Applying Privacy Preserving Data Mining to Intrusion Detection Systems

Clemens Frank

RWTH Aachen, Informatik 5
Lehrstuhl Prof. Decker

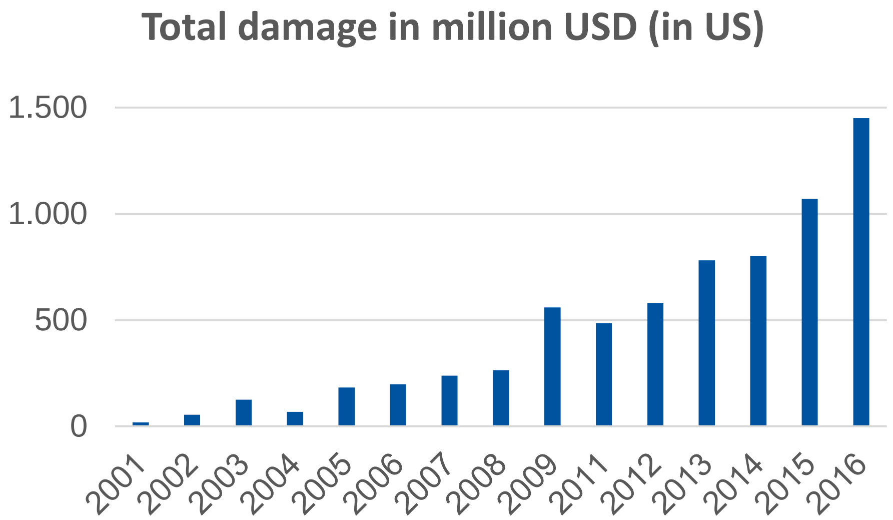
PETs4DS

RWTH AACHEN UNIVERSITY

Motivation

Cyberattacks on the Rise

- Huge increase in cyberattacks
- Cyberattack = any malicious action aiming at compromising confidentiality, integrity or availability of a computer system
- Global cost of damage due to cyber crime predicted to reach 6 trillion USD by 2021

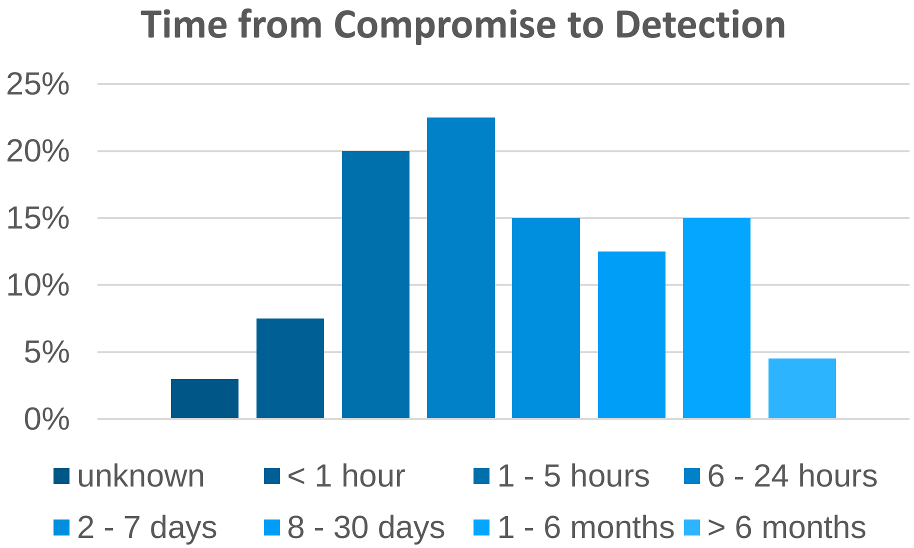


[1] IC3: total damage caused by reported cyber crime 2001-2017 - Published by statista.com

Motivation

Intrusion Detection Time

- Detection time takes more than 5 hours for two thirds of the cases
- Sharing of detection models between companies decreases detection time
- **Thesis Goal**:
  - Proof of concept
  - Explore trade-off between privacy preservation and performance (overhead and accuracy) in intrusion detection

