Expert Systems with Applications 42 (2015) 193–202

Contents lists available at ScienceDirect

Expert Systems with Applications

journal homepage: www.elsevier.com/locate/eswa

On the combination of genetic fuzzy systems and pairwise learning for improving detection rates on Intrusion Detection Systems

Salma Elhag [a], Alberto Fernández [b,*], Abdullah Bawakid [c], Saleh Alshomrani [c], Francisco Herrera [c,d]

[a] Department of Information Systems, King Abdulaziz University (KAU), Jeddah, Saudi Arabia
[b] Department of Computer Science, University of Jaén, Jaén, Spain
[c] Faculty of Computing and Information Technology - North Jeddah, King Abdulaziz University (KAU), Jeddah, Saudi Arabia
[d] Department of Computer Science and Artificial Intelligence, CITIC-UGR (Research Center on Information and Communications Technology), University of Granada, Granada, Spain

ARTICLE INFO

Article history:
Available online 11 August 2014

Keywords:
Intrusion Detection Systems
Genetic Fuzzy Systems
Pairwise learning
One-vs-One
Misuse detection

ABSTRACT

Security policies of information systems and networks are designed for maintaining the integrity of both the confidentiality and availability of the data for their trusted users. However, a number of malicious users analyze the vulnerabilities of these systems in order to gain unauthorized access or to compromise the quality of service. For this reason, Intrusion Detection Systems have been designed in order to monitor the system and trigger alerts whenever they found a suspicious event.

Optimal Intrusion Detection Systems are those that achieve a high attack detection rate together with a small number of false alarms. However, cyber attacks present many different characteristics which make them hard to be properly identified by simple statistical methods. According to this fact, Data Mining techniques, and especially those based in Computational Intelligence, have been used for implementing robust and accuracy Intrusion Detection Systems.

In this paper, we consider the use of Genetic Fuzzy Systems within a pairwise learning framework for the development of such a system. The advantages of using this approach are twofold: first, the use of fuzzy sets, and especially linguistic labels, enables a smoother borderline between the concepts, and allows a higher interpretability of the rule set. Second, the divide-and-conquer learning scheme, in which we contrast all possible pair of classes with aims, improves the precision for the rare attack events, as it obtains a better separability between a "normal activity" and the different attack types.

The goodness of our methodology is supported by means of a complete experimental study, in which we contrast the quality of our results versus the state-of-the-art of Genetic Fuzzy Systems for intrusion detection and the C4.5 decision tree.

© 2014 Elsevier Ltd. All rights reserved.

1. Introduction

In an era of communications, a lot of effort has been put on filtering out known malware, exploits and vulnerabilities within a network, which could compromise the confidentiality, integrity or availability of the system. Therefore, intrusion detection is an essential part of a complete security policy in information systems. Since a wide number of potential intrusions occur every day, Intrusion Detection Systems (IDS) have been research and developed for addressing these cyber-attack events (Axelsson, 1998). In conjunction with security audit mechanisms, which dynamically monitor logs and network traffic for gathering information of the system use, IDS have the function of analyzing this information and then applying detection algorithms to determine whether these events are symptomatic of an attack or constitute a legitimate use of the system (Denning, 1987).

When referring to IDS, two main categories are clearly emphasized (Debar, Dacier, & Wespi, 1999):

1. **Misuse detection**, which are based on a signature database of already known attacks (Lee & Stolfo, 2000).
2. **Anomaly detection** adopts a complementary procedure: they first define a profile for "normal behavior", and then attacks are detected as deviations from this normal profile (Patcha & Park, 2007).

The former types of IDS are very efficient, but they are limited to the information from which they were trained, i.e. new types of attack might be not identified. On the contrary case, anomaly detection could incur into more false positives, and they strongly depends on the continuity of the user for its "normal activity".

* Corresponding author. Tel.: +34 953 213016; fax: +34 953 212472.
E-mail addresses: salma53ster@gmail.com (S. Elhag), alberto.fernandez@ujaen.es (A. Fernández), abawakid@kau.edu.sa (A. Bawakid), sshomrani@kau.edu.sa (S. Alshomrani), herrera@decsai.ugr.es (F. Herrera).

http://dx.doi.org/10.1016/j.eswa.2014.08.002
0957-4174/© 2014 Elsevier Ltd. All rights reserved.

