Project AMSEL:

Automatically Collect and Learn To Detect Malware

Martin Apel, Michael Meier

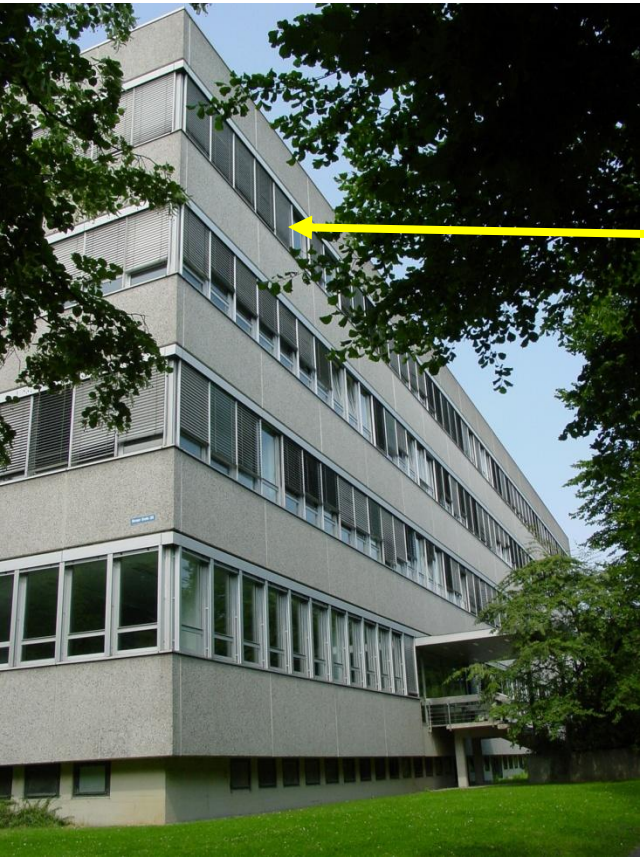
technische universität dortmund

Department of Computer Science

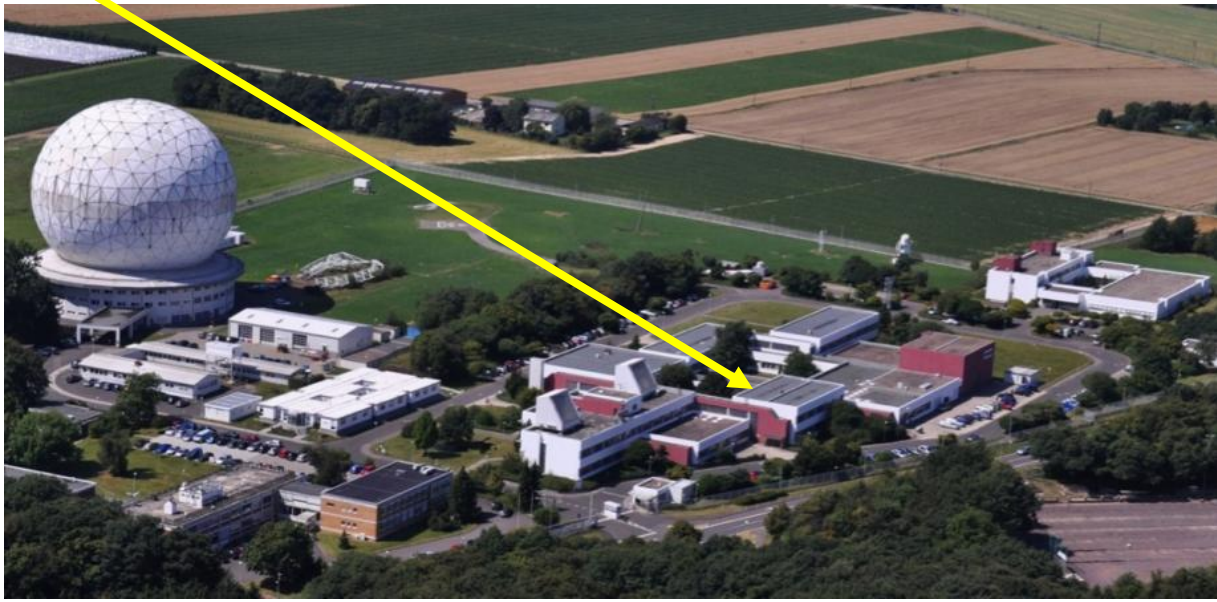
Chair VI – Information Systems and Security

universitätbonn

technische universität
dortmund



My Office



Fraunhofer
FKIE

Fraunhofer Cyber Defense Center Bonn

Cyber Defense Labs

Cyber Defense Lab Bonn

universitätbonn Institute of Computer Science 4 Communication and Distributed Systems

