Intrusion Detection Systems

Sokratis K. Katsikas

Dept. of Digital Systems
University of Piraeus
ska@unipi.gr

Agenda

- Overview of IDS
- Intrusion prevention using game theory
- Reducing false positives
- Clustering alerts from multiple sensors
- Conclusions

Credits

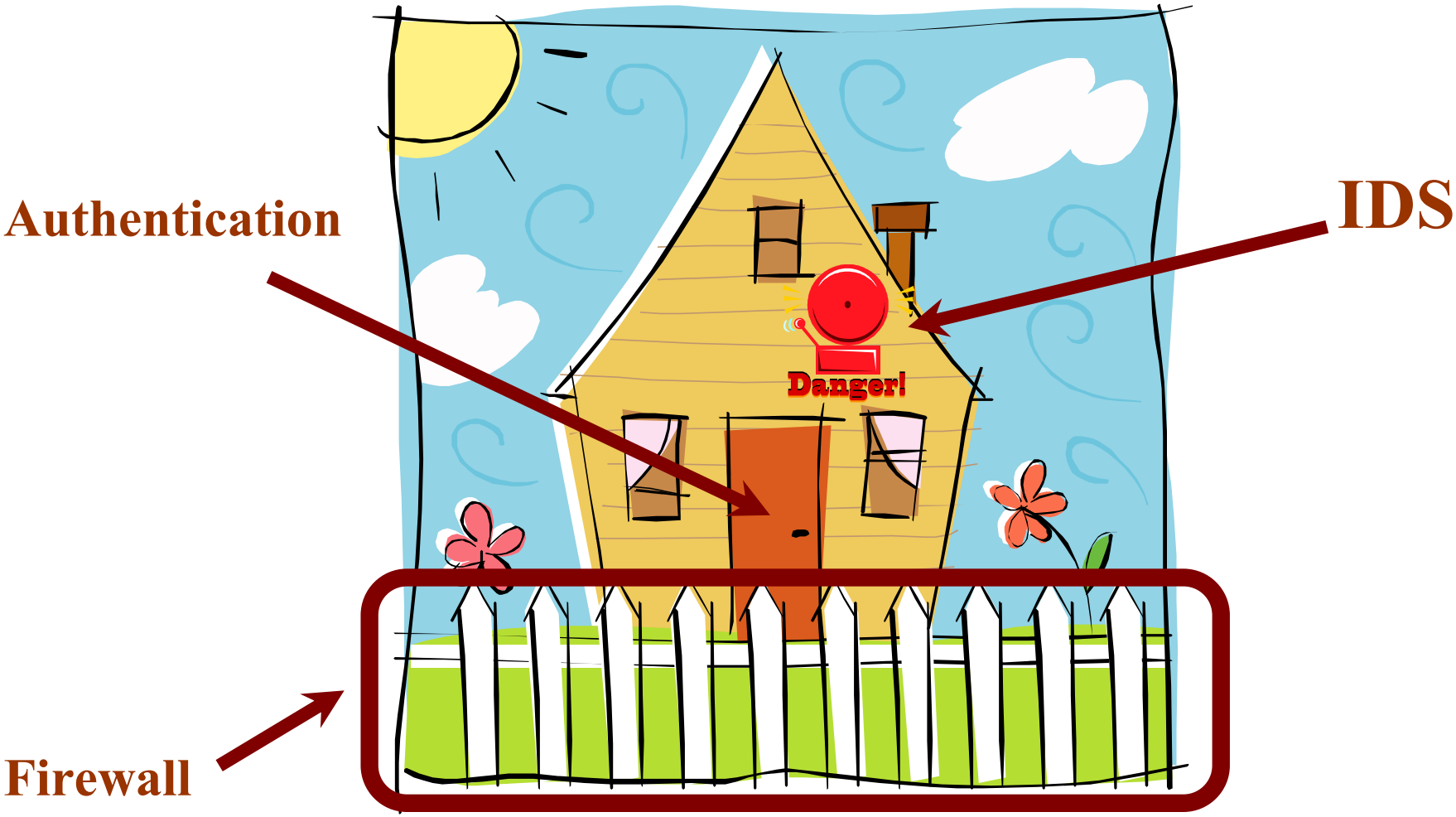
- Maria Papadaki, University of Plymouth, UK (Overview)
- Ioanna Kantzavelou, University of the Aegean, Greece (IPS using game theory)
- George Spathoulas, University of Piraeus, Greece (Reducing False Positives & Clustering alerts from multiple sensors).

OVERVIEW OF IDS

Intrusions and Intrusion Detection Systems

- An **intrusion** happens when somebody (hacker or cracker) attempts to break into or misuse your system (intrusion = attacks from the outside)
- "**Misuse**" is broad, and can reflect something severe such as stealing confidential data or something minor such as misusing email system for spam (misuse = an attack that originates from the internal network).
- **Intrusion Detection System** (IDS) is a software or hardware product that monitors the events occurring in a computer system or network and analyses them for signs of intrusions.
- **Intrusion Detection** is the art of detecting inappropriate, incorrect, or anomalous activity.

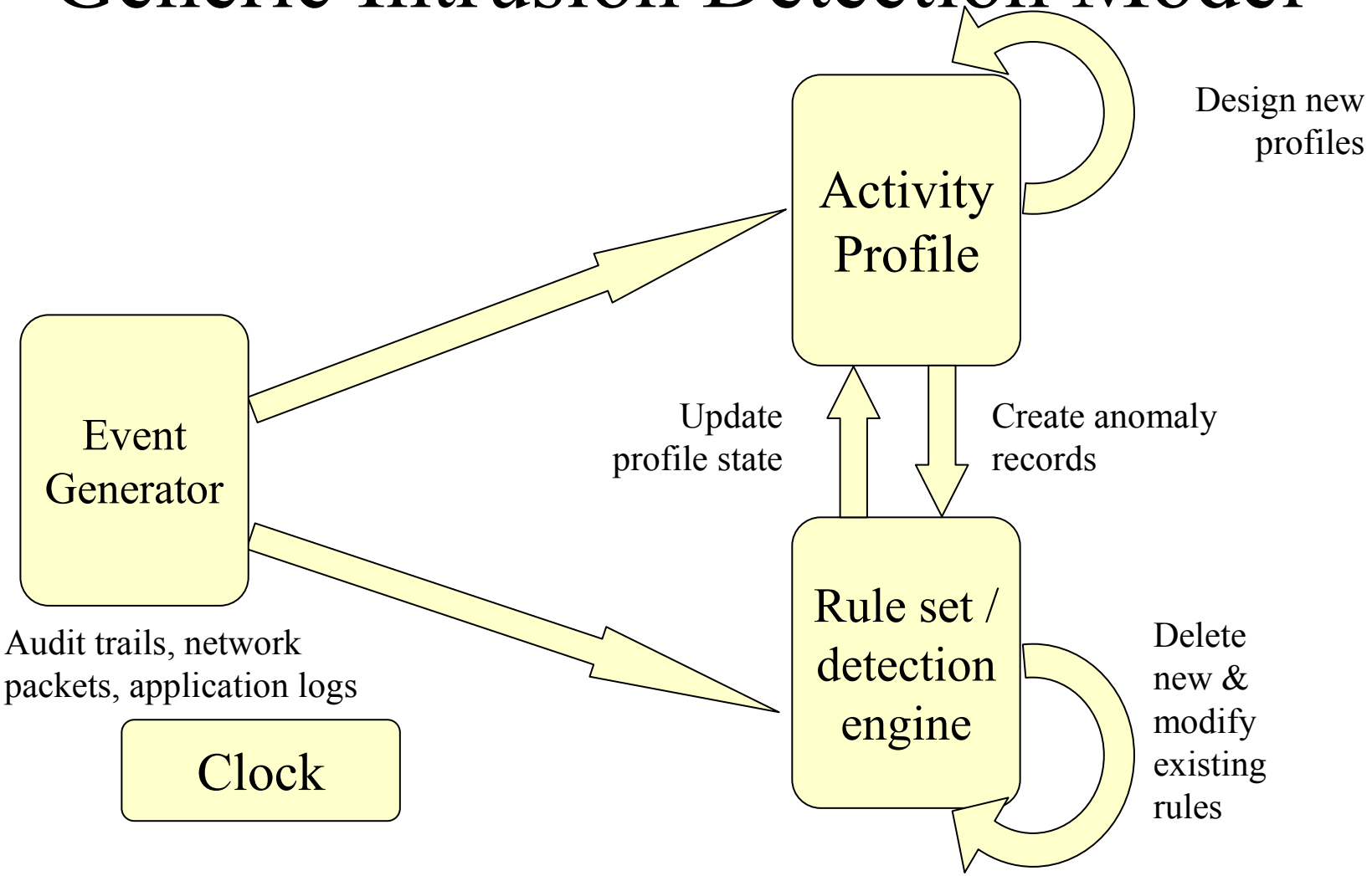
Consider your home



The origins of IDS

- 1980: James P. Anderson
  - Automated audit trail review
  - Automated collection of information for review by security personnel
  - Reduction of irrelevant records
- 1987: Dorothy Denning
  - Generic Intrusion Detection Model

Generic Intrusion Detection Model



Generic Intrusion Detection Model

- *Event generator*
  - Provides information about system activities. Events are derived from system audit trails, network traffic, and from application level systems.
- *Rule set*
  - The element that decides whether an intrusion has occurred. Events and state data are examined using rules, models, patterns and statistics in order to identify and flag intrusive behaviour.
- *Activity profile*
  - Maintains the state of the system or network being monitored. Variables in the profile are updated as events appear from the monitored data sources.

Is it any good?

- Early detection may prevent (or at least minimize the extent of) damage
- Existence of intrusion detection system serves as deterrent to potential intruders, thus preventing them
- Enables collection of information to be used to
  - strengthen the prevention facilities
  - prosecute (legally) intruder

Typical Intrusion Scenario

- Step 1: outside reconnaissance
- Step 2: inside reconnaissance
- Step 3: exploit
- Step 4: foot hold
- Step 5: profit

Anti-Intrusion Approaches

- **Prevention** precludes or severely handicaps the likelihood of a particular intrusion's success.
- **Preemption** strikes offensively against likely threat agents prior to an intrusion attempt to lessen the likelihood of a particular intrusion occurring later.
- **Deterrence** deters the initiation or continuation of an intrusion attempt by increasing the necessary effort for an attack to succeed, increasing the risk associated with the attack, and/or devaluing the perceived gain that would come with success.
- **Deflection** leads an intruder to believe that he has succeeded in an intrusion attempt, whereas instead he has been attracted or shunted off to a specially prepared, controlled environment for observation.

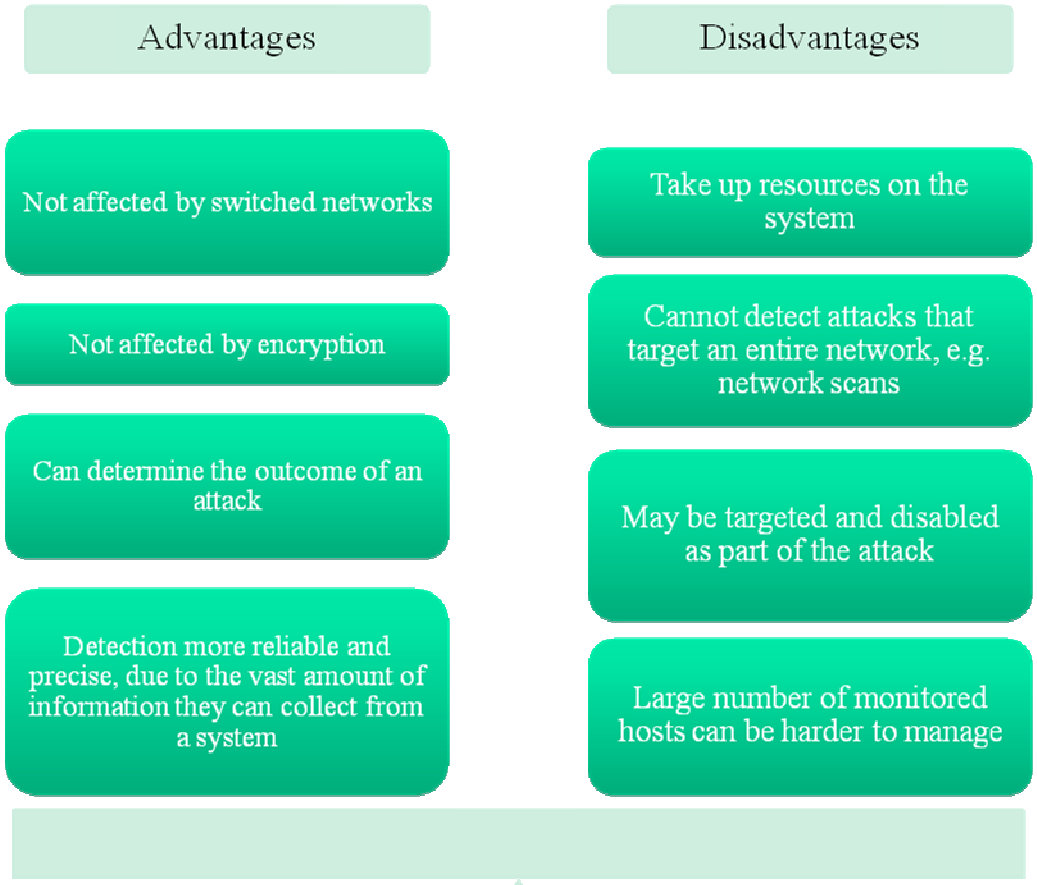
IDS Classification

- IDS can be classified according to three factors:
  - the source from which they collect data for analysis.
  - the detection mechanism by which the collected data is then analysed in order to detect potential intrusions.
  - the response mechanisms which are triggered as a result of generated alerts

