*IOSR Journal of Computer Engineering (IOSR-JCE)*
*e-ISSN: 2278-0661, p- ISSN: 2278-8727Volume 16, Issue 1, Ver. III (Jan. 2014), PP 01-07*
*www.iosrjournals.org*

# Detecting Anomaly IDS in Network using Bayesian Network

[1]Mrs.SumathyMuruganAsst. Prof [2]*Dr.M.SundaraRajan*Asst. Prof
[1]*Department of Computer Science*, Thiruthangal Nadar College, Chennai-51.
[2]*Department of Computer Science*, Government Arts College, Nandanam, Chennai-42.

***Abstract:*** *In a hostile area of network, it is a severe challenge to protect sink, developing flexible and adaptive security oriented approaches against malicious activities. Intrusion detection is the act of detecting, monitoring unwanted activity and traffic on a network or a device, which violates security policy. This paper begins with a review of the most well-known anomaly based intrusion detection techniques. AIDS is a system for detecting computer intrusions, type of misuse that falls out of normal operation by monitoring system activity and classifying it as either normal or anomalous .It is based on Machine Learning AIDS schemes model that allows the attacks analyzed to be categorized and find probabilistic relationships among attacks using Bayesian network.*

***Keywords:*** *Firewall; IDS; IDS types ; AIDS Techniques and Bayesian concept.*

## I. Introduction

Intrusion detection is the process of monitoring the events occurring in a computer system or network and analysing them for signs of possible incidents, which are violations or imminent threats of violation of computer security policies, acceptable use policies, or standard security practices [1]. Intrusion Detection Systems (IDS) are security tools that, like other measures such as antivirus software, firewalls and access control schemes, are intended to strengthen the security of information and communication systems.

CIDFCommon Intrusion Detection Framework, a working group created by DARPA in 1998 and integrated within IETF in 2000 and adopted the new acronym IDWG Intrusion Detection Working Group, defined a general IDS architecture based on the consideration of four types of functional modules.

Event Boxes:

It is composed of sensor elements that monitor the target system.

Database Boxes:

Store information for subsequent processing by analyze and response boxes.

Analysis Boxes:

Process modules for analyzing the events and detecting potential hostile behavior, So that some kind of alarm will be generated if necessary.

Response Boxes:

Main function of the block is the execution.

## II. Comparison with Firewalls and Antivirus

Mostly organization use firewall and antivirus software for the security. But still, some security gaps exist in the network from which intruders may generate intrusion. They both relate to network security; an Intrusion Detection System (IDS) differs from a firewall and antivirus.Firewall and antivirus software does not stop internal intrusion and as well as external attacks efficiently, as firewall (tunnelling process) and antivirus can be bypassed. Firewall generally works on static rules via which it filters traffic but never has ability to detect intrusion. Similarly, antivirus as signatures of malicious contents or anomalies, on the basis of which it simply compares packet pattern with its signature only. IDS, on the other hand, works in batch mode and detects intrusion after its first occurrence so that it counters for such attacks in future. [8].

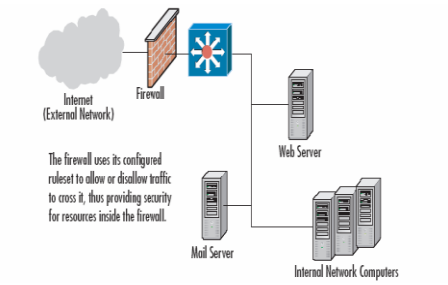


Figure 2. Basic Firewall Installation [9]

## III. Analysis Models

The Security implementation is achieved by two phase, namely

- ❖ BehaviouralAnalysis
- ❖ Knowledge Analysis

### 3.1. Behavioural Analysis:

Two types of behaviour,

- User Behaviour:

  Using this method, we need to recognize expected behaviour (legitimate use) or a severe behaviour deviation. The network must be correctly trained to efficiently detect intrusions.

- Node Behaviour:

The idea behind this approach is to measure a "baselines" of such as CPU utilization, disk activity, user logins, and file activity. It can detect the anomalies without having to understand the underlying cause behind the anomalies.

### 3.2. Knowledge Analysis:

Using this kind of intrusion detection is that we can add new rules without modifying existing ones. It is nothing but the set of rules which is formed from previous attacks.