Malware Analysis & Digita
- Malware Analysis
- Digital Forensics
- Honeypots/Honeynets
- Botnet Analysis & Countermea

Mobile and Sensor Netwo
- Secure Ad hoc Routing
- Secure Sensor Networks

Teaching and Training
- Student Education
- Professional Training

Cyber Defense Lab Wachtberg

Fraunhofer FKIE

Monitoring & Situational Awareness
- IDS for heterogeneous Networks
- Operational Picture & Situational Awareness
- Intrusion Response

Resource-efficient Cryptography
- Efficient Key Management
- Application Protection Protocols
- Network Protection Protocols

Secure Network Architectures
- Interoperable Coalition Architectures
- Multi-Level Security
- Gateway Concepts
- Protected Core Networking

Overview

- project overview
  - early warning systems
  - architecture of a malware early warning system
- experiences
  - early
    - detection delay
    - generalizing signatures
    - incompleteness
- summary

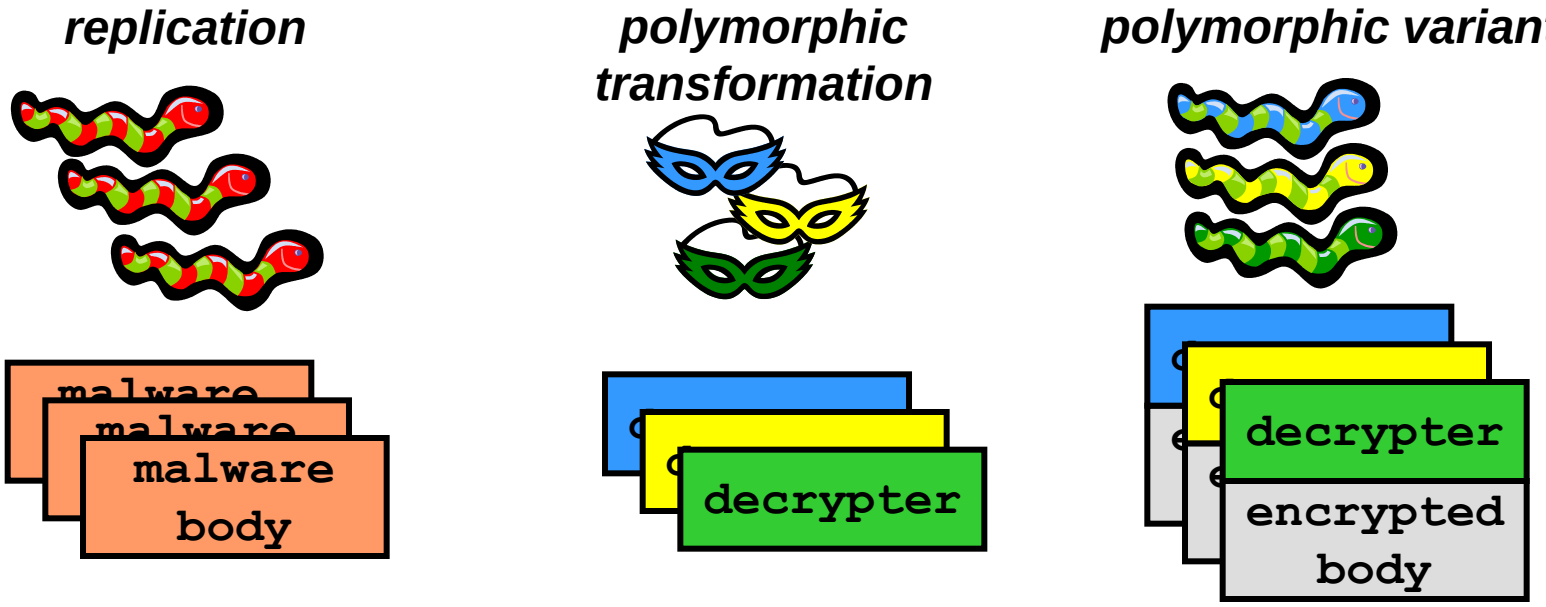
Early Warning Systems [1]

- aim at
  - detecting yet unclassified but potentially harmful system behavior
  - based on preliminary indications
  - establish hypotheses, predictions and advices in not yet completely understood situations
  - include two meanings of „early“
    - “fast”: start early in time in order to avoid/minimize damage
    - “incomplete”: process uncertain and incomplete information

[1] 08102 Manifesto -- Perspectives Workshop: Network Attack Detection and Defense. Dagstuhl, 2008.

