NOVEL AND ANOMALOUS BEHAVIOR DETECTION USING BAYESIAN NETWORK CLASSIFIERS

Salem Benferhat and Karim Tabia
CRIL - CNRS UMR8188, Université d'Artois, Rue Jean Souvraz SP 18 62307 Lens Cedex, France

Keywords: Bayesian classifiers, Intrusion detection, Anomaly approach, Novelty detection.

Abstract: Bayesian networks have been widely used in intrusion detection. However, most works showed that they are ineffective for anomaly detection since novel attacks and new behaviors are not efficiently detected. In this paper, we firstly analyze this problem due to inadequate treatment of novel and unusual behaviors and to insufficient decision rules which do not meet anomaly approach requirements. We accordingly propose to enhance the standard Bayesian classification rule in order to fit anomaly detection objectives and effectively detect novel attacks. We carried out experimental studies on recent and real $http$ traffic and showed that Bayesian classifiers using enhanced decision rules allow detecting most novel attacks without triggering significantly 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). Anomaly-based approaches build profiles representing normal activities and detect intrusions by comparing current system activities with learnt profiles. Every significant deviation may be interpreted as an intrusion since it represents an anomalous behavior. The main advantage of anomaly approaches lies their potential capacity to detect both known and novel attacks. However, there is no anomaly approach ensuring acceptable tradeoff between 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. 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). Examples of such classifiers are Bayesian networks which have been widely used in intrusion detection. In comparison with other classifiers, main advantage of Bayesian ones for detecting anomalous behaviors lies in using all the features and feature dependencies. 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. 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 Bayesian classifiers incapacity to correctly classify novel behaviors especially malicious ones. We first focus on how new behaviors affect and manifest through a given feature set. Then we explain why standard Bayesian classification rule fail in detecting these new events. 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

Salem S. and Tabia K. (2008).
NOVEL AND ANOMALOUS BEHAVIOR DETECTION USING BAYESIAN NETWORK CLASSIFIERS.
In *Proceedings of the International Conference on Security and Cryptography*, pages 13-20
DOI: 10.5220/0001923300130020

propose to enhance standard Bayesian classifiers with four new decision rules in order to improve detecting novel attacks involving abnormal behaviors. The main objective of enhancing standard Bayesian classifiers is to detect and identify both known and novel attacks. Experimental studies on recent 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 and recent $http$ traffic involving normal data and several Web attacks. Then we evaluated these classifiers on known and novel attacks as well novel normal behaviors.

The rest of this paper is organized as follows: Section 1 briefly presents Bayesian networks and Bayesian classification. In section 2, we introduce anomaly detection approach. We focus in section 3 on Bayesian classification problems in detecting novel attacks. Section 5 proposes enhancements to the standard Bayesian classification rule in order to improve detecting novel attacks. Experimental studies on $http$ traffic are presented in section 6. Section 7 concludes this paper.

2 BAYESIAN CLASSIFIERS

Anomaly detection can be viewed, to some extent, as classifiers which are mapping functions from a discrete or continuous feature space to a discrete set of class labels. Once a classifier is built on labeled training data, it can classify any new instance. Decision trees (Quinlan, 1986) and Bayesian classifiers (Friedman et al., 1997) are well-known classifiers.

Bayesian networks are powerful graphical models for representing and reasoning under uncertainty conditions (Jensen, 1996). They consist of a graphical component DAG (Directed Acyclic Graph) and a quantitative probabilistic one. The graphical component allows an easy representation of domain knowledge in the form of an influence network (vertices represent events while edges represent ”influence” relations between these events). The probabilistic component expresses uncertainty relative to relationships between domain variables using conditional probability tables.

Bayesian classification is a particular kind of Bayesian inference. 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:

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

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:

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

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. For computation complexity reasons, naive Bayes classifier (which the simplest form of Bayes networks) assumes that features are independent in the class variable context. This assumption leads to the following formula

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

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

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

Terms $Pa(a_i)$ denote parents of feature a_i. Note that learning naive Bayes classifiers requires only training data to compute the conditional probability tables since the structure is known. The other Bayesian classifiers require both structure and parameter learning.

3 ANOMALY DETECTION

Anomaly approaches build models or profiles representing normal activities and detect intrusions by computing deviations of current system activities form normal activity profile. Every significant deviation may be interpreted as an intrusion since it represents an anomalous behavior. Anomaly-based IDSs are efficient in detecting new attacks but cause high false alarm rates which may really encumber the application of anomaly-based IDSs in real environments. In fact, configuring anomaly-based systems to acceptable false alarm rates result in failure to detect most malicious activities. The main advantage of anomaly detection lies in it potential capacity to detect both new and unknown (previously unseen) as

well as known attacks. The capacity to detect unknown/new attacks is a key feature in IDSs effectiveness. 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.
In (Kumar and Spafford, 1994), authors maintain that intrusive activities used to extract signatures or train detection systems are a subset of anomalous behaviors and pointed out 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 is 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 events.
- 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 practice, when profiling normal activities for anomaly detection purposes, it is only a subset of normal activities which is profiled. This fact explains in part false alarm rates relative to anomaly-based IDSs. In (Kruegel et al., 2003), other problems causing high false alarm rates were identified such as simplistic individual anomaly scores aggregation. Similarly, building attack models or profiles involves only a subset of all intrusive activities and attack variants. This results in failure in detecting several new attacks and attack variants. For instance, feature extraction focuses on known attacks and normal events in order to differentiate between normal and intrusive audit events. Consequently, there will always be new attacks for which old feature sets do not catch new attacks evidence. In order to analyze the standard Bayesian classification incapacity to detect novel attacks, we particularly focus on how novel attacks involving new behaviors affect feature sets which provide input data to be analyzed.

