ON THE DETECTION OF NOVEL ATTACKS USING BEHAVIORAL APPROACHES

Benferhat Salem and Tabia Karim
CRIL-CNRS UMR8188, Université d'Artois, Rue Jean Souvraz 62307 Lens Cedex, France

Keywords: Intrusion detection, behavioral approaches, Bayesian networks.

Abstract: During last years, behavioral approaches, representing normal/abnormal activities, have been widely used in intrusion detection. However, they are ineffective for detecting novel attacks involving new behaviors. This paper first analyzes and explains this recurring problem due on one hand to inadequate handling of anomalous and unusual audit events and on other hand to insufficient decision rules which do not meet behavioral approach objectives. We then propose to enhance the standard classification rules in order to fit behavioral approach requirements and detect novel attacks. Experimental studies carried out on real and simulated *http* traffic show that these enhanced decision rules allow to detect most novel attacks without triggering higher false alarm rates.

1 INTRODUCTION

Intrusion detection aims at detecting any malicious action compromising integrity, confidentiality or availability of computer and network resources or services (Axelsson, 2000). Intrusion detection systems (IDSs) are either misuse-based (Snort, 2002) or anomaly-based (Neumann and Porras, 1999) or a combination of both the approaches in order to exploit their mutual complementarities (Tombini et al., 2004). Behavioral approaches, which are variants of anomaly-based approaches, build profiles representing normal and abnormal behaviors and detect intrusions by comparing system activities with learnt profiles. The main advantage of these approaches lies their capacity to detect both known and novel attacks. However, there is no behavioral approach ensuring acceptable tradeoff between novel attack detection and underlying false alarm rate.

Intrusion detection can be viewed as a classification problem in order to classify audit events (network packets, Web server logs, system logs, etc.) as normal events or attacks (Lee, 1999). During last years, several works used classifiers in intrusion detection (Kruegel et al., 2003)(Sebyala et al., 2002)(Valdes and Skinner, 2000) and achieved acceptable detection rates on well-known benchmarks such as KDD'99 (Lee, 1999), Darpa'99 (Lippmann et al., 2000). The recurring problem with the majority of classifiers is their high false negative rates mostly caused by the incapacity to correctly classify novel attacks (Elkan, 2000)(Lee, 1999). For instance, in (Benferhat and Tabia, 2005)(Barbará et al., 2001), decision trees and variants of Bayes classifiers are used to classify network connections and concluded that their main problem lies in their failure to detect novel attacks which they classify normal connections.

In this paper, we first analyze and explain the problem of high false negative rates and classifiers incapacity to correctly classify new malicious behaviors. We illustrate our approach with Bayesian classifiers. We first focus on how new behaviors affect and manifest through a given feature set. Then we explain why standard classification rules fail in detecting these new events. Different possibilities are considered and discussed. More precisely, we consider on one hand problems related to handling unusual and new behaviors and on other hand problems due to insufficient decision rules which do not meet anomaly detection requirements. After that, we propose to enhance standard classification rules in order to improve detecting novel attacks involving abnormal behaviors. Experimental studies on real and simulated *http* traffic are carried out to evaluate the effectiveness of the new decision rules in detecting new intrusive behaviors. Two variants of Bayesian classifiers using the enhanced classification rule are trained on real normal *http* traffic and several Web attacks. Then we evalu-

Salem B. and Karim T. (2008).
ON THE DETECTION OF NOVEL ATTACKS USING BEHAVIORAL APPROACHES.
In *Proceedings of the Third International Conference on Software and Data Technologies - PL/DPS/KE*, pages 265-272
DOI: 10.5220/0001894302650272

ated these classifiers on known and novel attacks as well new normal behaviors.