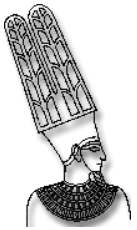
Malware Appearance

- malware propagates in polymorphic form
  - morphing/obfuscating tools generate programs of equal/similar functionality but different (“appearance”) feature instantiations
    - 30.000 new unique (wrt. “appearance” features) malware samples a day
      - polymorphic variants of a few malware types
    - would require to handle 30.000 new signatures a day



Idea of a Malware EWS

- automatically
  - collect malware
  - analyze malware behavior
  - generate signatures
    - n-gram based vectorization of behavior reports
    - manhattan distance
    - complete linkage clustering
    - shared sub-strings of clusters (Ukkonen's algorithm)
      - not shared with good pool
  - distribute and deploy signatures
  - report alerts centrally

AMUN 

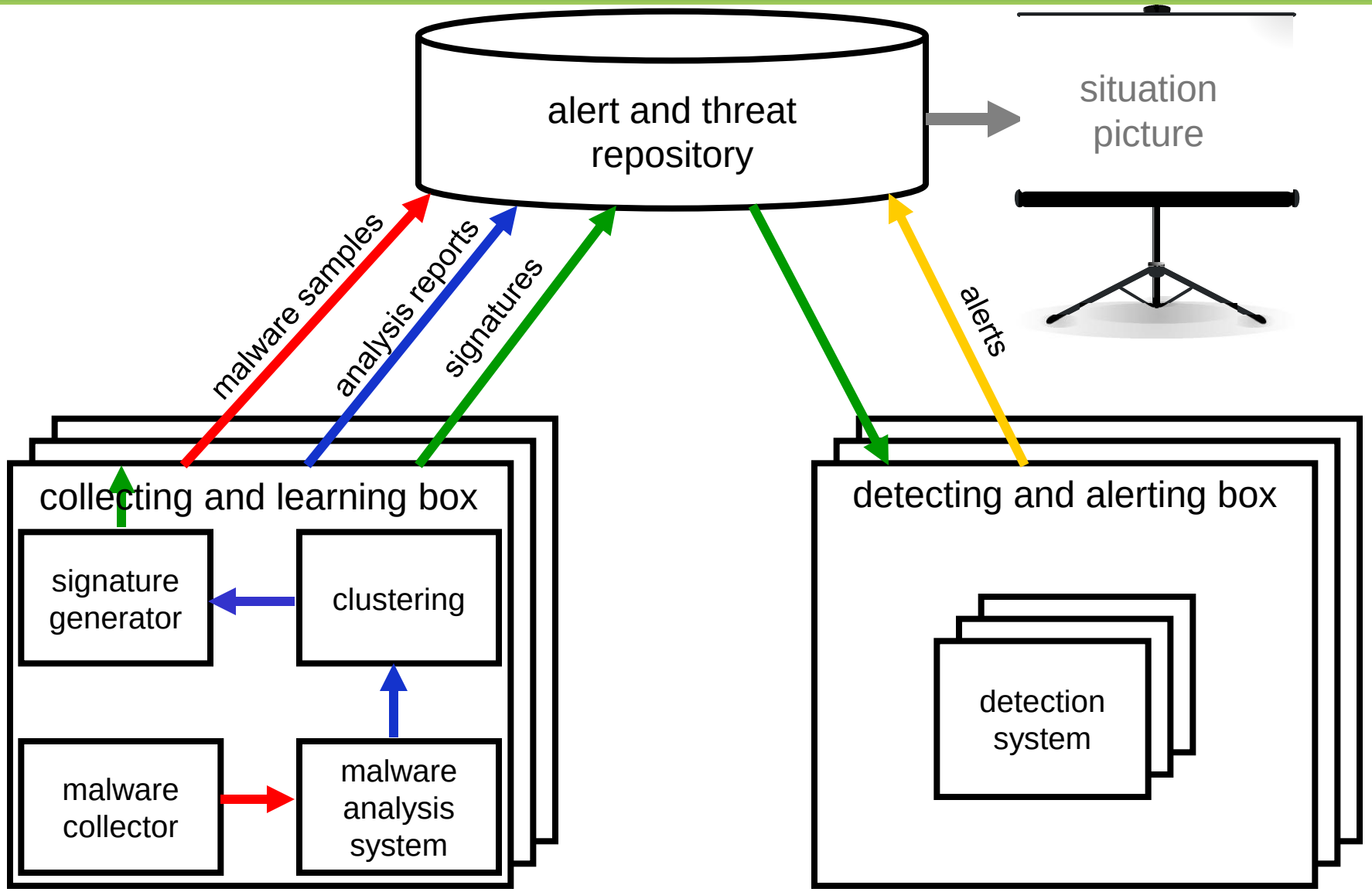
nepenthes - finest collection -

CWSandbox Sunbelt

Automatisch **M**alware **S**ammeln und **E**rkennen **L**ernen

automatically collect and learn to detect malware

Architecture



Early („fast“)

Definition: **Detection delay** describes the time elapsing between occurrence of a malware sample at a collecting and learning box and it‘s earliest possible detection at a detecting and alerting box.

Estimation of Detection Delay I

Example: known sample, signature available

alert and threat repository

malware samples

analysis reports

signatures

signatures

alerts