3.1 Novel Attacks' Impact on 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 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.
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.

From a theoretical point of view, the first three possibilities can be detected since intrusive behavior evidence had appeared in the feature set. For instance,

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

naive Bayes classifier can be used to detect new values appearing in features because this classifier uses all features. However, new value combinations require using attribute dependencies in order to be detected. A TAN, BAN or GBN classifiers (Friedman et al., 1997) can be suitable for detecting such anomalous evidence. As for anomalous sequence patterns, they can be detected by Bayesian classifiers if the feature set includes derived features properly summarizing past event sequences. However, anomalous audit event of fourth case can not be detected for lack of any anomalous evidence in the audit event. In practice, most novel attacks involving novel behaviors are flagged normal due to inadequate handling of novel and unusual events and insufficient decision rule.

4 WHY STANDARD BAYESIAN CLASSIFIERS FAIL IN DETECTING NOVEL ATTACKS

In intrusion detection, each instance to classify represents an audit event (network packet, connection, application log record, etc.).
Novel attacks often involve new behaviors. However, in spite of these anomalousness evidence in the feature set, Bayesian classifiers flag in most cases novel attacks as normal events. This failure is mainly due to the following problems:

1. *Inadequate Handling of Novel and Unusual Behaviors*: New and unusual values or value combinations are often involved by novel attacks. However, Bayesian classifiers handle such evidence inadequately regarding anomaly detection objectives. For instance, 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. An other problem with handling new and unusual events is floating point underflows which happen when multiplying several small probabilities.
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, Bayesian classifiers rely only on likelihood and prior probabilities to ensure classification. This strongly penalizes detection of new and unusual behaviors in favor of frequent and common behaviors. As we will see in experimental studies, standard Bayesian classifiers predict the major part of new normal/intrusive audit events as normal events (Benferhat and Tabia, 2005). Attacks often have specific signatures and may have slight variations. Consequently, a new (or very deviating) value in feature a_i will force the conditional probability $p(a_i/Attack_k)$ to zero (or an extremely negligible value in case of numeric features). Then the likelihood of the evidence will be negligible over all classes. This problem is even stressed by the weak a priori frequencies of some attack classes. As a consequence, classification will depend in this case on class prior probabilities. Given that normal training events often represent most training data (Lippmann et al., 2000)(Elkan, 2000), then new audit evidence will be classified normal. Furthermore, normal events are characterized by very large variance (Benferhat and Tabia, 2008b) because normal activities involve several users, applications, etc. This leads in most cases to a conditional probability of an attribute value in the normal class $p(a_i/Normal)$ greater than zero.

5 ENHANCING STANDARD BAYESIAN CLASSIFICATION RULE

In this section, we focus on enhancing Bayesian classifiers in order to effectively detect novel attacks.
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 most implementations, 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. In the following, we propose enhancements in order to better handle novel behaviors and effectively detect novel attacks.

5.1 Enhancing Bayesian Classification Rule to Exploit 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

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

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)<(∑(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, Rule 1 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.

5.2 Enhancing Bayesian Classification Rule to Exploit Zero Probabilities

As discussed in Section 4, anomalous audit events will 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 $Class = argmax_{c_k}(p(c_k/E))$
  ```

  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.

5.3 Enhancing Bayesian Classification Rule to Exploit Likelihood of Unusual Attacks

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 the 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 $Class = argmax_{c_k \in C}(p(c_k/E))$
```

This rule 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. For example, this threshold can fixed for attacks representing less that 1% of the training set. Then Rule 4 will be applied only for attacks representing less 1% of training instances.

Note that standard Bayesian classification rule (see Equation 1) is applied only if Rules 1, 2, 3 and 4 can not be applied. As for the priority for applying these rules, we must begin by zero probability rules (Rules 1 and 2) then normality/abnormality duality rule (Rule 3) and finally likelihood rule (Rule 4).

6 EXPERIMENTAL STUDIES

In this section, we provide experimental studies of our enhanced Bayesian classification rule on $http$ traffic including normal real data and several $http$ attacks. Before giving further details, we first present training and testing data sets then provide the experimentations' results.

