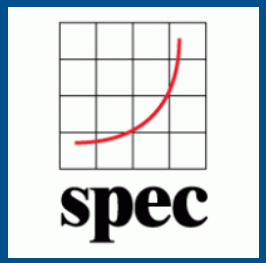
Julius-Maximilians-
UNIVERSITÄT
WÜRZBURG

Evaluation of Intrusion Detection Systems in Virtualized Environments Using Attack Injection

Aleksandar Milenkoski, Bryan D. Payne, Nuno Antunes, Marco Vieira, Samuel Kounev, Alberto Avritzer, Matthias Luft

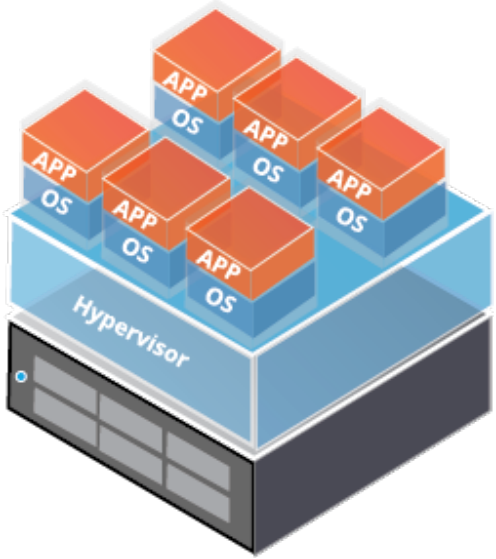


Outline

- Introduction
- Approach
  - Attack injection
- Case study
- Summary and future work