2 BEHAVIORAL APPROACH FOR BOTH NORMAL AND ABNORMAL BEHAVIORS

Behavioral approaches build models or profiles representing normal and abnormal activities and detect intrusions by computing deviations of current system activities form reference profiles. Behavioral approaches are like anomaly detection ones except that profiles are defined for both normal and abnormal behaviors while in anomaly approaches, only normal behaviors are profiled(Axelsson, 2000). For instance, every significant deviation from normal behavior profiles may be interpreted as an intrusion since it represents an anomalous behavior. The main advantage of anomaly approaches lies in their potential capacity to detect both new and unknown (previously unseen) attacks as well as known ones. This is particularly critical since new attacks appear every day and it often takes several days between the apparition of a new attack and updating signature data bases or fixing/correcting the exploit.
Within anomaly detection approaches, the detection of novel attacks has several negative side effects which concern triggering very high false alarm rates. This drawback seriously limits there use in real applications. In fact, configuring anomaly-based IDSs to acceptable false alarm rates cause their failure in detecting most malicious behaviors. In (Kumar and Spafford, 1994), authors pointed out that intrusive activities used to extract signatures or train detection systems are a subset of anomalous behaviors and identified four audit event possibilities with non zero probabilities:

- **Intrusive but Not Anomalous (False Negative).** They are attacks where input data do not catch any anomalous evidence. This usually due to feature extraction problem. Therefore, new attacks often require supplementary features and data in order to be detected.
- **Not Intrusive but Anomalous (False Positive).** Commonly called false alarms, these events are legitimate but new. Consequently, they significantly deviate from normal events profile. This problem requires updating normal profiles in order to integrate such new normal behaviors.
- **Not Intrusive and Not Anomalous (True Negative).** They correspond to known normal events.
- **Intrusive and Anomalous (True Positive).** Such events correspond to attacks where intrusive evidence is caught by input data.

In the following, we particularly focus on behavioral approaches failure to detect novel attacks involving abnormal behaviors and enhancing standard decision rules in order to detect these novel attacks.

3 WHY STANDARD CLASSIFICATION RULES ARE INEFFECTIVE FOR DETECTING NOVEL ATTACKS

Behavioral approaches can be viewed as classifiers which are mapping functions from a discrete or continuous feature space (observed variables $A_0 = a_0$, $A_1 = a_1$, .., $A_n = a_n$) to a discrete set of class labels C=$\{c_1, c_2,..,c_m\}$. Once a classifier is built on labeled training data, it can classify any new instance. The goal of classification is to maximize the generalization ability to correctly classify unseen instances. Decision trees(Quinlan, 1986) and Bayesian classifiers (Friedman et al., 1997) are well-known classification algorithms.
In intrusion detection, each instance to classify represents an audit event (network packet, connection, application log record, etc.). In order to analyze standard classification rules incapacity to detect novel attacks, we first focus on how novel attacks involving new behaviors affect feature sets which provide input data to be analyzed.

3.1 How Novel Attacks Affect Feature Sets

The following are different possibilities about how new anomalous events affect and manifest through feature sets:

1. ***New Value(s) in a Feature(s).*** A never seen[1] value is anomalous and it is due in most cases to a malicious event. For example, Web server response codes are from a fixed set of predefined values (ex. 200, 404, 500...). If a new response code or any other response is observed, then this constitutes an anomalous event. For instance, successful shell code attacks cause server response without a common code. Similarly, a

[1]By never seen value we mean new value in case of nominal features or very deviating value in case of numerical features.

new network service using a new and uncommon port number is probably intrusive since most back-door attacks communicate through uncommon ports while common services are associated with common port numbers.

2. ***New Combination of Known Values.*** In normal audit events, there are correlations and relationships between features. Then an anomalous event can be in the form of a never seen combination of normal values. For example, in some *http* requests, numerical values are often provided as parameters. The same values which are correctly handled by a given program, can in other contexts cause value misinterpretations and result in anomalous behaviors. Another example from network intrusion field is when a network service like *http* uses uncommon transport protocol like *UDP*. Both *http* and *UDP* are common protocols in network traffic. However, *http* service uses *TCP* protocol at the transport layer and never *UDP*, then *http* using *UDP* is anomalous event.
3. ***New Sequence of Events.*** There are several normal audit events which show sequence patterns. For example, in on-line marketing applications, users are first authenticated using *https* protocol for confidential data transfers. Then a user session beginning without *https* authentication is probably intrusive since the application control flow has not been followed. Such intrusive evidence can be caught by history features summarizing past events or by using appropriate mining anomaly sequence patterns algorithms.
4. ***No Anomalous Evidence.*** In this case, new anomalous events do not result in any unseen evidence. The underlying problem here is related to feature extraction and selection since not enough data is used for catching the anomalous evidence.