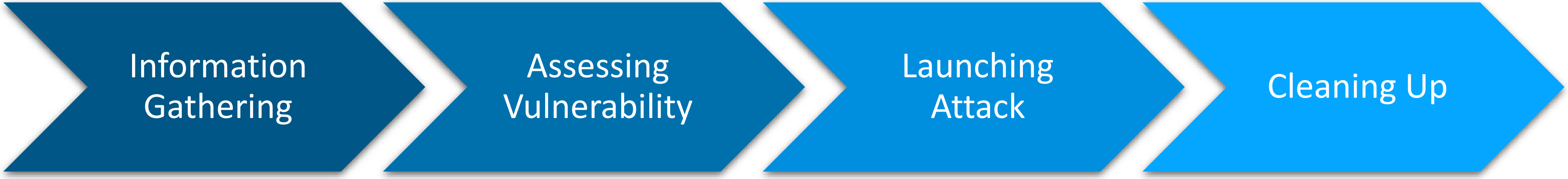


[2] The Show must go on – A SANS Survey by Matt Bromiley – Published 2017 by SANS Institute

Background

Cyberattack Cycle

1. Finding vulnerabilities (e.g. by port scans)
2. Attempting initial compromise of system
3. Launching actual attack (e.g. DoS, Trojan, Worm)
4. Hiding attack trace (e.g. by deleting log files)



Information Gathering

Assessing Vulnerability

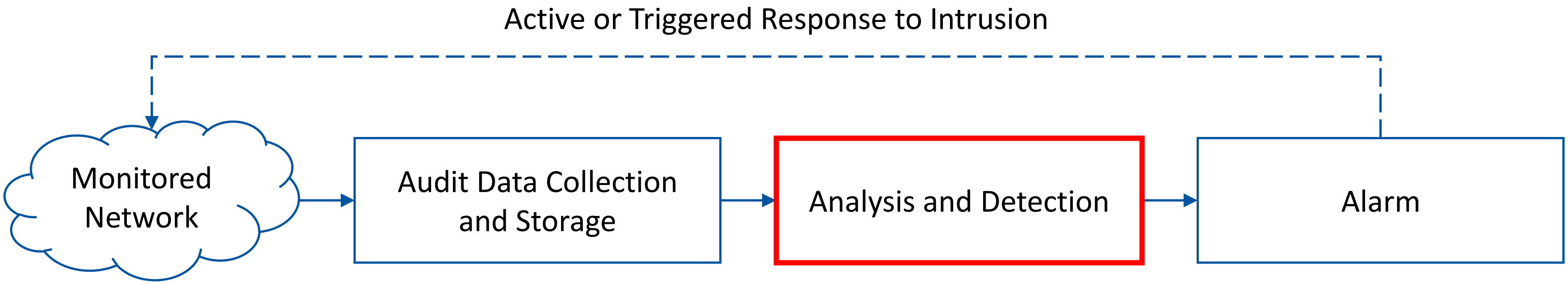
Launching Attack

Cleaning Up

Background

Intrusion Detection System

- Monitors network traffic
- Analyzes traffic based on reference data
- In case of attack detection → Triggers an alarm and/or automatic response to attack


Active or Triggered Response to Intrusion
Monitored Network
Audit Data Collection and Storage
Analysis and Detection
Alarm

