Design of an Intrusion Detection System Based on Bayesian Networks

MILAN TUBA, DUSAN BULATOVIC
Faculty of Computer Science
Megatrend University Belgrade
Bulevar umetnosti 29
SERBIA
tubamilan@ptt.rs dbulat@megatrend.edu.rs

Abstract: - This paper describes a structure of a standalone Intrusion Detection System (IDS) based on a large Bayesian network. To implement the IDS we develop the design methodology of large Bayesian networks. A small number of natural templates (idioms) are defined which make the design of Bayesian network easier. They are related to specific fragments of Bayesian networks representing the basic elements in reasoning about uncertain events. The idioms represent the graphical structure, without the probabilistic tables. The use of idioms speeds-up the development of Bayesian networks and improves their quality. Example network is constructed and examined. Such Bayesian network can represent an independent agent in a distributed system. Results are promising since with very limited computation and low sensitivity to the quality of prior knowledge, potentially dangerous situations are successfully detected and classified.

Key-Words: - Privacy, Security, Networks, Data protection, Bayesian network, Intrusion detection system (IDS)

1 Introduction

This paper is a continuation of [1]. Medical patience records are kept today in electronic form and should be accessible from anywhere, anytime. A patient may be in a distant country, far from hospitals, in the mountains, or on a ship, and may need emergency medical help which will be more successful if all the relevant medical records are accessible. Such situation require organization of medical records where they are as available as possible. On the other hand, medical patience records comprise sensitive, private information that has to be protected as well as possible. It is difficult to define an intrusion (inappropriate use of medical records) because, on one hand medical records have to be accessible to unknown users outside hospitals and, on the other hand, intruders can be legitimate users from hospitals. Sophisticated, automated Intrusion Detection System (IDS) is needed.

Aggregation of personal information from illegal information brokers can be disclosed by monitoring user's behavior and noting anomaly in the sense of suddenly accessing a mass of sensitive records.

2 Previous Work

The methodology of intrusion detection is very diverse [2] and is based on the statistical data analysis, expert systems, techniques using artificial intelligence, network traffic or various forms of the attack manifestation in the network. The resulting system can be either monolith, if it functions on the basis of data acquired from the network and control records, or distributed if it is composed of network agents dispersed in the entire network and mutually communicating or acting as stand-alone agents.

2.1 Statistical Approach

Statistical approach is one of the major approaches in anomaly intrusion detection systems. When this method is used, the anomaly detector observes the activity of subjects and generates profiles for them that captures their behavior. These profiles are updated regularly, with the older data appropriately aged. As input audit records are processed, the system periodically generates a value indicative of its abnormality.

Statistical methods give a straightforward way in detecting deviation from normal behavior and that is an advantage of anomaly intrusion detection that statistical techniques have applicability there. For example, data points that lie beyond a factor of the standard deviation on either side of the average might be considered anomalous. Or the integral of the absolute difference of two functions over time might be an indicator of the deviation of one function over the other.

Statistical intrusion detection systems also have several disadvantages. Even if statistical measures could be defined to capture the computer usage patterns unique to every user, by their very nature, these measures are insensitive to the order of the occurrence of events. That is, they miss the sequential interrelationships between events. For intrusions reflected by such an ordering of patterns, a purely statistical intrusion detection system will miss these intrusions.

Purely statistical intrusion detection systems also have the disadvantage that their statistical measures capturing user behavior can be trained gradually to a point where intrusive behavior is considered normal. Intruders who know that they are being so monitored can train such systems over a length of time. Thus most existing intrusion detection schemes combine both a statistical part to measure aberration of behavior, and a misuse part that watches for the occurrence of prespecified patterns of events.

It is also difficult to determine the right threshold above which an anomaly is to be considered intrusive. And, to apply statistical techniques to the formulation of anomalies, one has to assume that the underlying data comes from a quasi-stationary process, an assumption that may not always hold.

An open issue with statistical approaches in particular, and anomaly detection systems in general, is the selection of activities to monitor. It is not known exactly what is the subset of all possible activities that accurately predicts intrusive activities. Static methods of determining these activities are sometimes misleading because of the unique features of a particular system. Thus, it seems that a combination of static and dynamic determination of the set of activities should be done.

Statistical intrusion detection is supervised classification technique where user profiles are created based on each user's observed behavior. Unsupervised classification differs from supervised classification in that we don't assume the presence of an external teacher that can train the classifier. It is desirable to train the classifier in applications where the expected results are not known in advance.

2.2 Artificial Intelligence Approach

Artificial Intelligence is concerned with improving algorithms by employing problem solving techniques used by human beings, such as learning, or gaining the ability to perform tasks from examples and training. It is used in Intrusion Detection (ID) for anomaly detection, data reduction, or discovery of rules explaining audit data and user behavior classification. The network connection classification problem is also related to ID since intruders can create private communications services undetectable by other means.

