Intelligent Cyber-Security
Introduction to Cyber-Security (2)

Dr. Alexandru ARCHIP

Gheorghe Asachi Technical University of Iasi

Artificial Intelligence (MSc., second year) – 2025 – 2026

alexandru.archip@academic.tuiasi.ro

In brief...

1. Detection vs. Prevention
   - Detection & prevention techniques
2. Host/Network Intrusion Detection and Prevention Systems
   - Host vs. Network attacks
   - HID(P)S
   - NID(P)S
3. Threat Intelligence

Outline

Detection

Definition [1]

Detection is the process of monitoring and analysing computer system and/or network events for signs of possible suspicious activities.

- requires continuous data source monitoring
- *reactive* security mechanism
- *passive* by nature
  - issues *alerts* rather than actions
- may require large volumes of data

Prevention

Definition [1, 2]

Prevention refers to the set of potentially proactive measures, controls, and actions implemented to prevent a security incident.

- relies on *detection* to identify threats before issuing actions
- *potentially proactive* – depends on the actual implementation
- the term has a broader meaning, relating to both human and non-human factors:
  - identify weaknesses and apply patches
  - configure security response tools with appropriate response actions
  - educate and train developers and users, and raise security awareness

Detection & prevention techniques: a primary classification [1]

Signature-based (also known as *static* detection & prevention)

patterns and *rules* are developed for *known* attacks (e.g., a hash is computed over known opcode sequences or malicious parameters);

advantages: real-time response, highly efficient for *known* attacks;

disadvantages: even the *simplest* alterations of the *known* pattern may render these approaches useless.

Examples a remote login attempt targetting a user account who does not have such access rights;

an e-mail promising multimedia content (e.g., pictures of funny cats), but containing a binary executable file.

Detection & prevention techniques: a primary classification [1]

Anomaly-based (also known as *dynamic* detection & prevention)

normal behavioural patterns are modelled for different data sources such as users, hosts, network connections, and/or applications;

significant deviations are then monitored and analysed to identify potential attacks;

advantages: detect previously unknown threats by comparison with normal profiles;

disadvantages: high resource requirements; profiles may change over time and require adjustment.

Examples abnormal CPU or network usage of a monitored host;

uncommon responses for different web application requests (e.g., larger-than-usual response data); atypical request/response patterns in web application usage.

Outline

Host vs. Network attacks [3]

Host attacks

A *host attack* focuses on a specific host such as a server, a desktop, or a laptop.

Examples:

- installing malware (ransomware, spyware, etc.);
- exploiting vulnerabilities to gain control of a specific machine;
- using a keylogger to record every keystroke on a device.

Host Intrusion Detection (and Prevention) System

A *Host Intrusion Detection (and Prevention) System* is a security solution installed on an individual computer or server that monitors and analyses the system's activities to detect (and prevent) malicious behaviour.

Host vs. Network attacks [3]

Network attacks

A *network attack* focuses on the communication channels and the infrastructure connecting multiple devices.

Examples:

- intercepting communication between parties (*man-in-the-middle* attacks);
- overwhelming a network or a server with traffic to render it unavailable ((D)DoS).

Network Intrusion Detection (and Prevention) System

A *Network Intrusion Detection (and Prevention) System* is a security solution installed on strategically placed network devices or servers that monitors and analyses network traffic in real-time to detect (and prevent) malicious behaviour.

HID(P)S [3]

Typical deployments rely on software modules called *agents* that monitor:

- system and configuration files;
- various activity logs;
- (optionally) important content files.

An HIPS/HIDPS also includes an active component that issues system commands to prevent attacks.

> An antivirus may place a suspicious file in quarantine by removing certain access privileges from that file or by rendering specific binary components inactive (*disarming*).

HID(P)S [3]

An HID(P)S can detect (and prevent):

- system compromises and privilege escalation attacks;
- unauthorised application installation;
- alterations of critical system binaries and/or configuration files (e.g., modifications to `/etc/passwd` in Linux/Unix environments, `.dll` alterations in the W operating program);
- abnormal processes running on the monitored host;
- critical services that have been stopped or have failed to start.