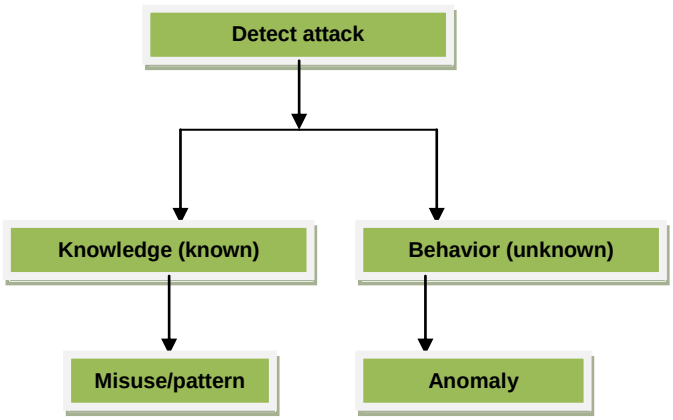


Figure 3 .Analysis of Attack

## IV. IDS Technologies

Several types of IDS technologies exist due to the variance of network configurations. Each type has advantages and disadvantage in detection, configuration, and cost.

- ❖ Network-Based Detection System
- ❖ Wireless Detection System
- ❖ Network Behaviour Anomaly Detection
- ❖ Host-Based Detection System

## V. Intruders Detection Types

Signature based detection is a pattern that corresponds to a known threat. Signature-based detection is the process of comparing signatures against observed events to identify possible incidents. *Anomaly based detection* IDS that looks at network traffic and detects data that is incorrect, not valid, or generally abnormal is called anomaly-based detection. *Stateful protocol inspection* is similar to anomaly based detection, but it can also analyze traffic at the network and transport layer and vender-specific traffic at the application layer, which anomaly-based detection cannot do.

### 5.1. Signature-Based Detection Vs Anomaly-Based Detection:

Depending on the type of analysis carried out (Analyze blocks in Figure. 1),intrusion detection systems are classified as either signature-based or anomaly-based.

Signature-based detectors (also denoted as misuse-based) seek defined patterns, or signatures, within the analyzed data. For this purpose, a signature database corresponding to known attacks is specified a priori. On the other hand, anomaly-based detectors attempt to estimate the ‘‘normal’’ behavior and generate an anomaly alarm whenever the deviation between a given observation at an instant and the normalbehavior exceeds a predefined threshold. Another possibility is to model the ‘‘abnormal’’ behavior of the system and to

raise an alarm when the difference between the observed behavior and the expected one falls below a given limit.

The main differences between these methodologies are inherent in the concepts of ‘‘attack’’ and ‘‘anomaly’’. [4]

- An attack can be defined as ‘‘a sequence of operations that puts the security of a system at risk’’.
- An anomaly is just ‘‘an event that is suspicious from the perspective of security’’

### 5.2. Signature-Based Intrusion Detection System (SIDS):

Most intrusion detection systems (IDS) are what are known as signature-based. This means that they operate in much the same way as a virus scanner, by searching for a known identity - or signature - for each specific intrusion event. And, while signature-based IDS is very efficient at sniffing out known of attack, it does, like anti-virus software, depend on receiving regular signature updates, to keep in touch with variations in hacker technique.

Signature based IDS can only ever be as good as the extent of the signature database, two further problems immediately arise.

(i) Firstly, it is easy to fool signature-base solutions by changing the ways in which an attack is made.

(ii) Secondly, the more advanced the signature database, the higher the CPU load for the system charged with analyzing each signature (beyond the max bandwidth packets may be dropped).

And, because of the hackers' tendency to continually test and probe, it is only a matter of time before they discover a way around even the most sophisticated signature-based intrusion detection systems.

### 5.3. Anomaly - Based Intrusion Detection System (AIDS):

Any organization wanting to implement a more thorough - and hence safer - solution, should consider what we call anomaly-based IDS. Anomaly testing requires more hardware spread further across the network than is required with signature based IDS. This is especially true for larger networks and, with high bandwidth connections, it is therefore necessary to install the anomaly sensors closer to the servers and network that are being monitored.