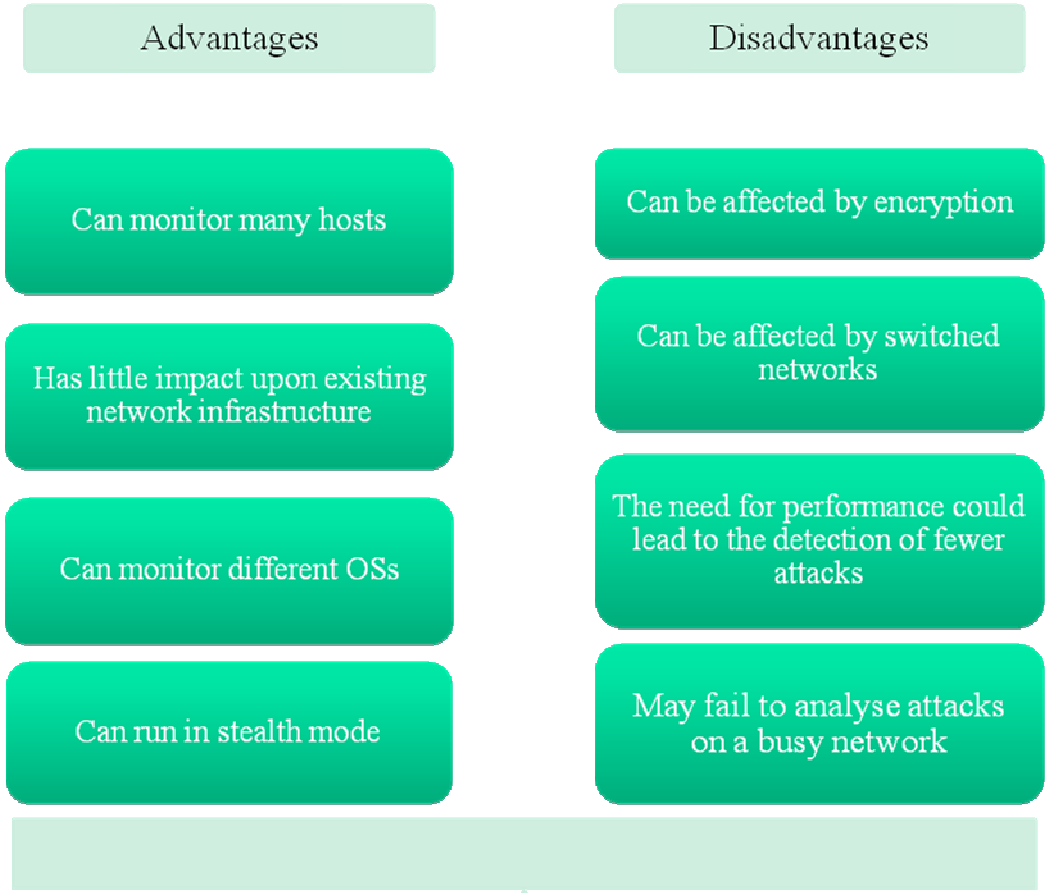
Data Source

- Application
  - Examines the behaviour of an application program, generally by analysing the application log files
- Host
  - Utilizes two types of information sources, operating system audit trails, and system logs
- Network
  - Examines network traffic. Often consist of single purpose sensors that run in stealth mode
- Hybrid
  - Combines two or more of the sources above

Host-based IDS



Network-based IDS



Detection Methods

Misuse Detection	• Based on the comparison of system activity to a predefined pattern of events that describes a known attack
Anomaly Detection	• Based on the assumption that misuse or intrusive behaviour deviates from historical norms
Specification-based Detection	• Based on rules that define the correct operation of a program/protocol
Hybrid Detection	• Combines detection models

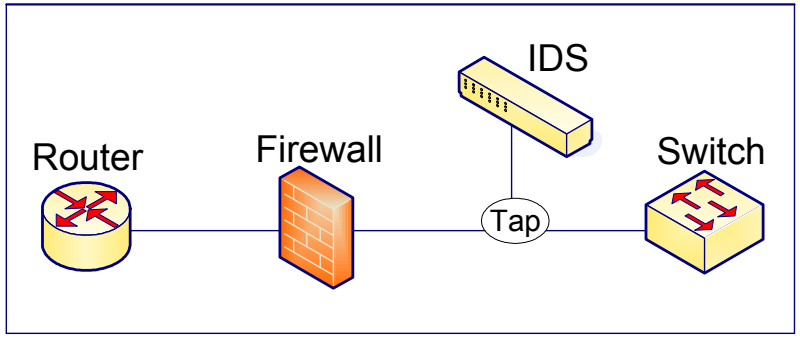
Comparing Anomaly with Misuse Detection

Anomaly Detection	Misuse Detection
☒Prone to false alarms	☑Better at detecting attacks without generating overwhelming number of false alarms
☒Often require extensive training periods	☑Do not require training
☑ Can often detect new and not previously known attacks	☒Must be updated with signatures of new attacks
☑ Can serve as source of information for misuse detectors	☒Could fail to detect variants of already defined attacks

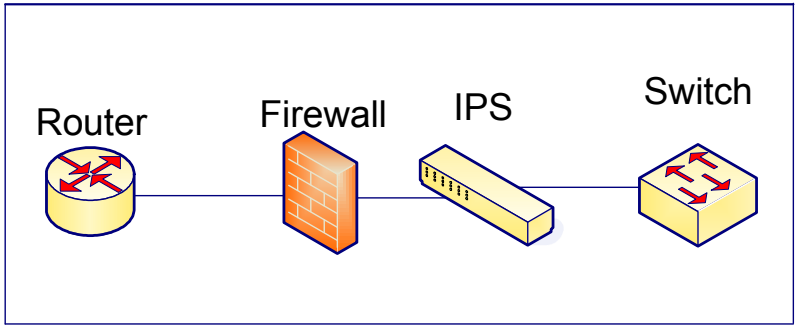
INTRUSION PREVENTION USING GAME THEORY

Countering Intrusions

IDS Mode



IPS Mode

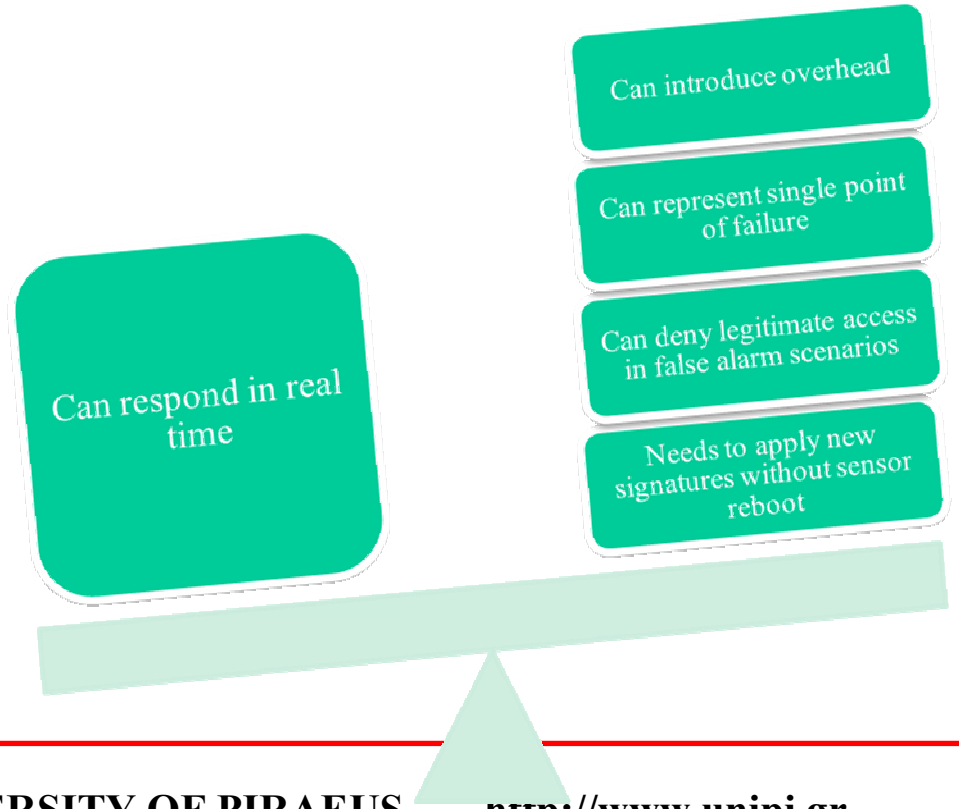


- Intrusion Detection Systems (IDS)
  - Monitor networks, looking for indications of malicious activity
  - Have mainly passive responses
- Intrusion Prevention Systems (IPS)
  - Positioned inline
  - Can respond in real-time
  - Proactively block detected attacks

Is Intrusion Prevention the answer?

Advantages

Challenges



The insider threat

- "*...a malfeasant user that falls in one of two categories: traitors and masqueraders*"
- The pie of damage caused by attacks is divided more or less equally between insiders (34%), outsiders (37%) and unknown (29%).
- Hard to detect insider attacks primarily because of insiders' privileges.

Insiders' activity

- Normally, in accordance with commitments and duties, but occasionally with mistakes that may cause damage.
- When attacking, the attack is based on a well-prepared plan.
- An insider either acts normally (N), or makes mistakes (M), or acts at a pre-attack phase (P) or attacks (A).

Approaches used so far

- Host-based user profiling (command sequence analysis)
- Network-based sensors
- Most prospective approach: User profiling that uncovers intentions.

Game theory

- Intrusion prevention is an interaction between a user and the Intrusion Prevention System (IPS) that protects a Target System (TS).
- The discipline that studies interactive situations is Game Theory.

Our approach

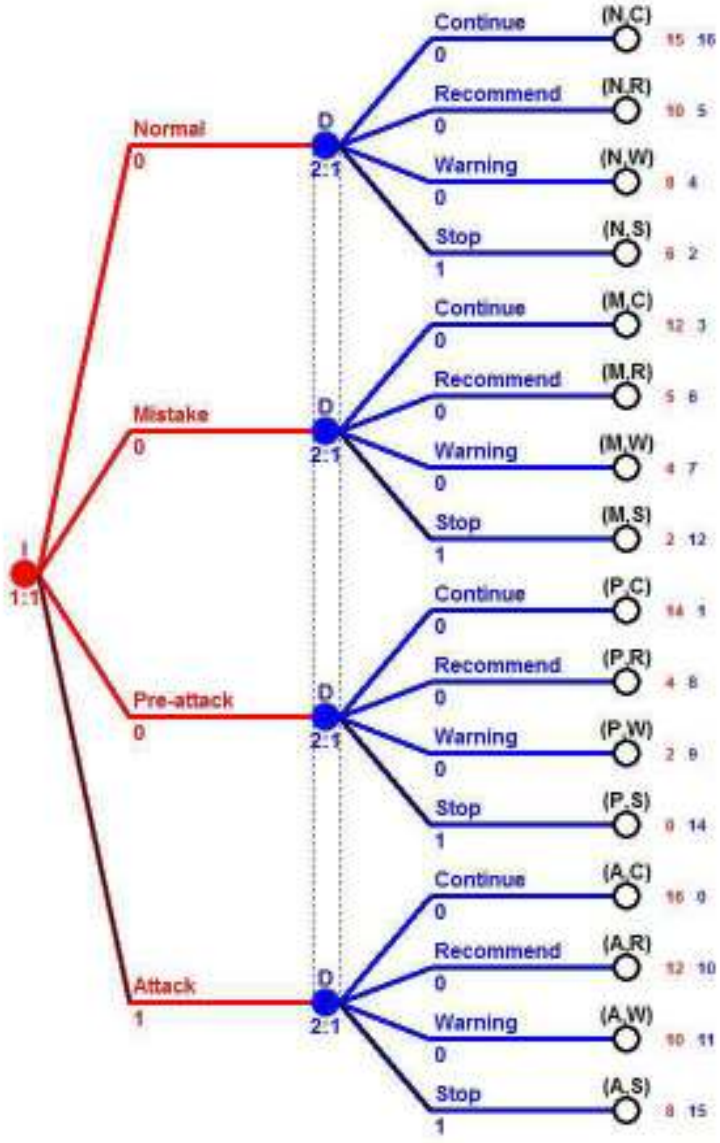
- A game (stage game) is constructed to model the interaction between user and IPS.
- An infinitely repeated game, based on the stage game, is constructed.
- The solutions to the stage game and to the repeated game are given and interpreted.

The outcome is…

- A mechanism to prevent insider intrusions by determining user intentions and thus predicting future behavior.

The game

- The interaction is modeled as a 2-player non-cooperative game, with two players: I and P.
- I is an insider with strategy set {N, M, P, A}.
- P is the IPS with strategy set {C, R, W, S}.


I
1:1
Normal
0
Mistake
0
Pre-attack
0
Attack
1
D
2:1
D
2:1
D
2:1
D
2:1
Continue
0
Recommend
0
Warning
0
Stop
1
(N,C)
15 16
(N,R)
10 5
(N,W)
8 4
(N,S)
6 2
(M,C)
12 3
(M,R)
5 6
(M,W)
4 7
(M,S)
2 12
(P,C)
14 1
(P,R)
4 8
(P,W)
2 8
(P,S)
0 14
(A,C)
16 0
(A,R)
12 10
(A,W)
10 11
(A,S)
8 15

Preferences and payoffs

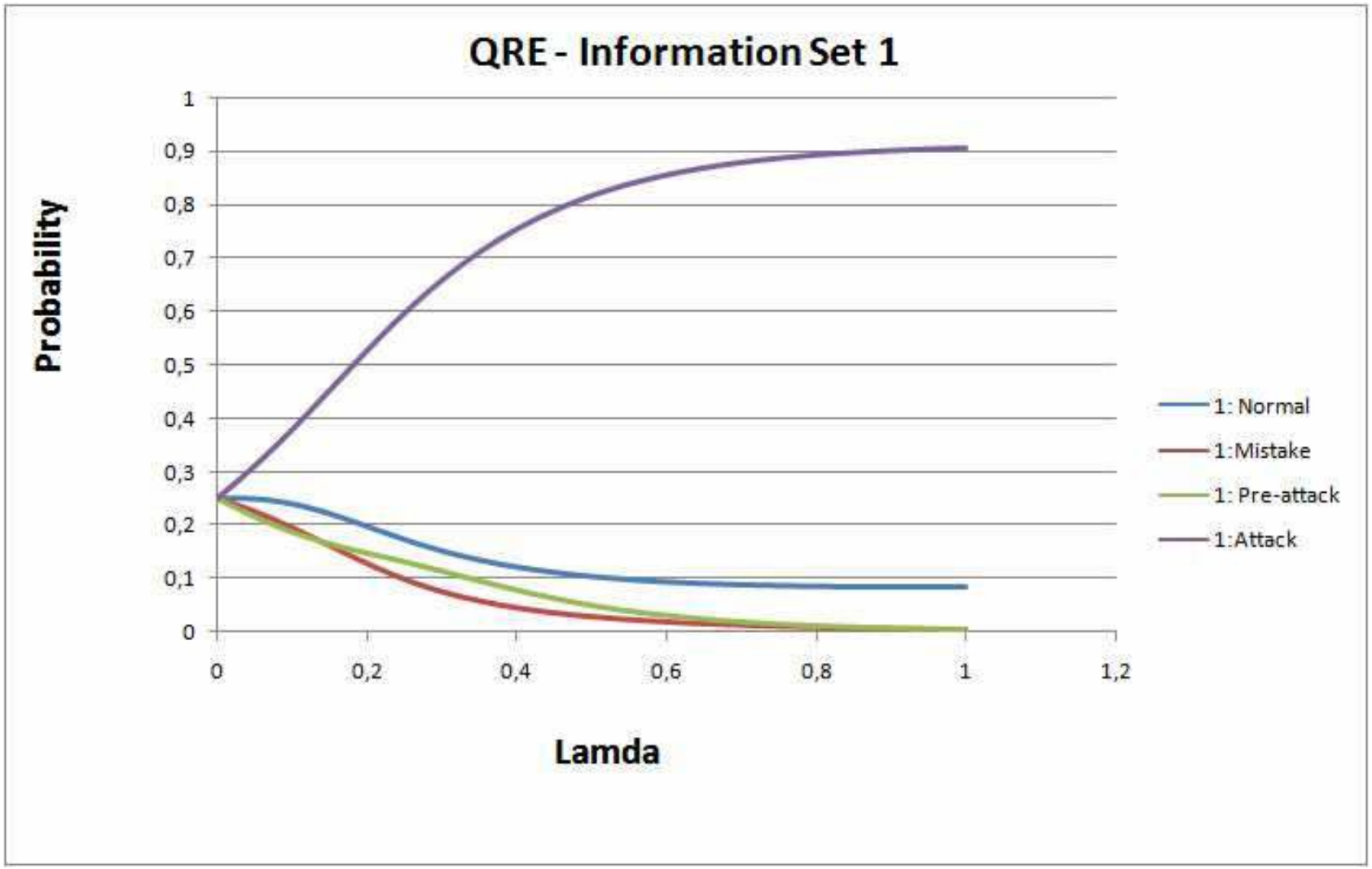
- I: (P,S) < (M,S) ~ (P,W) < (M,W) ~ (P,R) < (M,R) < (N,S) < (N,W) ~ (A,S) < (N,R) ~ (A,W) < (M,C) ~ (A,R) < (P,C) < (N,C) < (A,C)
- P: (A,C) < (P,C) < (N,S) < (M,C) < (N,W) < (N,R) < (M,R) < (M,W) < (P,R) < (P,W) < (A,R) < (A,W) < (M,S) < (P,S) < (A,S) < (N,C)