Throughout the years, a wide number of different approaches in the field of Data Mining have been proposed for the area of intrusion detection (Lee, Stolfo, & Mok, 2000). Among them, those based on Computational Intelligence techniques have achieved a high success according to their good properties to detect both known and unseen intrusion attacks and to recognize normal network traffic (Wu & Banzhaf, 2010; Guo et al., 2014).

Our aim in this paper is to develop a misuse detection system to automatically extract optimal classification rules from training data under two main premises. On the one hand, the learnt rule set must be capable of correctly identifying all types of attacks, including rare attack categories, which is a major challenge in the IDS research domain (Khor, Ting, & Phon-Amnuaisuk, 2012). On the other hand, the final model should be linguistically interpretable for human comprehension (Gacto, Alcalá, & Herrera, 2011).

For achieving these goals, we propose the use of linguistic Fuzzy Rule Based Classification Systems (FRBCSs) (Ishibuchi, Nakashima, & Nii, 2004) as baseline classifiers for the development of our proposal. Additionally, in order to enhance in a higher degree the recognition of the minority classes within the IDS, we consider the use of the fuzzy system in synergy with decomposition techniques (Lorena, Carvalho, & Gama, 2008).

This classification scheme is based on a "divide-and-conquer" strategy, in which the original multi-class problem is divided into binary subproblems, which are independently learned by different base classifiers whose outputs are then combined to classify an instance. Proceeding this way, the borderline areas among the classes are simplified and individual concepts can be better identified (Galar, Fernández, Barrenechea, Bustince, & Herrera, 2011).

The choice of FRBCSs is justified by two main reasons: first, the intrusion detection problem involves many numeric attributes, and models which are directly built on numeric data might cause high detection errors. Hence, small deviations in an intrusion might not be detected and small changes in the normal user profile will cause false alarms. Second, security itself includes fuzziness, as the boundary between the normal and abnormal behavior cannot be well defined.

Specifically, as fuzzy learning classifier we have considered the use of a robust FRBCS, i.e. the Fuzzy Association Rule-based Classification for High-Dimensional problems (FARC-HD) (Alcalá-Fdez, Alcalá, & Herrera, 2011). The inner procedure of this algorithm comprises an optimization stage carried out by means of Evolutionary Algorithms (Eiben & Smith, 2003). This type of hybridization is known as a Genetic Fuzzy System (GFS) (Cordón, Gomide, Herrera, Hoffmann, & Magdalena, 2004; Alcalá, Nojima, Ishibuchi, & Francisco, 2012). One of the main reasons for the success of this type of techniques is their ability to exploit the information accumulated about and initially unknown search space in order to bias subsequent searches into useful subspaces, i.e. their robustness (Herrera, 2008). However, to the best of our knowledge only few works on the topic have addressed the problem of IDS with this type of approach (Gomez & Dasgupta, 2001; Özyer, Alhajj, & Barker, 2007; Tsang, Kwong, & Wang, 2007; Abadeh, Mohamadi, & Habibi, 2011).

As pointed out previously, in this paper we will make use of the One-vs-One (OVO) methodology in which the binary subproblems are obtained by confronting all possible pair of classes (Hastie & Tibshirani, 1998). The usage of pairwise learning to deal with real-world applications is frequent, being a simple yet effective way of overcoming multi-class problems. Moreover, empirical results in those papers have shown that the usage of OVO can enhance the results of the direct application of the baseline classifiers with inherent multi-class support (Fürnkranz, 2002; Galar et al., 2011; Sáez, Galar, Luengo, & Herrera, 2014).

The validity of our approach will be tested using the standard KDDCUP'99 dataset (Lee & Stolfo, 2000). This way, the experimental results will be directly comparable with most of the Computational Intelligence approaches for intrusion detection. Specifically, for the evaluation of the goodness of our IDS proposal, we will contrast the experimental results versus the standard FARC-HD algorithm and several GFS approaches that have been developed for misuse detection. In particular, we have selected a multi-objective genetic fuzzy intrusion detection system (MOGFIDS) (Tsang et al., 2007), three different GFS schemes proposed by in Abadeh et al. (2011), and a GFS for boosting fuzzy association rules (Özyer et al., 2007). Finally, we will complement our comparison with the classical C4.5 decision tree (Quinlan, 1993).