Introduction

- Virtualization
  - Hypervisor & VMs

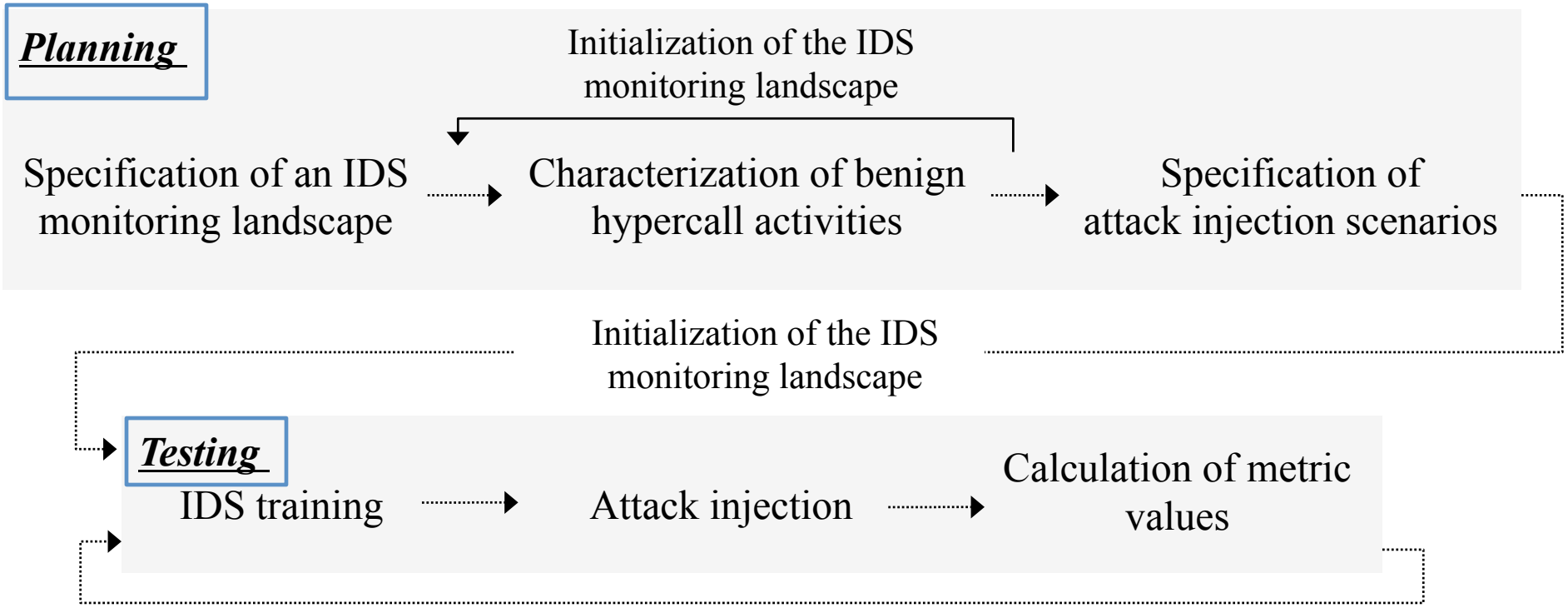


- Hypercalls
  - Identical to system calls
  - Critical attack surface of hypervisors [Rutkowska, J., Wojtczuk, R. @ BlackHat USA 2008]
- IDSes [*hypercall IDSes*]
  - Examples: Collabra, Xenini, CC Detector, OSSEC, ...
  - Components in the hypervisor, anomaly-based

Introduction (cont.)

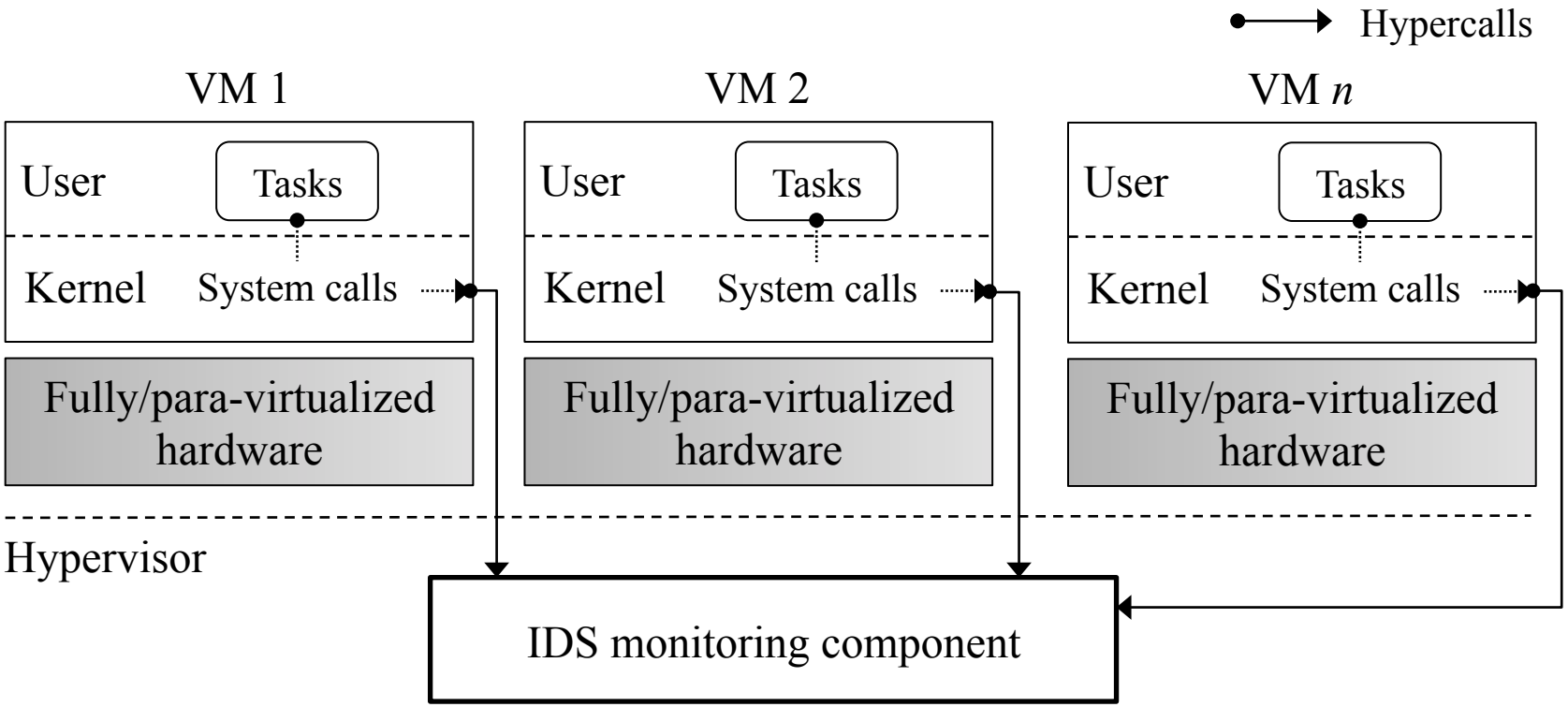
- How do we extensively evaluate the accuracy of hypercall IDSes?
  - There are no workloads: no traces, attack scripts targeting hypercall (hypervisor) weaknesses are extremely rare
- An approach for generating IDS evaluation workloads
  - Injection of malicious hypercall activities (e.g., attacks, covert channel operations) during regular operation of VMs
  - Live testing of hypercall IDSes
- Showcase of the feasibility of injecting (hypercall) attacks in virtualized environments for the purpose of evaluating IDSes

