Fuzzy logic on decision model for IDS

Agustín Orfila, Javier Carbó, Arturo Ribagorda
Carlos III University of Madrid,
Computer Science Department,
28911 Leganés, Madrid, Spain
Email: {adiaz, jcarbo, arturo}@inf.uc3m.es

Abstract—**Nowadays one of the main problems of Intrusion Detection Systems (IDS) is the high rate of false positives that they show. The number of alerts that an IDS launches are clearly higher than the number of real attacks. This paper tries to introduce a measure of the IDS prediction skill in close relationship with these false positives. So the prediction skill of an IDS is then computed according to the false positives produced. The problem faced is how to make an accurate prediction from the results of different IDS. The fraction of IDS over the total number of them that predicts a given event will determine whether such event is predicted or not. The performance obtained from the application of fuzzy thresholds over such fraction is compared with the corresponding crisp thresholds. The results of these comparisons allow us to conclude a relevant improvement when fuzzy thresholds are involved.**

I. Introduction

Over the last few years Intrusion Detection Systems have greatly evolved. In the early nineties preventive measures to defend networks such as firewalls were developed. After that, the need to be aware of when an attack has taken place crossing the firewall, and the need to take reactive measures, obliged researching and industry to develop Intrusion Detection Systems. These systems are still evolving and there are still a lot of problems for them to become a really automatic tool. Nowadays they are used in close combination with a site security officer (SSO).

There are two main ways of classifying IDS. One is based on what kind of system they watch over. The other is based on how they do it. According to the first there are two main types of IDS: Network IDS (NIDS) and Host-based IDS (HIDS) systems. The classification of an IDS based on how they work falls in one of the following categories: anomaly detection or misuse (signature) detection [1]. The first relies on detecting abnormal behaviour in systems from the one considered normal, the second tries to detect known intrusions characterized by a pattern.

For the purposes of this paper, no matter what kind of IDS we manage. Each of them is considered as a part of a probabilistic system and we only focus on its prediction capabilities. The existence of different intrusion detection techniques is the basis of our study.

For organizations interested in quantifying the IDS value prior to deploying it, their investment decision will depend on their ability to demonstrate a positive ROI (Return On Investment) [2]. ROI has traditionally been difficult to quantify because it is difficult to calculate risk accurately and statistics regarding security incidents are not always available. Actually, the calculation of risk does not work well. This happens because people tend not to tell the truth. The qualitative risk is often underestimated because if a security manager is aware of a high risk vulnerability, he will have to fix it. And if vulnerabilities are considered low risk there is no necessity to invest in IDS [3]. In this scenario the authors want to propose a model able to evaluate the skill of an IDS prediction in a different way.

This paper tries to quantify the improvement of using IDS technology in a system. What is becoming necessary for a specific IDS user is to quantify whether he should invest or not in this technology. To be precise the main aims of this paper are:

- to discuss the skill and value of multi-IDS predictions.
- to show that IDS fuzzy predictions are better than crisp ones.

II. The model

Two ways of evaluating the skill of an IDS prediction are exposed:

- Relative Operating Characteristic (ROC): it measures the generic skill of a prediction
- Decision model: it gives a user-specific measure of the prediction skill

A. The analysis of Relative Operating Characteristic

In order to assess the skill of a probabilistic prediction we use the Relative Operating Characteristic (ROC) [4] technique. It 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 ROC 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 giving the frequency that E occurred or not, and whether it was predicted or not (see Table I).

Based on these values, the 'hit-rate' (H) and 'false alarm rate' (F) for E are given by

0-7803-7810-5/03/$17.00 ©2003 IEEE

TABLE I

PREDICTION CONTINGENCY MATRIX

		Ocurrs	
		No	Yes
Prediction	No	α	β
	Yes	γ	δ

$$H = \delta/(\beta + \delta).$$
$$F = \gamma/(\alpha + \gamma). \quad (1)$$

Hit and false alarm rates for a probabilistic prediction can be defined as follows [5]. Suppose it is assumed that E will happen if the probability of the 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 $A_{ROC} = 1$, whilst a no skill-forecast for which the hit and false alarm rates are equal, will have $A_{ROC} = 0.5$. As discussed later, the ROC curve values are of direct use in assessing the user-specific value of a probabilistic prediction, based on a decision model analysis.

