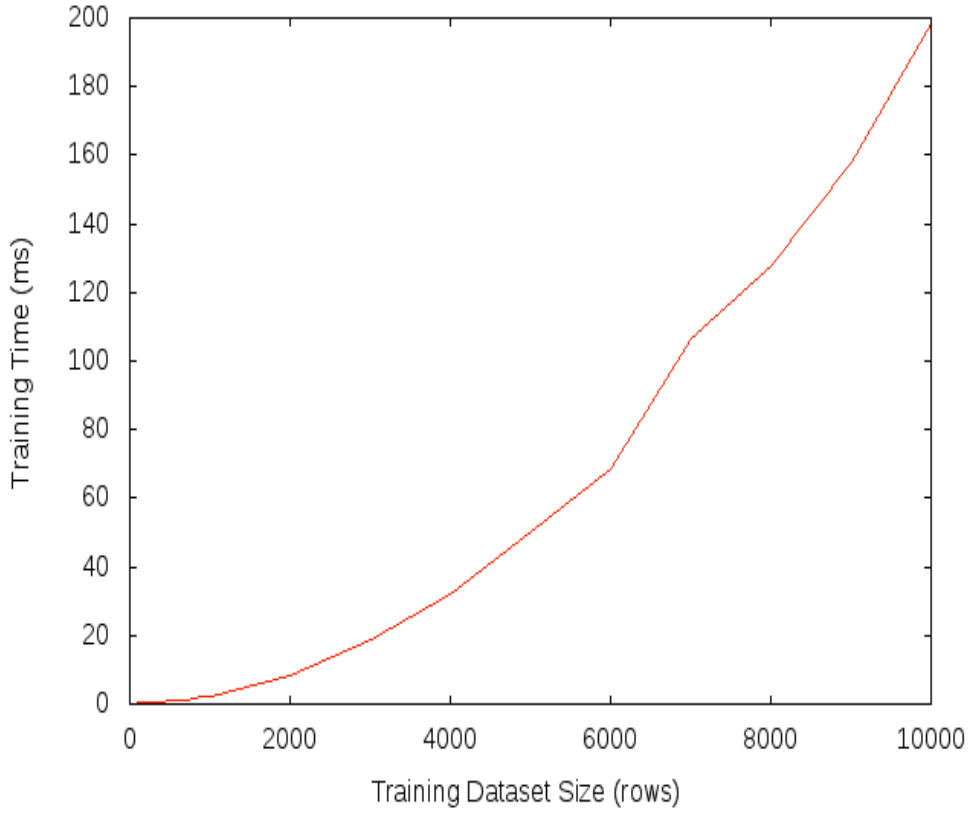
- Unlike signature-based systems we can detect malware variants with no prior experience of the variant.

Network vs. System Features

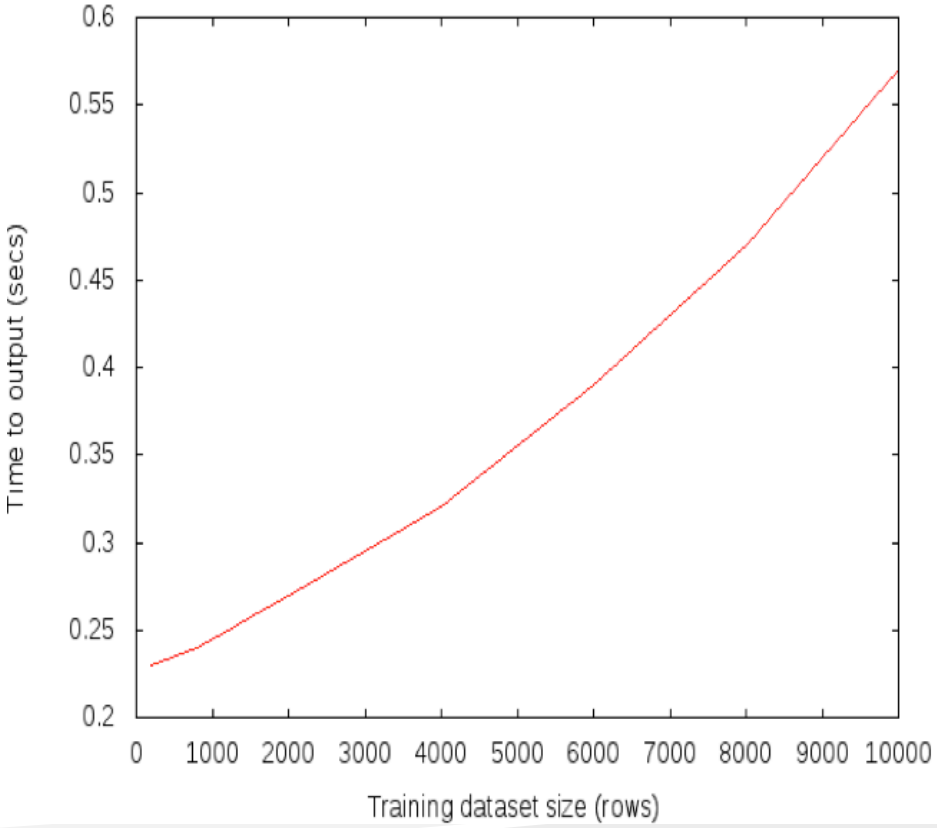


Training & Prediction time cost

Training (offline)



Prediction (online)



1 row = 3 seconds measurement sample

Conclusions / On-going work

- Average > 95% of overall detection accuracy for various malware types with reasonably low prediction time.
- *On-going work:*

1. Feature selection and anomaly clustering scheme.
2. Parallelized architecture.
3. Testbed/experiments expansion (new types of services, more physical nodes etc..)