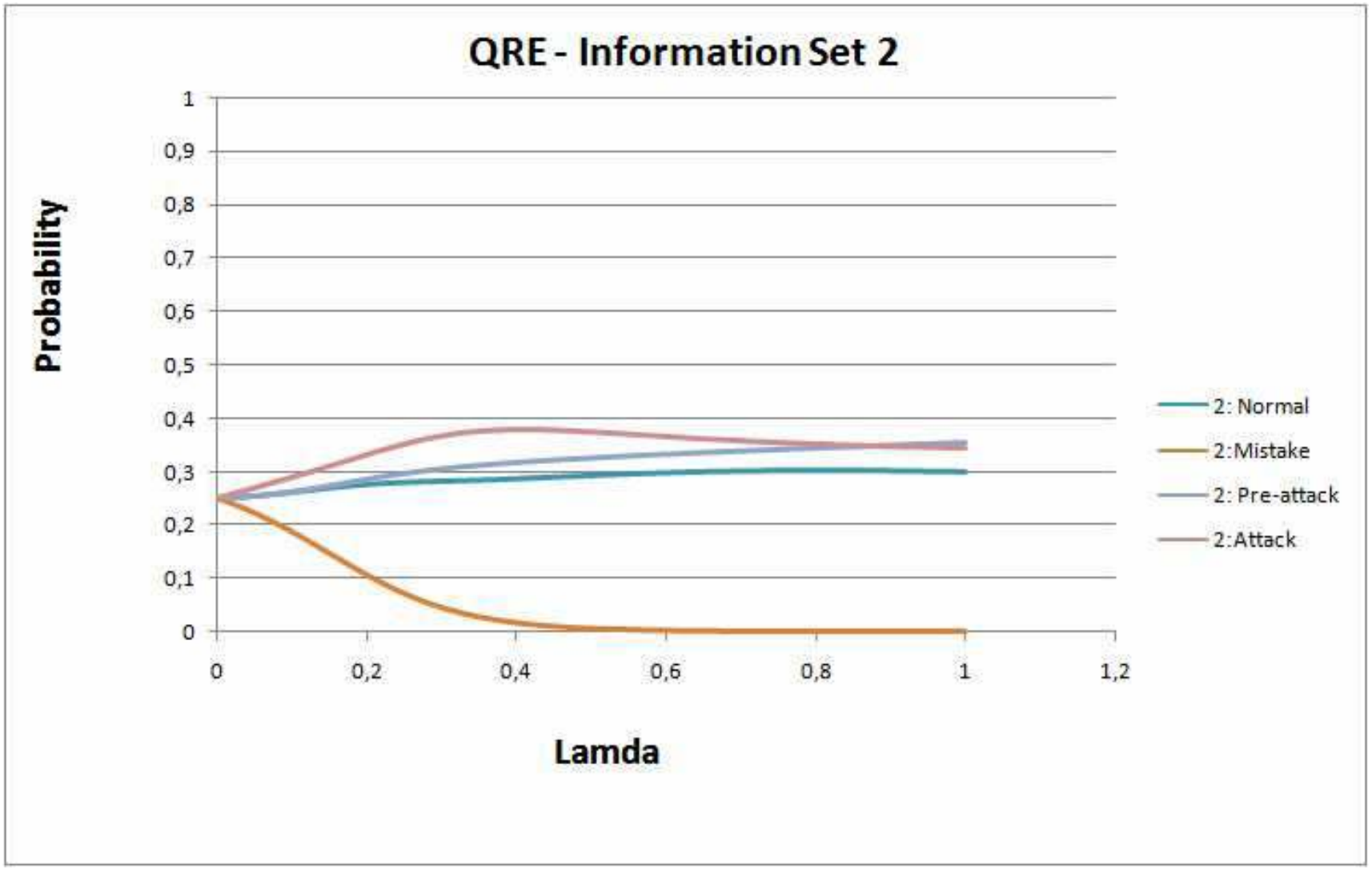
Solving the game

- Nash equilibrium: A set of players' decisions that results in an outcome such that no player has any reason to deviate from his choices, given that all players do the same.
- NE corresponds to (A,S), with payoffs (8,15).
- Another strategy profile (N,C) exists, with higher payoffs (15,16). This is both Pareto dominant (its outcome is higher) over (A,S) and Pareto efficient (no other strategy yields all players higher payoff).
- Thus, the NE would not definitely be the players' choice for ever.

Quantal Response Equilibria

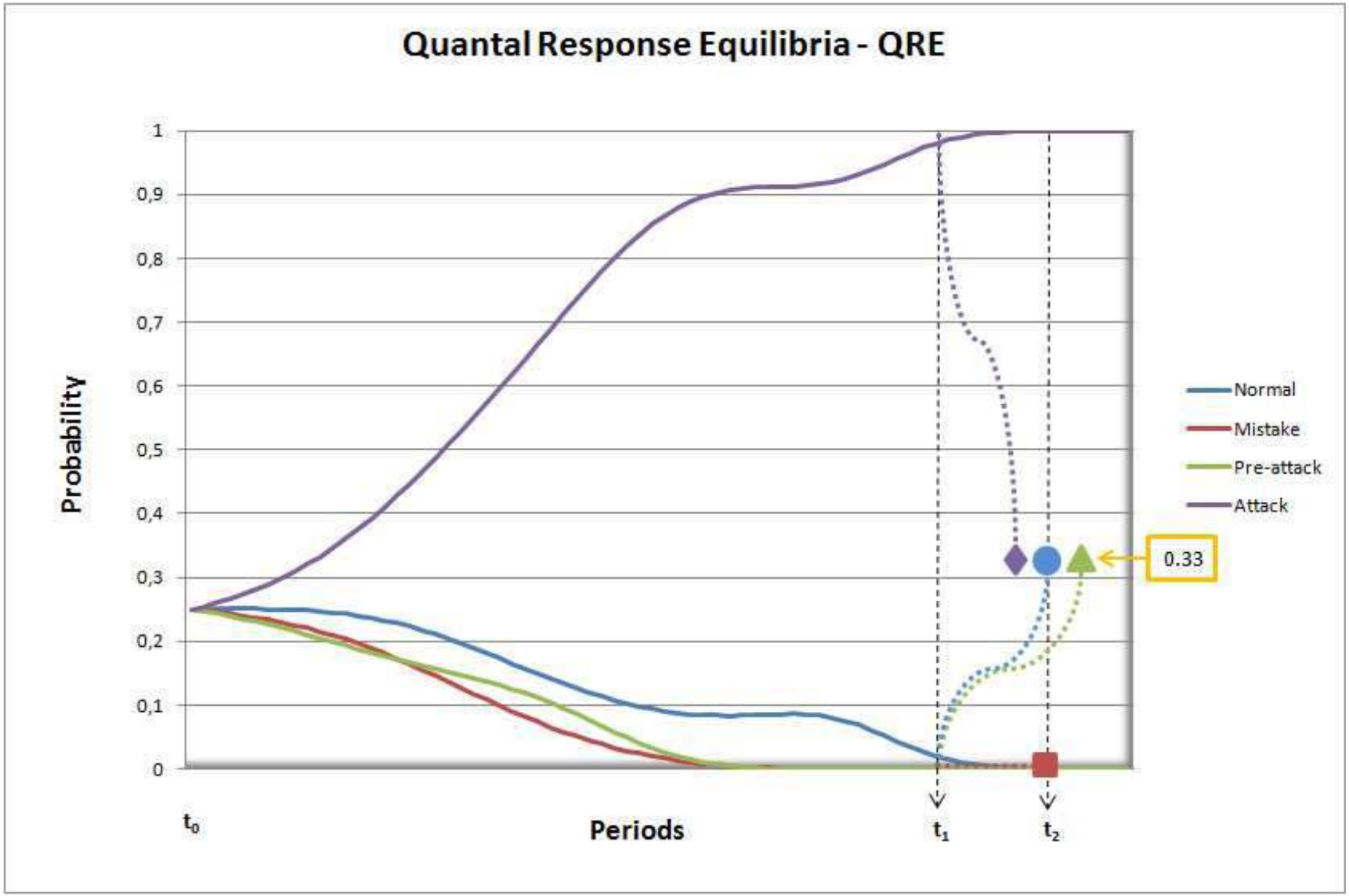
- A generalization of NE that give reasons why players deviate from the equilibrium path.
- QRE for one-shot game verifies NE.
- Re-design the game as one where I moves first, P responds and I moves again. This produces 4 NEs.
- To determine the actual set of moves, calculate QREs using equal prior strategy probabilities.


QRE - Information Set 1
Probability
1
0,9
0,8
0,7
0,6
0,5
0,4
0,3
0,2
0,1
0
0
0,2
0,4
0,6
0,8
1
1,2
Lamda
1: Normal
1:Mistake
1: Pre-attack
1:Attack


QRE - Information Set 2
Probability
1
0,9
0,8
0,7
0,6
0,5
0,4
0,3
0,2
0,1
0
0
0,2
0,4
0,6
0,8
1
1,2
Lamda
2: Normal
2:Mistake
2: Pre-attack
2:Attack

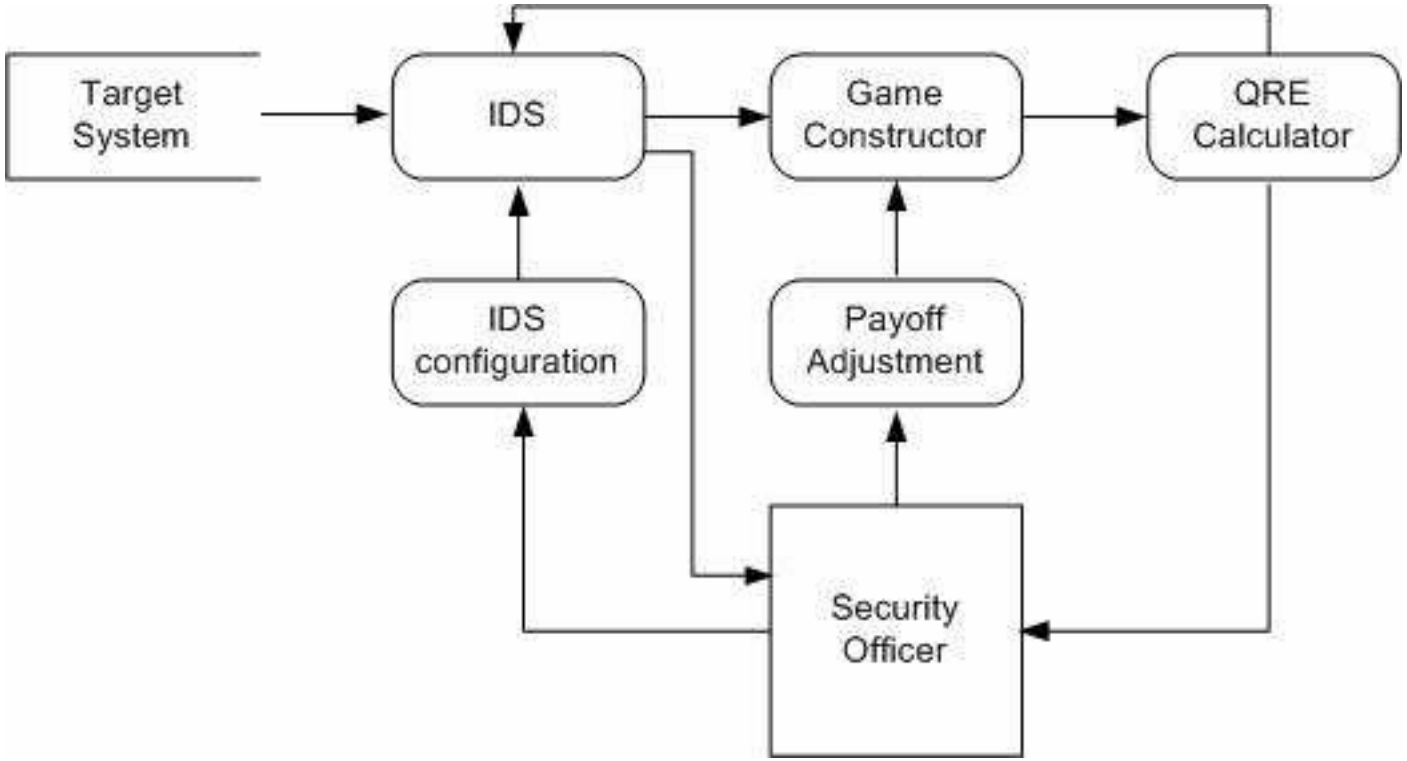
Interpreting the results

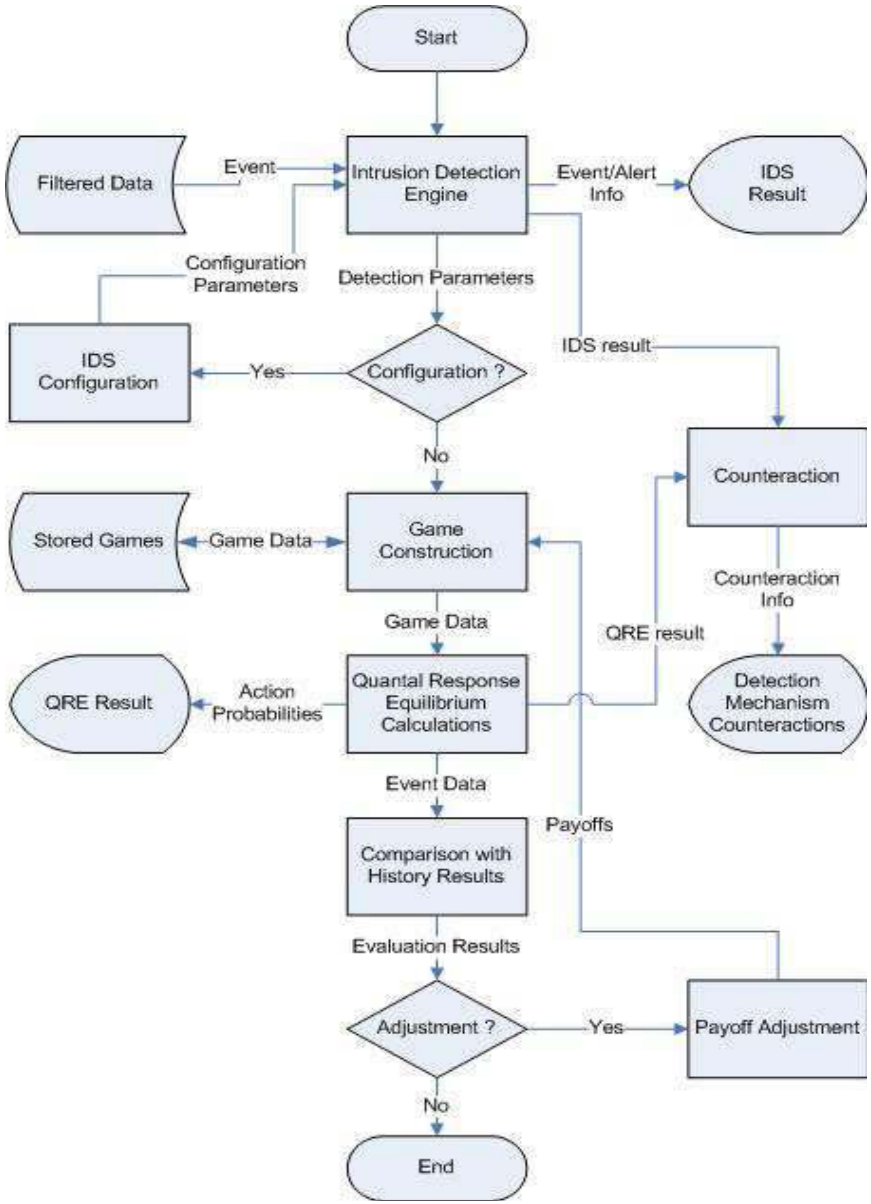
- Suppose that I moves first by electing A. If P responds with anything other than S, I has equal probabilities to choose between N, P or A as a second move.
- Hence, P maybe has a second chance.