Anomaly-based intrusion detection triggers an alarm on the IDS when some type of unusual behaviour occurs on your network. This would include any event, state, content, or behaviour that is considered to be abnormal by a pre-defined standard. Anything that deviates from this baseline of “normal” behaviour will be flagged and logged as anomalous. “Normal” behaviour can be programmed into the system based on offline learning and research or the system can learn the “normal” behaviour online while processing the network traffic.

Anomaly-based intrusion detection, on the other hand, takes a more generalized approach when looking for and detecting threats to your network. A baseline of “normal” behaviour is developed, and when an event falls outside that norm, it is flagged and logged.

The behaviour is a characterization of the state of the protected system, which is both reflective of the system health and sensitive to attacks. In this context, an anomaly-based method of intrusion detection has the potential to detect new or unknown attacks. Like the signature-based method, however, anomaly-based intrusion detection also relies on information that tells it what is normal and what isn’t. This is called a profile, and it is key to an effective anomaly-based intrusion detection system.

#### 5.3.1. Some examples of anomalous behaviour include:

- HTTP traffic on a non-standard port, say port 53
- Backdoor service on well-known standard port, e.g., peer-to-peer files sharing.
- A segment of binary code in a user password
- Too much UDP compared to TCP traffic
- A greater number of bytes coming from an HTTP browser than are going to it.

#### 5.3.2. Anomaly-based intrusion detection is useful for detecting these types of attacks:

- New buffer overflow attacks
- New exploits
- Variants of existing attacks in new environments (e.g., worms using different file names as they propagate)
- Intentionally stealthy attacks (e.g., to transform a shellcode)

## VI. AIDS TECHNIQUES

According to the type of processing related to the ‘‘behavioural’’ model of the target system, anomaly detection techniques can be classified into three main categories.

### 6.1. **Statistical-based AIDS techniques**

In the Statistical-based AIDS, the behaviour of the system is represented from a random viewpoint.

- Univariate models (independent Gaussian random variables)
- Multivariate models (correlations among several metrics)
- Time series (interval timers, counters and some other time-related metrics)

*Pros and Cons:*

Prior knowledge about normal activitynot required.
Accurate notification of malicious activities
Unrealistic quasi-stationary Process assumption

### 6.2. **Knowledge-based techniques**

In the Knowledge-based AIDS techniques try to capture the claimed behaviour from available system data (protocol specifications, network traffic instances, etc.).

- Finite state machines (states and transitions)
- Description languages
- Expert systems (rules-based classification)

*Pros and Cons:*

Flexibility and scalability
Difficult and time-consuming

### 6.3. **Machine learning-based AIDS schemes**

In the Machine learning AIDS schemes are based on the establishment of an explicit or implicit model that allows the patterns analyzed to be categorized.

- Bayesian networks (probabilistic relationships among variables)
- Markov models (stochastic Markov theory)
- Neural networks (human brain foundations)
- Fuzzy logic (approximation and uncertainty)
- Genetic algorithms (evolutionary biology inspired)
- Clustering and outlier detection (data grouping)

Pros and Cons:

Flexibility and adaptability.
High resource consuming.
Capture of interdependencies.
High dependency.

## VII. Assessment

Two key aspects concern the evaluation, and thus the comparison, of the performance of alternative intrusion detection approaches: these are the efficiency of the detection process, and the cost involved in the operation. Without underestimating the importance of the cost, at this point the efficiency aspect must be emphasized.

The main benefit of anomaly-based detection techniques is their potential to detect previously unseen intrusion events. However, and despite the likely inaccuracy in formal signature specifications, the rate of false positives (or FP), events erroneously classified as attacks in anomaly-based systems is usually higher due to the ever changing nature of networks, applications and exploits

### False Positives and Negatives

Four situations exist in this context, corresponding to the relation between the result of the detection for an analyzed event (‘‘normal’’ vs. ’’intrusion’’) and its actual nature (‘‘innocuous” vs. “malicious’’). These situations are:

- False Positive (FP), if the analyzed event is innocuous (or ‘‘clean’’) from the perspective of security, but it is classified as malicious;
- True Positive (TP), If the analyzed event is correctly classified as intrusion/malicious;
- False Negative (FN), if the analyzed event is malicious but it is classified as normal/innocuous; and
- True Negative (TN), if the analyzed event is correctly classified as normal/innocuous.

Step1: ***False positive rate*** is measured over normal data items. Suppose that *m* normal data items are measured and *n* of them are identified as abnormal.

