International Journal of Computer Science & Applications
Vol. 2, No. 1, pp. 1 - 6

Intrusion Detection Effectiveness Improvement by a Multiagent System

Agustín Orfila , Javier Carbó and **Arturo Ribagorda**[1]

Abstract. Recent studies about Intrusion Detection Systems (IDS) performance reveal that the value of an IDS and its optimal operation point depend not only on the Hit and False alarm rates but also on costs (such as those associated with making incorrect decisions about detection) and the hostility of the operating environment. An adaptive multiagent IDS is proposed in this paper and it is evaluated according to a promising metric that take into account all these parameters. This paper shows results of a prototype that clearly point out how multiagent technology can improve IDS effectiveness.

1 INTRODUCTION

The main task of any detection system is recognizing whether a specified condition is present or absent [19]. For IDS this condition is an intrusion attempt. The model to determine whether an intrusion is present or not could take many forms. Whatever the model is, a detector's performance can be described by its receiver operating characteristic (ROC) curve. ROC curve is a plot of the detection probability (H) versus false alarm rate (F). ROC analysis was originally introduced in the field of signal detection theory in the 50's [7].

The evaluation of intrusion detection systems (IDS) has begun to be an active topic over the last years [6] [8] [10] [14]. The 1998 and 1999 DARPA evaluations conducted by the Massachusetts Institute of Technology (MIT) Lincoln Laboratory [10] , [11] and McHugh's critique [14] of such experiments have shown that this area needs much more research and experimentation before a framework for the evaluation of IDS effectiveness can be widely accepted.

At present time, the work of [8] is the most complete study about a formal IDS effectiveness evaluation. They establish that the value of an IDS depend not only on the ROC curve but also on the costs and the hostility of the operating environment (as summarized by the probability of intrusion). They are the first in applying decision analysis techniques [4] [6] to the field of Intrusion Detection. Their method uses a decision tree that shows the relationship between the condition (if there is an intrusion or not), the detector's report (existence or not of an alarm), the probability of the detector to make erroneous reports, the prior intrusion probability and the action taken by the decision system based on the detector's report. The action will be the one that minimizes expected cost.

The decision system will respond or not depending on which action minimizes expected cost. Stolfo et al. [18] defined three sets of costs for IDS: damage, response and operational. Damage cost is due to detector's errors. Response cost corresponds to taking some action upon an intrusion when it is detected. The operational cost is the cost needed to run the IDS; it is not relevant to the decision problem considered here. [8] only considers damage cost. Our approach extends their model taking into account the response cost.

Multiagent systems have been applied to intrusion detection in the past. For instance, a methodology that uses intelligent agents to provide automated intrusion response [3] and an appropriate agent architecture [1] was proposed.

Agents look for a complete automation of complex processes acting in behalf of human users [12]. From the Artificial Intelligence point of view, agents are classified as reactive or deliberative according to the external or internal nature of the intelligent behavior. Deliberative agents often accomplish the so called BDI paradigm [16] that structures knowledge in three different levels of abstraction: beliefs, desires, and intentions.

Intelligent agents are supposed to adapt decision making through the cooperation with other agents [16]. Human-like typed messages usually model communication between agents, including performatives inspired in Speech Act Theory (for instance KQML [17]).

The remainder of this paper is organized as follows: section 2 explains the decision model proposed that includes response costs, section 3 describes our multiagent model approach and section 4 shows the results on our prototype using the metric introduced in section 2. Finally, section 5 finishes with the main conclusions.

2 DECISION MODEL ANALYSIS

The system to be analyzed can be in two possible states. With an intrusion (I) or without it (NI). The prior probability is represented by p. The estimation of prior probabilities is familiar to Bayesian statistics.

An IDS can launch an alarm (A) or not (NA). The ROC parameters are: the probability of an alarm given an intrusion, $P(A|I) = H$ and the probability of an alarm given no intrusion, $P(A|NI) = F$.