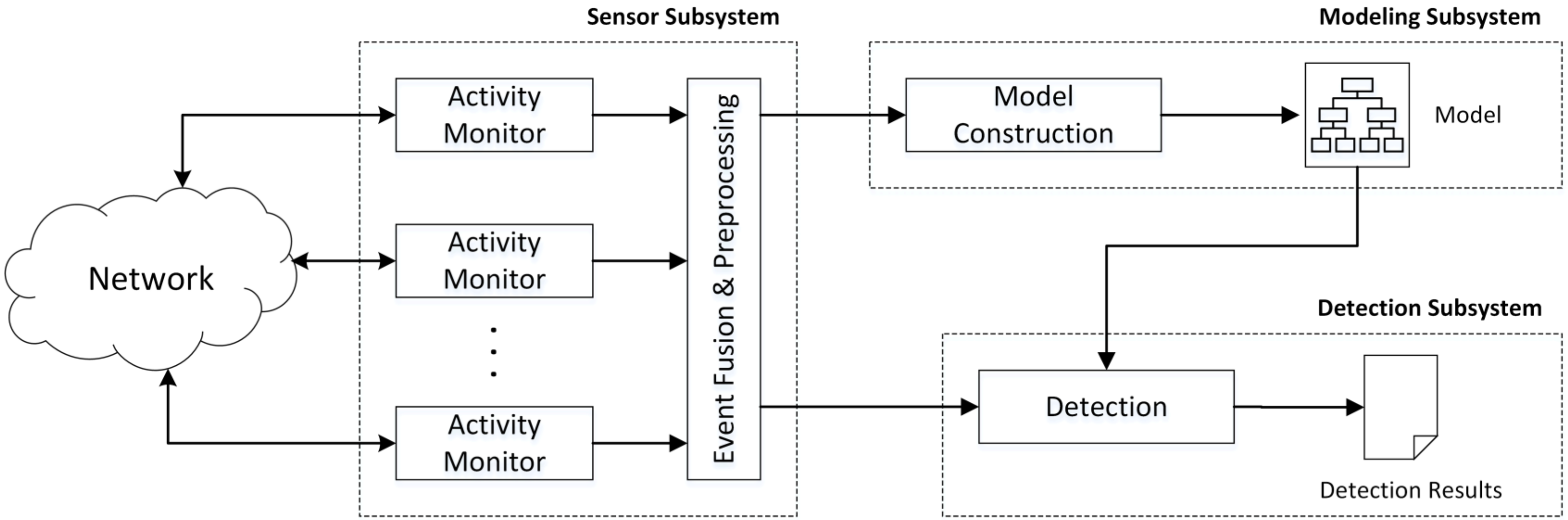
Background

Analysis and Detection

- Signature:
  - Extract signature for known attacks
  - Everything that shows signature behaviour is considered an intrusion
- Anomaly:
  - Generate baseline model of normal traffic
  - Everything that deviates from baseline model is considered an intrusion
- Hybrid:
  - Combination → Detect known attacks based on signature and unknown ones based on baseline model

Background

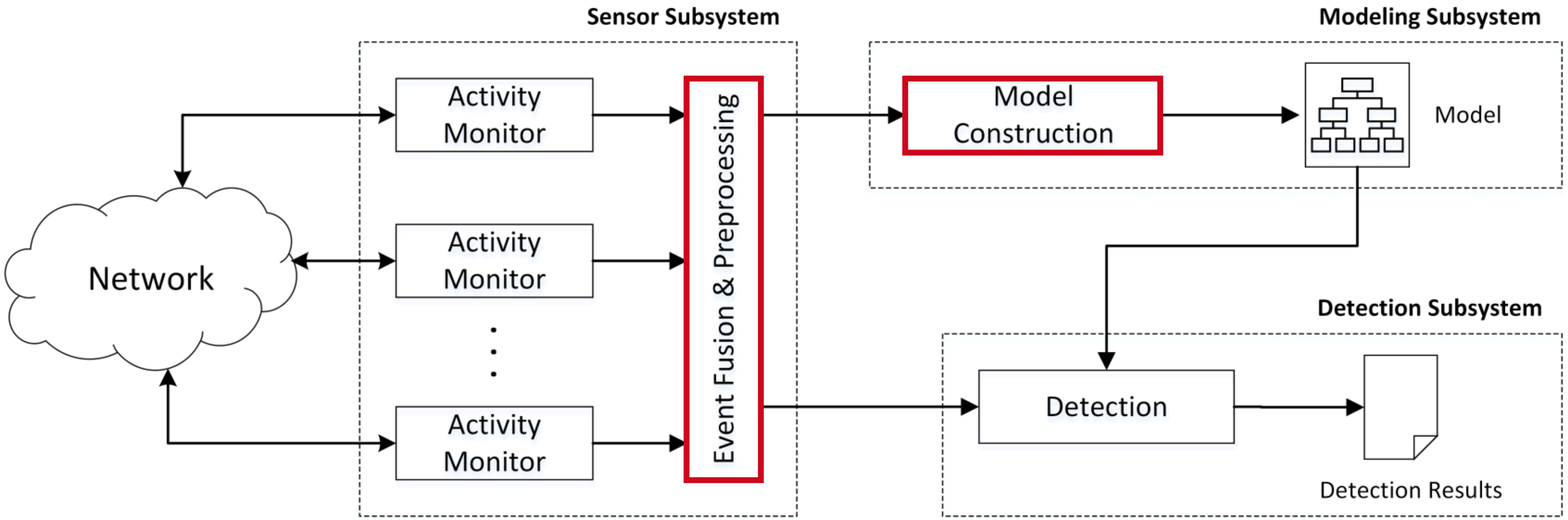
Anomaly-based Architecture



- Anomaly:
  - Generate baseline model of normal traffic
  - Everything that deviates from baseline model is considered an intrusion