In short, the main contributions of this work are enumerated below:

1. We consider the use of a GFS for the intrusion detection problem. This kind of soft computing technique provides two main advantages: (1) obtaining a better separability of the different types of alarms by means of the achievement of smoother borderline for the rules of the final system; (2) a higher interpretability of the obtained rule set for the better understanding of the working procedure of the system.
2. A pairwise learning approach is applied for addressing the classification of the multiple classes, i.e. normal behaviour and all intrusion alarms. By following a divide and conquer scheme, we are able to learn a better suited discrimination function for each pair of classes, thus improving the overall classification.
3. To the best of our knowledge, this is the first research paper that combines GFS and OVO for the IDS problem. Furthermore, the baseline classifier used, i.e. FARC-HD, excel from the algorithms of the state-of-the-art as it is well-suited for high dimensional problems.
4. Finally, the goodness of this new methodology is shown by means of its high performance when contrasted versus several GFS algorithms developed for IDS, and with the C4.5 decision tree. We must stress the good behaviour of our approach especially for the minority classes.

In order to carry out the study, this manuscript is organized as follows. First, Section 2 introduces the preliminary concepts for this paper, i.e. the context of IDS, some basic notions on FRBCSs and the related work. Next, Section 3 introduces our proposal for the development of a combined approach between GFS and pairwise learning for the improvement on misuse detection. Then, the experimental framework including the features of the KDD-CUP'99 dataset, metrics of performance, algorithms for comparison and their parameters, are presented in Section 4. The analysis of the results is shown throughout Section 5. Finally, Section 6 summarizes and concludes the work.

2. Preliminaries: Intrusion Detection Systems and fuzzy rule based classification systems

Prior to the description of our proposal, we must introduce some preliminary concepts which will help to understand the context of this work and the features of the solution which is to be develop. According to this, we will first present a brief review on IDS (Section 2.1), and then we will recall some basic concepts on FRBCSs (Section 2.2). Finally, Section 2.3 merges the former topics, and study those works on IDS related to fuzzy systems in general and GFS in particular.

2.1. Intrusion Detection Systems

Any information system should accomplish three main principles for guarantee a correct access to the data, namely

confidentiality, integrity and availability. Unfortunately, all networks could be the object of unauthorized accesses so that a strong security policy must be established for avoiding this violation of the prior principles (Chebrolu, Abraham, & Thomas, 2005). The technology developed for this aim is known as IDS, which dynamically monitors logs and network traffic, applying detection algorithms to identify these potential intrusions within a network (Axelsson, 1998; Denning, 1987). In particular, IDS can be split into two categories according to the detection methods they employ, including (1) misuse detection and (2) anomaly detection.

Misuse detection systems use an established set of known attack patterns, and then monitor the net trying to match incoming packets and/or command sequences to the signatures of known attacks (Lee & Stolfo, 2000). Hence, decisions are made based on the prior knowledge acquired from the model. Starting from a wide collection of cyber-attacks results in an extremely efficient system, comprising low false alarm rates. Additionally, the system administrator could reliably determine which attacks the system is experiencing immediately upon installation. This fact is the main advantage and, at the same time, the main drawback of this kind of system: maintaining a database for all of the possible attacks against a network is a tedious, if not impossible task in a modern computer network environment, limiting its accuracy when faced with the challenge of detecting new intrusive activities.

On the contrary, anomaly detection methods seek to overcome this problem by defining a "normal" behavioral model, and assuming that any deviation from this profile is considered to be an attack (Patcha & Park, 2007). Therefore, good detection results can be obtained from novel attacks. Additionally, learned profiles of normal activity are customized for every system, making it quite difficult for an attacker to know with certainty what activities it can carry out without getting detected. However, anomaly detection systems also present several technical challenges. First of all, the complexity of developing a system of these characteristics is higher than in the case of misuse detection. Furthermore, a higher percentage of false alarms are raised, together with the problem of accurately determining which kind of alarm has been triggered.

2.2. Introduction to FRBCSs

Any classification problem consists of m training patterns $x_p = (x_{p1}, \ldots, x_{pn}, C_p), p = 1, 2, \ldots, m$ from M classes where x_{pi} is the ith attribute value ($i = 1, 2, \ldots, n$) of the pth training pattern.