In principle, the three first possibilities can be detected since intrusive behavior evidence had appeared in the feature set. However, anomalous audit event of fourth case can not be detected for lack of any anomalous evidence in the audit event.

3.2 Why Novel Attacks Cause False Negatives

Novel attacks often involve new behaviors. However, in spite of these anomalousness evidence in the feature set, most classifiers flag novel attacks as normal events. This failure is mainly due to the following problems:

1. *Inadequate Handling of New and unusual Events*. New and unusual values or value combinations are often involved by novel attacks. However, most classifiers handle such evidence inadequately regarding anomaly detection objectives. For instance, in Bayesian classifiers (Friedman et al., 1997), new values cause zero probabilities which most implementations replace with extremely small values and rely on remaining features in order to classify the instance in hand. Decision trees (Quinlan, 1986), which are very efficient classifiers, often use few features in order to classify audit events. Therefore, if the anomalous evidence (abnormal values or value combinations) appears in a feature which is not used, then this evidence is not used. Other techniques suffer from other problems such as incapacity to handle categorical features which are common in intrusion detection (Shyu et al., 2005).
2. *Insufficient decision rules*: The objective of standard classification rules is to maximize classifying previously unseen instances relying on known (training) behaviors. However, unseen behaviors which should be flagged abnormal according to anomaly approach, are associated with known behavior classes. For instance, when decision trees classify instances, new behaviors such as new values are assigned to the majority class at the test where the new value is encountered. As for Bayesian classifiers, they rely only on likelihood and prior probabilities to ensure classification. This strongly penalize detection of new and unusual behaviors in favor of frequent and common behaviors. Given that normal data often represent major part of training data sets (Lee, 1999)(Lippmann et al., 2000), standard classification rules fail in detecting novel attacks involving new behaviors and flag them in most cases normal events.

In the following, we will basically focus on Bayesian networks as an example of behavioral approaches. This choice is motivated by the fact that they are among most effective techniques and because of their capacity to handle both numeric and categorical features which are common in intrusion detection (Shyu et al., 2005). Furthermore, in comparison with other classifiers, main advantage of Bayesian ones for detecting anomalous behaviors lies in using all the features and feature dependencies.

4 ENHANCING STANDARD CLASSIFICATION RULES

In order to overcome standard classification rules drawbacks, we propose enhancing them in order to fit behavioral approach requirements.

4.1 Standard Bayesian Classification

Bayesian classification is a particular kind of Bayesian inference (Friedman et al., 1997). Classification is ensured by computing the greatest a posteriori probability of the class variable given an attribute vector. Namely, having an attribute vector A (observed variables $A_0 = a_0$, $A_1 = a_1$, .., $A_n = a_n$), it is required to find the most plausible class value c_k ($c_k \in$ C=$\{c_1, c_2,..,c_m\}$) for this observation. The class c_k associated to A is the class with the most a posteriori probability $p(c_k/A)$. Then Bayesian classification rule can be written as follows:

$$Class = argmax_{c_k \in C}(p(c_i/A)) \quad (1)$$

Term $p(c_i/A)$ denotes the posterior probability of having class c_i given the evidence A. This probability is computed using Bayes rule as follows:

$$p(c_i/A) = \frac{p(A/c_i) * p(c_i)}{p(A)} \quad (2)$$

In practice, the denominator of Equation 2 is ignored because it does not depend on the different classes. Equation 2 means that posterior probability is proportional to likelihood and prior probabilities while evidence probability is just a normalizing constant. Naive Bayes classifier assumes that features are independent in the class variable context. This assumption leads to the following formula

$$p(c_i/A) = \frac{p(a_1/c_i) * p(a_2/c_i)..p(a_n/c_i) * p(c_i)}{p(A)} \quad (3)$$

In the other Bayesian classifiers such as TAN (Tree Augmented Naive Bayes), BAN (Augmented Naive Bayes) and GBN (General Bayes Network) (Friedman et al., 1997), Equation 2 takes into account feature dependencies in computing conditional probabilities as it is denoted in Equation 4.

$$p(c_i/A) = \frac{p(a_1/Pa(a_1)) * .. * p(a_n/Pa(a_n)) * p(c_i)}{p(A)} \quad (4)$$

Note that terms $Pa(a_i)$ in Equation 4 denote parents of feature a_i.

