

---

## Techniques for intrusion detection

---

## Techniques used in intrusion detection

### Techniques:

- Statistical analysis
- Neural Networks and machine learning
- Rule-based and signature techniques
- State-transition analysis
- Immune systems based techniques
- Data mining

---

## Statistical analysis approach

- The user or system behaviour (set of attributes) is measured by a number of variables over time.
- Examples of such variables are:
  - user login, logout,
  - number of files accessed in a period of time,
  - usage of disk space, memory, CPU, etc.
- The frequency of updating can vary from minutes to months.
- The system stores *mean* values for each variable used for detecting exceeds that of a predefined threshold.
- More sophisticated model: short- and long-term user profiles.

---

## Neural networks and machine learning

- Neural networks use their learning algorithms to learn about the relationship between input and output
- The main purpose is to learn the behaviour of actors in the system (e.g., users, processes).
- The advantage of using neural networks over statistics:
  - a simple way to express nonlinear relationships between variables;
  - learning about relationships automatically.
- The disadvantage: NNs are still a computationally intensive technique.
- It has been shown that NNs can be used to predict behaviour of users and super-users (roots);
- Other machine learning techniques may be used

## Rule-based (expert) systems

- Rule-based (expert) systems: work on a previously defined set of rules describing an attack.
- All security related events are translated in terms of **if-then-else rules**.
- Inference engine maybe used to infer conclusions.
- **Examples:** Wisdom & Sense and ComputerWatch (developed at AT&T).

COMP 522

## Signature analysis

- The method (as well rule-based systems) is based on the attack knowledge.
- An attack scenario can be described, for example, as a *sequence* of audit events that a given attack generates or patterns of searchable data that are captured in the audit records.
- Detection is accomplished by using *common text string matching mechanisms*.

COMP 522

## Signature based method

Example of a signature: alert icmp \$EXTERNAL\_NET any -> \$HOME\_NET any (msg:"MISC large ICMP"; dsiz: >800; reference:arachnids,246; classtype:bad-unknown; sid:499;)

Alarm will be raised if a ICMP packet incoming from the external network,m associated with any port and having a size more than 800 bytes

### Advantages:

- very low false alarm rate;
- simple algorithms, easy implementation.

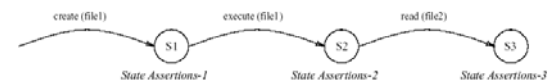
### Disadvantages:

- difficulties in updating information on new types of attacks;
- Unable to detect unknown attacks (knowledge based)

COMP 522

## State-transition analysis

- an attack is described with a set of goals and transitions that must be achieved by an *intruder* to compromise a system.
- Transitions are represented on state-transition diagrams.



- Inference engine matches current audit records with attacks descriptions: event driving, forward chaining

COMP 522

# Computer Immunology

---

- Analogies with immunology => technique that constructs a model of *normal behaviour of (UNIX) network services*, rather than that of individual users.
- S.Forest at al., **Sense of Self for UNIX Processes**, 1996
- This model consists of *short sequences of system calls* made by the processes.
- The assumption is *attacks that exploit flaws in the application code are likely to take unusual execution paths*.
- The model consists of two parts :
  - audit data representing the appropriate behaviour of services;
  - knowledge base with all the known “good” sequences of system calls.

# Computer Immunology

---

- Stored patterns are used for monitoring of system calls to check whether the sequence generated is listed in the knowledge base; if not — an alarm is generated.

**Example:** pattern of normal behaviour:

**open, read, mmap, mmap, open, getrlimit, mmap, close** (system calls in UNIX)

Information about all sequences of the **length, say, 3** may be stored in knowledge base;

If the input sequence of calls is

**open, read, mmap, open, open, getrlimit, mmap, close**

then 4 mismatches will be found

# Computer Immunology

---

## Advantages:

- Potentially very low false alarm rate (if knowledge base complete enough)
- Updating the knowledge base can be done on-line

## Disadvantages :

- if an attacker uses legitimate actions to gain unauthorized access, no alarm will be generated
- Arguments of system calls are not taken into account

# User intention identification

---

- This technique models normal behaviour of users by the set of high-level tasks they have to perform on the system.
- These tasks are taken as series of actions, which in turn are matched to the appropriate audit data.
- If mismatch is encountered, an alarm is produced.

# Data mining

---

- set of techniques that use the process of extracting *previously unknown* information from large stores of data.
- Data mining techniques are suitable for large volumes of audit data (off-line processing).
- DM is less useful for stream analysis of network traffic.
- Based on analysis of co-occurrences and correlations of event records, DM may produce
  - Decision trees
  - Association rules

as a concise representations of discovered relationships

# Data mining

---

## Example of the rule, generated by DM:

**if** for the past 2 seconds, the *count of connections to the same dst host* is greater than 4; **and** the *the percentage of those that have the same service* is greater than 75%; **and** the *percentage of those that have the “S0” flag* is greater than 75%, **then** there is a *syn flood* attack.

(taken from **Real Time Data Mining-based Intrusion Detection**,  
Wenke Lee et al.)