Approach


Planning
Initialization of the IDS monitoring landscape
Specification of an IDS monitoring landscape
Characterization of benign hypercall activities
Specification of attack injection scenarios
Initialization of the IDS monitoring landscape
Testing
IDS training
Attack injection
Calculation of metric values

Approach: Planning

- **Step 1**: Specification of an IDS monitoring landscape
  - System architecture and hardware
  - Virtualization mode
  - Workloads [repeatable]



Approach: Planning (cont.)

- **Step 2**: Characterization of benign hypercall activities
  - *Step 2.1.* Initialization of the IDS monitoring landscape
  - *Step 2.2*. Estimate time t_s of steady-state of the *detection-relevant property (DRP)*
    - Growth curve and target σ over time t
      - Example: 15 new hypercall sequences over 100 secs.
  - *Step 2.3.* Calculation of relevant statistics given time t_{max}
    - Number of occurences of each DRP variation [while progressing towards steady-state]
    - Average rate of occurrence of the DRP [after steady-state has been reached]

Approach: Planning (cont.)

- **Step 3**: Specification of attack injection scenarios
  - *Step 2.1.* Attack contents
    - Example: Specification of malicious hypercall sequences
    - Specification of „regular“ or IDS evasive „mimicry“ attacks
  - *Step 2.2.* Attack injection times
    - Example: Specification of IDS evasive „smoke-screen“ attacks
  - *Step 2.3.* Re-initialize the IDS monitoring landscape

Are hypercall activities repeatable?

