# Adaptive Agents applied to Intrusion Detection

Javier Carbó, Agustín Orfila, and Arturo Ribagorda

Carlos III University of Madrid,
Computer Science Department,
28911 Leganés, Madrid, Spain
{jcarbo,adiaz,arturo}@inf.uc3m.es

**Abstract.** This paper proposes a system of agents that make predictions over the presence of intrusions. Some of the agents act as predictors implementing a given Intrusion Detection model, sniffing out the same traffic. An assessment agent weights the forecasts of such predictor agents, giving a final binary conclusion using a probabilistic model. These weights are continuously adapted according to the previous performance of each predictor agent. Other agent establishes if the prediction from the assessor agent was right or not, sending him back the results. This process is continually repeated and runs without human interaction. The effectiveness of our proposal is measured with the usual method applied in Intrusion Detection domain: Receiver Operating Characteristic curves (detection rate versus false alarm rate). Results of the adaptive agents applied to intrusion detection improve *ROC* curves as it is shown in this paper.

## 1 Introduction

Research on intrusion detection topic is very active. Several models have been proposed and they are still evolving. There are also problems with the automation of such intrusion detection models. In fact, nowadays these models are used in close combination with a human supervisor (so called Site Security Officer) [14].

One of the main questions about intrusion detection models or systems (in advance, IDS) is to know how effective they are at detecting intrusions while raising as fewer false alarms as possible. In order to evaluate the effectiveness [15] of IDS the use of Receiver Operating Characteristic (ROC) has been largely proposed [16],[17],[15],[18]. ROC curves plot detection rate versus false alarm rate. A ROC curve shows the false alarm rate incurred by choosing a particular detection rate. Some IDS can be tuned to accept a higher false alarm rate in order to detect more attacks. IDS based on signature detection, however, produce a binary output of 1 for a declared attack and 0 for a normal session. That is why they are represented by a single point on the ROC curve [17] (sometimes a line joining this point to (0,0) and another line joining it to (1,1) is plotted). ROC technique has also been used in other fields where detection rates are object of study such as signal detection, speaker identification, medical risk prediction

and meteorology [19][20][21][22][23]. The bigger the area under the ROC curve is ($A_{ROC}$) the bigger the IDS effectiveness is [17].

In this scope agents may play an important role: agents intend a complete automation of complex processes acting in behalf of human users [11]. Several definitions and approaches to agent term have led to some confusion. 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. In this way, deliberative agents often accomplish the so called BDI paradigm [12] where knowledge is structured in three different levels of abstraction: beliefs, intentions and desires.

Intelligent agents presumably can adapt decision making through the cooperation with other agents [10]. Communication between agents is usually modeled through human-like typed messages including performatives inspired in Speech Act Theory (for instance, KQML [13]).

Particularly, agents have been applied to this dominion in the past [2]. For instance, a methodology [3], an architecture [1] were proposed, and also some IDS were implemented as agent systems using genetic algorithms [5], and neurofuzzy controllers [4].

## 2 The Analysis of Relative Operating Characteristic

In order to asses the skill of a probabilistic prediction (based on several IDS agents) we use the Receiver Operating Characteristic (ROC) in a slightly different way [23]. In our case study, *ROC* measures the success and false alarm rates of an ensemble; made by assuming an event $E$ will occur if it is predicted with a probability exceeding some specified probability threshold $p_t$. The difference with usual ROC approximation in intrusion detection is that the threshold we vary is not the false alarm threshold, but the probability of occurrence the "event". Indirectly, this is like tuning the false alarm threshold of the probabilistic system.

This definition is based on the notion that a prediction of an event $E$ is assumed if $E$ is predicted by at least a fraction $p = p_t$ of ensemble members, where the threshold $p_t$ is defined a priori.

Let us consider first a deterministic (single model) prediction of $E$ (either that it will occur or that it will not occur). Over a sufficiently large sample of independent predictions, we can form the prediction contingency matrix (Table 1) giving the frequency that $E$ occurred or not, and whether it was predicted or not.

Based on these values, the hit rate($H$) and false alarm rate ($F$) for a deterministic prediction are given by

$$\begin{aligned} H &= \delta/(\beta + \delta). \\ F &= \gamma/(\alpha + \gamma). \end{aligned} \tag{1}$$

Hit and false alarm rates for a probabilistic prediction can be redefined as follows [24]. Suppose it is assumed that $E$ will happen if the probability of the

**Table 1.** Prediction contingency matrix

| | | Ocurrs | |
|---|---|---|---|
| | | No | Yes |
| Prediction | No | $\alpha$ | $\beta$ |
| | Yes | $\gamma$ | $\delta$ |

prediction $p$ is greater than $p_t$ (and will not if $p < p_t$ ). By varying $p_t$ between 0 and 1 we can define $H = H(p_t)$, $F = F(p_t)$.