There are several approaches to the attack detection using artificial intelligence. **Expert system** codes the knowledge about attacks and consolidates the facts derived from the events registered in a control record. Furthermore, the rules are coded to specify the premises of attacks in ***if*** part. When all premises on the left side of a rule are satisfied, the rule is fired and the action on its right side is executed. The rules must be modified manually and can contain a probabilistic or statistical component. Expert systems do not have usually built-in the sequential manipulation of data. In other words, the rule base corresponding to the left side is not recognized as sequential by the system. To achieve the natural ordering of facts, one of the possible approaches is to test the constraints of ordering for every available pair after the generation of the rule set corresponding to the left side. The limitations of expert systems in attack detection are thatonly known weaknesses can be detected.

2.3 Classifier systems

Classifier systems learn how to classify future patters applying the data set used to train them [3], [4], [5]. One example of such a system is the **neural network**. The neural network is trained to predict the next user action if a window of *n* previous actions or commands is given. Neural networks represent highly connected networks trained on a set of representative user commands, with hope that the network will correctly classify next examples. After the training period, the neural network tries to find the coincidence between actual user commands and profile of a user already present in the network. Every incorrectly predicted event is, in fact, a measure of the user discrepancy from an established profile.

However, there exist certain problems in the application of neural networks in IDS. Small training window may result in giving false positive data (there is no attack), but too large window may cause false negative data (there is an attack). Second, the network topology is defined only after a significant number of training examples. Finally, the training must be conducted in a real environment, thus giving to an intruder the opportunity to train himself the network. This last disadvantage constraints the application of neural networks in IDS.

Another example of classifier systems is based on the **inductive approach**. The decision-making tree is constructed to separate data (examples) into two or more groups. Afterwards, every group is split into new subgroups until no decomposition is further possible.

2.4 Genetic Programming

A paradigm of **genetic programming**, represents a solution to searching for the most appropriate computer program producing the desired output for a given input. In this paradigm, the population of computer programs grows according to Darwin's principle of survival using a suitable genetic recombination operator. Instead of a large monolith IDS module, a finer and more subtle approach is used: a series of autonomous agents are created, working independently each other and the system. An agent is defined as a system trying to achieve a set of goals in a complex and dynamic environment. In the context of attack detection, every agent tries to detect the anomalies due to attack under the continuously variable conditions. In other words, an agent represents IDS. If an IDS can be decomposed into a lot of functional entities which may work as stand-alone, every entity could represent a particular agent. In this way, one obtains a large number of IDS which may run simultaneously. The advantage of using the genetic programming observed through the described model of autonomous agents are the efficiency, fault tolerance, resistance to degradation, possibility of extension and increase of IDS where the agents are used.

IDS designed as a set of small agents has certain advantages over a monolith one. A clear analogy can be made with human immunological system. Human immunological system is made of a large number of white grains distributed in the organism. They must detect and attack all strange and suspicious elements before they jeopardize the organism.

A noticed disadvantage of this approach is the overloading of central processor and network traffic due to a large number of agents. Furthermore, it is required a considerable time to train the agents; if an intruder succeeds to infiltrate into the training procedure, the whole system will be jeopardized. However, an interesting possibility is opened when IDS contains a large number of agents - agents can disconnect a suspicious link instead of reporting an attack The key point of this model of attack detection is the choice of a real scenario of attack used in the agent training. That problem should be a subject of research by experts in the field of data and system protection.

2.5 Bayesian Networks

Bayesian networks represent a new approach to detection and prevention of attacks in computer networks [6]. The application of Bayesian networks in IDS solves the majority of problems present in previously discussed methods.

Bayesian Networks allow the representation of causal dependencies between random variables in graphical form and permit the calculation of the joint probability distribution of the random variables by specifying only a small set of probabilities, relating only to neighboring nodes. This set consists of the prior probabilities of all the root nodes (nodes without parents) and the conditional probabilities of all the non-root nodes given all possible combinations of their direct predecessors. Bayesian networks, which are DAGs (directed acyclic graph) with arcs representing causal dependence between the parent and the child, permit absorption of evidence when the values of some random variables become known, and provide a computational framework for determining the conditional values of the remaining random variables, given the evidence.

Besides, Bayesian networks offer significant advantages which are not possible to implement using other methods [9], [10], [11], [12]. Relations between events are not given on the basis of expert knowledge, but represent mutual relations between events in the domain under consideration. Network training is not possible in a real environment; thus, systems for detection and prevention of attacks are not exposed to training by an intruder. Due to interrelationships of independent IDS, the previous knowledge on attacks in the entire network is synthesized and integrated into a unique system. Events used to estimate the probability of attacks are analyzed at the network location where they happened; in this way, an unnecessary communication and processing overloading are avoided, and the problem of incompatible various control records (generated at different computer systems) does not exist.