collecting and learning box

signature-generator

clustering

malware collector

malware analysis system

detecting and alerting box

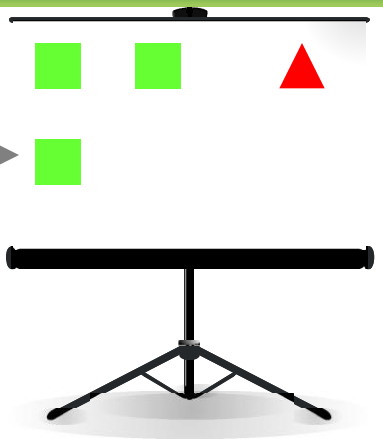
e-mail Scanner

execution interceptor

Estimation of Detection Delay II

Example: unknown sample, no signature available

alert and threat repository



malware samples

analysis reports

signatures

signatures

alerts

collecting and learning box

signature generator

clustering

malware collector

malware analysis system

detecting and alerting box

e-mail scanner

execution interceptor

Summary - Detection Delay

process	main influence	component	delay (min:sec)
malware collector	sample download	collecting and learning box	0:10
malware analysis	analysis time limit	collecting and learning box	2:00
clustering and signature generation	number of locally available samples GoodPool size	collecting and learning box	0:30 2:00
signature distribution		alert and threat repository	0:10
		sum	**4:50**

Cases: Detection Delay

	sample	signature	sum (delay)
known malware	known	available	≤ 0:10[a]
new malware	unknown	not available	4:50
new variant	**unknown** unknown	**generalizing** not generalizing	≤ **0:10**[b] 4:50[c]
„fresh malware“	known	not available	≤ 4:50[d]

[a] signature available, but possibly not yet distributed globally

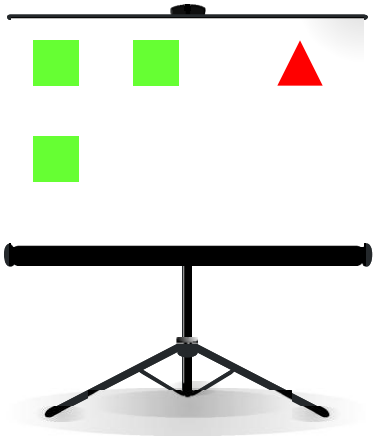
[b] available signature covers variant, but possibly not yet distributed globally