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.

B. Decision-model analysis

Although A_{ROC} gives an objective measure of skill for ensemble predictions, it is difficult to say what constitutes a threshold of useful skill. This is not surprising since usefulness is a user-specific concept. In an attempt to define 'usefulness' objectively, we consider here a simple decision model [6] [7].

Consider a potential prediction user who can take some precautionary action depending on the likelihood that E will occur. Taking precautionary action incurs a cost C irrespective of whether or not E occurs. However, if E 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 E is given in the decision-model contingency matrix, as shown in Table II.

TABLE II

DECISION-MODEL CONTINGENCY MATRIX. IF AN EVENT HAPPENS AND NO ACTION IS TAKEN THEN A LOSS L IS INCURRED. IF AN ACTION IS TAKEN WE INCUR IN A COST C EITHER THE EVENT HAPPENS OR NOT

		Ocurrs	
		No	Yes
Take action	No	0	L
	Yes	C	C

The decision maker wishes to pursue a strategy that will minimize expenses over a large number of cases.

If only information on the frequency $\bar{o}(= (\beta + \delta))$ is available, there are two basic options: either always or never take precautionary action. Always taking action incurs a cost C on each occasion, whilst never taking action incurs a loss L only on the proportion $\bar{o}$ of occasions when E occurs, giving an expense $\bar{o}L$.

The purpose of the current study is to analyze if the decision maker would reduce expenses beyond what could be achieved using frequency information alone and to quantify this reduction. The use of fuzzy logic will improve the results.

Consider first a deterministic prediction system with characteristics described by the prediction-model contingency matrix in Table I. Then, using the prediction and decision contingency values, the user's expected mean expense M (per unit loss) is

$$M = \frac{(\beta L + (\gamma + \delta)C)}{L}. \quad (2)$$

This can be written in terms of the hit rate H and false alarm F using (1), so

$$M = F\frac{C}{L}(1 - \bar{o}) - H\bar{o}(1 - \frac{C}{L}) + \bar{o}. \quad (3)$$

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

$$M_{per} = \bar{o}\frac{C}{L}. \quad (4)$$

To calculate the mean expense per unit loss knowing only the frequency, suppose first the decision maker always protects, then $M = \frac{C}{L}$ (equivalent to using a prediction system where the event is always predicted and for which $H = 1$ and $F = 1$). Conversely, if the decision maker never protects then $M = \bar{o}$ (equivalent to using a prediction system where the event is never predicted and for which $H = 0$ and $F = 0$). So if the decision maker knows only the frequency $\bar{o}$, M can be minimized by either always or never taking precautionary action, depending on whether $\frac{C}{L} < \bar{o}$, or $\frac{C}{L} > \bar{o}$ respectively. Hence, the mean expense per unit loss associated with the knowledge of frequency only is

$$M_{fre} = min(\frac{C}{L}, \bar{o}). \quad (5)$$

We define the value of forecast information to be a measure of the reduction in M over M_{fre}, normalized by the maximum possible reduction associated with a perfect deterministic forecast, ie

$$V = \frac{(M_{fre} - M)}{(M_{fre} - M_{per})}. \quad (6)$$

For a predictive system which is no better than frequency, $V = 0$; for a perfect deterministic system $V = 1$.

As has been previously discussed a multimodel ensemble prediction gives hit and false alarm rates $H = H(p_t)$, $F = F(p_t)$, as a function of probability thresholds p_t. Hence V is defined for each p_t ie $V = V(p_t)$. Using (3) (4) and (5)

$$V(p_t) = \frac{min(\frac{C}{L}, \bar{o}) - F(p_t)\frac{C}{L}(1 - \bar{o}) + H(p_t)\bar{o}(1 - \frac{C}{L}) - \bar{o}}{min(\frac{C}{L}, \bar{o}) - \bar{o}\frac{C}{L}}. \quad (7)$$

For a given $\frac{C}{L}$ relationship, the optimal value is

$$V_{opt} = \max V(p_t). \quad (8)$$

III. Crisp decision over ROC

The main goal of this section is to explain how our model works and to show the benefits that good results would give us. No real experiments have been made yet but these results are, at least, a possibility.