Approach: Testing

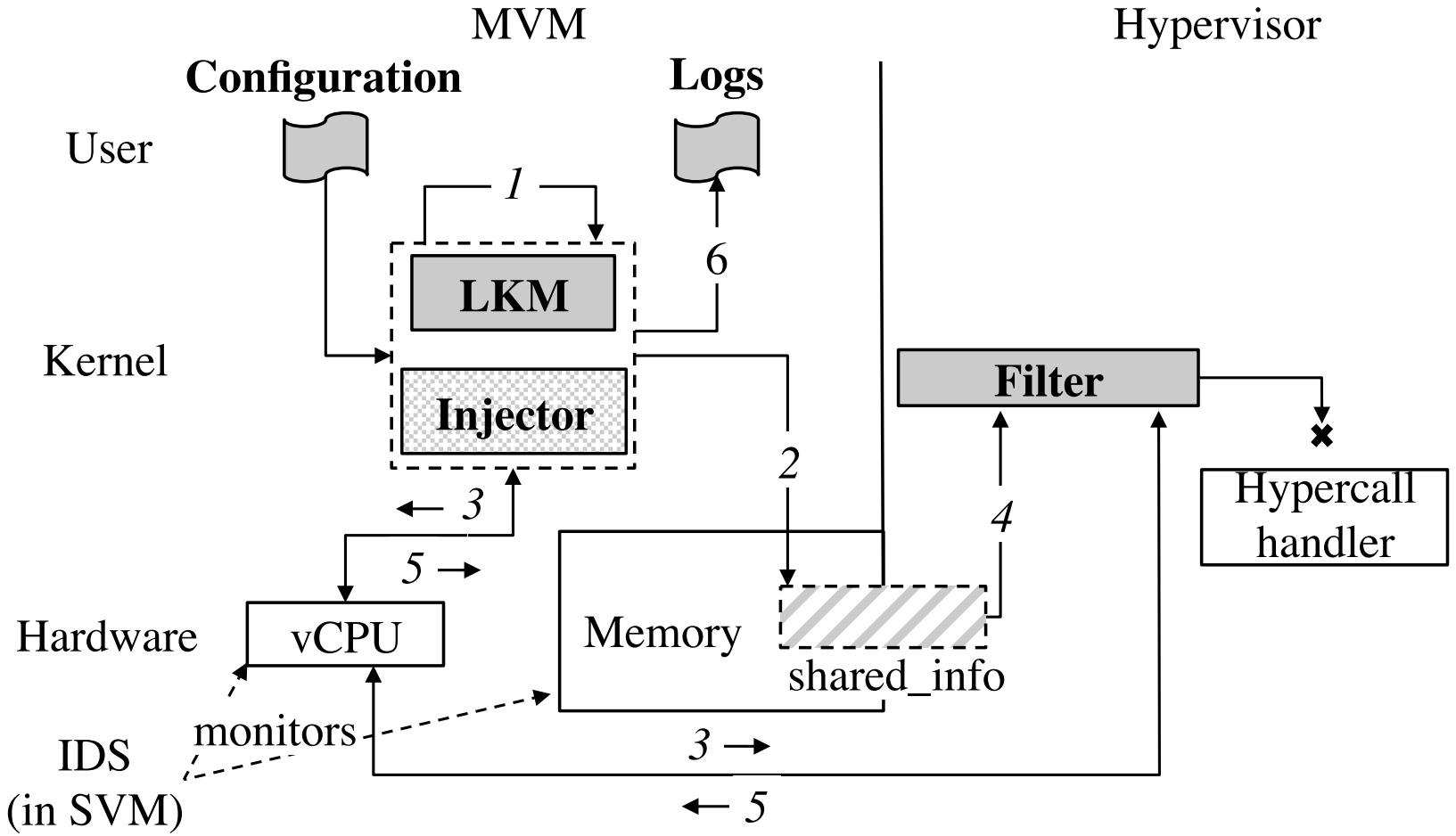
- **Step 1**: IDS training until time t_s
- **Step 2**: *Attack injection* until time t_{max}
- **Step 3**: Calculation of metric values

Until statistically accurate metric values are calculated

Are hypercall activities repeatable?

Attack Injection

- hInjector
  - Publicly available at https://github.com/hinj/hInj



Attack Injection (cont.)

