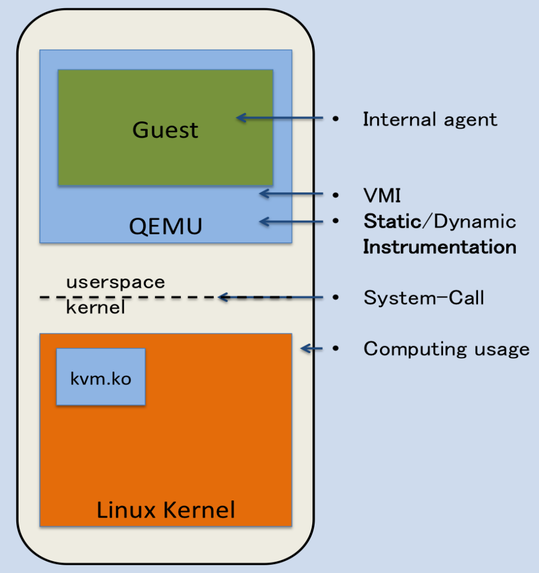
Leveraging static probe instrumentation data for VMM-based Anomaly Detection System

Background & Motivation

Aside from its rapid development and decreasing cost, cloud computing has not been fully embraced by organizations and industries around the world because of their concern over security. One main security threat in virtualization environments of cloud computing is the guest VMs. The challenge for monitoring public IaaS is how to collect a clear view of the guest VM's behavior without interfering guest OS's operation.
In this research, we introduce a novel observation point, static instrumentation data, and study its applicability for VM-based Anomaly Detection System.



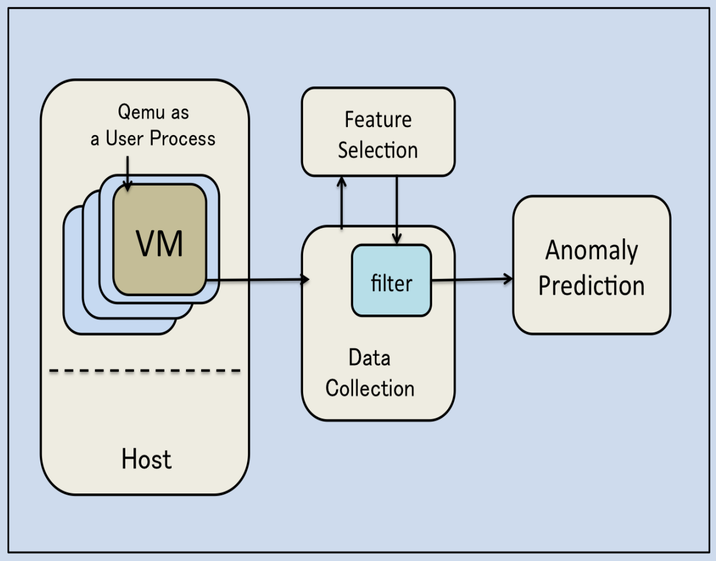
Framework

DATA COLLECTION
We add multiple trace-points inside the VMM. Every time a function within VMM get called, information about the trace-point related to that function is written into a log file. Information in the log file will be used to detect whether a certain VM behavior is normal or not. We use "ust" (user space tracer) backend from LTTng user space tracing (LTTng-UST) library.

FEATURE EXTRACTION
To minimize noise and computation requirement, we reduce the number of trace-points to be collected using Linear Discrimination Analysis. After applying LDA to our Normal dataset, we have five trace-points:

* tap_send
* bdrv_aio_flush
* memory_region_ops_write
* qemu_deliver_packet
* virtqueue_fill

ANOMALY DETECTION
We use semi-supervised learning approach and apply One-Class Support Vector Machine (OC-SVM) as prediction engine.