Bayesian classifiers have been widely used in intrusion detection. For instance, in (Valdes and Skinner, 2000), naive Bayes classifier is used to detect malicious audit events while in (Kruegel et al., 2003), authors use Bayesian classification in order to improve the aggregation of different anomaly detection model outputs.

4.2 Enhancing Standard Bayesian Classification for Anomaly Detection

Bayesian classification lies on posterior probabilities given the evidence to classify (according to Equations 1 and 2). The normality associated with audit event E (observed variables $E_0 = e_0$, $E_1 = e_1$, .., $E_n = e_n$) can be measured by posterior probability $p(Normal/E)$. This measure is proportional to the likelihood of E in *Normal* class and prior probability of *Normal* class.

In practice, normality can not be directly inferred from probability $p(Normal/E)$ because this probability is biased. For instance, major Bayesian classifier implementations ignore denominator of Equation 2 while zero probability and floating point underflow problems are handled heuristically. Assume for instance that a never seen value had appeared in a nominal feature e_i. Then according to Equation 2, the probability $p(e_i/c_k)$ equals zero over all classes c_k. In practice, it is an extremely small value that is assigned to this probability. The strategy of assigning non zero probabilities in case of new values is to use remaining features and prior probabilities in order to classify the instance in hand. The other problem consists in floating point underflow which is caused by multiplying several small probabilities each varying between 0 and 1. This case is often handled by fixing a lower limit when multiplying probabilities.

4.2.1 Using Normality/Abnormality Duality

Anomaly-based systems flag audit events "Normal" or "Abnormal" according to a computed normality degree associated with each audit event. Having two scaled functions computing respectively normality and abnormality relative to audit event E then these two functions are dual. Namely, this propriety can be formulated as follows:

$$Normality(E) + Abnormality(E) = constant \quad (5)$$

The intuitive interpretation of this propriety is more an event is normal, less it is abnormal. Conversely, less normal is the event, it is more abnormal. Translated in probability terms, Equation 5 gives the following propriety:

$$P(Normal/E) + P(Abnormal/E) = 1 \quad (6)$$

Term $P(Normal/E)$ (resp. $P(Abnormal/E)$) denotes the probability that audit event E is normal (resp. abnormal). Bayesian classifiers associate a probability distribution with the instance to classify (audit event) and return the class having the utmost posterior probability. Let us assume for instance that training data

involve normal data (with class label *Normal*) and several attack categories (labeled $Attack_1$, $Attack_2$,.., $Attack_n$). Consider the case when $p(Normal/E)$ is greater than all posterior probabilities $p(Attack_1/E)$, ..., $p(Attack_n/E)$. In this case, standard Bayesian rule, will return *Normal* class according to Equation 1. However, if

$$p(Normal/E) < (p(Attack_1/E) + .. + p(Attack_n/E))$$

Then according to Equation 6, the probability that audit event E is abnormal is 1-($p(Normal/E)$). Intuitively, this audit event should be flagged anomalous. We accordingly propose to enhance standard Bayesian rule as follows:

Rule 1

If $p(Normal/E) < (\sum(p(c_k \neq Normal/E))$
then $Class = argmax_{c_k \in C}(p(c_k \neq Normal/E)$
else $Class = argmax_{c_k \in C}(p(c_k/E))$

Rule 1 enhances standard Bayesian classification rule in order to take into account normality/abnormality duality relative to audit events. Unlike standard Bayesian classification rule, Equation 7 first compares normality with abnormality relative to audit event E and returns *Normal* only when the posterior probability $p(Normal/E)$ is greater than the sum of posterior probabilities $p(Attack_i/E)$. When abnormality is more important, this rule returns the attack having the utmost posterior probability.