Quantal Response Equilibria - QRE
Probability
1
0,9
0,8
0,7
0,6
0,5
0,4
0,3
0,2
0,1
0
Normal
Mistake
Pre-attack
Attack
0.33
t0
Periods
t1
t2

Applicability

- Option 1: Develop a new IPS
- Option 2: Use an existing event detector (part of an existing IDS) and develop the remaining parts.
- A game-based prevention algorithm is needed.


Target System
IDS
Game Constructor
QRE Calculator
IDS configuration
Payoff Adjustment
Security Officer


Start
Filtered Data
Event
Intrusion Detection Engine
Event/Alert Info
IDS Result
Configuration Parameters
Detection Parameters
IDS Configuration
Yes
Configuration ?
IDS result
No
Counteraction
Stored Games
Game Data
Game Construction
Counteraction Info
Game Data
QRE result
QRE Result
Action Probabilities
Quantal Response Equilibrium Calculations
Detection Mechanism Counteractions
Event Data
Payoffs
Comparison with History Results
Evaluation Results
Adjustment ?
Yes
Payoff Adjustment
No
End

In a nutshell

- The systematic construction of a game model, where an insider and an IDS interact, reveals interesting findings from the combination of Game Theory with Intrusion Detection.
- The players do not play utterly rationally.
- We are able to determine how an insider will move next, and suggest reactions to an IDS against this behavior to protect the system.

REDUCING FALSE POSITIVES

The inherent problem of Intrusion Detection

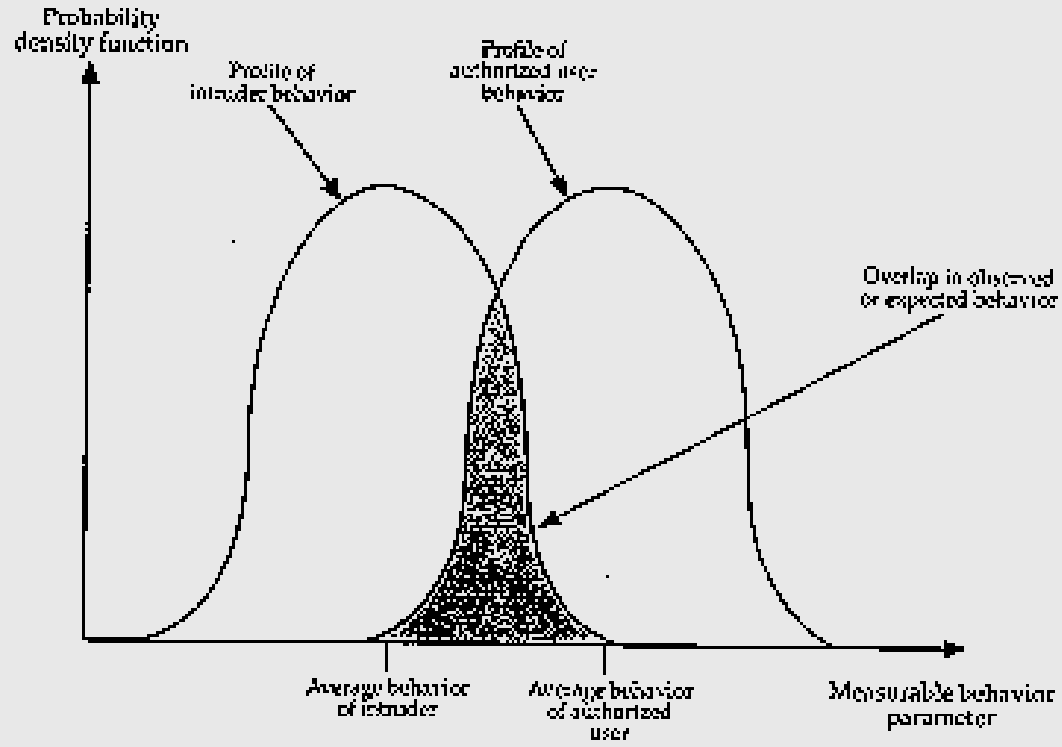


FIGURE 6.4. Profiles of Behavior of Intruders and Authorized Users

Classes of alerts

- The IDS produces an alert for a real intrusion (true positive, TP)
- The IDS produces an alert for normal activity (false positive, FP)
- The IDS does not produce an alert for a real intrusion (false negative, FN)
- The IDS does not produce an alert for normal activity (true negative, TN)

Some observations

- Real alerts (true positives) are usually observed in batches of alerts, which present similarities regarding their source or destination IP addresses.
- Real alerts are observed in higher signature-related frequency compared to the mean signature-related frequency that corresponds to their signatures.
- For a given network, every signature has a specific probability of producing false positives, which depends on the network topology.