System using Bayesian network offers an unique advantage over other systems when one calculates the influence of newly produced events on the other observed events; accordingly, all data and rules used in other systems can be built into IDS based on Bayesian networks. Bayesian networks provide a full compatibility of corresponding software products without respect to platform used for execution; this fact can speed-up the development and application of standalone and distributed IDS.

3 Some Design Aspects of Large Bayesian Networks

Bayesian network represents the central component of the system which, when initialized by needed data, gives an estimate of the probability that an attack is going on. As a result of Bayesian network activity, the following steps can be undertaken:

- automatic activation of attack prevention
- reporting an attack in due course
- signaling the possibility of an attack
- proposing the countermeasures to the system administrator

Data processed by Bayesian network are elicited events which, according to the designer estimates, should be observed in order to bring the right conclusions on the possibility of certain type of attack. The interdependence of events is represented by directed paths connecting events, and the state conditional probabilities of events are described by the corresponding tables. Once a network is described in a graphical and tabular form, any state change of some observed event will trigger the calculation of the changed probabilities of other observed events; the purpose is to obtain the probability that an attack on computer or local network is going on.

The result of the propagation of changed probabilities of certain events observed by Bayesian network can be an automatic activation of some mechanism for attack prevention such as: breaking TCP connections, traffic redirection or disabling user activity. If the probability of an attack is significantly enlarged but not in proportion to be considered as an attack, the network will generate a report about the event and warning to the system administrator. The component **Bayesian Management Information Base** (BMIB) is used as a data storage on attacks and events which caused them.

To implement the idea of IDS, it is necessary to develop the design methodology of large Bayesian networks. Building Bayesian network for solving the problem of larger complexity requires the solution to the following problems:

- Building the graph structure (topology)
- Defining the probabilistic tables for each graph node.

One of the approaches to the implementation of large Bayesian networks which will be presented here is similar to the design of large systems [13], [14], and consists of the following steps:

- Development of tool for design of small components representing the basis of Bayesian network
- Development of tool for interconnecting these components in such a way to provide an efficient control of network complexity.

Toward this goal, a small number of natural templates are defined which make the design of Bayesian network easier. These templates are called idioms. They are related to specific fragments of Bayesian networks representing the basic elements in reasoning about uncertain events. The idioms represent the graphical structure, without the probabilistic tables. The use of idioms speeds-up the development of Bayesian networks and improves their quality. The following idioms are identified:

- *Definition/Synthesis idiom* – models the synthesis, that is, the combination of several nodes into one for the sake of better organization of Bayesian networks; it models, too, the probabilistic and deterministic relations between variables
- *Cause-effect idiom* – models the uncertainty of some causal processes with effects which can be observed
- *Measurement idiom* – models the accuracy uncertainty of some measuring instruments
- *Induction idiom* – models the uncertainty of inductive reasoning based on the population of similar or replaceable members
- *Reconciliation idiom* – models the result reconciliation of alternative measurement methods or prediction systems.

Idioms serve as the template library in the design process of Bayesian networks. We present here only those idioms which, by authors opinion, can be of use in design of IDS. They do not represent the final list of all type of reasoning in all domains. The designer may identify and define new idioms.

3.1 Definition/Synthesis idiom

Definition/Synthesis idiom is shown in Fig. 1; it models all cases where a synthesized node is determined by the values of super-nodes using some combinational rules.

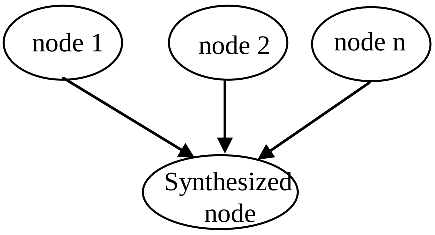


Fig. 1 Definition/Synthesis idiom

First case. Definitional relation between variables. In this case, synthesized node is defined through nodes: *node 1, node 2, ..., node n* (where *n* can be any integer). For example, if node C is defined by relation C=A|B, the Definition/Synthesis idiom would have the form as in Fig. 2.

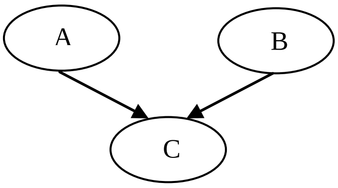


Fig. 2 Definition/Synthesis idiom for relation C = A|B

Second case. Combining different nodes into one with the goal to reduce the combinatorial explosion (separation).

It is possible to connect some nodes of interest into a synthesized node instead to connect them to super-nodes in order to make easier the determination of probabilities and to reduce the effect of combinatorial explosion. For example, if there are four variables, A, B, C, D, with four states each, then the number of probability values needed to fill the table of conditional probabilities for node A, p(A|B, C, D), amounts 256. Instead of it, this table can be decomposed into two tables for p(A|B, S) and p(S|C, D) by introducing node S as a synthesized node, Fig. 3. Now, it is needed to model the tables of conditional probabilities for nodes S and A, what requires only 64 values instead of 256.