The ROC curve is a plot of $H(p_t)$ against $F(p_t)$. A measure of skill is given by the area under the ROC curve ($A_{ROC}$). A perfect deterministic forecast will have $A_{ROC} = 1$, while a no skill-forecast for which the hit and false alarm rates are equal, will have $A_{ROC} = 0.5$.

In our case study the event $E$ is "an intrusion". We will have different models (different IDS) that try to detect intrusions. These models must deal with the same traffic and must not take any action on it.

## 3 The role of Agents

The proposed system has three different types of agents: predictor, assessor and manager. In the scenario considered, there are several predictor agents and only one assessor agent and one manager agent. But it would make sense to use several assessor agents rather than just one, since not every predictor would be asked by each assessor agent, and they would also adopt different weighting criteria.

The main role of a predictor agent consists of suggesting if there is an intrusion or not when an assessor agent asks him for a prediction. On the other hand, major goal of assessor agents is giving proper weights to predictor agents according to the previous level of success, and afterwards, making a binary decision based on such weighted references. Finally the manager agent calculates H and F and, at last, it communicates the results to the assessor agent. The manager agent knows if there was an intrusion or not because the experiment is done under a training environment. The interactions between agents are typed as KQML messages. You can see an illustrative example of such interactions in figure Fig. 1.

All of this agents adopt a BDI-like architecture, where abstract desires became in concrete goals when external perceptions would be sensed. Each of these goals has an associated generic plan composed of a sequence of atomic intentions.

The intelligence of agents relies on how predictions are weighted according to the success of previous predictions in order to make a suggestion. Nevertheless plans of predictor and manager agents show a straightforward behavior rather than the adaptive reasoning of assessor agents.

Our proposal in this publication involves the use of the ratio H-F as a weight in the aggregated sum of predictor agents. Therefore, the reputation of certain predictor agent would increase if the number of hit rates became higher, and

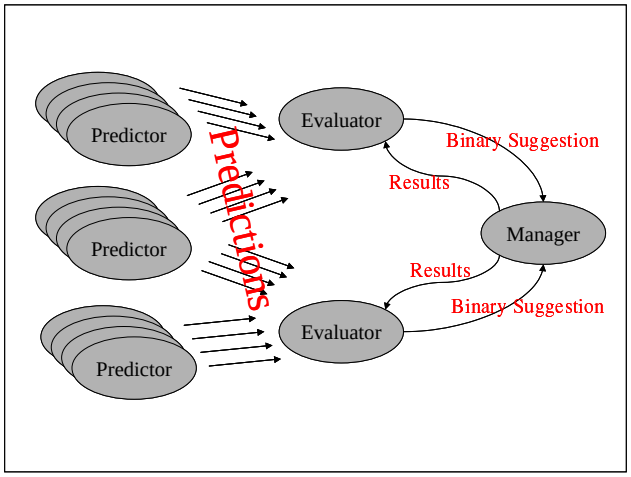


**Fig. 1.** Interactions between agents

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:

$$Weight_i = \frac{(H_i - F_i + 1) \div 2}{\sum_j (H_j - F_j + 1) \div 2} \tag{2}$$

$$Suggestion = \sum_i Weight_i \times Prediction_i \tag{3}$$

## 4 Simulated Results

The main goal of this section is to explain how our model of agents works and to show the benefits of dynamic adaptation of agents to improve the decision making. The level of success of this adaptive behavior is compared with agents evaluating all the predictions with the same behavior (weight= 1 / number of predictors). No real data of traffic was able to be used in the experiments, but a possible simulation of such data were tested.

Let us analyze the experimental setup: Many predictor agents (IDS models), eight in our example, are considered. The events to predict are the same event $E$ for all of them, and it consists of an intrusion. A possible scenario is the one in Table 2.

An assessor agent suggests if there has been an intrusion or not depending on how many models has predicted the event (Prob. of table 2) vs. a weighted sum. This suggestion is made through the comparison with a certain threshold $p_t$.