Consider a potential IDS which can take some precautionary action depending on the likelihood that an intrusion will occur. Taking precautionary action incurs a cost C irrespective of whether the intrusion occurs or not. However, if the intrusion occurs and no action has been taken, then a loss L is incurred. The expense associated with each combination action/inaction and occurrence/non-occurrence of the intrusion is given in the decision-model contingency matrix shown in Table 1. Figure 1 represents the decision tree with the sequence of actions (squares) and uncertain events (circles) that describes the detector's operation and the responses that can be taken according to the detector's report. The costs shown corresponds to the consequences of the action taken.

Action nodes (squares) are under the decision system control which decides what branch to follow. Event nodes (circles), are not

[1] Universidad Carlos III de Madrid, Leganés, Spain email: {adiaz, jcarbo, arturo}@inf.uc3m.es

Table 1. Decision model contingency matrix.

		Occurs	
		No	Yes
Take action	No	0	L
	Yes	C	C

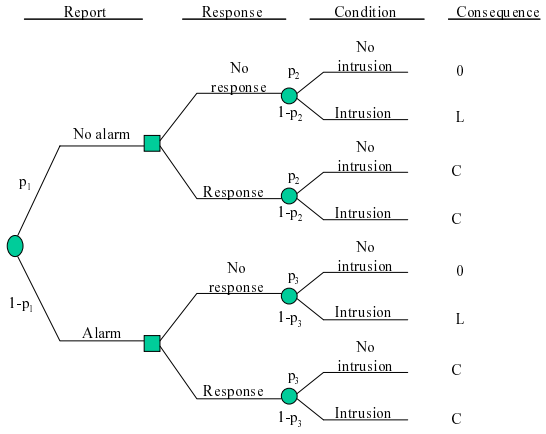

Figure 1. Decision tree of the detector's expected cost

under the decision system control but depend on uncertainty. A probability distribution represents the uncertainty about which branch will follow an event node.

Each combination of actions and events is characterized by its cost. There is a probability of occurrence associated to each uncertain event. There are three probabilities specified in the tree:

- p_1: the probability that the detector reports no alarm.
- p_2: the conditional probability of no intrusion given that the detector reports no alarm.
- p_3: the conditional probability of no intrusion given that the detector reports an alarm.

The last two probabilities account for both possible detector errors, falsely reporting that there is an intrusion (p_3) and falsely reporting that there is no intrusion ($1 - p_2$).

The expected cost is determined for each event node by taking the sum of products of probabilities and costs for all of the node's branches. The expected cost at a decision node is the lowest expected cost from among the node's outgoing branches.

An operation point is defined by a pair (F,H). The probabilities of the detector's report are calculated by applying the formula of total probability:

$$p_1 = P(NA) = P(NA|NI) \cdot P(NI) + P(NA|I) \cdot P(I) = (1-F) \cdot (1-p) + (1-H) \cdot p \quad (1)$$

$$1 - p_1 = P(A) = P(A|NI) \cdot P(NI) + P(A|I) \cdot P(I) = F \cdot (1-p) + H \cdot p \quad (2)$$

The probabilities of the system's state depending on the detector's rate are calculated by applying Bayes Theorem [5]:

$$p_2 = P(NI|NA) = \frac{P(NA|NI) \cdot P(NI)}{P(NA)} = \frac{(1-F) \cdot (1-p)}{p1} = \frac{(1-F) \cdot (1-p)}{(1-F) \cdot (1-p) + (1-H) \cdot p} \quad (3)$$

$$1 - p_2 = P(I|NA) = \frac{P(NA|I) \cdot P(I)}{P(NA)} = \frac{(1-H) \cdot p}{p_1} = \frac{(1-H) \cdot p}{(1-F) \cdot (1-p) + (1-H) \cdot p} \quad (4)$$

$$p_3 = P(NI|A) = \frac{P(A|NI) \cdot P(NI)}{P(A)} = \frac{F \cdot (1-p)}{1-p_1} = \frac{F \cdot (1-p)}{F \cdot (1-p) + Hp} \quad (5)$$

$$1 - p_3 = P(I|A) = \frac{P(A|I) \cdot P(I)}{P(A)} = \frac{H \cdot p}{1-p_1} = \frac{H \cdot p}{F \cdot (1-p) + H \cdot p} \quad (6)$$

The expected cost of each response is calculated by taking the sum of the products of the probabilities and costs for the node following the response.