***False positive rate is defined as n/m.***

**Step 2: Detection rate** is measured over abnormal data items. Suppose that *m* abnormal data items are measured, and *n* of them are detected.

***Detection rate is defined as n/m.***

The nemesis of anomaly-based detection has been the false positive. A detection system cannot be perfect (even if it uses a human expert). It produces false positive (it thinks it has detected a malicious event, which in fact is legitimate) and has false negative (it fails to detect actual malicious events).

### *Often there is a trade-off between the two:*

When one puts the threshold very low to avoid false negative, one often ends up with a higher rate of false positive.

If a detector has a false positive probability of 1%, this does not imply that if it raises a flag it will be a false alert only 1% of the time (and 99% probability that it detected an actual malicious event). It means that when it analyzes random legitimate events 1% of the time it will raise a flag. If the detector analysis 10,000 events, it will flag 100 legitimate events. If out of the 10,000 events one was malicious, it will raise an additional flag, making its total 101.Out of the 101 events detected, 1 was malicious and 100 were legitimate. In other words, out of the 101 alerts only one is real and 100 out of 101, i.e. (Figure 4) more than 99% of the time the alert was a false positive.

.

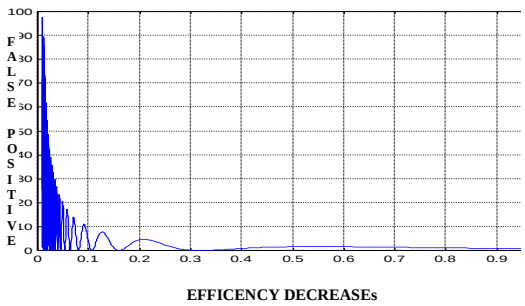


Figure 4. Increases in FP Alert

False positives are a problem in any kind of test: no test is perfect, and sometimes the test will incorrectly report a positive result.

*For example,* if a test for particular intrusion detection is performed on a host or network, then there is a chance that the test will return a positive result even if the host or network does not have the problem. The problem lies, however, not just in the chance of a false positive prior to testing, but determining the chance that a positive result is in fact a false positive.

## VIII. Bayesian Concept

Graphically, a Bayesian network consists of a set of nodes, representing individual random variables, and directed arcs from “parent" nodes to “child" nodes indicate conditional dependence. Cycles are not allowed, so that the resulting structure is an acyclic directed graph. The complete joint distribution of the set of nodes is defined by a conditional probability distribution for each node relative to its parent nodes. Bayes' theorem can be used to make inferences about the underlying (unobserved) state of the modelled system from observations. This structure provides a means to decompose and rapidly compute conditional and marginal probability distributions, generating likelihoods of specific events or condition. [10]

The two random variables are represented as two nodes in the network. Given an ordered stream of input events S ={e1; e2;. } the task of the event classification mechanism is to decide for each $e_i$ S whether it is normal or anomalous.

**Using Bayes' theorem:**

[7]As we will demonstrate, **using Bayes' theorem**, if a condition is rare, then the majority of positive results may be false positives, even if the test for that condition is otherwise reasonably accurate.

By definition, $P(A \mid B) = \frac{P(A \cap B)}{P(B)}$

$P(B \mid A) = \frac{P(A \cap B)}{P(A)}$

From $P(B \mid A)$ : $P(A \cap B) = P(A)P(B \mid A)$

This can be substituted into $P(A \mid B)$ to give you: $P(A \mid B) = \frac{P(A)P(B \mid A)}{P(B)}$

This is a simplified form of the *conditional probability* of A, given B, and it is Bayes Theorem.
In general, for more events, $A_1$, $A_2$… that partition the sample space, with $P(A_i) > 0$, this is the result:

$$P(A_i \mid B) = \frac{P(A_i)P(B \mid A_i)}{\sum_j P(B \mid A_j)P(A_j)}$$

A to be event "**positive test**"
B to be event **"actually infected person"**

**Demonstration:**

Suppose that a test for a particular system has a very high success rate:

*(1) if a tested system has the attack, the test accurately reports this, a 'positive', 99% of the time (or, with probability 0.99), and*
*(2) if a tested system does not have the attack, the test accurately reports that, a 'negative', 95% of the time (i.e. with probability 0.95).*

Suppose also, however, that only 0.1% of the alerts have that attack (i.e. with probability 0.001).We now have all the information required to calculate the probability that, given the test was positive, that it is a false positive.