Conceptual Approach

Privacy Preserving Intrusion Detection



1. Event Fusion & Preprocessing:
   - Modify data to satisfy privacy guarantees
   - Before or after merging

2. Model Construction:
   - Adapt model construction algorithm

   → Use **Privacy Preserving Data Mining** algorithm

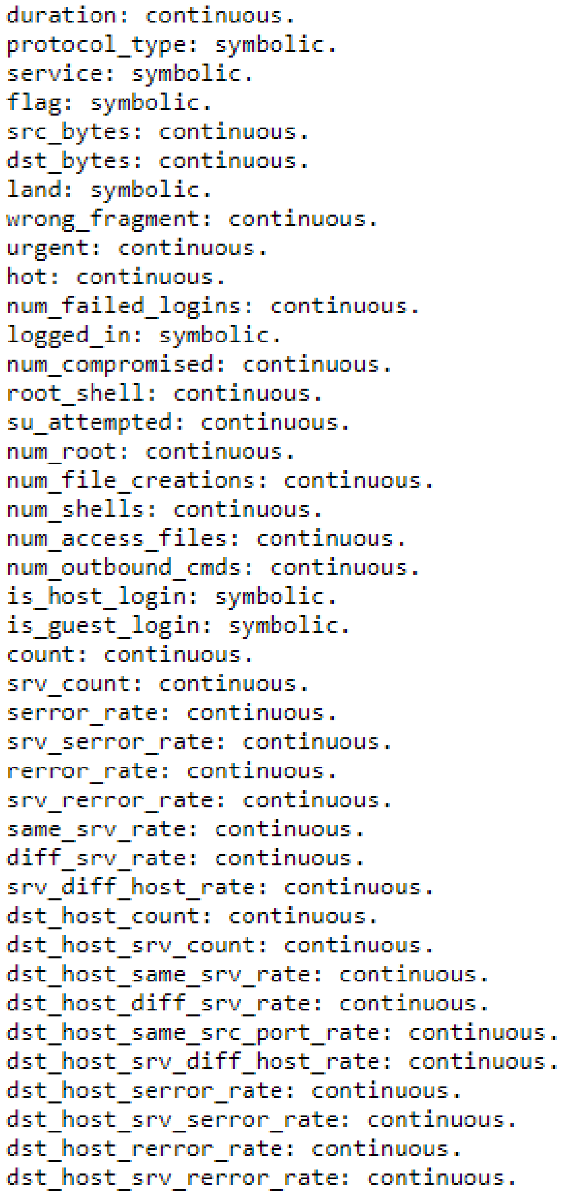
Conceptual Approach

Privacy Preserving Data Mining

- Extract hidden knowledge from data while preserving individuals’ privacy
- Classification of PPDM:
  - *Data distribution*: centralized vs. decentralized
  - *Data modification*: data swapping, perturbing, ..
  - *Data mining algorithm*: association rule mining, clustering, k-nn classification, ..
  - *Data or rule hiding*: raw vs. aggregated
  - *Privacy preservation*: Heuristic-based, cryptography-based, reconstruction-based

Implementation

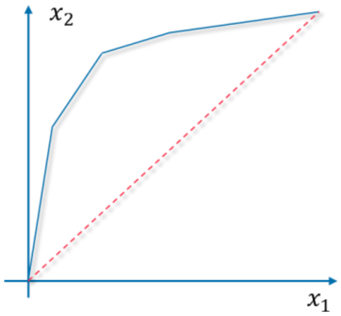
- Used Techonology:
  - Python:
    - NumPy (basic mathematical library)
    - Scikit-learn (machine-learning library)
- Datasets:
  - Contain basic, content and traffic features
  - DARPA KDD Cup 1999 (benchmark)
  - NSL-KDD
  - TUIDS (real-life)



```
duration: continuous.
protocol_type: symbolic.
service: symbolic.
flag: symbolic.
src_bytes: continuous.
dst_bytes: continuous.
land: symbolic.
wrong_fragment: continuous.
urgent: continuous.
hot: continuous.
num_failed_logins: continuous.
logged_in: symbolic.
num_compromised: continuous.
root_shell: continuous.
su_attempted: continuous.
num_root: continuous.
num_file_creations: continuous.
num_shells: continuous.
num_access_files: continuous.
num_outbound_cmds: continuous.
is_host_login: symbolic.
is_guest_login: symbolic.
count: continuous.
srv_count: continuous.
serror_rate: continuous.
srv_serror_rate: continuous.
rerror_rate: continuous.
srv_rerror_rate: continuous.
same_srv_rate: continuous.
diff_srv_rate: continuous.
srv_diff_host_rate: continuous.
dst_host_count: continuous.
dst_host_srv_count: continuous.
dst_host_same_srv_rate: continuous.
dst_host_diff_srv_rate: continuous.
dst_host_same_src_port_rate: continuous.
dst_host_srv_diff_host_rate: continuous.
dst_host_serror_rate: continuous.
dst_host_srv_serror_rate: continuous.
dst_host_rerror_rate: continuous.
dst_host_srv_rerror_rate: continuous.
```