The results of the expected costs are shown in Table 2.

The expected cost given the detector's report is the expected cost of the least costly response given the report. So the expected cost given no alarm is:

$$\frac{Min\{L \cdot (1-H) \cdot p, C \cdot (1-F) \cdot (1-p) + C \cdot (1-H) \cdot p\}}{p_1} \quad (7)$$

Similarly, the expected cost given an alarm is :

$$\frac{Min\{L \cdot H \cdot p, C \cdot F \cdot (1-p) + C \cdot H \cdot p\}}{1-p_1} \quad (8)$$

The expected cost of operating at a given point on the ROC curve is the sum of the products of probabilities of the detector's report and the expected costs conditional on the reports. The expected cost of operating at an operating point is:

$$Min\{L \cdot (1-H) \cdot p, C \cdot (1-F)(1-p) + C \cdot (1-H) \cdot p\} + Min\{L \cdot H \cdot p, C \cdot F \cdot (1-p) + C \cdot H \cdot p\} \quad (9)$$

The expected cost per unit loss (M) is:

$$M = Min\{(1-H)p, C/L(1-F)(1-p) + C/L(1-H)p\} + Min\{Hp, C/LF(1-p) + C/LHp\} \quad (10)$$

It is important to mention that our formulation includes the possibility that a decision is made to respond or not, regardless of the detector's report. This makes this model stronger than others [9] [13] [15].

Table 2. Expected costs of responses conditional on the detector's report.

	No Response	Response
No alarm	$L \cdot (1-p_2) = \frac{L \cdot (1-H) \cdot p}{p_1}$	$C \cdot p_2 + C \cdot (1-p_2) = C \cdot (1-F) \cdot (1-p) + \frac{C \cdot (1-H) \cdot p}{p_1}$
Alarm	$L \cdot (1-p_3) = \frac{L \cdot H \cdot p}{1-p_1}$	$C \cdot p_3 + C \cdot (1-p_3) = C \cdot F \cdot (1-p) + \frac{C \cdot H \cdot p}{1-p_1}$

For a perfect deterministic forecast H=1, F=0, hence

$$M_{per} = Min\{p, C/L \cdot p\} \tag{11}$$

To calculate the expected cost per unit loss knowing only the prior probability of intrusion, suppose first the decision maker always protects (equivalent to using a prediction system where the event is always predicted and for which H = 1 and F = 1), then

$$M_{fre1} = Min\{p, C/L\} \tag{12}$$

Conversely, if the decision maker never protects (equivalent to using a prediction system where the event is never predicted and for which H = 0 and F = 0) then:

$$M_{fre0} = Min\{p, C/L\} \tag{13}$$

So if the decision maker knows only the prior probability of intrusion p, M can be minimized depending on whether $C/L < p$, or $C/L > p$. Hence, the mean expense per unit loss associated with the only knowledge of the probability of intrusion is:

$$M_{fre} = Min\{p, C/L\} \tag{14}$$

We define the value of an IDS prediction to be a measure of the reduction in M over M_{fre}, normalized by the maximum possible reduction associated with a perfect deterministic forecast, i.e.:

$$V = \frac{M_{fre} - M}{M_{fre} - M_{per}} \tag{15}$$

For a predictive system which is no better than the one based on the probability of intrusion, V = 0; for a perfect deterministic system V = 1. For a parametric IDS that depends on a threshold p_t, the hit and false alarm rates will also depend on it $H = H(p_t)$, $F = F(p_t)$. Hence V is also defined for each p_t, ie $V = V(p_t)$.

For a given C/L relationship, the optimal value is $V_{opt} = maxV(p_t)$.