Let us analyze the probabilistic approach. Many models (IDS), eight in our example, are now available. Consider again the same event E that consists on an intrusion. We can predict E happens or not with a certain probability p depending on how many models has predicted the event. We make a prediction periodically taking into account we predict the event will happen if its probability is over a certain threshold p_t. A possible scenario is the one in Table III.

TABLE III

Possible measures on the event E (intrusion). 1 represents the IDS has launched an alert on the event E (predicts that E happens). 0 represents it does not. If the event E occurs it is represented by 1. If it does not it is represented by 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

In this way, for a threshold $p_t = 0.2$, Table IV shows predictions and facts over thirty days and Table V illustrates the corresponding prediction contingency matrix. Table VI shows the main figures involved in the calculation of the economic value and Fig. 1 plots $V(0.2)$ for different $\frac{C}{L}$ relationships.

TABLE IV

Thirty 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

Let us show the figures for the different thresholds p_t.

- With $p_t = 0$, the values are: $H = 1$, $F = 1$ and $\bar{o} = 0.367$.
- With $p_t = 0.2$ the predictions are: $H = 0.909$, $F = 0.684$ and $\bar{o} = 0.367$.

TABLE V

Prediction contingency matrix

$p_t = 0.2$		Ocurrs	
		No	Yes
Prediction	No	0.200	0.033
	Yes	0.433	0.333

TABLE VI

Economic value V for different $\frac{C}{L}$ relationships ($p_t = 0.2$)

C	L	$\frac{C}{L}$	M	M_{per}	M_{fre}	$V(0.2)$
5	50	0.1	0.110	0.037	0.100	-0.158
10	50	0.2	0.187	0.073	0.200	0.105
15	50	0.3	0.263	0.110	0.300	0.193
18	50	0.36	0.309	0.132	0.360	0.222
20	50	0.4	0.340	0.147	0.367	0.121
25	50	0.5	0.417	0.183	0.367	-0.273
30	50	0.6	0.493	0.220	0.367	-0.864
35	50	0.7	0.570	0.257	0.367	-1.848
40	50	0.8	0.647	0.293	0.367	-3.818
45	50	0.9	0.723	0.330	0.367	-9.727

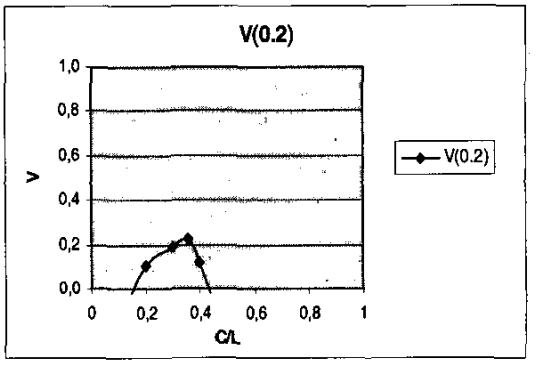

Fig. 1. Economic value against $\frac{C}{L}$ relationship for threshold $p_t = 0.2$. There is value for the interval [0.143,0.367]. The maximum is reached at $\frac{C}{L} = \bar{o} = 0,367$

- With $p_t = 0.4$, we obtain: $H = 0.545, F = 0.263$ and $\bar{o} = 0.367$.
- With $p_t = 0.6$, we obtain: $H = 0.545, F = 0.158$ and $\bar{o} = 0.367$.
- With $p_t = 0.8$, we obtain: $H = 0.273, F = 0.053$ and $\bar{o} = 0.367$.
- And at last, with $p_t = 1$, we obtain: $H = 0, F = 0$ and $\bar{o} = 0,367$.

Finally, Fig. 2 shows the corresponding A_{ROC} and Fig. 3 shows the individual graphs V against $\frac{C}{L}$ for every p_t. $V_{opt}(p_t)$ peaks for users with $\frac{C}{L}$ next to 0.367 which is the frequency value $\bar{o}$. At this point $H(p_t) - F(p_t) = V(\bar{o})$. The envelope function V_{opt} shows value for all users with $\frac{C}{L}$ between 0.143 and 0.750. This illustrates the benefit of probabilistic predictions over deterministic ones. The probabilistic approach gives us a wider range $\frac{C}{L}$ 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

it.