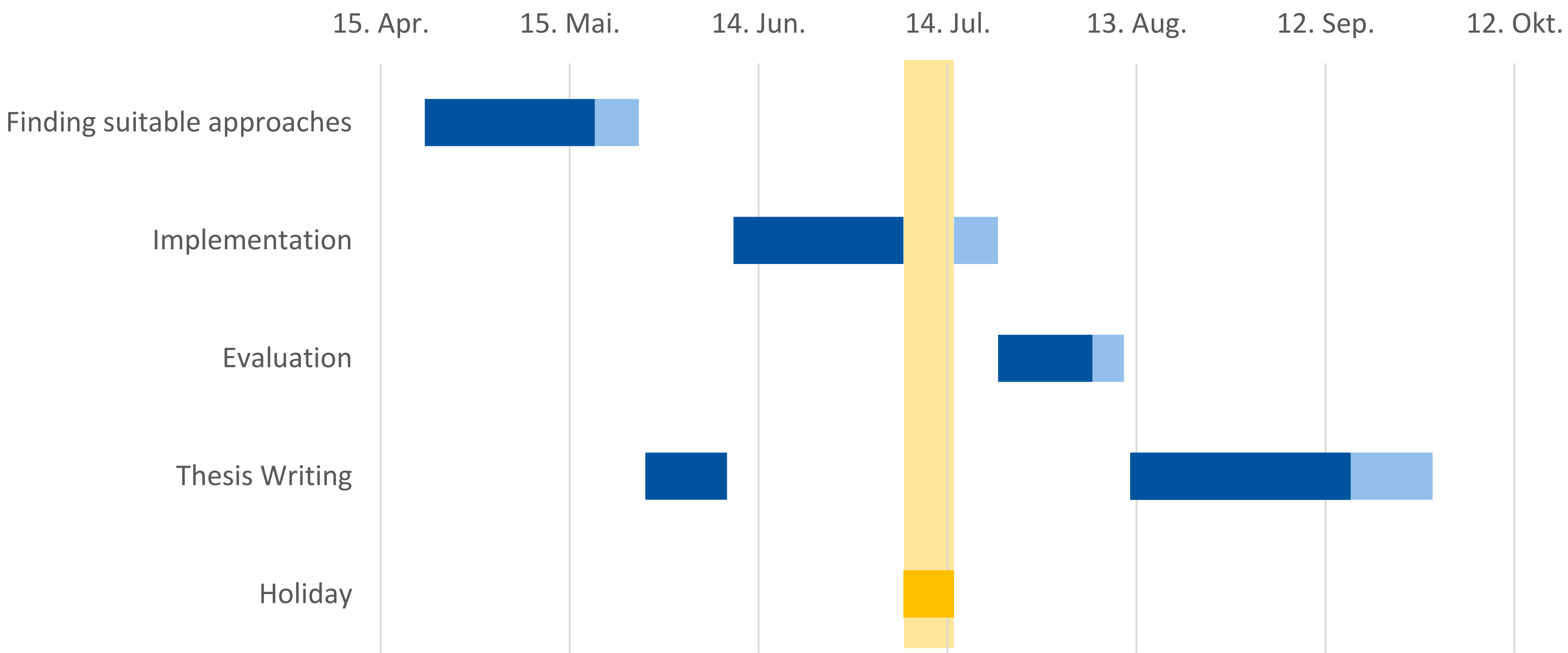
Evaluation

- Use of benchmark datasets (KDDcup99, TUIDS)
- Evaluation:
  - **Accuracy**: Detection Rate vs. False Positive Rate (ROC curve)
  - **Performance**: Computational overhead w.r.t. time
  - **Privacy**: Input and output metric



- Comparison to state-of-the-art intrusion detection approaches

Timeline


15. Apr.
15. Mai.
14. Jun.
14. Jul.
13. Aug.
12. Sep.
12. Okt.
Finding suitable approaches
Implementation
Evaluation
Thesis Writing
Holiday