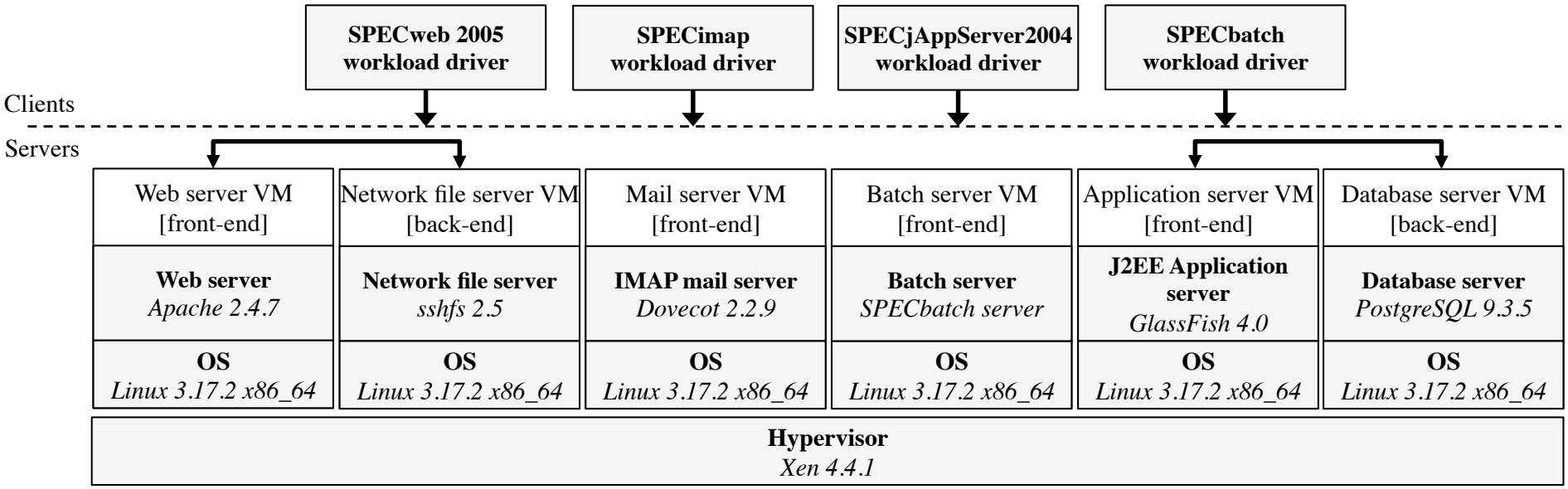
- Design criteria for ***realistic*** and ***practically feasible*** IDS evaluation
  - ***Injection of realistic attacks*** [35 PoCs, new attacks can be easily configured]
  - ***Injection during regular system operation***
  - ***Non-disruptive attack injection***
  - ***Low performance overhead***

Case Study

- IDS under test: Xenini [Maeiro et al. 2011]
  - DRP: Sequence of hypercalls of length *n* [n=10]
  - Calculates anomaly scores between 0 and 1 and fires an alert if a given threshold *th* is exceeded
- Scenarios
  - [Scenario #1] Evaluate the attack detection accuracy of Xenini for th in [0.1; 0.5]
  - [Scenario #2] Evaluate Xenini‘s ability to detect IDS evasive attacks --- „mimicry“ and „smoke screen“ attacks

Case Study: Planning

- **Step 1**: Specification of an IDS monitoring landscape
- The SPEC_sc2013 environment



Case Study: Planning (cont.)

- **Step 2**: Characterization of benign hypercall activities

	Run 1 [t=100 sec.; σ=15]		Run 2 [t=100 sec.; σ=15]	
Server VM	t_s *(sec.)*	*r (occ./sec.)*	t_s *(sec.)*	*r (occ./sec.)*
Web	5350	19644.5	5357	19627.3
Network file	5343	10204.9	5360	10231.3
Mail	5391	3141.5	5382	3148.7
Batch	5315	633.4	5330	623.8
Application	5361	31415.9	5377	31437.5
Database	5285	27294.9	5273	27292.3