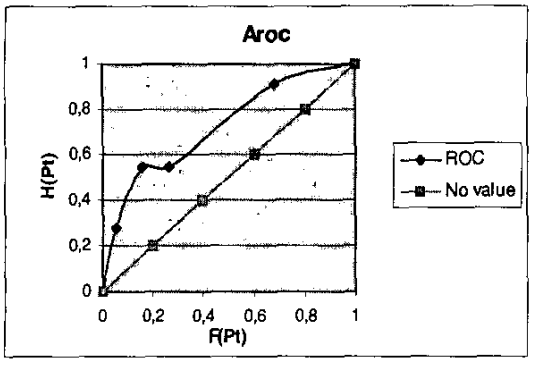

Fig. 2. A_{ROC} is the area under the ROC curve (represented in black). Each point on the curve corresponds to a threshold $(F(p_t), H(p_t))$. The curve is interpolated based on these points

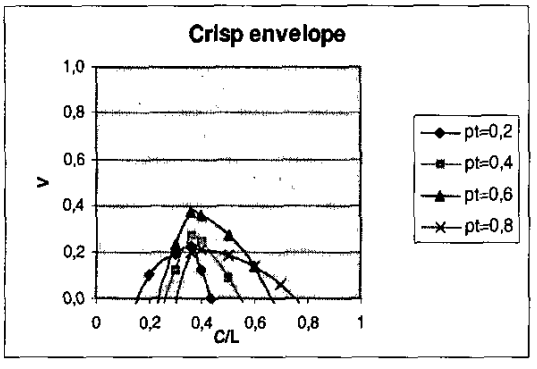

Fig. 3. Envelope. This figure shows the result of combining the value curves for each threshold. The final envelope curve (multi-model) is better than single model curves because the interval with positive value is wider and its optimum is the maximum of the single models

In practice, the user will know the $\frac{C}{L}$ relationship for his system. He will have the value of the intrusion frequency ($\bar{o}$). With this data, the user will run the model. It will give the probability of having an intrusion. Depending on where is its corresponding point in the graph of value he will know if the IDS prediction improves results on the frequency or not. Let us imagine that, in our example, $\frac{C}{L} = \bar{o}$. This would give user the maximum value for $p_t = 0.6$. In this case, the user can say that with a probability of 60% his prediction has a value of almost 40% of the perfect prediction. In this way it will be appropriate for him to take some action (for example, unplugging the system, changing a rule in the firewall...).

IV. Fuzzy decision over ROC

Fuzzy logic aims to give sound mathematical foundations to vague and imprecise reasoning typical of humans [12]. Fuzzy controllers have shown good performance where preconditions for a reactive behaviour are uncertain, or not clearly defined [9][10]. They are also often used in systems where state transitions should be softened making decisions with fuzzy boundaries [11][8].

Applied to our IDS model, using crisp thresholds to decide whether an event is predicted or not, generates isolated partitions where the percentages around the borderline were ignored. Instead of using a crisp threshold p_t, a fuzzy threshold, to some extent, may take these cases into account. Therefore, the final decision would probably became sounder (as it should be reflected in the envelope curve) using fuzzy sets as thresholds.

A fuzzy set associates a truth level between 0 and 1 per each possible value of a given dominion in order to model vague concepts like tall, old, hot, etc. as fuzzy sets. Fuzzy sets are often represented in a piece-wise way, for instance through four or less pairs of values in the form (x-point, truth level) such as: $\{(x_1,0)(x_2,1)(x_3,1)(x_4,0)\}$, or $\{(x_1,0)(x_2,1)(x_3,1)\}$ or even $\{(x_1,0)(x_2,1)\}$. So statements about this concepts are asserted with some certainty factor obtained from the membership level of the given value to a vague concept represented by a fuzzy set. For instance, the conditions to deduce the statement 'event E is predicted' may be represented as $\{(0.25,0)(0.55,1)(1,1)\}$. Then, if 4 of 8 IDS models predict such event, the resulting prediction would be asserted with a certainty factor (CF) of 0.83 (see Fig. 4).