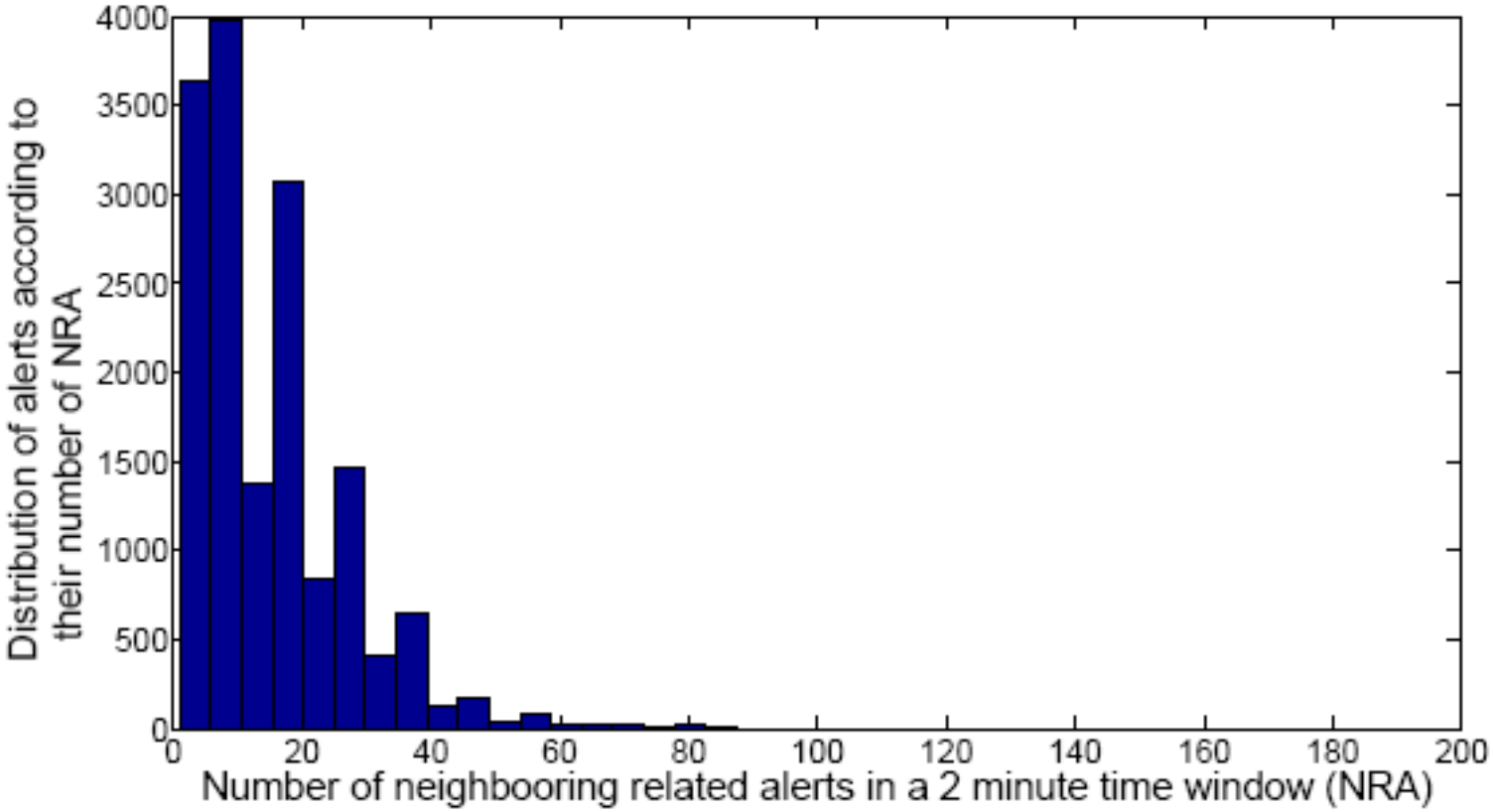


Figure 1: Distribution of Neighboring Related Alerts for False Positives

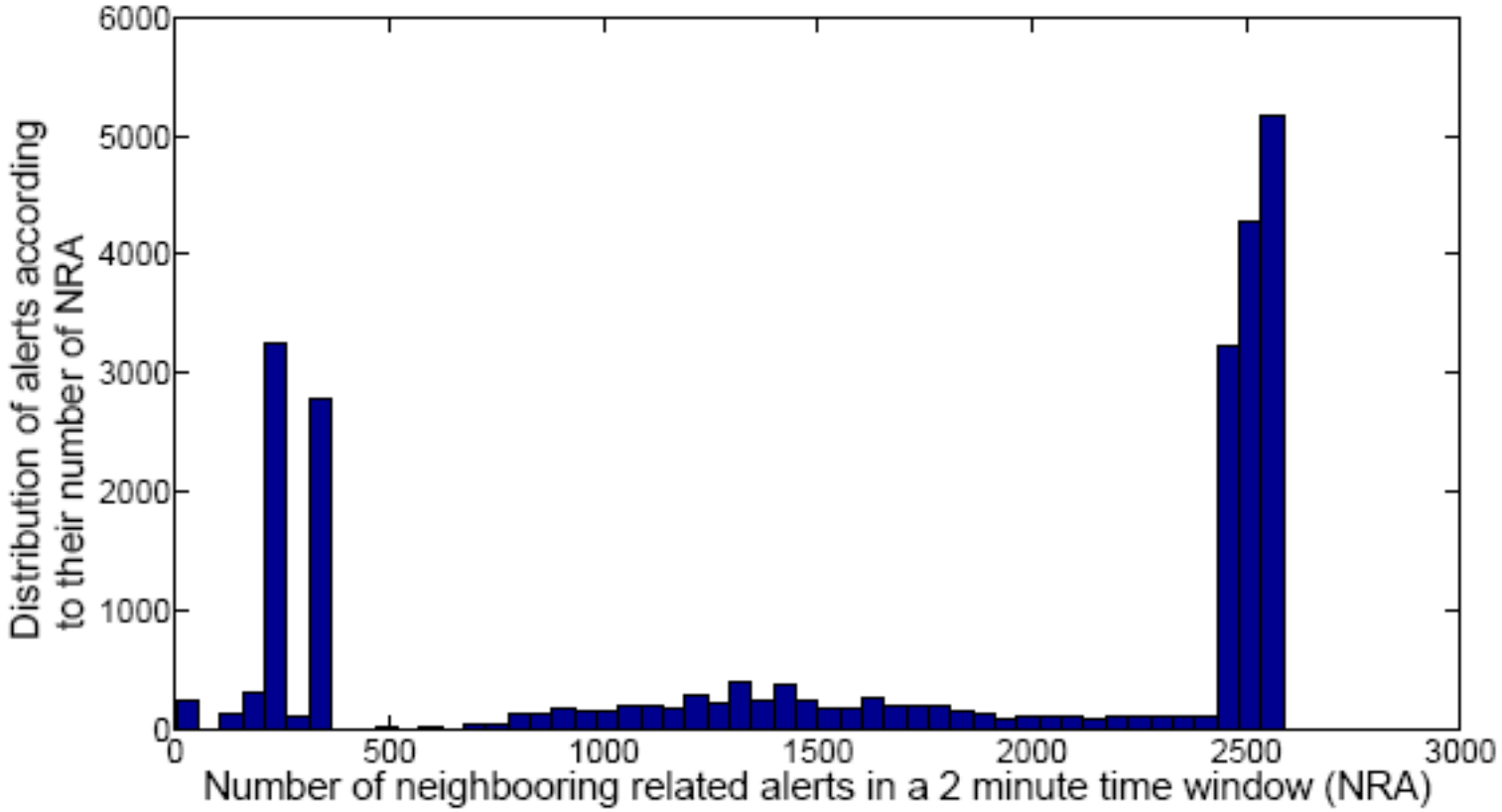


Figure 2: Distribution of Neighboring Related Alerts for True Positives

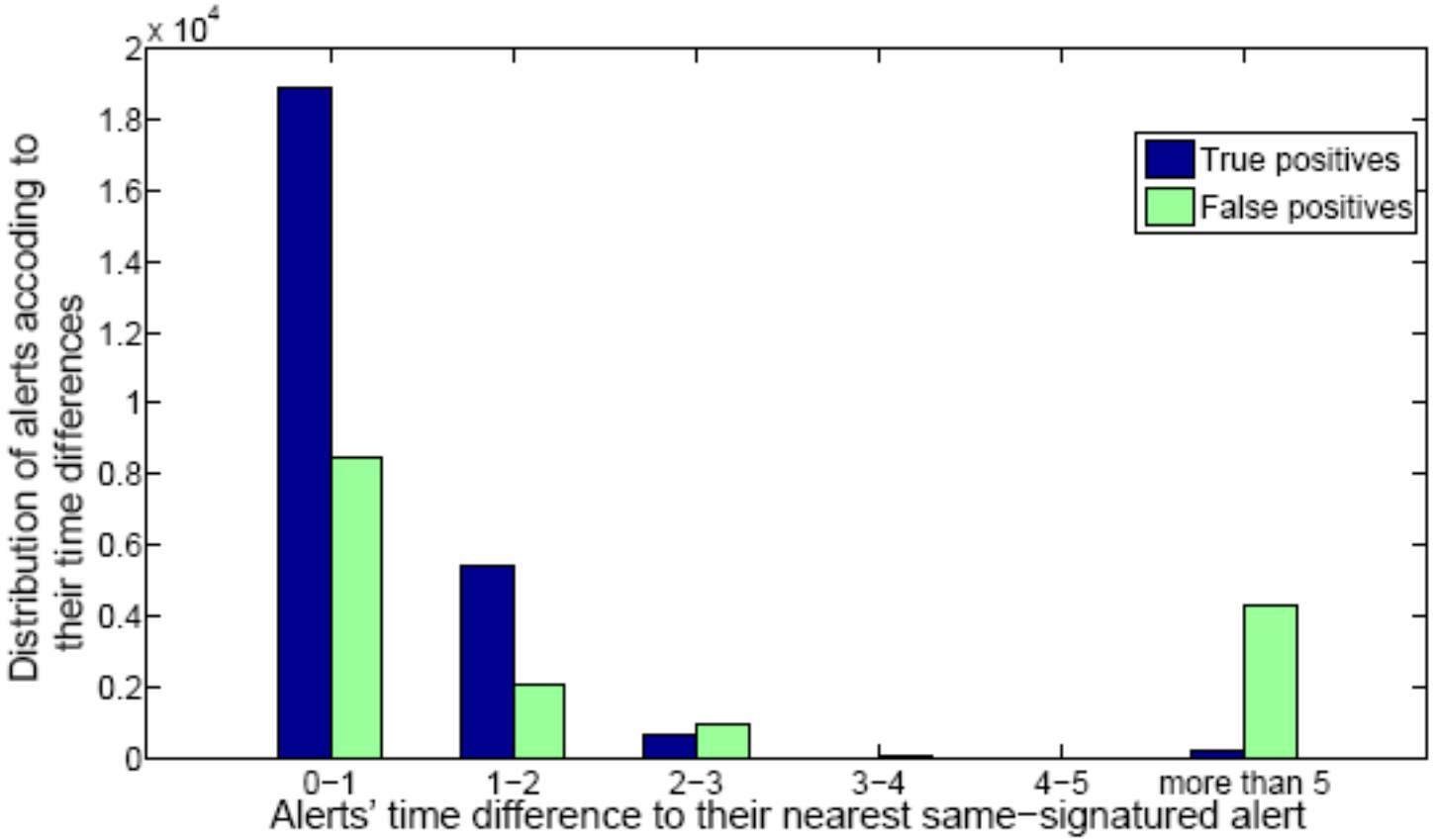


Figure 3: Distribution of alerts according to the time difference

False positives

- FP dangerous: frequent signatures in the attack-free week of the dataset
- Non FP dangerous: the opposite
- FP dangerous are the main cause of FP alerts.

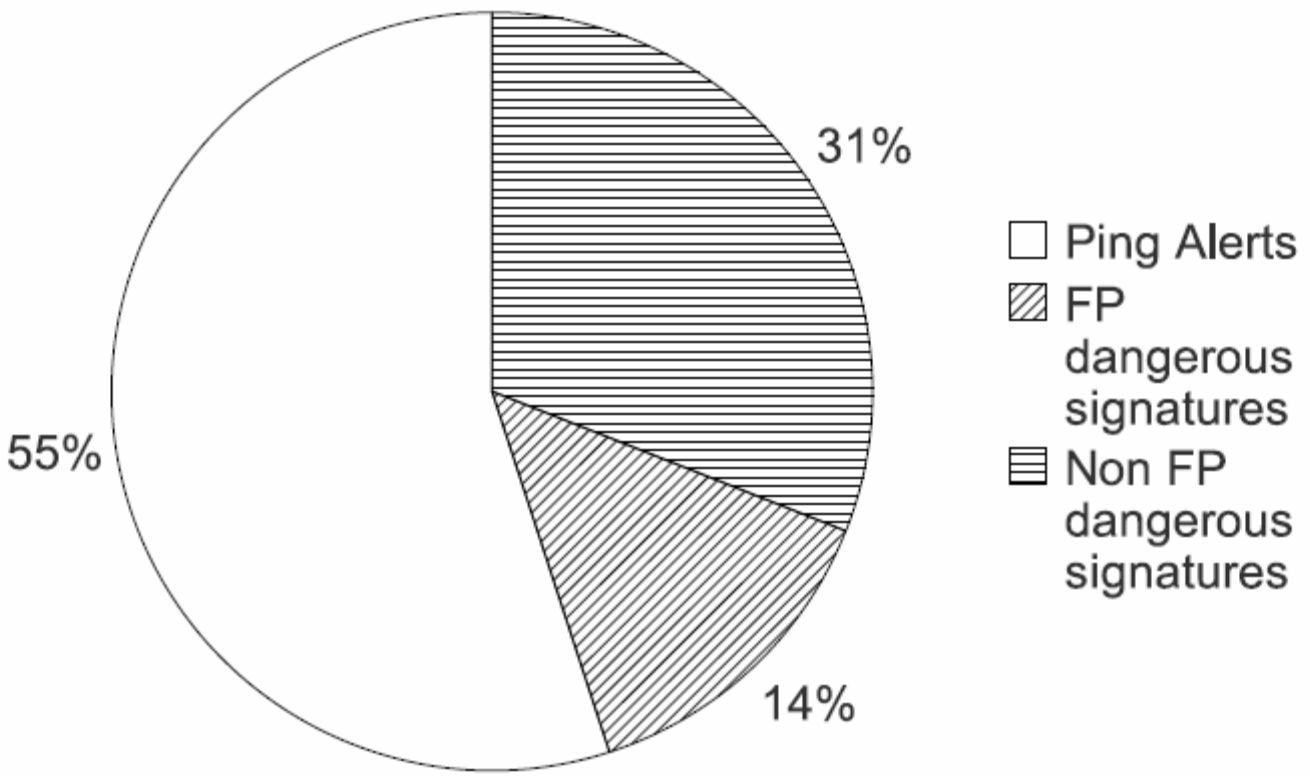


Figure 4: Distribution of TP alerts according to their signature

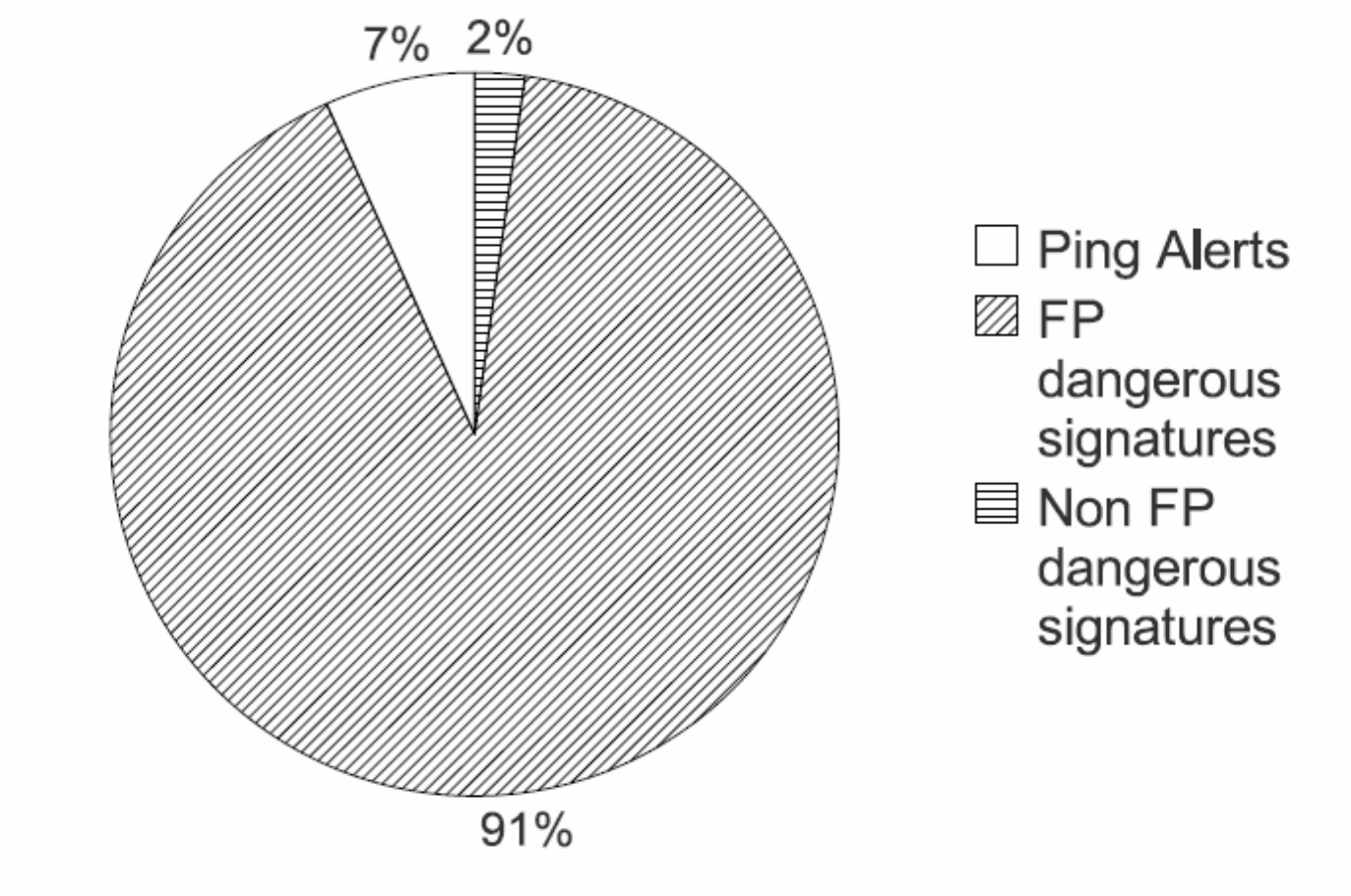


Figure 5: Distribution of FP alerts according to their signature

The filter

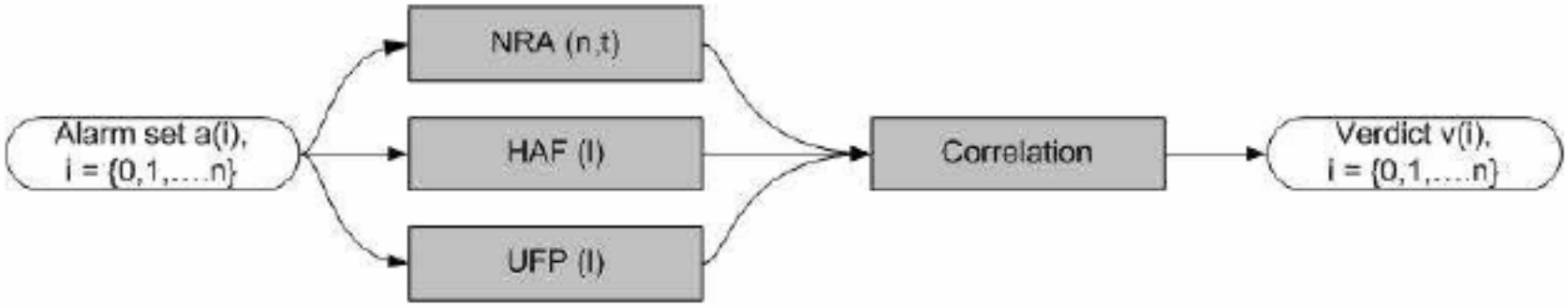


Figure 6: Filter architecture

Neighboring Related Alerts

- The NRA component is based on the observation that TPs appear in batches with similarities in the source and destination IP addresses.
- The NRA is configured by two parameters, namely t0 and n0. Parameter t0 defines the size of the time window which is used to count neighbors. Parameter n0 is used as a threshold for converting the number of neighbors to belief.
- The result of the NRA component is an array which contains the beliefs that alerts are TPs.

High Alert Frequency

- The HAF component is based on the observation that TPs are characterized by signature-related frequency which is significantly higher than the average signature-related frequency of their signature.
- The HAF is configured by a single parameter ℓ, which is used as the threshold for determining if an alert is true or false.
- The result of the HAF component is an array which contains the beliefs that alerts are TPs.

Usual False Positives

- The UFP is based on the probability that an alert is a FP, given its signature. This probability can be easily extracted from an attack-free period.
- The idea behind the UFP is to calculate the frequencies for each signature in an attack-free period.
- The result of the UFP component is an array which contains the belief that alerts are TPs.

Combination

- A combined belief can be produced by picking the maximum, the average or the minimum of the three beliefs. Then the combined belief must be compared against a threshold belief.
- If the combined belief is greater than the threshold value, then the alert is considered as true.
- If it is less than the threshold value, then the alert is considered as false.