Potential pitfall: HID(P)S are highly dependent on the operating system of the monitored host.

HID(P)S – examples

Common antivirus solutions

- a simple type of HIDPS;
- monitor files and file systems for changes, and perform various actions such as deleting or quarantining suspicious files;
- (particularly advanced solutions) issue alerts and exchange data with authorised parties to improve detection.

Wazuh [4]

- endpoint security solution providing configuration assessment, malware detection and file integrity monitoring;
- integrated with threat intelligence platforms to enhance log data analysis and vulnerability detection;
- it includes Ollama [5] modules (Llama 3 [6] being the preferred LLM) to enhance threat detection.

NID(P)S [3]

Typical deployments include:

- a number of *sensors* to monitor packet traffic;
- one or more servers for NID(P)S management.

A *sensor* may be deployed as:

- an inline sensor – a device inserted into the network segment so that the traffic passes through it;
- a passive sensor (the most common approach according to [3]) – the actual traffic is copied over to the equipment and monitored offline.

Traffic analysis is performed either at the *sensor*, at the management server, or through a combination of the two.

NIPS/NIDPS include a traffic-altering component that provides active responses by blocking or allowing traffic.

NID(P)S – examples

Advanced firewall solutions

- might be considered as a primary form of NID(P)S (despite being focused on policy enforcements);
- application-layer solutions monitor traffic content (also called Next-Gen Firewalls);
- proxy-like firewalls may be regarded as inline sensors.

Snort [7]

- provides real-time network traffic analysis and logging;
- implements rule-based intrusion detection and prevention;
- includes a machine learning based detection engine (*SnortML*).

NID(P)S – examples

Suricata [8]

- provides high-performance network monitoring capabilities;
- offers comprehensive protocol parsing techniques, including deep packet inspection for protocols such as HTTP, DNS and TLS;
- offers both signature-based and anomaly-based detection features.

Outline

1. MITRE ATT&CK
2. Common Vulnerabilities and Exposures (CVE)
3. MISP Threat Intelligence Platform
4. OpenCTI Platform by Filigran

ML/AI models for cyber-security

- High-quality training data are a mandatory requirement for effective ML/AI models.
- IDS, when used in combination with offensive techniques, provide an excellent data source.
  - This proactive approach is essential to increase overall security.
- IDS may be enhanced by descriptive ML models:
  - clustering techniques provide a new understanding of data through the identified groups;
  - frequent pattern mining and correlation analysis may detect previously unknown attacks.
- IPS are a key application for AI models:
  - suitable inference techniques could enhance IPS resilience against unknown threats;
  - so-called *0-day attacks* could be pre-emptively mitigated.

Bibliography

1. Scarfone, K., & Mell, P. (2010). *Intrusion detection and prevention systems*. In P. Stavroulakis & M. Stamp (Eds.), *Handbook of Information and Communication Security* (Chapter 9). National Institute of Standards and Technology. https://doi.org/10.1007/978-3-642-04117-4_9, [NIST Source Link]
2. National Institute of Standards and Technology. (n.d.). *Intrusion prevention*. In *Glossary of Key Information Security Terms*. National Institute of Standards and Technology. https://csrc.nist.gov/glossary/term/intrusion_prevention
3. Buțincu, C., & Mironeanu, C. (2024). *Intrusion Detection and Prevention Systems*, MSc. studies in Cybersecurity, lecture notes (Romanian version)
4. Wazuh, Inc. (founded 2015) — https://wazuh.com
5. Ollama, Michael Chiang & Jeffrey Morgan (founded 2023) — https://ollama.com
6. Meta, "Llama 3" (released 2024) — Meta Llama 3 announcement

Bibliography

7. Cisco Systems (2021), *Snort 3: Open Source Network Intrusion Detection and Prevention System*
8. Open Information Security Foundation (OISF, 2025). *Suricata: High-performance open-source network analysis and threat detection software*. https://suricata.io