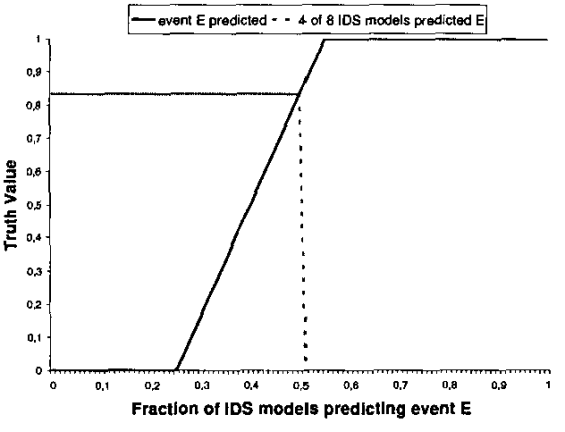

Fig. 4. Application of a fuzzy threshold when half of the IDS models predict an event E

Based on this fuzzy interpretation of the threshold used to decide whether an event is predicted or not from the percentage of IDS models that predict such event, six fuzzy sets were defined to compare with the corresponding crisp thresholds: $p_t = 0, 0.2, 0.4, 0.6, 0.8, 1$.

$$p_t = 0, \rightarrow \{(0,0)(0.12,1)(1,1)\}$$
$$p_t = 0.2, \rightarrow \{(0.08,0)(0.32,1)(1,1)\}$$
$$p_t = 0.4, \rightarrow \{(0.28,0)(0.52,1)(1,1)\}$$
$$p_t = 0.6, \rightarrow \{(0.48,0)(0.72,1)(1,1)\}$$
$$p_t = 0.8, \rightarrow \{(0.68,0)(0.92,1)(1,1)\}$$
$$p_t = 1, \rightarrow \{(0.88,0)(1,1)\}$$

Piece-wise definition of fuzzy thresholds

The fuzzy sets defined with these values may be represented with a trapezium over the domain of possible fractions of IDS models that can predict the given event E (from 0, no IDS model predict event E, to 1, where all models predict event E), as shown in Fig. 5.

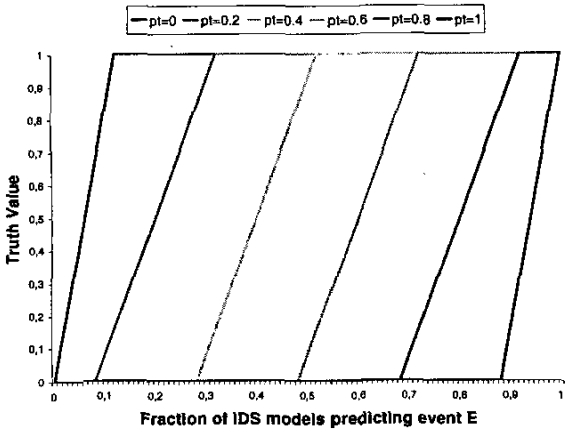

Fig. 5. Graphical definition of the fuzzy thresholds used

Other fuzzy sets could be defined to test the IDS model, and even, to improve our results, since the improvement in the decision making come from the width of the range where truth values are between 0 and 1. But, in order to propose a coherent group of fuzzy sets, they were defined in such a way that they have a wide range while these ranges almost do not overlap.

Assumed such fuzzy sets, α, β, γ, and δ values of Table I, can be computed from the frequency of success of the predictions according to the certainty factor CF obtained from the application of these fuzzy thresholds (CF and CF) instead of binary values obtained from crisp thresholds (1 and 0).

Fig. 6 shows the resulting A_{ROC} obtained from the application of the fuzzy sets mentioned above. In this figure we observe that this A_{ROC} covers a greater area, and therefore, we can conclude that better results are achieved using fuzzy thresholds.

On the other hand, Fig. 7 shows the corresponding envelope obtained with these fuzzy thresholds. Comparing these results on the illustrative example with those obtained with crisp thresholds, the envelope optimum is reached at a fuzzy threshold $p_t = 0.4$ with value 0.384. The envelope optimum for the crisp model was reached at a threshold $p_t = 0.6$ with value 0.374. The $\frac{C}{L}$ interval in which we have economical value (benefits) is also improved with the fuzzy model. This interval is now $[0.109, 0.750]$ while with crisp thresholds, it was $[0.143, 0.750]$.

Finally, Table VII and Table VIII show (in percentage) a measure of the improvement obtained from the fuzzy approach related to the enlargement of the area under ROC curve, and the extra width and height of the envelope obtained with fuzzy thresholds over the envelope obtained with crisp thresholds.