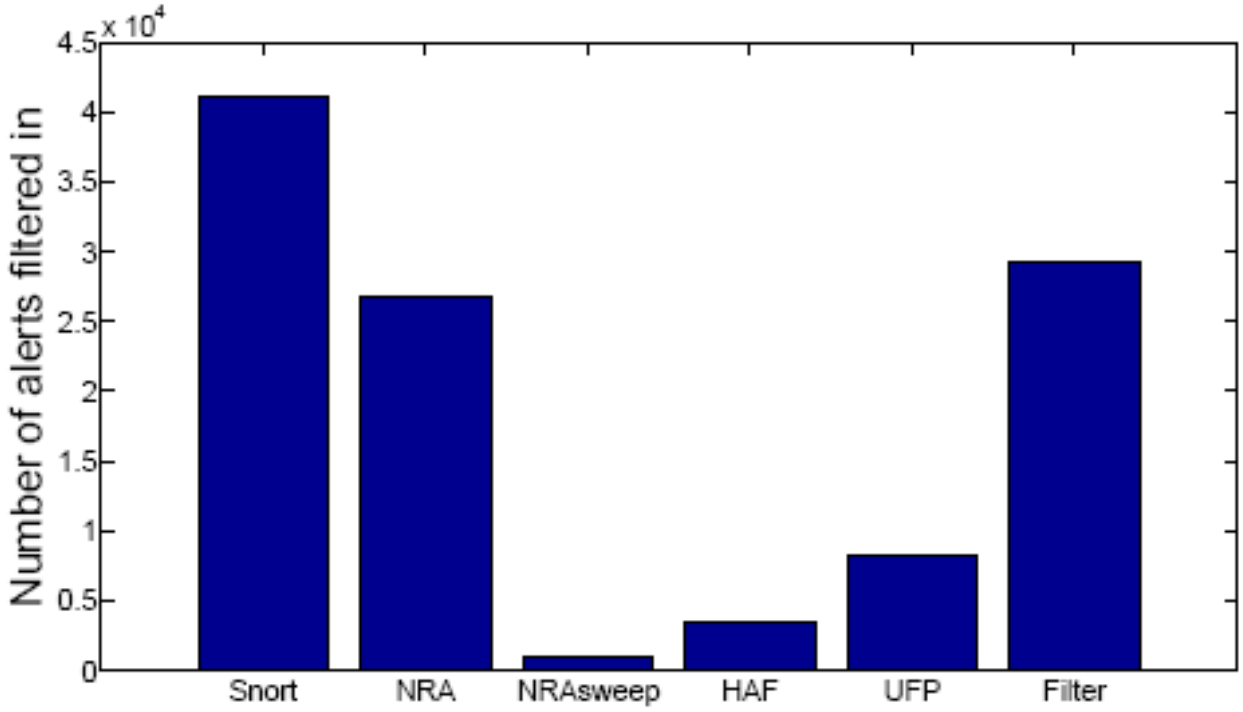


Figure 7: Number of alerts filtered in

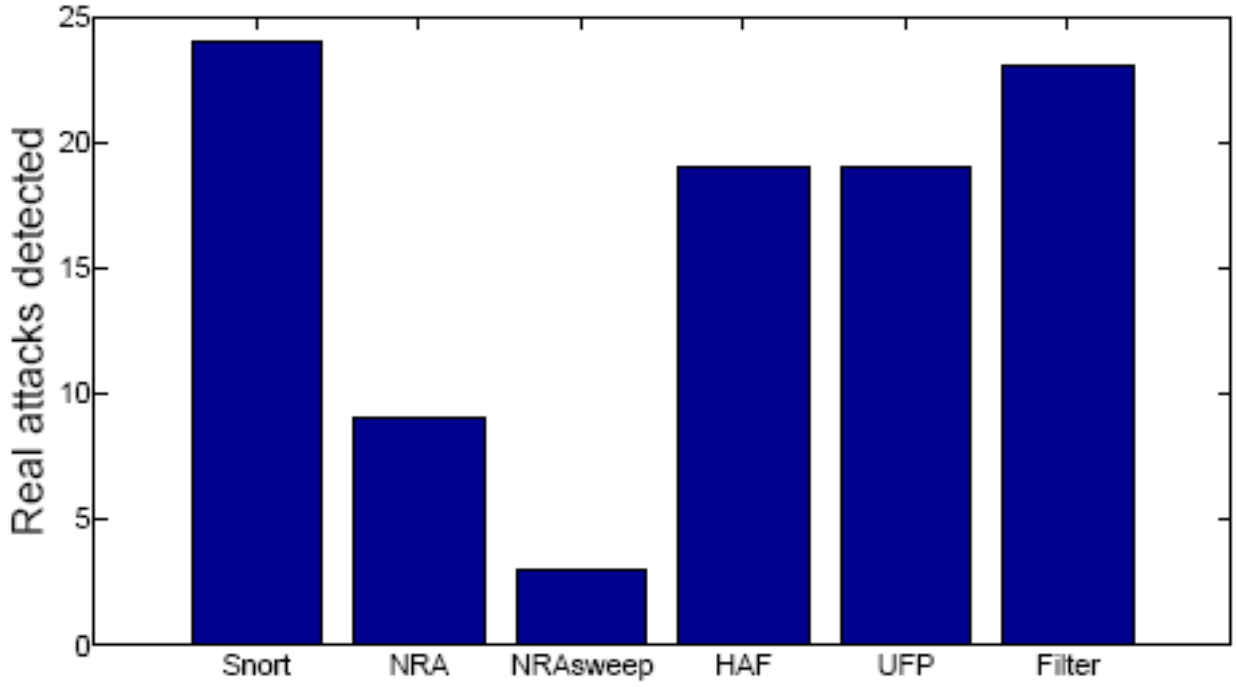


Figure 8: Number of attacks detected

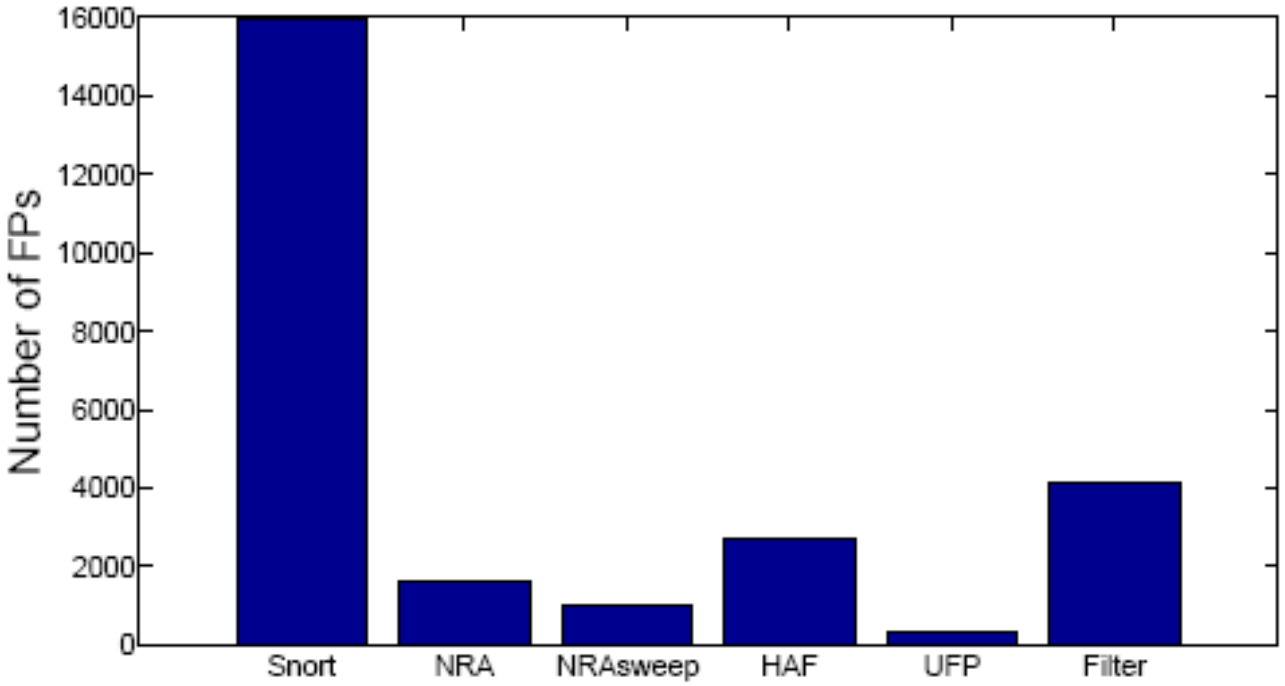


Figure 9: Number of FPs filtered in

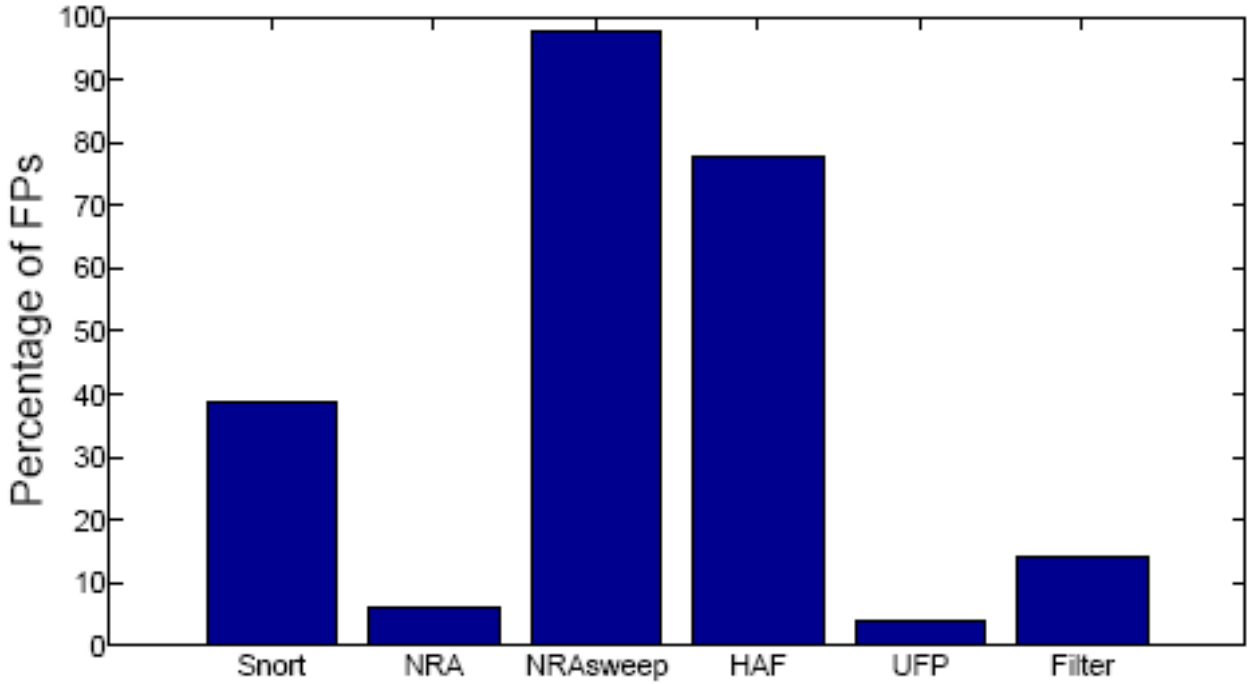


Figure 10: Percentage of FPs filtered in

In a nutshell

- The number of alerts was reduced by 29%.
- The number of FPs was reduced by 74%, while their percentage was reduced by 63%.
- These reductions were achieved while only one out of 24 real attacks detected by Snort was missed.