This metric is very useful because it includes all the relevant parameters involved in the evaluation of the IDS effectiveness. A similar metric was proposed in [15] but it did not manage the possibility that a decision is made contrary to the detector's report.

3 THE MULTIAGENT SYSTEM

The proposed multiagent system intends to completely automate intrusion detection in a distributed way. We assume that several intrusion detection techniques can be applied on the same traffic data, and the performative of those techniques on a specific traffic data can not be known a priori. We will show that with cooperative autonomous agents it is possible to compute a dynamic evaluation of the predictive ability of several IDS that would lead to more successful intrusion detections.

In our system, agents play different roles: predictor, evaluator and manager. Although several combinations of these three types of agents are possible, and they have been tested with different intentions [2], we consider a particular setup according to the intrusion detection problem: several predictor agents and just one evaluator agent and one manager agent.

- The main role of a predictor agent is to guess if there is an intrusion or not when an evaluator agent asks it for a prediction. Each of them implements a specific Intrusion Detection technique over a shared traffic data.
- On the other hand, the goal of evaluator agents is giving proper weights to predictor agents. Each weight is computed according to the previous level of success of the corresponding predictor. Evaluator agents will communicate the result (a binary decision based on such weighted references) to the manager agent. Evaluator agent are in charge to update dynamically the weights, and the way this updating is computed is the key factor for improving IDS effectiveness. It would also make sense to use several evaluator agents rather than only one, in order to make them able to adopt different weighting criteria.
- Finally the manager agent has a posteriori information about if there was really an intrusion or not. An automated post-morten analysis should be computed by this type of agent. The manager runs under a training environment in order to make the system learn. Therefore, the manager agent is able to calculate H, F and the value V, and it will communicate the results to the evaluator agent.

An illustrative example of agents interactions is shown in Figure 2.

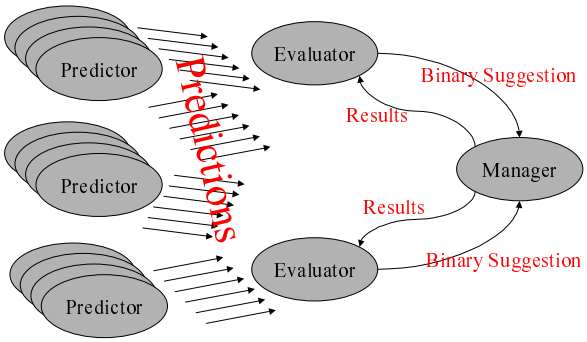

Figure 2. Interactions between agents

Since agents are supposed to be acting in behalf of humans, we can see predictor agents as operators of IDS, evaluator agents as ex-

perts that weight the results from operators, and manager agents as post-morten analyzers. Agents follow then a human-like reasoning process based on the BDI architecture, and agent communications emulate human dialogs through KQML-typed messages like those shown in Table 3

Agents were built ad-hoc in java, following the abstract execution cycle of [16]. The BDI-like architecture of our agents consists of abstract desires that are transformed into explicit goals when external perceptions are sensed. Each of these goals has an associated generic plan composed of a sequence of atomic intentions:

- The generic plan of prediction desire is just a sequence of two intentions: *waits* until prediction request, *computes* prediction and *gives* prediction.
- The plan corresponding to evaluator agents is formed of the next ordered set of intentions: *wait* until decision request, for each predictor: { *asks* for prediction, *waits* for prediction }, *computes* new decision applying weights on predictions, *gives* decision to manager, *waits* for HVF from manager, for each predictor: { *computes* new weight of predictor }.
- The plan of manager agents consists of a repeating cycle: *asks* evaluator for decision, *waits* for decision from evaluator, *computes* HVF, *gives* HVF to evaluator

The intelligence of the agent system relies on the computations of evaluator agents related to the weights of predictors according to the success of previous predictions. The final suggestion from evaluator agents directly depend on these weights. Nevertheless, predictor and manager agents show a straightforward behavior rather than the adaptive reasoning of evaluator agents. Adaptation is achieved from the dynamic changes operated in the updating computations of weights.