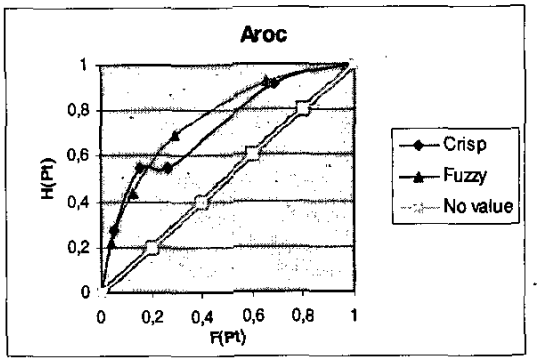

Fig. 6. Fuzzy ROC vs Crisp ROC. Better area is achieved by the fuzzy approximation

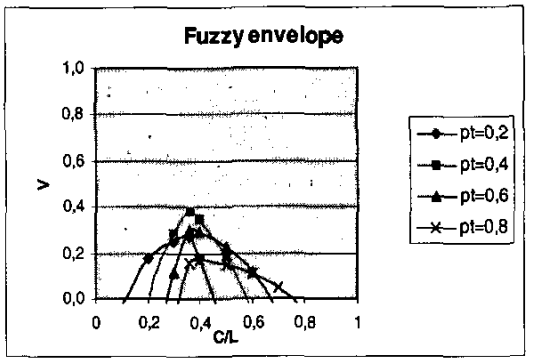

Fig. 7. Fuzzy envelope. The interval showing value is wider than the crisp case. The optimum is also higher

TABLE VII

ENLARGEMENT OF AREA UNDER ROC CURVE OBTAINED THROUGH THE USE OF FUZZY THRESHOLDS

	$CrispROC$	$FuzzyROC$
comparing with 'No Value' curve	28.07%	45.17%
comparing with 'Crisp ROC' curve	-	13.35%

TABLE VIII

IMPROVEMENT OF ENVELOPE THROUGH THE USE OF FUZZY THRESHOLDS

	extra width	extra height
fuzzy envelope vs. crisp envelope	11.21%	4.38%

V. CONCLUSIONS

The approach of this paper is motivated by the need to develop a general methodology to evaluate the skill (and usefulness) of user-specific IDS predictions. The relative operating characteristic (ROC) gives the predicted hit rate and false alarm rate for an event E, made by assuming that E will occur if it is predicted with a probability that exceeds some specified threshold p_t (and that E will not occur if it is predicted with a probability that does not exceed p_t). A plot of hit rate versus false alarm rate for varying p_t is known as a ROC curve, and the area under the ROC curve, A_{ROC}, is a measure of the

skill of a probabilistic forecast system. Therefore, the larger area under ROC curve obtained from the application of fuzzy thresholds, implies that this decision model will provide more accurate predictions than using crisp thresholds p_t. Despite of this results were obtained from an illustrative simulation, we can expect a similar improvement when both decision models would be tested with real data. It is likely to get good results if the measures on $H(p_t)$ and $F(p_t)$ give us a curve with an $A_{ROC} > 0.5$ and if we have a wide enough sample.

The hit and false alarm rate are fundamental parameters in one simple assessment of the user-specific value of the predictive system. The analysis is based on a simple and idealized decision model [6]. We imagine a user who has to decide whether or not to take some form of precautionary action, at cost C, on the basis that if E occurs, a loss L will be incurred. For example, in the event of an attack, if hackers activity has increased in a certain period of time, a company can invest more money in security measures. This precautionary action would imply a cost to the company. Knowing only the frequency of occurrence of certain attacks, and the cost/loss ratio $\frac{C}{L}$, the user can decide to either always or never take precautionary action. The forecast system can be said to be of value if the user's mean expense is less than the expense based on the frequency. Frequency information is of no value to users whose cost/loss ratio $\frac{C}{L}$ is next to $\bar{o}$. (For these users, the cost of always or never taking precautionary action is about the same). So this economical analysis of frequency and cost/loss ratio was applied over the use of both models, and it also concluded that fuzzy logic on decisions models for IDS improves the expected benefits of a predictive system. Although it is very dependant on the particular case, it is necessary an evaluation of C and L. We need models to estimate both values. For these estimations some studies has been made [13][3][2].