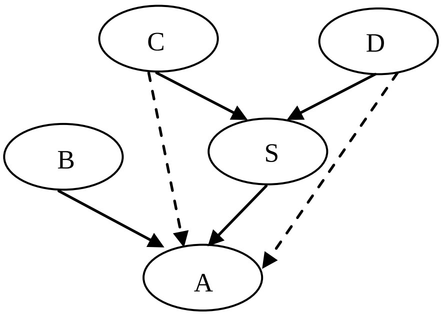


Fig. 3 Introducing synthesized node

This technique of shortening the combinatorial space is called the separation process. Here, the synthesized node S separates super-nodes C and D from node B.

Third case: Follow the organizational principles established by the experts for the node organization using the hierarchy.

One of the first questions imposed in the design of Bayesian networks is the possibility to organize the variables using some hierarchical structure through certain valid organizational principle. There are practical reasons why this is necessary - for example, if we want to observe some nodes as nodes of the same types or to have the same kind or the same degree of influence on some other node.

The hierarchy of synthesized nodes can be specified to model even some informal organizational principles required by the experts. This kind of hierarchy can model the attributes or sub-attributes of some complex structure. One example is shown in Fig. 4. Here, the experts defined variable A to be composed of two sub-attributes B and C. On the other hand, C is a composed attribute and can be decomposed into sub-attributes C1, C2, C3 and C4.

The directions of arcs at synthesized idiom do not designate the causality because it would not have sense (the causal links cannot be established towards other idioms). Instead of it, the links point to directions describing how sub-attributes define attributes or how attributes define super- attributes.

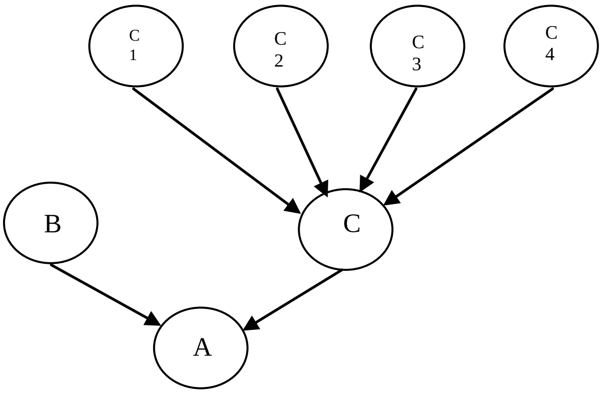


Fig. 4 Hierarchy of synthesized nodes

Definition/Synthesis idiom can be used in IDS in all cases when it is possible to describe the conditions of attack by some rules defined in advance. In this way, it is possible to cover the functionality of the attack detection; otherwise, it must comprise the systems based on rules.

3.2 Cause-effect idiom

Cause-effect idiom is used to model the causal process through relations between their causes (those events or facts which represent the inputs to the process) and their effects (those events or facts which represent the outputs from the process). Fig. 5 shows the basic cause-effect idiom. The direction of arc designates the direction of causality, that is, where inputs cause some changes on outputs through the causal process.

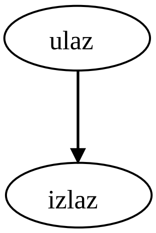


Fig. 5 Cause-effect idiom

The corresponding causal process is not represented in Fig. 5. This is not necessary because the role of the corresponding causal process in Bayesian network is represented by tables of conditional probabilities associated with input and output. This information gives all what is necessary to know about the uncertainty of relation between the cause and the effect.

The cause-effect idiom can be used in IDS in all those cases when we can estimate that some event will represent an attack with a given probability. In the circumstances of notified events, after the probability propagation in Bayesian network, one can see the real influence of a specific event on the probability of attack.

3.3 Measurement idiom

The measurement idiom represents the uncertainty which we have about the observed object. The difference between this idiom and the cause-effect idiom lies in the fact that a node in measurement idiom contains the estimated value of another node but not the attribute of a particular node.

Fig. 6 shows the measurement idiom. The directions of arcs are explained in a straightforward way. The true value must exist before the estimated one in order to perform the measurement procedure. The measuring instrument interacts (physically or functionally) with the object of measurement and gives certain result. The result is considered more or less exact, depending on the measurement circumstances and prejudices on the instrument accuracy.

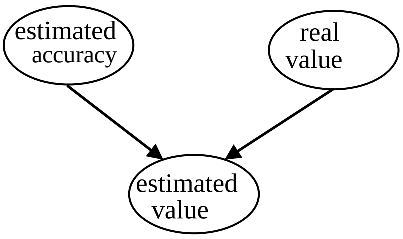


Fig. 6 Measurement idiom

The true value of an attribute is measured by a measuring instrument with known estimated accuracy, and as a result one obtains the estimated value of the attribute.

The measurement idiom can be used to explain false positive results; it is not intended do model the sequence of repeated experiments to get the real value. Also, it should not be used to model the inference which can be derived on the basis of other similar entities. For this purpose, the induction idiom is more appropriate.