Let D be the event that the system has the attack, and P be the event that the test returns a positive result.

***The probability of a true positive is***

$$P[D|P] = \frac{0.99 * 0.001}{0.99 * 0.001 + 0.05 * 0.999} \simeq 0.019$$

The probability of a false positive is about **(1 - 0.019) = 0.981.**

Despite the apparent high accuracy of the test, the incidence of the attack is so low (one in a thousand) that the vast majority of host whose test positive (98 in a hundred) do not have the attack. Nonetheless, this is 20 times the proportion before we knew the outcome of the test! The test is not useless, and retesting may improve the reliability of the result. In particular, a test must be very reliable in reporting a negative result when the host or network does not have the intrusion attack, if it is to avoid the problem of false positives.

In mathematical terms, this would ensure that the second term in the denominator of the above calculation is small, relative to the first term. For example, if the test reported a negative result in system without the intrusion

attack with probability **0.999**, then using this value in the calculation yields a probability of a ***false positive of roughly 0.5.***

## Occurrence of Event Possibility

We have set up possible event .Each of which we assumed, occurs some numbers of times. Thus if there are n distinct possible event **X1,X2,.....Xn**, and the event occurred frequency **N1,N2,.....Nn**. Nowthe probability of event X is symbolized by **P(X).**Probability of an event X lies between **0≤P(X) ≤1.**

Now measured entropy, represented by H is calculated with the help of given formula:

$$H = -\sum_i p_i(\log_2 p_i)$$

Where $p_i$ the probability of event.

**It is clear that low FP and FN rates, together with high TP and TN rates, will result in good efficiency values.**

## IX. Conclusion

The anomaly based detection is based on defining the network behavior. The network behavior is in accordance with the predefined behavior, then it is accepted or else it triggers the event in the anomaly detection. The accepted network behavior is prepared or learned by the specifications of the network administrators. The important phase in defining the network behavior is the IDS engine capability to cut through the various protocols at all levels. The Engine must be able to process the protocols and understand its goal. Though this protocol analysis is computationally expensive, the benefits it generates like increasing the rule set helps in less false positive alarms.

## X. Future Work

The enhancement of anomaly detection is defining its rule set. The efficiency of the system depends on how well it is implemented and tested on all protocols. Rule defining process is also affected by various protocols used by various users. Apart from these, custom protocols also make rule defining a difficult job. For detection to occur correctly, the detailed knowledge about the accepted network behavior need to be developed. But once the rules are defined and protocol is built then anomaly detection systems works well.

## References


[1] Anomaly Based Intrusion Detection and Artificial Intelligence -Benoît Morel, Carnegie Mellon University, United States.
[2] Intruders Detection System-Tools The Information Assurance Technology Analysis Center (IATAC) - Department of Defense (DoD)
[3] Guide to Intrusion Detection and Prevention Systems –National Institute of Standard & Technologies-US Department
[4] www.Sience Direct.com-Journals
[5] Anomaly-Based Network Intrusion Detection-Computer Science and Telecommunications Faculty, University of Granada,
[6] Deciphering Detection Techniques: Part II Anomaly-Based Intrusion Detection-By Dr.Fengmin Gong, Chief Scientist, McAfee Network Security Technologies Group
[7] Icjrc2010.files.wordpress.com/2010/08/**derivation**-of-**bayes-theorem**
[8] Saira Beg etal.Feasibility of Intrusion Detection System with High Performance Computing: A Survey- International Journal for Advances in Computer Science
[9] Vera Marinova-Boncheva, "A Short Survey of IntrusionDetection Systems", Problems of Engineering Cyberneticsand Robotics, 58, 2007. http://www.iit.bas.bg/PECR/58/23-30.pdf
[10] Integrating Correlated Bayesian Networks Using Maximum Entropy-Kenneth D. Jarman Applied Mathematical Sciences, Vol. 5, 2011, no. 48, 2361 – 2371
[11] Network intrusion detection using Naïve bayes- Mrutyunjaya Panda1 and Manas Ranjan Patra2IJCSNS International Journal of Computer Science and Network Security, VOL.7 No.12, December 2007