[c] available signature must be extended

[d] signature generation still running

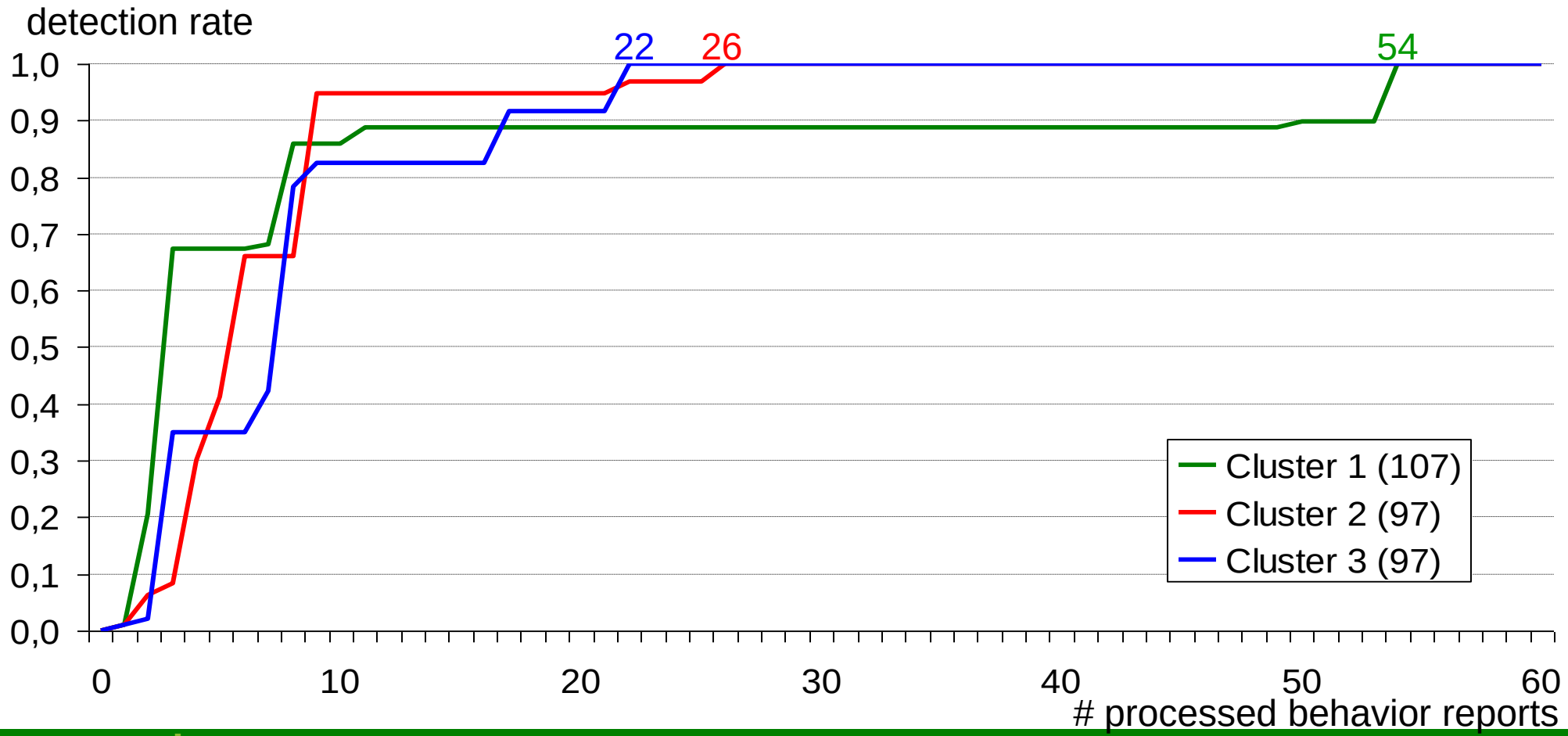
Conclusion on Detection Delay

- reference value for detection delay: **4 min 50 sec**
- generalizing signatures allow early detection
  - case „new variant“
- results of individual collecting and learning steps are available for the situation picture immediately



Detection Rate and Generalization of Signatures

- How many behavior reports of a cluster (of size N) need to be processed in order to create a signature with detection rate 1 (wrt. N cluster elements)?



Early (incomplete)

- honeypot assumption
  - all samples collected by malware collectors are malware
- incompleteness due to dynamic analysis
  - assumption: malware show malicious behavior during analysis
  - limited analysis time (about 2 minutes)
  - only simulated user interaction during malware analyse
- incompleteness of GoodPool

Time for Demo?

- Sample #10911
- Report
- Cluster
- Signature

Summary

- architecture of an automatic EWS
- focus of our ongoing research
  - new approaches for malware collection and analysis
  - clustering of malware behavior
  - generating behavior signatures
  - balancing conflicting availability and confidentiality requirements

Thank You!

- **Contact**

Michael Meier
TU Dortmund / Informatik 6
michael.meier@udo.edu

michael.meier@cs.uni-bonn.de
michael.meier@fkie.fraunhofer.de

http://ls6-www.cs.tu-dortmund.de/~meier/