The measurement idiom can be applied in IDS in those situations when we want to correct the supposed value of a variable on the basis of the accuracy estimation of the method used to make the conjecture.

3.4 Induction idiom

The induction idiom, Fig. 7, is intended to model the statistic reasoning process where a series of similar entities is used to infer something about the future entity with the similar attributes. The induction idiom has two components:

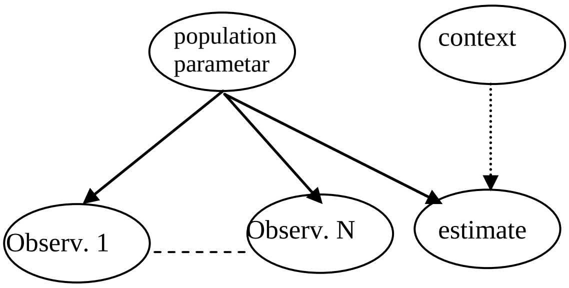


Fig. 7 Induction idiom

Induction idiom models Bayesian update of population parameters, where the population entities can be considered as mutually exchangeable.

Induction idiom allows the experts to adjust the estimate obtained if one expects that the observed entity differs from the population, that is, it is not exchangeable due to the context change.

Each observation, Fig. 7, is used to estimate the population parameter which should contain the characteristics of the whole population. It is used later as previous knowledge for next observation. Finally, the distribution of the population parameter can be used to estimate the attribute of the entity observed.

The context effect is modelled by a particular node, marked as context; it is used to adjust the population estimate in agreement with the relevance of historical parameters to the population of interest. If the historical data are very different from the current context, a probabilistic table with uniform distribution for node prognosis may be established. If the historical data are similar, the probabilistic table will be similar to table derived through Bayesian learning.

3.5 Reconciliation idiom

The reconciliation idiom is intended to reconcile the independent sources of information about the unique attribute or unique entity; the sources of information are obtained as the results of different measurements or predictions, i.e., the different Bayesian networks. The reconciliation idiom is shown in Fig. 8.

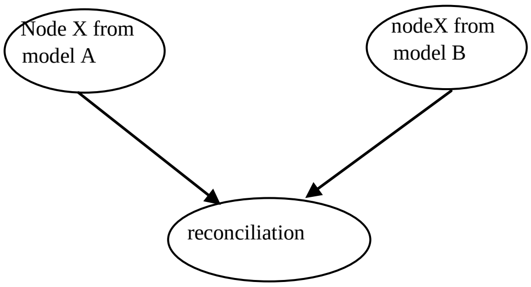


Fig. 8 Reconciliation idiom

The node of interest, node X, is estimated by two independent procedures, model A and model B. The reconciliation idiom is a Boolean node with values true or false. When a node is set to value true, value X from model A is equal to value X from model B. In this way, the transfer of information about the event from model A is transferred to model B. However, there is a constraint – if the information from model A and model B is contradictory, then the conclusion about the reconciliation, clearly, cannot be derived.

The next example of the reconciliation idiom is typical in the field of security/reliability. Suppose that there are two models for the estimation of the quality of module testing to detect the security misses:

- Prediction on the basis of causal factors represented by the instantiation of cause-effect idioms.
- Inference on the basis of sub-attributes of testing quality, which represent the partial observation of quality testing, are represented by the instantiation of the definition-synthesized idiom.

The relevant process produces an instantiated idiom shown in Fig. 9.

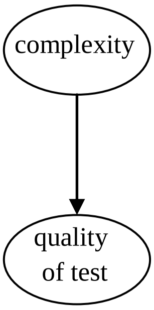


Fig. 9 Instantiation of reconciliation idiom

One of the possible implementations of the reconciliation idiom in IDS could be in DIDS. DIDS consists of the set of Bayesian networks which, independently each other, estimate the probability of the local attack. At the higher hierarchical level, Bayesian network obtains as inputs all these estimates which are used to bring the conclusion about the probability of the network attack.

During the design of Bayesian network from the scratch, the next sequence of steps is recommended to provide the choice of right idiom:

1. Make the list of entities and their attributes considered as relevant for Bayesian network
2. Perform the analysis of relationships between entities and attributes; this is used to determine the subsets of attributes and entities belonging to certain groups
3. Make choice of idiom on the basis of the reasoning type applied:
   - Cause-effect idiom – for causal reasoning
   - Measurement idiom – for causal reasoning based on observation
   - Induction idiom – for statistical reasoning or reasoning through analogy using the historical data to come to cognition about new, unknown case

- Definition-synthesized idiom – for definitional reasoning
- Reconciliation idiom – for performing the reconciliation process of two different Bayesian networks giving alternative data for the same event.

4 Example Definition

As an illustration of the proposed solution, we will present here the possibility of attack detection on the privacy of medical data. The corresponding Bayesian network is shown in Figure 10.