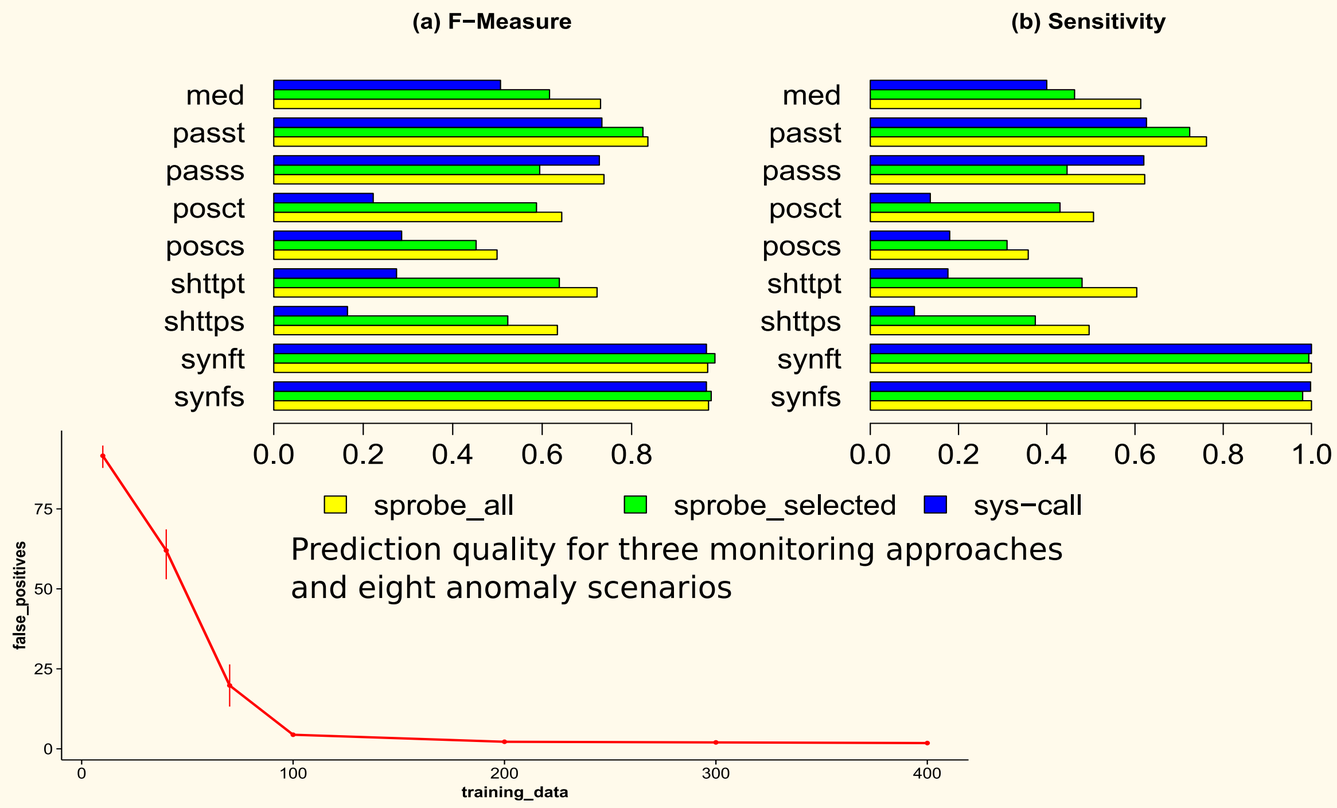


Evaluation

SETUP

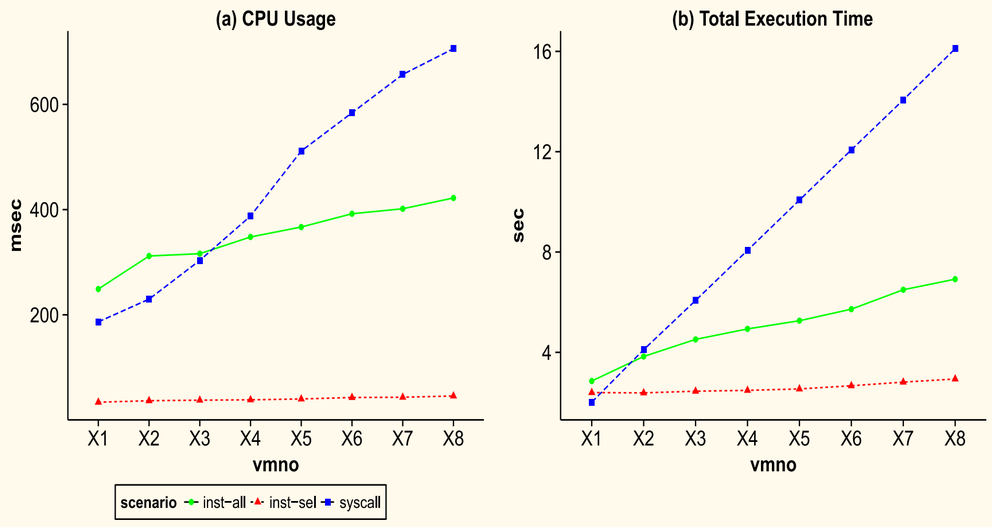
Physical	One host with eight VM. All nodes using Ubuntu 14.04 LTS. VMM is Qemu-KVM	
Normal	An online auction site emulated using RUBiS.	
Anomalous (source/target)	• Synchronous Packet Flood attack • SlowHTTP attack	• Portscan attack • Password Brute-Force attack
Data Collection	• *Static probe using LTTng (Linux Trace Toolkit: next generation) tool* • System-call using 'strace' tool	
Anomaly Prediction	OC-SVM modules from Scikit-Learn library	

RESULTS

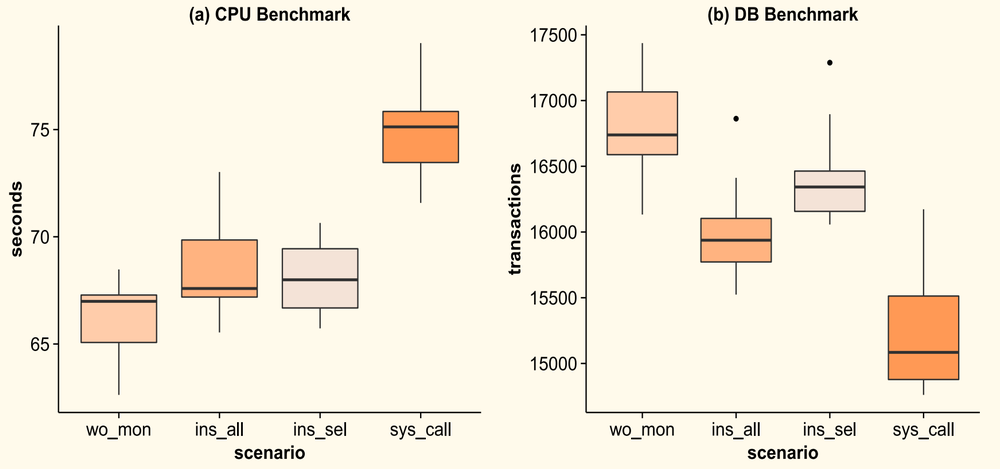


Prediction quality for three monitoring approaches and eight anomaly scenarios

Percentage of false positives as the number of training data increased



Performance statistics of a process to capture one unit data as the number of VM increased



The impact of monitoring process for VM's performance

Conclusion

Overall effectiveness and efficiency of the VMM-based ADS using static probe instrumentation data is better than VMM-based ADS using system-call data.

ady.paundu.ak9@is.naist.jp