CLUSTERING ALERTS FROM MULTIPLE SENSORS

Multiple sensors

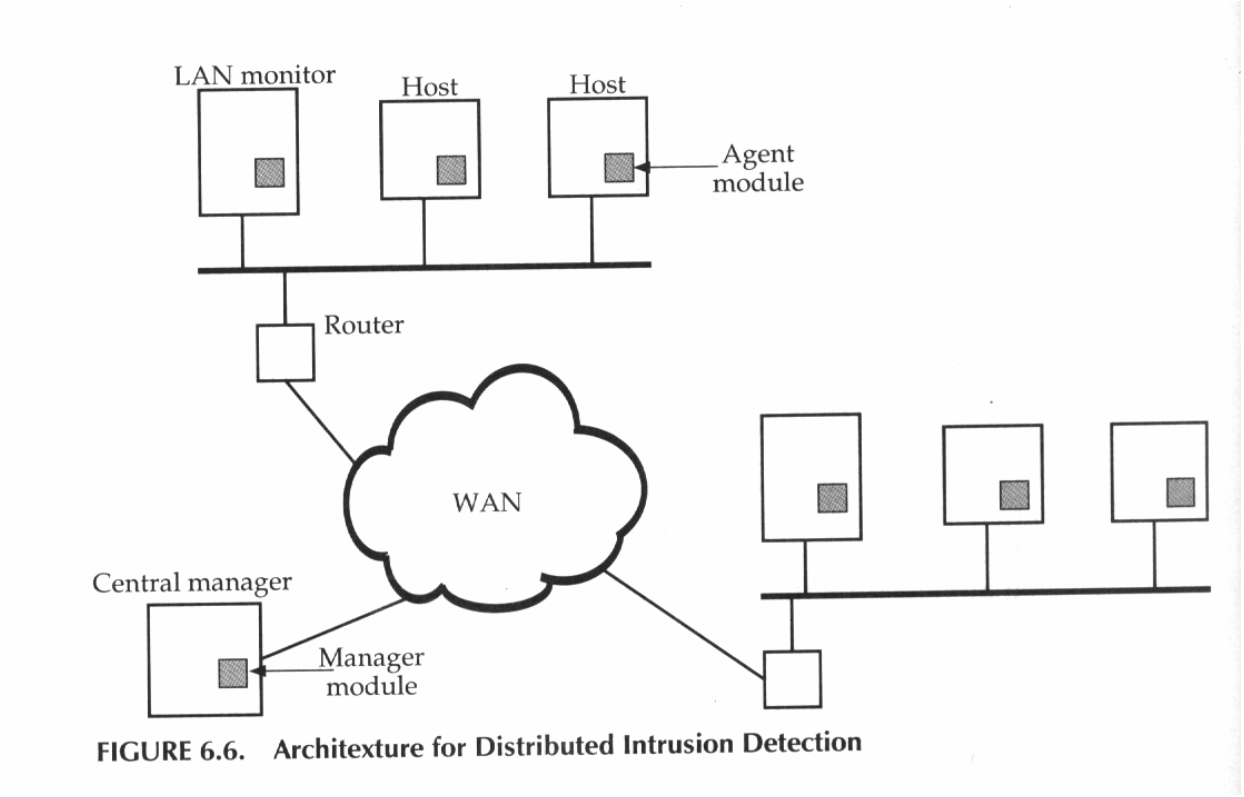


FIGURE 6.6. Architexture for Distributed Intrusion Detection

The architecture

Figure 1: Diagram of the three phases of the system

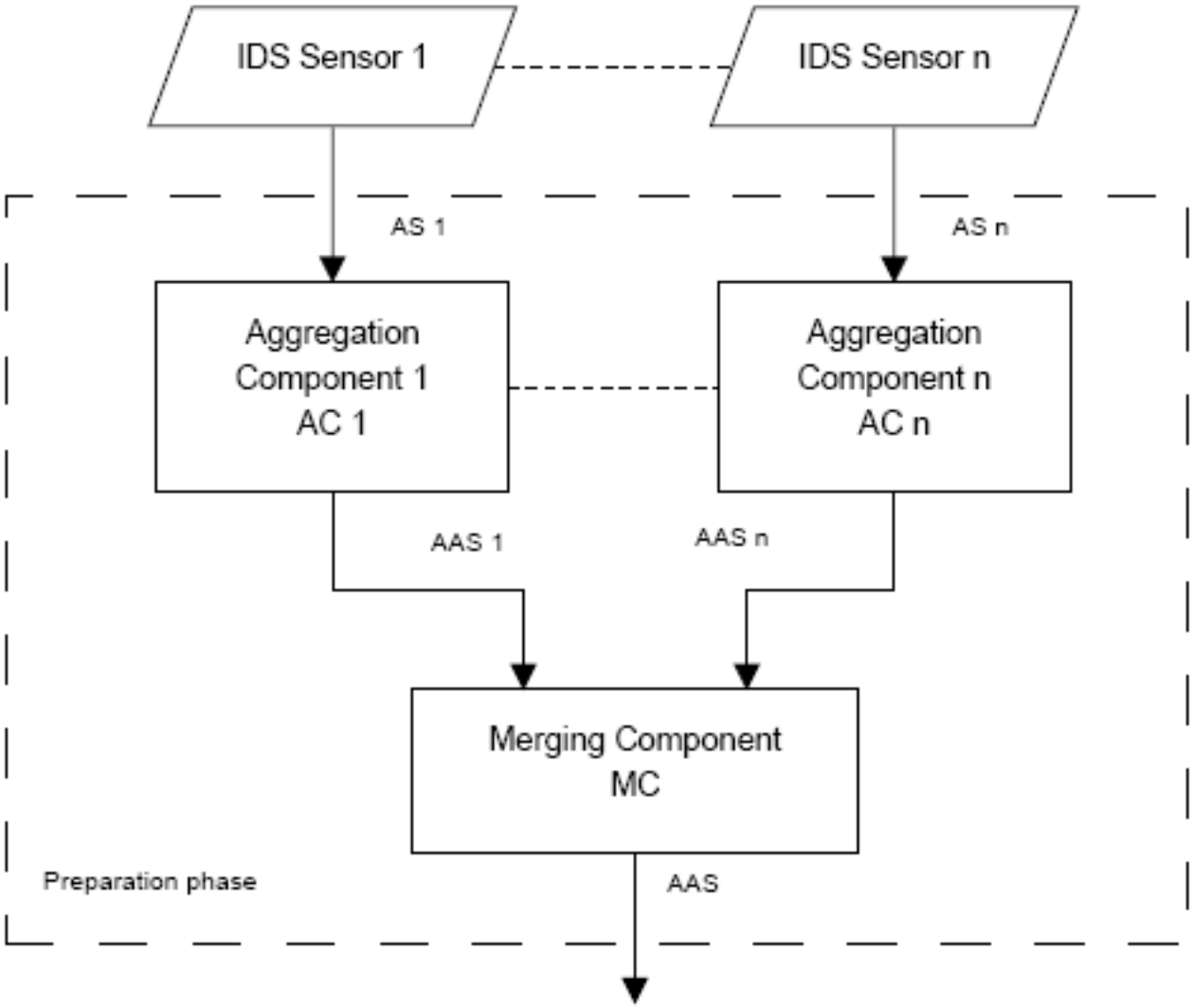


Figure 2: Preparation phase

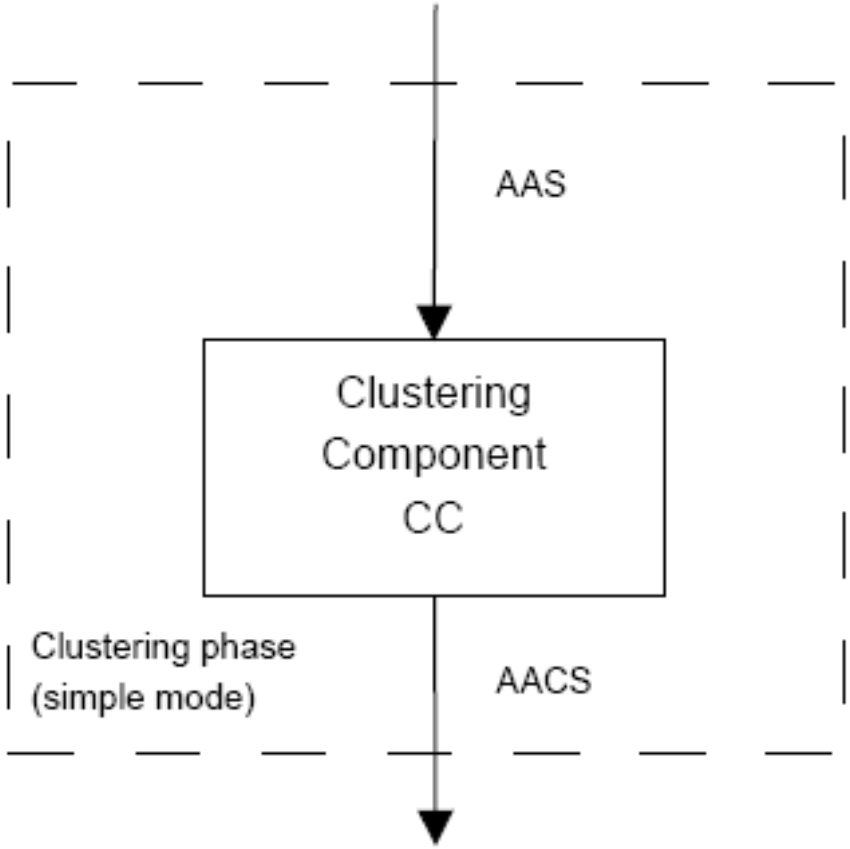


Figure 3: Clustering phase (simple mode)

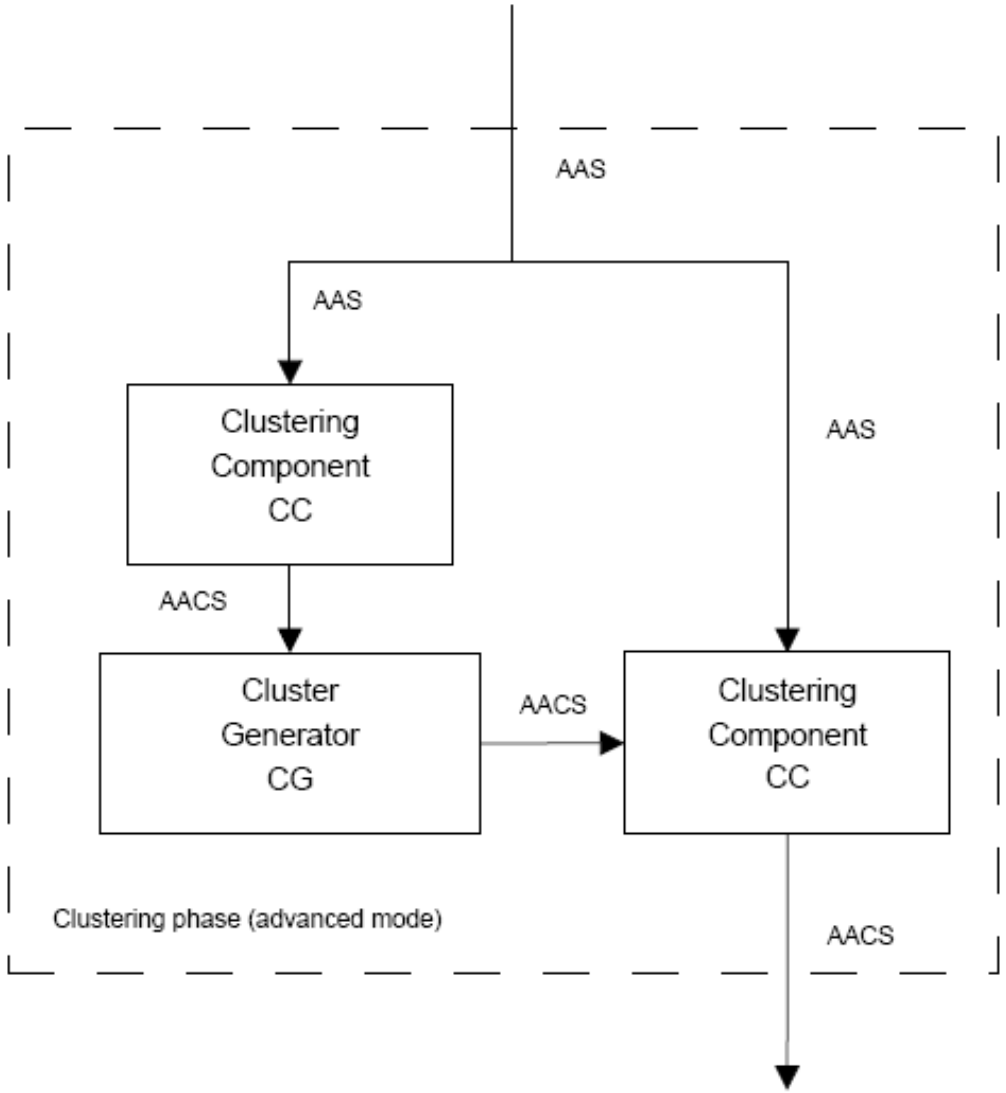


Figure 4: Clustering phase (advanced mode)

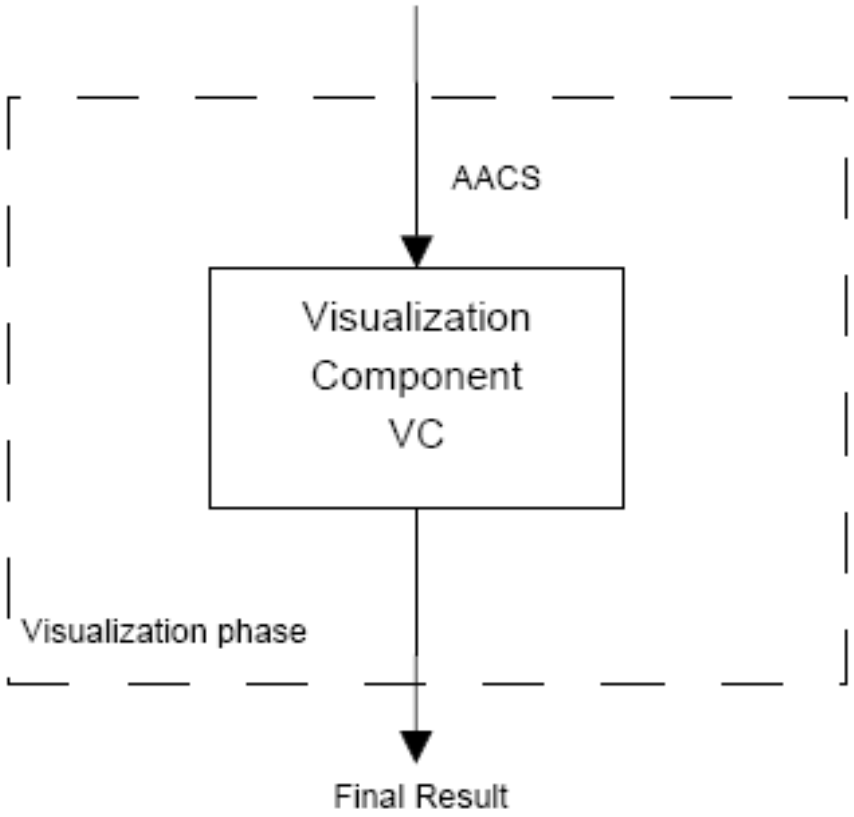


Figure 5: Visualization phase

Data formats

Data format	Acronym	Arguments
Alert	A	AID,ACL,T,SIP,DIP
Aggregated alert	AA	AID,ACL,ST,ET,SIP,DIP,NA
Aggregated alerts' cluster	AAC	AID,ACL,ST,ET,SIP,DIP,NA,AR

Table 1: Data Formats

An AS consists of all the alerts (A) that a sensor produces for a given period of time. The fields of these alerts are :

- The attack id (AID)
- The attack class (ACL)
- The time-stamp (T)
- The source IP (SIP)
- The destination IP (DIP)

Aggregation component

- Replaces a set of alerts related to the same security event with one alert.
- Criterion: same SIP, DIP and AID and close in time.
- Brute-force complexity is N2.
- By segmenting according to AID and by using indexing the complexity becomes linear.

The merging component

- Merges two or more AASs
- The result is a merged (among all sensors) AAS.
- Indexing again reduces brute-force complexity (m*n) to 20%.

The clustering component

- Creates clusters of similar AAs by checking similarity
- Similarity is computed by combing similarity values for T, AID, SIP and DIP.

The visualization component

- Produces a high level view of the clusters produced by the clustering component.
- Each cluster is depicted as an ellipse.

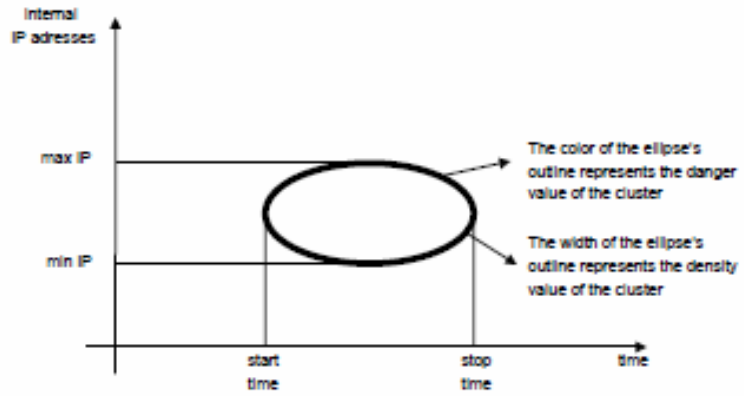


Figure 7: Visualization component analysis

The cluster generator

- Reconstructs missed events be examining neighbouring alerts using missing data theory.
- Looks for silent time windows.

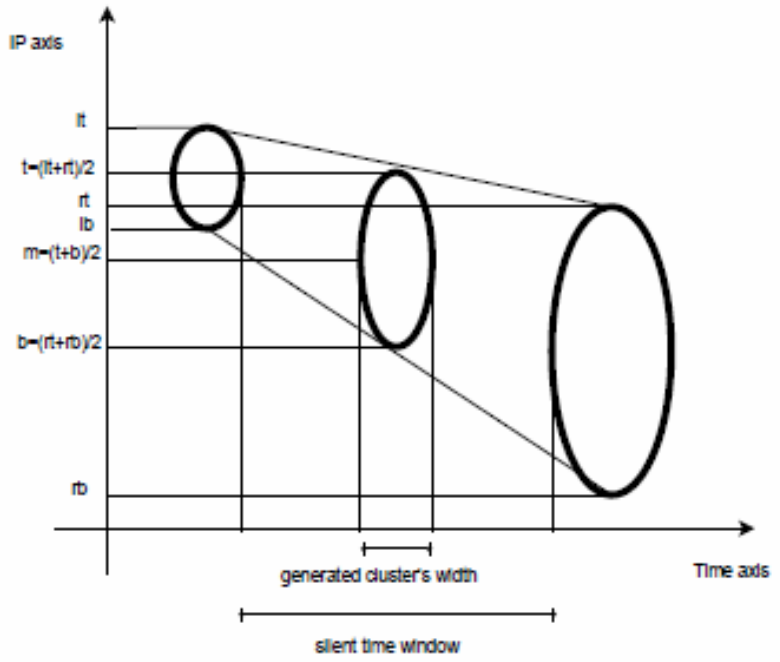


Figure 9: Geometry of clusters generation

Test scenarios

Testing Scenarios	
Name	Description
All in	Alert sets are used as obtained from Snort sensors
Phase 1 out	Alerts of Phase 1 are dropped from alert sets
Phase 2 out	Alerts of Phase 2 are dropped from alert sets
Phase 3 out	Alerts of Phase 3 are dropped from alert sets
Phase 4 out	Alerts of Phase 4 are dropped from alert sets
Phase 5 out	Alerts of Phase 5 are dropped from alert sets