The probabilistic tables of conditional dependencies are filled on the basis of our previous knowledge about conditional probabilities of particular events. The physical meaning of particular variables is the following:

Aids – access to records with diagnosis of Aids

External – external access to other block or other hospital

Medical_Staff – access to records by medical staff

Non_Medical_Staff – access to records by nonmedical staff

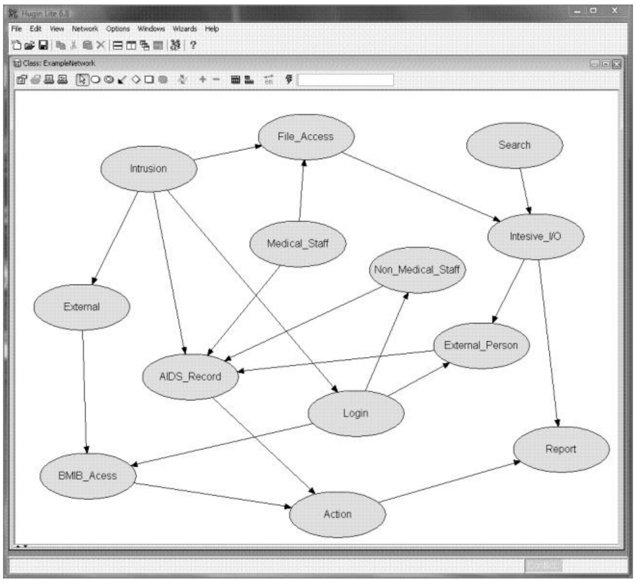


Fig. 10. Bayesian network for specific attack

External_Person – access to records by external person

Intrusion – designates an attack

BMIB_Access – access to BMIB (Bayesian Management Information Base) where the data relevant to network security are stored and read , using SNMP

File_Access – designates the usual access activity to files or intensive file access indicating higher activity of medical staff or possibly a potential attack

Login – counts the number of unsuccessful login to the system

Report – depending on the value of this variable, the administrator of the system must be informed about the events in the system

Action – designates undertaking automatic protection measures; it is supposed that an attack is going on or an attack can follow

Search – file searching; if it is unusually high, it may indicate a trial of attack

Intensive_I/O – intensive input/output activity, higher than usual, as a consequence of previous activities (File_Access or Search)

The conditional probabilities of all variables are given in the Table 1. It is necessary to fill large number of tables with probabilities and one of main criticisms of this method is that it is difficult to "know" all these probabilities. However, the same criticism is valid for neural networks and number of weights and thresholds, with added problem that prior knowledge cannot be used and training increases chances for intruder. Probabilities for Bayesian network are combination of theoretical and empirical knowledge, training and subjective estimates [15], [16].

Medical_Staff	
Yes	0.2
No	0.8

Search	
Normal	0.8
Increased	0.2

Intrusion	
Yes	0.012
No	0.988

Non_Medical_Staff			
Login	Normal	Larger	Very Large
Yes	0.01	0.5	0.8
No	0.99	0.5	0.2

Intesive_I/0				
Search	Normal		Increased	
File_Access	Intensive	Normal	Intensive	Normal
Yes	1	0	1	0.7
No	0	1	0	0.3

Report				
Action	Yes		No	
Intesive_I/0	Yes	No	Yes	No
Yes	0.75	0.75	0.5	0.01
No	0.25	0.25	0.5	0.99

File_Access				
Intrusion	Yes		No	
Medical_Staff	Yes	No	Yes	No
Intensive	0.99	0.7	0.2	0.2
Normal	0.01	0.3	0.8	0.8

Action				
AIDS_Record	Yes		No	
BMIB_Access	Yes	No	Yes	No
Yes	0.88	0.88	0.7	0.001
No	0.12	0.12	0.3	0.999

External_Person			
Intesive_I/0	Yes		
Login	Normal	Larger	Very Large
Yes	0.01	0.2	0.5
No	0.99	0.8	0.5
	No		
	Normal	Larger	Very Large
	0.001	0.02	0.1
	0.999	0.98	0.9

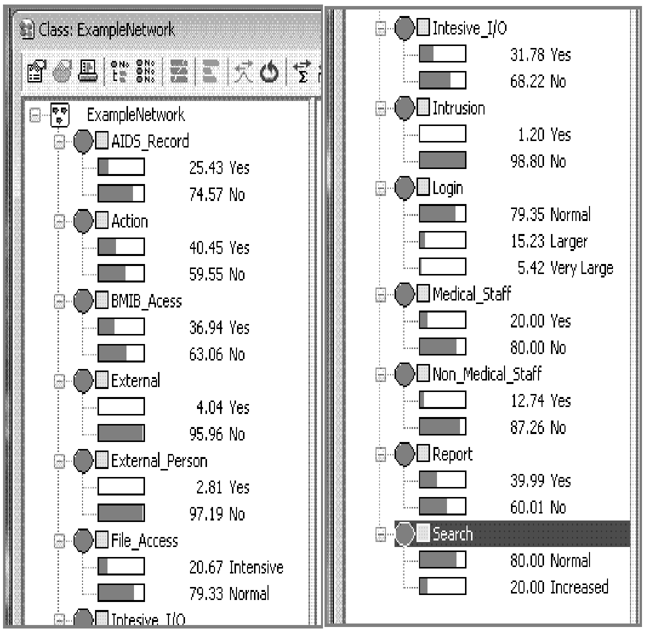
Login		
Intrusion	Yes	No
Normal	0.26	0.8
Larger	0.34	0.15
Very Larger	0.4	0.05