4.2.2 Using Zero Probabilities as Abnormal Evidence

Anomalous audit events can affect the feature set either by new values, new value combinations or new audit event sequences. Then classifying anomalous events strongly depends on how zero probability and floating point underflow problems are dealt with. However, since a zero probability is due to new (hence anomalous) value, then this is anomalousness evidence. The underlying interpretation is that instance to classify involves a never seen evidence. Then anomaly approach should flag this audit event anomalous. Similarly, an extremely small a posteriori probability can be interpreted as a very unusual event, hence anomalous. Then, standard Bayesian classification rule can accordingly be enhanced in the following way:

- If there is a feature e_i where probability $p(e_i/c_k)$ equals zero over all training classes, then this is a new value (never seen in training data). Enhanced Bayesian classification rule can be formulated as follows:

Rule 2

If $\exists\ e_i,\ \forall k, p(e_i/c_k) = 0$ **then** $Class = New$
else $Class = argmax_{c_k \in C}(p(c_k/E))$

- New intrusive behaviors can be in the form of unseen combination of seen values. In this case, feature dependencies must be used in order to reveal such anomalousness. Since new value combinations will cause zero conditional probabilities, then this anomalous evidence can be formulated as follows:

Rule 3

If $\exists\ e_i,\ p(e_i/Pa(e_i)) = 0$ **then** $Class = New$
else $Class = argmax_{c_k \in C}(p(c_k/E))$

Note that when building Bayesian classifiers, structure learning algorithms extract feature dependencies from training data. Then there may be unseen value combinations that can not be detected if the corresponding dependencies are not extracted during structure learning phase.

4.2.3 Using Likelihood of Rare Attacks as Abnormal Evidence

When training classifiers, some attacks have often very small frequencies in training data sets. The problem with such prior probabilities is to strongly penalize the corresponding attacks likelihood. This problem was pointed out in (Ben-Amor et al., 2003) where authors proposed simple duplication of weak classes in order to enhance their prior probabilities. An alternative solution is to exploit likelihood of audit events as if training classes (*Normal*, $Attack_1$, .., $Attack_n$) were equiprobable. Assume for instance intrusive audit event E is likely to be an attack (for example, likelihood $p(E/Attack_j)$ is the most important). Because of the negligible prior probability of $Attack_j$, posterior probability $p(Attack_j/E)$ will be extremely small while $p(Normal/E)$ can be significant since *Normal* class prior probability is important. Then we can rely on likelihood in order to detect attacks with small frequencies:

Rule 4

If $\exists\ Attack_j,\ \forall k, p(E/Attack_j) >= p(E/c_k)$ **and**
$p(Normal/E) > P(Attack_j/E)$ **and** $p(Attack_j) < \varepsilon$
then $Class = Attack_j$
else $Class = argmax_{c_k \in C}(p(c_k/E))$

Rule 4 is provided in order to help detecting anomalous events with best likelihood in attacks having extremely small prior probabilities ($p(Attack_j) < \varepsilon$). It will be applied only if the proportion of instances of $Attack_j$ in training data is less than threshold ε fixed by the expert.
Note that the standard classification rule (see Equation 1) is applied only when Rules 1, 2, 3 and 4 can not be applied.

5 EXPERIMENTAL STUDIES

In this section, we provide experimental studies of our enhanced Bayesian classification rule on *http* traffic including normal real data and several Web attacks. Before giving further details, we first present training and testing data then we report evaluation results of naive Bayes and Tree Augmented naive Bayes (TAN), the two Bayesian classifiers we used in the following experimentations.

5.1 Training and Testing Data Sets

We carried out experimentations on a real *hhttp* traffic collected on a University campus during 2007. We extracted *http* traffic and preprocessed it into connection records using only packet payloads. Each *http* connection is characterized by four feature categories:

- **Request General Features** providing general information on *http* requests. Examples of such features are request method(s), request length, etc.
- **Request Content Features** searching for particularly suspicious patterns in *http* requests. The number of non printable/metacharacters, number of directory traversal patterns, etc.
- **Response Features** extracted by analyzing *http* response to a given request. These features can reveal the success or failure of an attack and can reveal suspicious *http* content in the response, in which case Web clients are targeted by a possible attack.
- **Request History Features** providing statistics about past connections given that several Web attacks such as flooding, brute-force, Web vulnerability scans perform through several repetitive connections. Examples of such features are the number/rate of connections issued by same source host and requesting same/different URIs.