Our proposal in this publication involves the use of the ratio $H_i - F_i$ as a weight in the aggregated sum of predictor agents, although we intend to apply a more complex computation in the future. Therefore, the reputation of certain predictor agent would increase if the number of hit rates became higher, and if the number of false alarm rates remains in a low level. After updating these weights with the last results, all of them are normalized, and the corresponding equations result.

$$weight_i = \frac{H_i - F_i + 1}{\sum_j (H_j - F_j + 1)} \tag{16}$$

$$suggestion = \sum_j (Prediction_j \cdot weight_j) \tag{17}$$

4 EXPERIMENTS AND RESULTS

The main goal of this section is to explain how our model of agents works showing the benefits that dynamic adaptation using agents can bring to the decision making problem. The value computed through this adaptive reasoning is compared with agents that evaluates all the predictions with the same behavior ($weight = 1/numberofpredictors$). The prototype was not analyzed against real traffic but a possible simulation of such data were tested.

Let us explain the experimental setup: many predictor agents (IDS models), eight in our example, are considered. The event to predict is the same for all of them, and it consists of an intrusion. We have tested our prototype against thirty connections. There is a complete agent cycle for each connection. The eight predictor agents made their forecast responding to the evaluator request. It makes its suggestion based on the predictions and send it to the manager. Then it waits for the manager results. When it receives it, re-weights its agents and ask the predictors for their forecasts over the second connection. This time the suggestion will not be just an average sum of predictors.

In order to evaluate our prototype's effectiveness we have made a couple of experiments. First the model is run without any adaptive behavior and the predictions are just based on how many models has predicted the event (average sum). This figure is compared to a certain threshold p_t. The prediction is positive (an intrusion is present in the system) if the average sum of predictors' suggestions is greater than the threshold. The second experiment tests the adaptive agent proposal. The evaluator agent suggests if there has been an intrusion or not weighting each predictor each time according to its past success. If the weighted sum is over the threshold then the suggestion confirms that there is an intrusion.

Five different thresholds uniformly distributed were considered in both experiments ($0.2, 0.4, 0.6, 0.8, 1.0$). These are the results:

With the lowest and the highest thresholds the hit rate (H), and the false alarm rate (F) are the same in both experiments:

- For $p_t = 0.0 : H = 1.000; F = 1.000$
- For $p_t = 0.2 : H = 0.909; F = 0.684$
- For $p_t = 0.8 : H = 0.273; F = 0.053$
- For $p_t = 1.0 : H = 0.000; F = 0.000$

The suggestions from the evaluator agent for $p_t = 0.6$ are:

- With a constant evaluation of predictor agents (average sum): $H = 0.545; F = 0.158$
- With an adaptive evaluation of predictor agents (dynamic weights): $H = 0.545; F = 0.107$

From these data we observe a similar number of hit rates and a slightly lower number of false alarms with an adaptive evaluation.

At last, the suggestions from the evaluator agent for $p_t = 0.4$ are:

- With a constant evaluation of predictor agents (average sum): $H = 0.545 F = 0.263$
- With an adaptive evaluation of predictor agents $H = 0.909 F = 0.421$

From these data, we can observe more hit rates in the adaptive evaluation than in the constant evaluation, but it also appears to be more false alarms. So at first glance it does not show clearly which alternative is better.

The metric of value introduced in section 2 shows the results are promising. Adaptive approach clearly improves the detection system behavior as can be seen comparing Figures 3 and 4. Let us explain first how to read value graphs. V_{opt} peaks for users with C/L next to the probability of intrusion (0.367). At this point $H(p_t) - F(p_t) = V(p)$. The envelope function in Figure 3 shows value for all intrusions with C/L between 0.143 and 0.750. This illustrates the benefit of probabilistic predictions over deterministic ones. The probabilistic approach gives us a wider range C/L for which users have economic value. The value of the curve for a deterministic forecast would be no better than that of a single $V(p_t)$ curve, since a deterministic prediction has a single hit and false alarm rate associated with.