Table 3: Testing platform scenarios

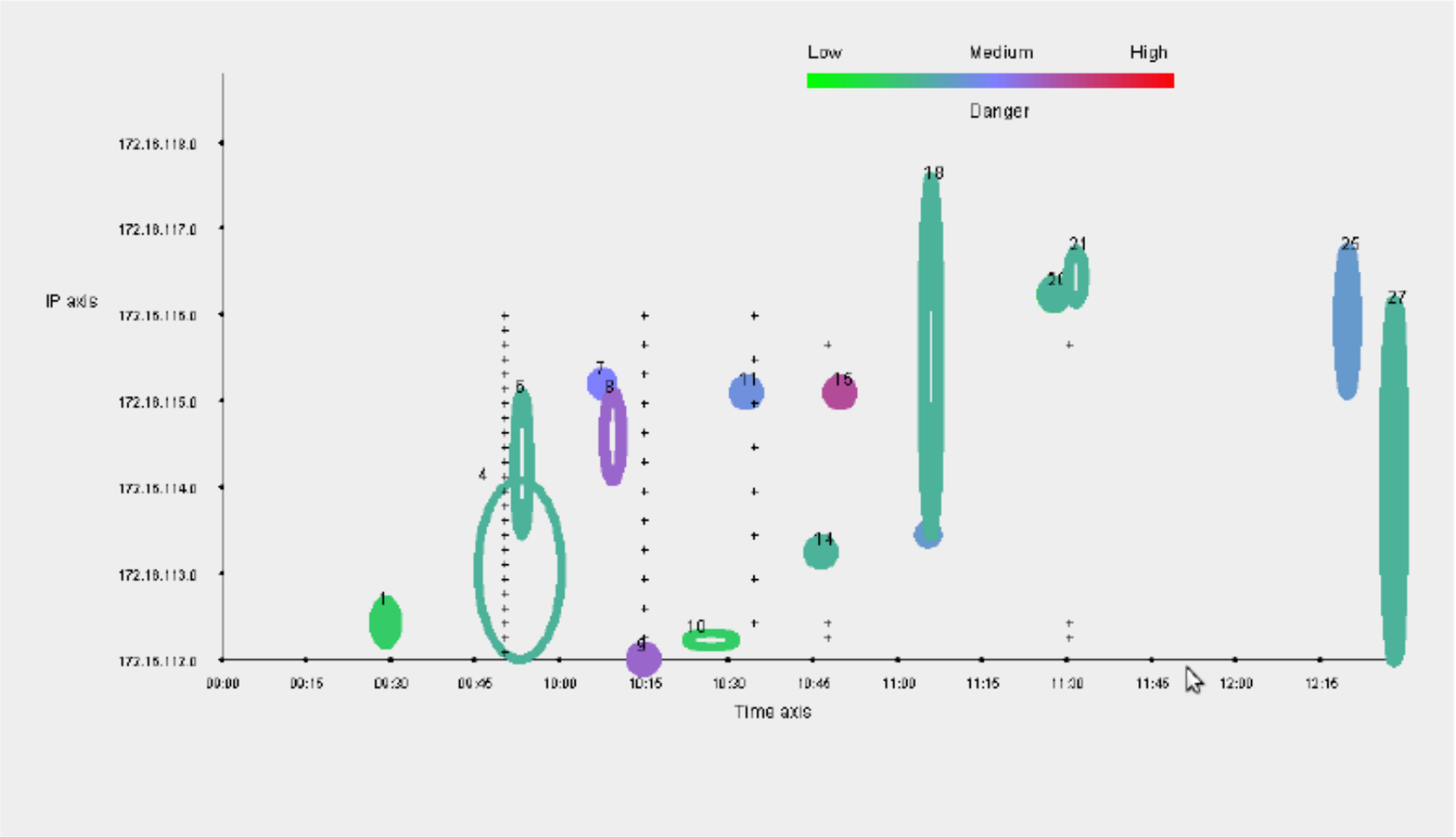
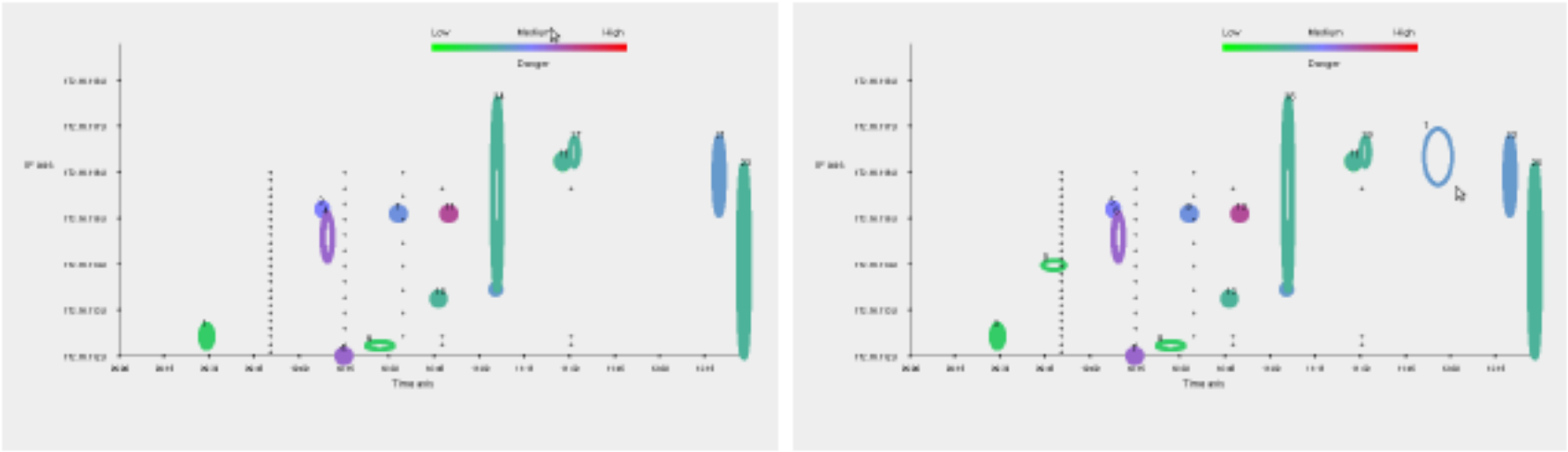


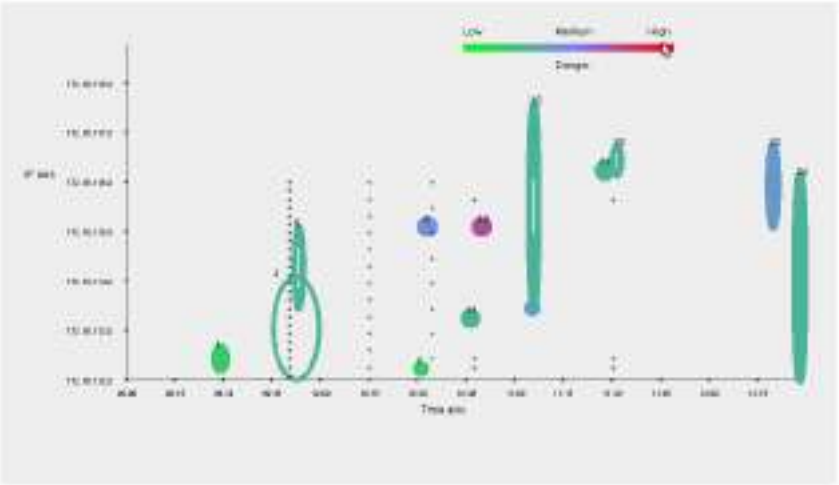
Figure 10: Results for scenario "All in"



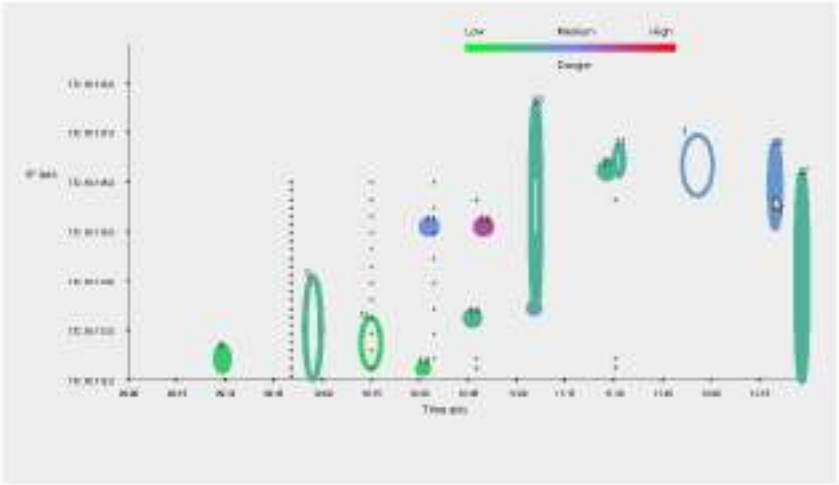
(a) Simple mode

(b) Advanced mode

Figure 11: Results for scenario "Phase 1 out"

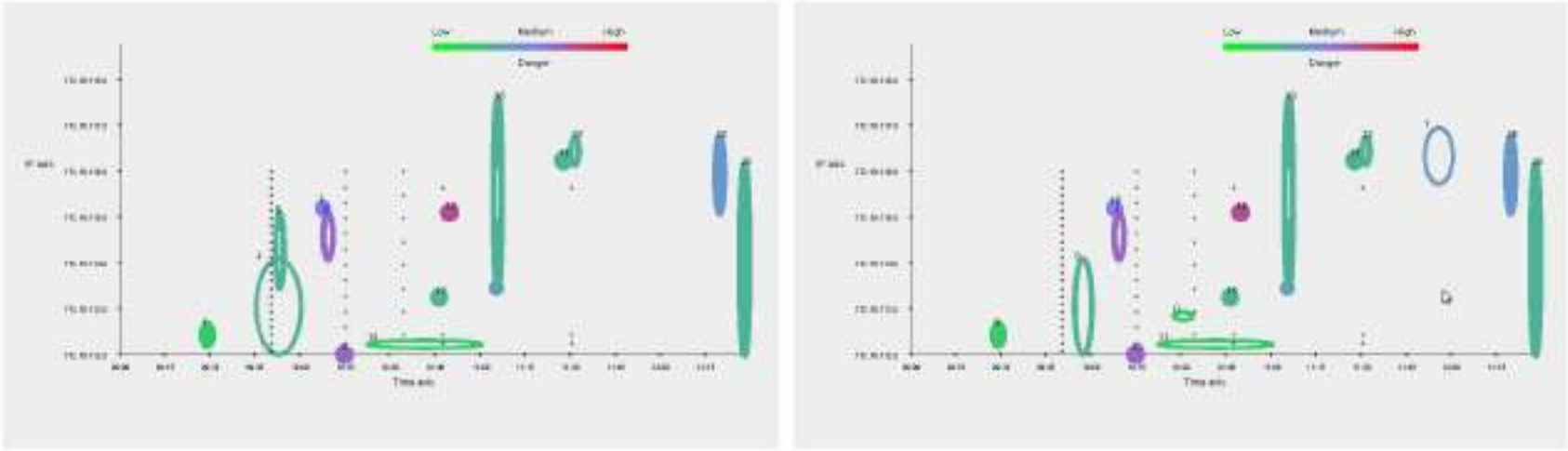


(a) Simple mode



(b) Advanced mode

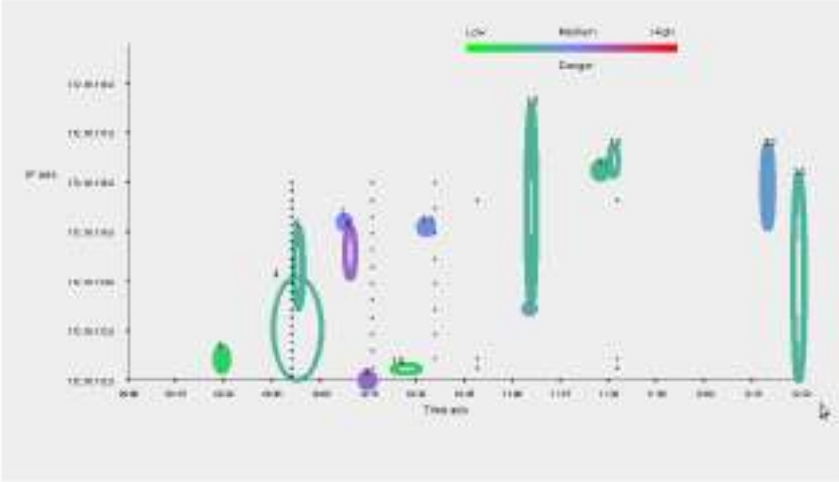
Figure 12: Results for scenario "Phase 2 out"



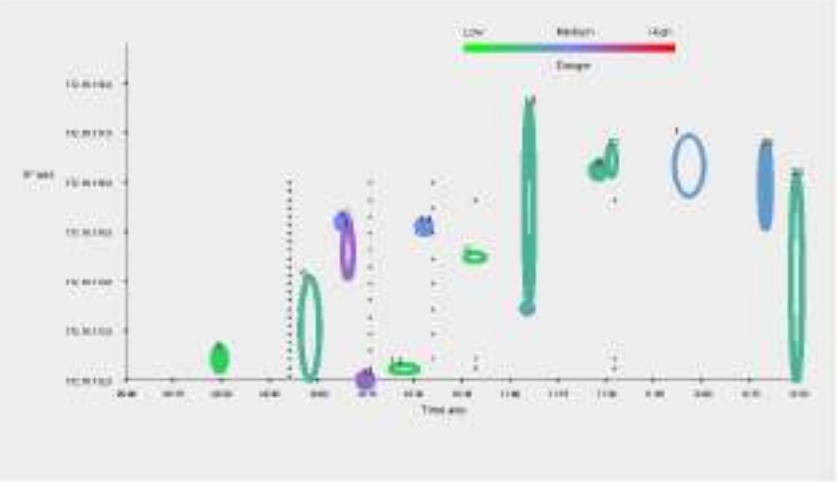
(a) Simple mode

(b) Advanced mode

Figure 13: Results for scenario "Phase 3 out"

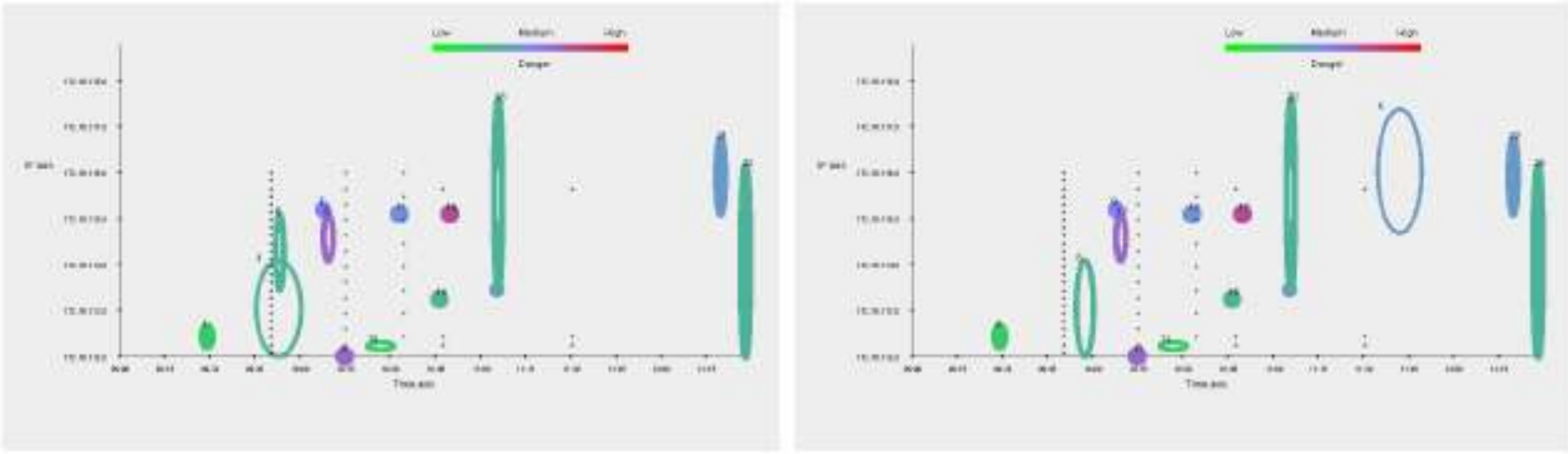


(a) Simple mode



(b) Advanced mode

Figure 14: Results for scenario "Phase 4 out"



(a) Simple mode

(b) Advanced mode

Figure 15: Results for scenario "Phase 5 out"

In a nutshell

- The system is a complete solution for high level interpretation of the low level alerts produced by multiple intrusion detection sensors. It achieves :
  - Discarding multiple identical alerts produced by specific low level events
  - Merging of the alerts produced by multiple IDS sensors (located in different parts of the network)
  - Creation of clusters that represent high level actions of the intruder
  - Generation of artificial clusters that approximate missed events
  - Visualization of the end result in a meaningful for the user manner

Conclusions

- A fascinating research area
- Several challenging open research problems

Thank you!