External		
Intrusion	Yes	No
Yes	0.9	0.03
No	0.1	0.97

AIDS_Record								
Intrusion	Yes							
External_Perso	Yes				No			
Non_Medical_	Yes		No		Yes		No	
Medical_Staff	Yes	No	Yes	No	Yes	No	Yes	No
Yes	0.8	0.9	0.5	0.99	0.6	0.95	0.5	0.99
No	0.2	0.1	0.5	0.01	0.4	0.05	0.5	0.01
	No							
	Yes				No			
	Yes		No		Yes		No	
	Yes	No	Yes	No	Yes	No	Yes	No
	0.85	0.7	0.75	0.9	0.85	0.7	0.85	0.0001
	0.15	0.3	0.25	0.1	0.15	0.3	0.15	0.9999

Table 1. Conditional probabilities for example

Figure 11. shows apriori probabilities before any events are registered. For probability calculation and propagation after registered events software Hugin Lite 6.28 [17] was used. Without such software calculations even for modest networks are difficult



Class: ExampleNetwork

ExampleNetwork
AIDS_Record
25.43 Yes
74.57 No
Action
40.45 Yes
59.55 No
BMIB_Acess
36.94 Yes
63.06 No
External
4.04 Yes
95.96 No
External_Person
2.81 Yes
97.19 No
File_Access
20.67 Intensive
79.33 Normal
Intesive_I/O

Intesive_I/O
31.78 Yes
68.22 No
Intrusion
1.20 Yes
98.80 No
Login
79.35 Normal
15.23 Larger
5.42 Very Large
Medical_Staff
20.00 Yes
80.00 No
Non_Medical_Staff
12.74 Yes
87.26 No
Report
39.99 Yes
60.01 No
Search
80.00 Normal
20.00 Increased

Fig. 11 Prior probabilities

5 Example Solution

Now it is possible to make different assumptions about possible attack scenarios to check behavior of the Bayesian network regarding attack detection and prevention.

If we assume that intensive file access was noted (Table 2, line 1), after the probabilities are propagated, report will be generated but no action will be taken and event will not be registered as significant for security.

Trigger=1	Report	Action	BMIB_Access	Intrusion
File_Access	**61**	42	38	4
Intensive_I/O	**60**	42	38	3
Search	**53**	41	37	1
Login	**61**	**75**	**72**	9
Search & External	**64**	**70**	**82**	27
Medical_Staff	**63**	**79**	37	1
Non_Medical_Staff	**61**	**75**	**57**	5
External_Person	**67**	**79**	**60**	9
Medical_Staff & External	**66**	**81**	**82**	27
Non_Medical_Staff & External	**67**	**83**	**87**	**59**
External_Person & External	**70**	**84**	**88**	**76**

Table 2. Results

Similarly, if **Intensive_I/O** or **Search** happen isolated (only one of them) only report will be generated, Table 2, lines 2 and 3.

But if **Login** (unsuccessful logins count) happens, beside report, an action will be taken (for example disconnecting terminal from which suspicious login attempts are coming) and BMIB record will be generated for later analysis. However, all this events will not be classified yet as an intrusion, Table 2, line 4.

Next scenario shows that intensive file search but combined with access to sensitive data somewhere else in the network will trigger an action for intrusion prevention, BMIB record (and, as usually, a report), Table 2, line 5.

Last six examples deal with more serious attacks. An exceptional capability of attack detection Bayesian network will demonstrate in cases when it is needed to estimate simultaneous influence of several variables.

Isolated events **Medical_Staff**, **Non_Medical_Staff** or **External_Person** will generate report and cause action (some mild action, like more careful monitoring) but no intrusion alert. **Non_Medical_Staff** or **External_Person** will also generate BMIB record, Table 2, lines 6, 7 and 8.
Case when simultaneously medical person accesses records and there are external accesses is still below intrusion threshold, Table 2, line 9. But, combined access by non-medical or external person with external activity will alert intrusion, where the second one is more pronounced, Table 2, lines 10 and 11.

6 Conclusion

On the basis of the obtained results in chosen examples (attacks to privacy), one can conclude that IDS is functioning equally efficiently in the detection process as well as in the decision process about the actions which will be undertaken to prevent the initiated attack. To make a decision, the current and historical data relevant to security are used.