The interval with positive value for average sum approach is $[0.143, 0.750]$ and for our adaptive agents proposal is $[0.083, 0.750]$. V_{opt} is also higher for the weigthed approach since the maximum is 0.485 for $p_t = 0.4$ and it is 0.374 in the case of the average sum.

Table 3. KQML performatives of predictors, evaluators and managers.

from	to	speech act	what	about
Manager	Evaluator	ask	decision	traffic data id
Evaluator	Predictor	ask	prediction	traffic data id
Predictor	Evaluator	tell	prediction	traffic data id
Evaluator	Manager	tell	decision	traffic data id
Manager	Evaluator	tell	H, F, V	

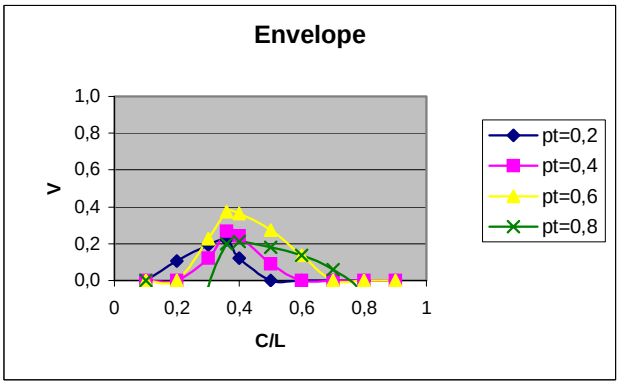

Figure 3. Average sum envelope. Economic value versus C/L relationship

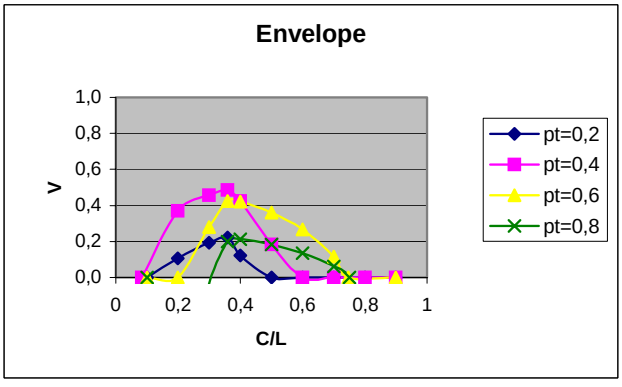

Figure 4. Adaptive agents envelope. Economic value versus C/L relationship

In practice, the site security officer (SSO) will estimate the C/L relationship for his system. He will also be able to estimate the probability of intrusion (p). Let us imagine that, in our example, the C/L relationship is equal to p. This would give us the maximum value for $p_t = 0.4$. In this case, the SSO can say that with a probability of 40% his prediction has a value of $48,5\%$ of the perfect prediction. Based on this analysis it could be appropriate for him to respond taking some action (for example, unplugging the system, changing a rule in the firewall, etc.).

5 CONCLUSION

This paper proposes a multiagent system that cooperates in order to detect intrusions. Some of the agents implement IDS models, other evaluates the predictions made by the firsts and finally a third kind of agent considers the evaluator suggestion establishing the IDS effectiveness. The dynamic weights involved in the adaptive evaluation performed to generate a final suggestion from several different intrusion detection models showing a better performance than classical approach based on the average sum of the predictions received.

The improvement has been measured introducing a promising metric that takes into account the response costs. The recent introduction of decision making techniques to intrusion detection reveals the necessity of formal robust metrics that considers all the parameters involved in the task.

Future work will include testing our prototype with real data instead of synthetic data. This is not an easy task because of the problems to experiment in real networks and the suspicious results based on simulated traffic [14].

This paper have shown that adaptive behavior of agents can be really useful in the intrusion detection field due to the very changing environment that is faced and the need of automated responses.

ACKNOWLEDGEMENTS

The research reported in this paper is partly supported by CICYT (TIC2001-5108-E) project.