Five different thresholds uniformly distributed were considered in both experiments $(0.2, 0.4, 0.6, 0.8, 1.0)$. In this way, comparing the average sum and

**Table 2.** Possible measures on the event E (intrusion). 1 represents the predictor agent (IDS model) has launched an alert on the event $E$ (predicts that $E$ happens). 0 represents it does not. If the event $E$ is finally verified by the manager agent, it is represented by 1. If it does not it is represented by a 0

| Day | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 | 9 | 10 | 11 | 12 | 13 | 14 | 15 |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| IDS1 | 0 | 0 | 0 | 0 | 0 | 0 | 1 | 1 | 1 | 1 | 0 | 0 | 1 | 1 | 0 |
| IDS2 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 1 | 0 | 0 | 0 | 1 | 0 | 0 |
| IDS3 | 1 | 0 | 1 | 1 | 0 | 1 | 1 | 1 | 1 | 1 | 0 | 1 | 1 | 1 | 0 |
| IDS4 | 0 | 0 | 0 | 1 | 0 | 1 | 0 | 1 | 0 | 0 | 0 | 1 | 0 | 0 | 0 |
| IDS5 | 0 | 0 | 0 | 1 | 1 | 0 | 0 | 1 | 0 | 1 | 1 | 1 | 0 | 1 | 0 |
| IDS6 | 0 | 0 | 1 | 1 | 0 | 1 | 0 | 0 | 0 | 0 | 0 | 0 | 1 | 0 | 0 |
| IDS7 | 0 | 1 | 0 | 0 | 0 | 1 | 0 | 0 | 0 | 0 | 0 | 1 | 1 | 0 | 1 |
| IDS8 | 0 | 1 | 0 | 1 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 1 | 0 | 0 | 0 |
| Prob | 0.125 | 0.25 | 0.25 | 0.625 | 0.125 | 0.5 | 0.25 | 0.5 | 0.375 | 0.375 | 0.125 | 0.625 | 0.625 | 0.375 | 0.125 |
| Occurs | 1 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 1 | 0 | 1 | 1 | 1 | 0 |
| | | | | | | | | | | | | | | | |
| Day | 16 | 17 | 18 | 19 | 20 | 21 | 22 | 23 | 24 | 25 | 26 | 27 | 28 | 29 | 30 |
| IDS1 | 1 | 0 | 0 | 0 | 0 | 0 | 1 | 1 | 1 | 1 | 0 | 0 | 1 | 1 | 0 |
| IDS2 | 1 | 0 | 0 | 0 | 0 | 1 | 0 | 1 | 1 | 0 | 0 | 0 | 1 | 0 | 0 |
| IDS3 | 1 | 0 | 1 | 1 | 0 | 1 | 1 | 1 | 1 | 1 | 0 | 1 | 1 | 1 | 0 |
| IDS4 | 1 | 0 | 0 | 1 | 0 | 1 | 0 | 1 | 0 | 0 | 0 | 1 | 0 | 0 | 0 |
| IDS5 | 1 | 0 | 0 | 1 | 1 | 1 | 0 | 1 | 0 | 1 | 1 | 1 | 1 | 1 | 0 |
| IDS6 | 1 | 0 | 1 | 1 | 0 | 1 | 0 | 1 | 0 | 0 | 0 | 0 | 1 | 0 | 0 |
| IDS7 | 1 | 1 | 0 | 0 | 0 | 1 | 0 | 1 | 0 | 0 | 0 | 1 | 1 | 0 | 1 |
| IDS8 | 0 | 1 | 0 | 1 | 0 | 1 | 0 | 0 | 0 | 0 | 0 | 1 | 1 | 0 | 0 |
| Prob | 0.875 | 0.25 | 0.25 | 0.625 | 0.125 | 0.875 | 0.25 | 0.875 | 0.375 | 0.375 | 0.125 | 0.625 | 0.875 | 0.375 | 0.125 |
| Occurs | 1 | 0 | 0 | 0 | 0 | 0 | 0 | 1 | 0 | 1 | 0 | 1 | 1 | 1 | 0 |

threshold $p_t = 0.2$ for instance, we can observe the suggestions made in table 3.

**Table 3.** Ten days measures on the event $E$. Prediction (based on a threshold $p_t = 0.2$) and what really happened is binary represented

| Day | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 | 9 | 10 | 11 | 12 | 13 | 14 | 15 |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| $p_t = 0.2$ | 0 | 1 | 1 | 1 | 0 | 1 | 1 | 1 | 1 | 1 | 0 | 1 | 1 | 1 | 0 |
| Occurs | 1 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 1 | 0 | 1 | 1 | 1 | 0 |
| | | | | | | | | | | | | | | | |
| Day | 16 | 17 | 18 | 19 | 20 | 21 | 22 | 23 | 24 | 25 | 26 | 27 | 28 | 29 | 30 |
| $p_t = 0.2$ | 1 | 1 | 1 | 1 | 0 | 1 | 1 | 1 | 1 | 1 | 0 | 1 | 1 | 1 | 0 |
| Occurs | 1 | 0 | 0 | 0 | 0 | 0 | 0 | 1 | 0 | 1 | 0 | 1 | 1 | 1 | 0 |

With the lowest and the highest thresholds the hit rate (H), and the false alarm rate (F) values 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 assessor 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 assessor 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 (dynamic weights):
  $H = 0.909$ $F = 0.421$

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. But Fig. 2 representing the $A_{ROC}$ curves of both possible evaluations shows that an adaptive evaluation of agents include a greater area than the constant evaluation.