Bayesian networks simulate the causality between the events in the domain; however, it is possible to enter all relations which can be expressed through the rule base as well as the knowledge acquired additionally on the basis of experience. By simulating various scripts of attacks, as it is done in the examples illustrating the methods, one can verify the functionality of IDS even in the situations which cannot be predicted by rules, i.e., defined previously by experts.

Bayesian reasoning in making decision about uncertain events, applied to detection and prevention of attacks on computers and computer networks is a reliable and efficient approach. Its implementation can be significantly relaxed by the existence of available software algorithms for the calculation of the probability propagation through Bayesian networks, which can be then provided with the data in a particular environment. This process can be significantly speeded-up by the efforts done with the purpose to standardize the messages and procedures of data exchange between intrusion detection systems.

Examples that are presented here, and many additional ones, show adjustability and suitability of Bayesian network for intrusion detection. Work is in progress to further develop this system and to implement this Bayesian network as an independent agent in a distributed system.

References:

[1] M. Tuba, D. Bulatovic and O. Miljkovic: On Suitability of Bayesian Classification for Aggregation Control and Intrusion Detection, *Proceedings of the 29th International Conference of the Romanian Medical Informatics Society MEDINF 2007,* pp. 233-239.

[2] T. Lunt: A Survey of Intrusion Detection Techniques, *Computers&Security*, 12(4), pp. 405-418, June 1993.

[3] S. Segrera and M.N. Moreno: Application of Multiclassifiers in Web Mining for a Recommender System, *WSEAS Transactions on Information Science & Applications*, Vol. 3, No.12, 2006, pp. 2471-2476.

[4] Y. Yu, Y. Yu, H. Lin, and C. Chen: Mining on Web Logs for Recommendation, *WSEAS Transactions on Computers*, Vol.5, No.9, 2006, pp. 1818-1822.

[5] Ruey-Shun Chen 1, Yung-Shun Tsai, K.C. Yeh, D.H. Yu, and Yip Bak-Sau: Using Data Mining to Provide Recommendation Service, *WSEAS Transactions on Information Science & Applications*, Vol. 5, No. 4, 2008, pp. 459-474.

[6] S. L. Scott: A Bayesian paradigm for designing intrusion detection systems, *Computational Statistics & Data Analysis*, Volume 45, Issue 1, 2004, pp. 69-83

[7] J. Ma and K. Sivakumar: Privacy-Preserving Bayesian Network Parameter Learning, *WSEAS Transactions on Information Science & Applications*, Vol. 3, No.1, 2006, pp.1-6.

[8] A. Heni, M.N. Omri, and A.M. Alimi: Fuzzy Knowledge Representation Based on Possibilistic and Necessary Bayesian Networks, *WSEAS Transactions on Information Science & Applications*, Vol. 3, No.2, 2006, pp.224-231.

[9] A. A. Sebyala, T. Olukemi, L. Sacks: Active Platform Security through Intrusion Detection Using Naïve Bayesian Network for Anomaly Detection, *Proceedings of London Communications Symposium 2002,* http://www.ee.ucl.ac.uk/ lcs/papers2002/LCS116.pdf

[10] C. Kruegel, D. Mutz, W. Robertson and F. Valeur: Bayesian event classification for intrusion detection, *Computer Security Applications Conference Proceedings* , Dec. 2003, pp. 14- 23

[11] S. Chebrolu, A. Abraham and J. P. Thomas: Feature deduction and ensemble design of intrusion detection systems, *Computers & Security*, Volume 24, Issue 4, June 2005, pp. 295-307

[12]H. Tu, J. Levchuk, and K. R. Pattipati, .Robust action strategies to induce desired effects,. *IEEE Transactions on Systems, Man & Cybernetics - Part A: Systems and Humans*, Vol. 34, No. 5, pp. 664-680, September 2004.

[13]Wenhui Liao, Weihong Zhang, and Qiang Ji, A Factor Tree Inference Algorithm for Bayesian Networks and its Application, *The 16th IEEE International Conference on Tools with Artificial Intelligence*, Nov., 2004.

[14]K. B. Laskey and S. M. Mahoney. Network engineering for agile belief network models, *IEEE Transactions on Knowledge and Data Engineering*, Vol. 12, pp. 487–498, 2000.

[15]M. J. Druzdzel and L. C. Van der Gaag. Building Bayesian networks: Where do the numbers come from? *IEEE Transactions on Knowledge and Data Engineering*, Vol. 12, pp. 481–486, 2000.

[16]Laskey, K. B.: Sensitivity Analysis for Probability Assessments in Bayesian Networks, *IEEE Transactions on Systems, Man and Cybernetics*, Vol. 25, No. 6, pp. 901-909.

[17] Hugin Lite 6.28 software for Bayesian networks, www.hugin.dk

Acknowledgment: The research was supported by the Ministry of Science, Republic of Serbia, Project no. 144007