REFERENCES

[1] J. Balasubramaniyan, J. O. García-Fernández, D. Isacoff, E. H. Spafford, and D. Zamboni. An architecture for intrusion detection using autonomous agents. Dept. of Comp. Sciences, Purdue University, Technical Report 98-05, 1998.

[2] J. Carbó, J.M. Molina, and J. Dávila, 'Trust management through fuzzy reputation', *Int. Journal of Cooperative Information Systems*, **12**(1), 135–155, (2003).

[3] C. Carver, J. Hill, J. Surdu, and U. Pooch, *A methodology for using Intelligent Agents to provide Automated Intrusion Response*, Procs. of the IEEE Systems, Man, and Cybernetics Information Assurance and Security Workshop, IEEE Computer Society Press, New York, U.S., 2000.

[4] M. de Groot, *Optimal Statistical Decisions*, McGraw-Hill, New York, U.S., 1970.
[5] M de Groot, *Optimal Statistical Decisions*, McGraw-Hill, New York, U.S., 1970.
[6] R. Durst, T. Champion, B. Witten, E. Miller, and L. Spagnuolo, 'Testing and evaluating computer intrusion detection systems', *Communications of the ACM*, **42**(7), 53–61, (1999).
[7] J.P. Egan, *Signal detection theory and ROC-analysis*, Academic Press, 1975.
[8] J.E. Gaffney and J.W. Ulvila, *Evaluation of Intrusion Detectors: A Decision Theory Approach*, The IEEE Symposium on Security and Privacy, 2001.
[9] J. Hancock and P. Wintz, *Signal Detection Theory*, McGraw-Hill, 1966.
[10] R. Lippmann, D. Fried, I. Graf, J. Haines, K. Kendall, D. McClung, D. amd Weber, S. Webster, D. Wyschogrod, R. Cunningham, and M. Zissman, *Evaluating Intrusion Detection Systems: The 1998 DARPA Offline Intrusion Detection Evaluation*, Proceedings of the DARPA Information Survivability Conference and Exposition, IEEE Computer Society Press, California, U.S., 2000.
[11] R.P. Lippmann and J. Haines, 'Analysis and results of the 1999 darpa off-line intrusion detection evaluation', *Lecture Notes in Computer Science*, **1907**, 162–182, (2000).
[12] P. Maes, 'Agents that reduce work and information overload', *Communications of the ACM*, **37**(7), 31–40, (1994).
[13] A. Martin, M. Przybocki, G. Doddington, and D. Reinolds, 'The nist speaker recognition evaluation- overview, methodology, systems, results, perspectives', *Speech Communications*, **31**, 225–254, (2000).
[14] J. McHugh, 'Testing intrusion detection systems: A critique of the 1998 and 1999 darpa intrusion detection system evaluations as performed by liconln laboratory', *ACM Transactions on Information and System Security*, **3**(4), 262–2944, (2000).
[15] A. Orfila, J. Carbó, and A. Ribagorda, *Fuzzy logic on Decision Model for IDS*, Proceedings IEEE Int. Conf. on Fuzzy Systems, IEEE Computer Society Press, Missouri, U.S., 2003.
[16] A.S. Rao and M.P. Georgeff, *An abstract architecture for rational agents*, 439–449, Proceedings of the 3rd Int. Conf. on Principles of Knowledge Representation and Reasoning, 1992.
[17] R.G. Smith and R. David, 'Frameworks for cooperation in distributed problem solving', *IEEE Trans. On Systems, Man and Cybernetics*, **11**(1), 61–70, (1995).
[18] S. Stolfo, W. Fan, W. Lee, A. Prodromidis, and P. Chan, *Cost-Based Modeling for Fraud and Intrusion Detection: Results from the JAM Project*, Proceedings of DISCEX 2000, IEEE Computer Society Press, California, U.S., 2000.
[19] J.A. Swets, R.M. Dawes, and J. Monahan, 'Psychological science can improve diagnostic decisions', *Psychological Science in the Public Interest*, **1**(1), (2000).