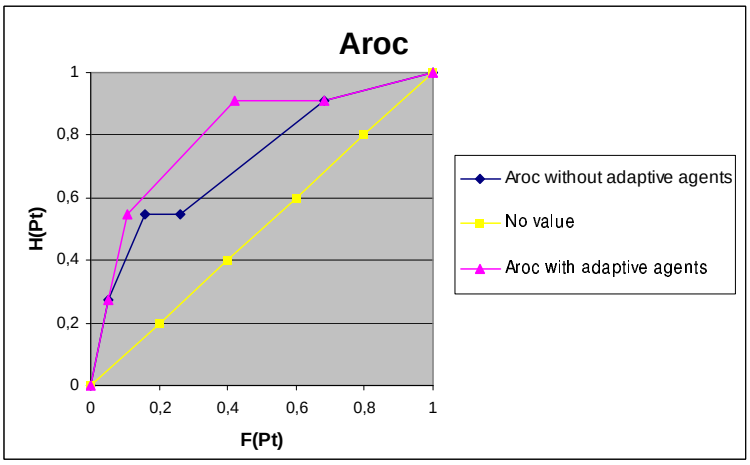


**Fig. 2.** $A_{ROC}$ of adaptive agent approach vs. average one. Area corresponding to adaptive agent model is bigger, showing a better effectiveness of the model

Mathematically, the area of both alternatives is:

- With a constant evaluation of predictor agents (average sum): 0.75
- With an adaptive evaluation of predictor agents (dynamic weights): 0.81

So an improvement of 7.25% is achieved with the adaptive behaviour of agents proposed in this paper.

## 5 Conclusions and Future Work

In this paper, we have proposed a system of different types of agents cooperating in order to detect intrusions. One of these types implements IDS models, other evaluates the predictions from them, and finally a third kind of agent considers the suggestions and emulates the results with which future evaluations will be more accurate. The dynamic weights involved in the adaptive evaluation performed to generate a final suggestion from several different Intrusion Detection models showed a better performance than classical approach based on the average sum of the predictions received.

Furthermore, the multiagent system applied to this dominion, allow us different future directions of research. For instance, the use of a fuzzy threshold applied over the average sum of predictions causes a relevant improvement in the overall performance of Intrusion Detection task [6]. Such fuzzy threshold may be used also as adapted dynamically in order to obtain even better results.

In the future, we intend to use the architecture of agents proposed to compute and represent the reputation of predictor agents as a fuzzy set in the same way that AFRAS [9]. In order to do such fuzzy evaluation we will use an analysis of

the economical costs involved in the precautionary action taken by the manager agent. This reputation mechanism showed a fast convergence, while sensitivity to sudden changes avoids a high level of deception [8]. Finally, it could be interesting to try with real data, and also to test several assessor agents analyzing different traffic data.

