KATHOLIEKE UNIVERSITEIT
LEUVEN

FACULTEIT
INGENIEURSWETENSCHAPPEN

Master of
Mathemtical
Engineering

Master's thesis
Henri De Plaen

Promotor
Pr. dr. ir. Bart Preneel

Academic year
2017-2018

Privacy-friendly machine learning algorithms for intrusion detection systems

Context

- As computers are used more and more (including for sensible data), and the connection between them increases as well, there is a constant need for better intrusion detection systems.
- Machine learning algorithms can increase performance of a lot of existing applications, but they need a significant dataset.
- The amount of data needed increases as well, but remains sensible in the case of intrusion detection systems, hence the need for encryption.
- *« Privacy-friendly collaborative network intrusion detection system »*

Detection

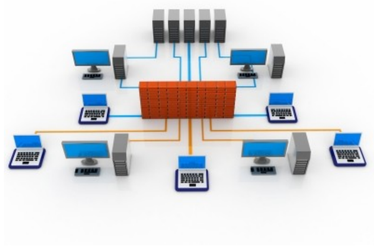
- Signature based
  - Advantages: accuracy and time
  - Disadvantages: only known intrusion types are detected
- Anomaly based
  - Advantages: new intrusion types can be detected
  - Disadvantages: malicious activity disguised as normal traffic can pass through
- **Machine learning (classification)**

Intrusion detection systems

- Network Intrusion Detection System (NIDS)
  - Advantages: detects attack before it occurs
  - Disadvantages: needs to be implemented on the network
- **Host Intrusion Detection System (HIDS)**
  - Advantages: collects broader data type
  - Disadvantages: needs to be implemented on each machine and only detects after the intrusion
- Hybrid Intrusion Detection System
  - Advantages: much more effective
  - Disadvantages: huge implementation necessary, not privacy-friendly

Implementation

- Build on existing application, as a plug-in
- Feed with constant new data, provided from other analysis tools
- User-friendly
- Off-line



Working

- Secure multi-party computation
  - Idea: different people jointly compute a function over their imput that remains secret
  - Two types:
    - Garbled
      - Boolean
      - Constant number of rounds

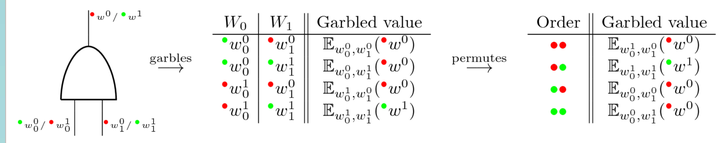


Figure 2: Our garbled AND gate using the permute-and-point method with colors.

    - **Arithmetic**
      - Based on somwhat homomorphic encryption
      - Allows for addition and multiplication
      - Ex: Samir secret sharing
  - SPDZ library (n-party, arithmetic)
- Machine learning
  - A lot of existing datasets
  - Different existing algorithms
- Further optimizations
  - Hybrid Intrusion Detection Systems
  - Speed and complexity optimizations (research on homomorphic encryption)