6.1 Training and Testing Data Sets

We carried out experimentations on a real $http$ traffic collected on a University campus during 2007. Note that this traffic includes inbound $http$ connections to the university Web server and outbound $http$ connections of inside university users requesting outside Web servers. We extracted $http$ traffic and preprocessed it into connection records using only packet payloads. Each $http$ connection is characterized by four feature categories(Benferhat and Tabia, 2008a):

Request general features providing general information on $http$ requests. Examples of such features are request method, 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. are few examples of request content features.

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. Examples of these features are response code, response time, etc.

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).

Attacks of Table 1 are categorized according to the vulnerability category involved in each attack. Regarding attacks effects, attacks of Table 1 include denial of service attacks, vulnerability 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 normal real $http$ 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.

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%
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%
Other input validations	44	0.044%	4	0.01 %
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%**

Table 2: Evaluation of naive Bayes (NB) and TAN classifiers using standard/enhanced Bayesian classification rules on *http* traffic.

	Standard Bayesian rule		Enhanced Bayesian rule	
	NB	TAN	NB	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%
Value misinterpretation	100%	0.00%	100%	100%
Other input validation	75.0%	100%	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%**

6.2 Brief Description of Naive Bayes and TAN Classifiers

Naive Bayes classifier is the simplest form of Bayesian networks. Its graphical component only includes two node types: (1) A unique parent node called root which is associated to the hidden variable in classification problems, and (2) a child node for every observed variable (attribute). Note that naive Bayes assumes that child nodes are independent in their parent context. It is a simplifying assumption which is not true in many real world problems but useful for reducing computational complexity. In order into relax this problematic assumption, other Bayesian classifiers represent some of feature dependencies. For instance TAN classifier is a naive Bayes classifier augmented by allowing child node dependencies to form a tree (Friedman et al., 1997). We use naive Bayes classifier in order to evaluate the ability to detect anomalous events causing new feature values while TAN classifier is used for detecting new value combinations as TAN classifiers allow child node dependencies.

6.3 Standard vs Enhanced Bayesian Classification Rule on *http* Traffic

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

Note that enhanced classification rule evaluated in Table 2 uses normality/abnormality duality and zero probabilities (see Rule 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 PCC of *Normal* class in Table 2). Furthermore, TAN classifier correctly detects and identifies all known and novel attacks.

Figure 1 reports results of enhanced naive Bayes using likelihood rule (see Rule 4) with a threshold fixed to different values.

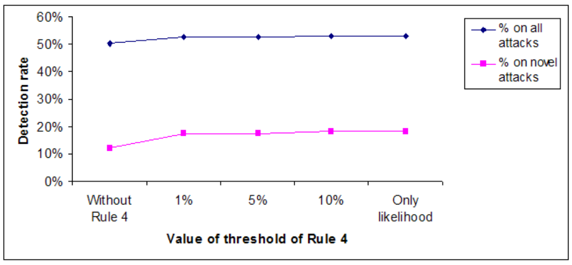

Figure 1: Naive Bayes evaluation using different thresholds for Rule 4.

Figure 1 shows that novel attacks detection rates can be improved by exploiting likelihood of attacks having small prior probabilities. For instance, fixing the threshold of Rule 4 to 1% significantly improves detection rates of several attacks since the detection of these attacks was strongly penalized by their frequencies in training data.
Results of Table 2 and Figure 1 show that significant improvements can be achieved in detecting novel attacks by enhancing standard classification rules in order to meet anomaly detection requirements.
Note that we carried out other experimentations[2] on Darpa'99 data set (Lippmann et al., 2000) and concluded that our enhancements allow significantly improving the detection of novel attacks.

7 CONCLUSIONS

In this paper, we proposed enhancements to the standard Bayesian classification rule in order to effectively detect both known and novel attacks. We firstly analyzed Bayesian classifiers failure to detect most novel attacks which they flag normal behaviors. Accordingly, we proposed to enhance standard Bayesian classification rule in order to meet anomaly detection objectives. Our enhancements aim at better handling novel and unusual behaviors and providing a Bayesian classification rule which better fits anomaly detection requirements. More precisely, we proposed enhancements to exploit normality/abnormality duality relative to audit events as well as zero probabilities caused by anomalous evidence occurrence and likelihood of attacks having extremely small prior probabilities. Experiments on recent $http$ traffic involving real data and several Web attacks showed the significant improvements achieved by the enhanced classification rule in comparison with the standard one.

ACKNOWLEDGEMENTS

This work is supported by MICRAC project (http://www.irit.fr/MICRAC/).

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.

Benferhat, S. and Tabia, K. (2008a). Classification features for detecting server-side and client-side web attacks. In *23rd International Security Conference*, Italy.

Benferhat, S. and Tabia, K. (2008b). Context-based profiling for anomaly intrusion detection with diagnosis. In *ARES2008 : Third International Conference on Availability, Reliability and Security*, Barcelona, Spain.

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.

Jensen, F. V. (1996). *An Introduction to Bayesian Networks*. UCL press, London.

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. *Computer Networks*, 34(4).

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*.

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.

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

[2] Because of the limit on the number of pages, we cannot report the results on Darpa'99 data set