30 repetitions: Std. dev of 8.036 for t_s and 15.95 for r

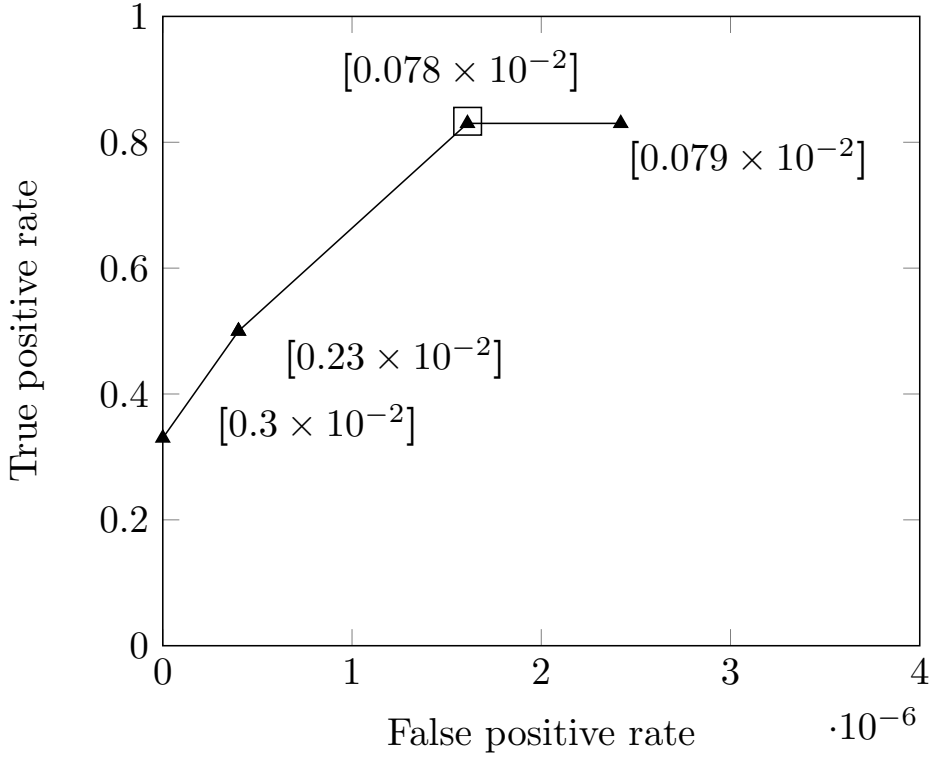
Case Study: Planning (cont.)

- **Step 3**: Specification of attack injection scenarios
  - *[Scenario #1]*
    - Injection from the web and mail server VM using LKM
    - [Attack content] CVE-2012-5525, CVE-2012-3495, CVE-2012-5513, CVE-2012-5510, CVE-2013-4494, and CVE-2013-1964
    - [Attack injection time] 10 secs. of separation between each attack
  - *[Scenario #2]*
    - Injection from the database server VM using LKM
    - [Attack content/injection time] „Mimicry“ and „smoke screen“ versions of the attacks above

Case Study: Testing

- *[Scenario #1]*
  - **Step 1**: IDS training until time t_s = *5391* sec.
  - **Step 2 and 3**: Attack injection and calculation of metric values

Targeted vulnerability	Detected
CVE-2012-3495	✔
CVE-2012-5525	x
CVE-2012-5513	✔
CVE-2012-5510	✔
CVE-2013-4494	x
CVE-2013-1964	x



Case Study: Testing (cont.)

- *[Scenario #2]*
  - **Step 1**: IDS training until time t_s = 5285 sec.
  - **Step 2 and 3**: Attack injection and calculation of metric values

Targeted vulnerability	Anomaly scores		
	Unmodified	„Mimicry“	„Smoke screen“
CVE-2012-3495	1.0	0.17	0.25
CVE-2012-5513	0.32	0.107	0.28
CVE-2012-5510	1.0	0.14	0.31
CVE-2013-4494	0.21	0.14	0.14
CVE-2013-1964	0.25	0.14	0.14

Summary and Future Work

- We provided an approach and a tool for evaluating IDSes in virtualized environments
  - hInjector is publicly available at https://github.com/hinj/hInj
  - First work on injecting attacks in virtualized environments [Pham et al. 2011], [Le et al. 2008]

- Future work
  - Injection of attacks that involve operations functionally similar to hypercalls (e.g., KVM ioctl calls)
  - Evaluation studies involving a variety of IDSes
    - .... and mandatory access control systems – XSM-FLASK
  - Establish a continuous effort on updating hInjector‘s attack library

Attack Injection

- Virtualized Debian 8.0 on top of Xen 4.4.5
- The *mmuext_op* hypercall executed 40000 times
- Injector manipulated the second parameter of *mmuext_op*