## References

1. Balasubramaniyan, J.S., Garcia J.O., Isacoff D., Spafford E., Zamboni D.,: An Architecture for Intrusion Detection using Autonomous Agents. Procs. of the $14^{th}$ Annual Computer Security Applications Conf., pp. 13-24. IEEE Computer Society, December 1998.
2. Vigna G., Cassell B., Fayram D.,: An Intrusion Detection System for Aglets. $6^{th}$ Int. Conf. on Mobile Agents, Barcelona, Spain, October 2002.
3. Carver C., Hill J., Surdu J., Pooch U.,: A methodology for using Intelligent Agents to provide Automated Intrusion Response. Procs. of the IEEE Systems, Man, and Cybernetics Information Assurance and Security Workshop, West Point, NY, June 6-7, 2000.
4. Dasgupta, D., Brian H.,: Mobile Security Agents for Network Traffic Analysis. Procs. DARPA Information Survivability Conf. adn Exposition II, IEEE Society Press, Anaheim, California, June 2001.
5. Crosbie, M., Spafford G.,: Active Defense of a Computer System using Autonomous Agents. Technical Report No. 95-008, Purdue University, U.S., June 1995.
6. Orfila A., Carbo J., Ribagorda A.: Fuzzy logic on Decision Model for IDS. Procs. IEEE Int. Conf. on Fuzzy Systems, St. Louis, May 2003.
7. Baldwin, J.F.,: A calculus for mass assignment in evidential reasoning. Advances in Dempster-Shafer Theory of Evidence, M. Fedrizzi, J. Kacprzyk, R.R. Yager, eds., John Wiley, 1992.
8. Carbo, J., Molina J.M., Davila, J.,: Trust management through fuzzy reputation. Accepted for Int. Journal of Cooperative Information Systems, to appear.
9. Carbo, J., Molina J.M., Davila J.,: A fuzzy model of reputation in multiagent system. Procs. $5^{th}$ Int. Conf. on Autonomous Agents, Montreal, June 2001.
10. Smith R.G., David R.,: Frameworks for cooperation in distributed problem solving. IEEE Trans. On Systems, Man and Cybernetics, vol. 11, number 1, pp.61-70, June 1995.
11. Maes, P.,: Agents that reduce work and information overload. Communications of the ACM, vol. 37, number 7, pp. 31-40, 1994.
12. Rao, A.S., Georgeff, M.P.,: BDI-agents from theory to practice. Procs. $1^{st}$ Int. Conf. on Multiagent Systems (ICMAS'95), San Francisco, June 1995.
13. Finin, T., McKay R., Fritzson, R., McEntire R.,: KQML: an information and knowledge exchange protocol. Procs. Int. Conf. on Building and Sharing of Very Large-Scale Knowledge Bases, December 1993.
14. Axelsson, S.,: Intrusion-detection systems: A taxonomy and survey. Technical Report 99-15, Department of Computer Engineering, Chalmers University of Technology,SE-41296, Goteborg, Sweden, March 2000.
15. Axelsson, S.,: The base rate fallacy and its implications for the difficulty of intrusion detection. In $6^{th}$ ACM conference on computer and communications security.Kent Ridge Digital Labs, Singapore, 1-4 November 1999, pp. 1-7

16. Lippman, R.P., Fried, D.J., Graf, I., Haines, J.W., Kendall, K.R., McClung, D., Weber, D., Webster, S.E., Wyshhogrod, D., Cunningham, R.K, Zissman, M.A.: Evaluating Intrusion detection systems: the 1998 DARPA Off-line Intrusion Detection Evaluation. Proceedings of the 2000 DARPA information survivality Conference and Exposition (DISCEX), Vol.2, IEEE Press, January 2000
17. Durst, R., Champion, T., Witten, B., Miller, E., Spagnolo, L. : Testing and evaluating computer intrusion detection systems. Communications of the ACM, 42(7), 1999, pp.53-61
18. Gomez, J., Dasgupta, D.: Evolving Fuzzy Classifiers for Intrusion Detection. Proceedings of the 2002 IEEE. Workshop on Information Assurance. United States Military Academy, West Point, NY June 2001
19. Swets, J.A: The Relative Operating Characteristic in Psychology. Science,182,1973,pp. 990-1000
20. Egan, J.P: Signal detection theory and ROC-analysis. Academic Press, 1975
21. Martin,A., Doddington,G., Kamm,T., Ordowski,M.,Przybocki,M.: The DET Curve in Assessment of Detection Task Performance. Proceedings EuroSpeech 4. 1998, pp. 1895-1898.
22. Lippmann, R.P., Shahian,D.M.:Coronary Artery Bypass Risk Prediction Using Neural Networks. Annals of Thoracic Surgery, 63. 1997. pp. 1635-1643.
23. Stanski, H.R., Wilson, L.J., Burrows, W.R. Survey of common verification methods in meteorology. World Weather Report No. 8. World Meteorological Organization. Geneva.
24. Palmer, T.N., Brankovic, C., and Richardson, D.S. A Probability and Decision-Model Analysis of PROVOST Seasonal Ensemble Integrations. Research Department. Technical Memorandum No.265. Nov 1998.
25. Murphy,A.H. A new vector partition of the probability score. J. Appl. Meteor.1973.
26. Katz, R.W., Murphy, A.H. Forecast value: prototype decision-making models. In Economic value of weather and climate forecasts. Eds. Cambridge University Press. 1997.
27. Wenke, L., Wei, F., Miller, M., Stolfo. S., Zadok, E. Toward Cost-Sensitive Modeling for Intrusion Detection and Response. North Carolina State University. Computer Science