To summarize, this paper tried to propose a model that can quantify the usefulness that a multi-model IDS prediction can bring to the user. Relevant better results were achieved using fuzzy logic instead of crisp logic. The main handicap of the present work is the absence of real network data to apply both models (fuzzy and crisp). Some studies have been made simulating networks [14][15] but the settings of the experiments remain controversial [16]. This is a common problem of intrusion detection evaluation, since publicly accessible data are very hard to obtain, and therefore experiments are hardly verifiable in real domains.

Future work may involve a weighted mean of IDS models according to their past accuracy, instead of the average mean used in this paper. It could also be interesting to study the possible application of a multi agent system to coordinate and evaluate such IDS models working in parallel.

REFERENCES

[1] S. Axelsson, "Intrusion detection systems: a survey and taxonomy". Department of Computer Engineering. Chalmers University of Technology. Sweden. 2000.

[2] D. Kinn, and K. Timm, "Justifying the expense of IDS, part one: an overview of ROIs for IDS". http://online.securityfocus.com/infocus/1608 July 18, 2002.

[3] S. Northcutt, and J. Novak, "Network Intrusion Detection An Analyst's Handbook". Second Edition. New Riders Publishing. 2001.

[4] H.R. Stanski, L.J. Wilson, and W.R. Burrows, "Survey of common verification methods in meteorology". World Weather Report No. 8. World Meteorological Organization. Geneva.

[5] T.N Palmer, C. Brankovic, and D.S. Richardson, "A probability and decision-model analysis of PROVOST seasonal ensemble integrations". Research Department. Technical Memorandum No.265. Nov 1998.

[6] A.H. Murphy, "A new vector partition of the probability score". J. Appl. Meteor. 1973.

[7] R.W. Katz, and A.H. Murphy, "Forecast value: prototype decision-making models". In Economic value of weather and climate forecasts. Eds. Cambridge University Press. 1997.

[8] J. Carbó, J.M. Molina, and J. Dávila, "A fuzzy model of reputation in agent-mediated electronic commerce". Advances in Fuzzy Systems and Evolutionary Computation, pp. 22-28. World Scientific and Engineering Society Press. 2001.

[9] D. Dubois, H.T. Nguyen, H. Prade, and M. Sugeno, "The real contribution of fuzzy systems". Fuzzy Systems, Kluwer Academic Publishers, eds. pp. 1-14. 1998.

[10] S. Farinwata, D. Filev, and R. Langari, "Fuzzy control", eds. John Wiley and Sons Ltd. 2000.

[11] J.M. Molina, F. J. Jimnez, and J.R. Casar, "Fuzzy reasoning for multisensor management". IEEE International Conference on Systems, Man and Cybernetics, pp 1398-1403. Vancouver, Canada. 1995

[12] L.A. Zadeh, "Fuzzy sets". Information and Control, n8, pp. 338-353. 1965.

[13] L. Wenke, F. Wei, M. Miller, S. Stolfo, E. Zadok, "Toward cost-sensitive modeling for intrusion detection and response". Journal of Computer Security,Vol. 10 n1,2. 2002.

[14] R.P. Lippmann, D.J. Fried, I. Graf, J. Haines, K. Kendall, D. McClung, D. Webber, S. E. Webster, D. Wyschogrod, R.K. Cunningham, and M. A. Zissman, "Evaluating intrusion detection systems: The 1988 DARPA off-line intrusion detection evaluation". Proceedings of the on DARPA Information Survivability Conference and Exposition (DISCEX'00), Hilton Head, South carolina, Jan. 25-27). IEEE Computer Society Press, Los Alamitos, CA, 12-26. 2000.

[15] R.P. Lippmann, and J. Haines, "Analysis and results of the 1999 DARPA off-Line intrusion detection evaluation" in Recent Advances in Intrusion Detection, Third International Workshop, RAID 2000. Toulouse, France, October 2000, Proceedings. H. Debar, L.Me and S.F. Wu, eds. Springer Verlag. Lecture Notes in Computer Science, Vol.1907, pp.162-182. 2000.

[16] J. McHugh, "Testing intrusion detection systems: A critique of the 1998 and 1999 DARPA intrusion detection system evaluations as performed by Lincoln Laboratory". ACM Transactions on Information and System Security, Vol.3, No.4, pp. 262-294. November 2000.