Note that in order to label the preprocessed *http* traffic (as normal or attack), we analyzed this data using Snort(Snort, 2002) IDS as well as manual analysis. As for other attacks, we simulated most of the attacks involved in (Ingham and Inoue, 2007) which is to our knowledge the most extensive and uptodate open Web-attack data set. In addition, we played vulnerability scanning sessions using w3af(Riancho, 2007). Note that attack simulations are achieved on a simulation network using the same platform (same Web server software and same operating system) and same Web site content.

Table 1: Training/testing data set distribution.

	Training data		Testing data	
Class	**Number**	**%**	**Number**	**%**
Normal	55342	55.87%	61378	88.88 %
Vulnerability scan	31152	31.45%	4456	6.45 %
Buffer overflow	9	0.009%	15	0.02%
Input validation	44	0.044%	4	0.01 %
Value misinterpretation	2	0.002%	1	0.00%
Poor management	3	0.003%	0	0.00%
URL decoding error	3	0.003%	0	0.00%
Flooding	12488	12.61%	3159	4.57 %
Cross Site Scripting	**0**	**0.00%**	**6**	**0.0001 %**
SQL injection	**0**	**0.00%**	**14**	**0.001 %**
Command injection	**0**	**0.00%**	**9**	**0.001 %**
Total	**99043**	**100%**	**69061**	**100%**

Attacks of Table 1 are categorized according to the vulnerability category involved in each attack. Regarding attack effects, attacks of Table 1 include DoS attacks, Scans, information leak, unauthorized and remote access (Ingham and Inoue, 2007). In order to evaluate the generalization capacities and the ability to detect new attacks, we build a testing data set including real normal *http* connections as well as known attacks, known attack variations and novel ones (attacks in bold in Table 1).
Note that new attacks included in testing data either involve new feature values or anomalous value combinations:

- *Attacks Causing New Values.* New attacks involving new values are SQL injection, XSS and command injection attacks. These attacks are characterized by never seen values in training data.
- *Attacks Causing New Value Combinations*. New attacks involving new value combinations are buffer overflow attacks including shell code and shell commands and vulnerability scans searching particularly for SQL injection points. In training data, vulnerability scans search for vulnerabilities other than SQL injection points. Similarly, in training data, there are shell command injection attacks and buffer overflow attacks without shell codes or shell commands.

Note that SQL injection attacks include two insufficient authentication attacks performing through SQL injection. These attacks cause simultaneously anomalous new values and value combinations.

5.2 Experiments on Standard/Enhanced Bayesian Classification Rule

Table 2 compares results of standard then enhanced naive Bayes and TAN classifiers built on training data and evaluated on testing one.

Table 2: Evaluation of naive Bayes and TAN classifiers using standard/enhanced Bayesian classification rules.

	Standard Bayesian rule		**Enhanced Bayesian rule**	
	Naive Bayes	**TAN**	**Naive Bayes**	**TAN**
Normal	98.2%	99.9%	91.7%	97.8%
Vulnerability scan	15.8%	44.1%	100%	100%
Buffer overflow	6.7%	20.2%	80%	100%
Input validation	75.0%	100%	100%	100%
Value misinterpretation	100%	0.00%	100%	100%
Flooding	100%	100%	100%	100%
Cross Site Scripting	**0.00%**	**0.00 %**	**100%**	**100%**
SQL injection	**0.00%**	**0.00%**	**100 %**	**100%**
Command injection	**0.00%**	**0.00 %**	**100 %**	**100%**
Total PCC	**92.87%**	**96.24%**	**96.45%**	**98.07%**

Note that enhanced classification rule evaluated in Table 2 uses normality/abnormality duality and zero probabilities (see Rules 1, 2 and 3).

- *Experiments on standard Bayesian classification rule*: At first sight, both classifiers achieve good detection rates regarding their PCCs (Percent of Correct Classification) but they are ineffective in detecting novel attacks (attacks in bold in Table 2). Confusion matrixes relative to this experimentation show that naive Bayes and TAN classifiers misclassified all new attacks and predicted them *Normal*. However, results of Table 2 show that TAN classifier performs better than naive Bayes since it represents some feature dependencies. Furthermore, testing attacks causing new value combinations of seen anomalous values (involved separately in different training attacks) cause false negatives. For instance, testing vulnerability scans are not well detected since they involve new value combinations.
- *Experiments on enhanced Bayesian classification rule*: Naive Bayes and TAN classifiers using the enhanced rule perform significantly better than with standard rule. More particularly, both the classifiers succeeded in detecting both novel and known attacks. Unlike naive Bayes, enhanced TAN classifier improves detection rates without triggering higher false alarm rate (see correct classification rate of *Normal* class in Table 2. Furthermore, TAN classifier correctly detects and identifies all known and novel attacks.

Results of Table 2 show that significant improvements can be achieved in detecting novel attacks by enhancing standard classification rules in order to meet behavioral approach requirements.

6 CONCLUSIONS

The main objective of this paper is to overcome one of the main limitations of behavioral approaches. We proposed how to enhance standard classification rules in order to effectively detect both known and novel attacks. We illustrated our enhancements on Bayesian classifiers in order to improve detecting novel attacks involving abnormal behaviors. More precisely, we have proposed four rules relying on normality/abnormality duality relative to audit events, zero probabilities caused by anomalous evidence occurrence and likelihood of attacks having extremely small prior frequencies. Experiments on $http$ traffic show the significant improvements achieved by the enhanced decision rule in comparison with the standard one. Future work will address handling incomplete and uncertain information relative to network traffic audit events.

REFERENCES

Axelsson, S. (2000). Intrusion detection systems: A survey and taxonomy. Technical Report 99-15, Chalmers Univ.

Barbará, D., Wu, N., and Jajodia, S. (2001). Detecting novel network intrusions using bayes estimators. In *Proceedings of the First SIAM Conference on Data Mining*.

Ben-Amor, N., Benferhat, S., and Elouedi, Z. (2003). Naive bayesian networks in intrusion detection systems. In *ACM*, Cavtat-Dubrovnik, Croatia.

Benferhat, S. and Tabia, K. (2005). On the combination of naive bayes and decision trees for intrusion detection. In *CIMCA/IAWTIC*, pages 211–216.

Elkan, C. (2000). Results of the kdd'99 classifier learning. *SIGKDD Explorations*, 1(2):63–64.

Friedman, N., Geiger, D., and Goldszmidt, M. (1997). Bayesian network classifiers. *Machine Learning*, 29(2-3):131–163.

Ingham, K. L. and Inoue, H. (2007). Comparing anomaly detection techniques for http. In *RAID*, pages 42–62.

Kruegel, C., Mutz, D., Robertson, W., and Valeur, F. (2003). Bayesian event classification for intrusion detection.

Kumar, S. and Spafford, E. H. (1994). An application of pattern matching in intrusion detection. *Tech. Rep. CSD–TR–94–013, Department of Computer Scien'ces, Purdue University, West Lafayette*.

Lee, W. (1999). *A data mining framework for constructing features and models for intrusion detection systems.* PhD thesis, New York, NY, USA.

Lippmann, R., Haines, J. W., Fried, D. J., Korba, J., and Das, K. (2000). The 1999 darpa off-line intrusion detection evaluation. *Comput. Networks*, 34(4):579–595.

Neumann, P. G. and Porras, P. A. (1999). Experience with EMERALD to date. pages 73–80.

Quinlan, J. R. (1986). Induction of decision trees. *Mach. Learn.*, 1(1).

Riancho, A. (2007). w3af - web application attack and audit framework.

Sebyala, A. A., Olukemi, T., and Sacks, L. (2002). Active platform security through intrusion detection using naive bayesian network for anomaly detection. In *Proceedings of the London Communications Symposium 2002*.

Shyu, M.-L., Sarinnapakorn, K., Kuruppu-Appuhamilage, I., Chen, S.-C., Chang, L., and Goldring, T. (2005). Handling nominal features in anomaly intrusion detection problems. In *RIDE*, pages 55–62. IEEE Computer Society.

Snort (2002). Snort: The open source network intrusion detection system. http://www.snort.org.

Tombini, E., Debar, H., Me, L., and Ducasse, M. (2004). A serial combination of anomaly and misuse idses applied to http traffic. In *ACSAC '04: Proceedings of the 20th Annual Computer Security Applications Conference (ACSAC'04)*, pages 428–437, Washington, DC, USA. IEEE Computer Society.

Valdes, A. and Skinner, K. (2000). Adaptive, model-based monitoring for cyber attack detection. In *Recent Advances in Intrusion Detection*, pages 80–92.