In this work we use fuzzy rules of the following form for our FRBCSs:

$$\begin{aligned} &\text{Rule } R_j : \text{If } x_1 \text{ is } A_{j1} \text{ and } \ldots \text{ and } x_n \text{ is } A_{jn} \\ &\text{then Class} = C_j \text{ with } RW_j \end{aligned} \tag{1}$$

where R_j is the label of the jth rule, $x = (x_1, \ldots, x_n)$ is an n-dimensional pattern vector, A_{ji} is an antecedent fuzzy set, C_j is a class label, and RW_j is the rule weight (Ishibuchi & Yamamoto, 2005). We use triangular MFs as antecedent fuzzy sets.

When a new pattern x_p is selected for classification, then the steps of the fuzzy reasoning method (Cordón, del Jesus, & Herrera, 1999) are as follows:

1. **Matching degree**, that is, the strength of activation of the if-part for all rules in the Rule Base with the pattern x_p. A conjunction operator (t-norm) T, is applied in order to carry out this computation:

$$\mu_{A_j}(x_p) = T(\mu_{A_{j1}}(x_{p1}), \ldots, \mu_{A_{jn}}(x_{pn})), \quad j = 1, \ldots, L \tag{2}$$

2. **Association degree**. To compute the association degree of the pattern x_p with the M classes according to each rule in the Rule Base. To this aim, a combination operator h, is applied to combine the matching degree with the rule weight (RW). In our case, this association degree only refers to the consequent class of the rule (i.e. $k = Class(R_j)$).

$$b_j^k = h(\mu_{A_j}(x_p), RW_j^k), \quad k = 1, \ldots, M; \quad j = 1, \ldots, L \tag{3}$$

3. **Pattern classification soundness degree for all classes**. We use an aggregation function f, which combines the positive degrees of association calculated in the previous step.

$$Y_k = f(b_j^k, j = 1, \ldots, L \quad \text{and} \quad b_j^k > 0), \quad k = 1, \ldots, M \tag{4}$$

4. **Classification**. We apply a decision function F over the soundness degree of the system for the pattern classification for all classes. This function will determine the class label l corresponding to the maximum value.

$$F(Y_1, \ldots, Y_M) = \arg\max(Y_k), \quad [k = 1, \ldots, M] \tag{5}$$

Where L denote the number of rules in the Rule Base and M the number of classes of the problem.

2.3. Related work for fuzzy systems in IDS

The ultimate goal of IDS is to achieve a high attack detection rate along with a low false alarm rate, being this a serious challenge to be overcome. For this reason, both misuse detection and anomaly detection system make use of Data Mining techniques to aid in the processing of large volumes of audit data and the increasing complexity of intrusion behaviors (Zhu, Premkumar, Zhang, & Chu, 2001; Peddabachigari, Abraham, Grosan, & Thomas, 2007). In particular, Soft Computing and Computational Intelligence techniques have become essential pieces for addressing this problem (Wu & Banzhaf, 2010).

In the introduction of this paper, we stressed the good properties related to the use of fuzzy logic for the development of IDS. For this reason, throughout the years many approaches have been proposed and analyzed aiming to take advantage of these fuzzy systems. One of the first techniques was the Fuzzy Intrusion Recognition Engine (FIRE) (Dickerson & Dickerson, 2000; Dickerson, Juslin, Koukousoula, & Dickerson, 2001). This approach employ the well known C-means algorithm for defining the fuzzy sets and their membership functions, and then authors determine their own hand-encoded rules for malicious network activities, which was probably the main limitation of this work.

Regarding GFS, to the best of our knowledge few works have been published in the specialized literature that address this area. For example, in Gomez and Dasgupta (2001) a genetic programming algorithm evolves tree-like structure of chromosomes (rules) whose antecedents are composed of triangular membership functions. Multiple objective functions are defined, which are then combined into a single fitness function by means of user-defined weights. The hitch here is that these weights cannot be optimized dynamically for different cases.

A deep study of different architectures for GFS have been developed in Abadeh, Habibi, and Lucas (2007); Abadeh et al. (2011). In these works, fuzzy rules are expressed in the same way as presented in Section 2.2. Then, authors analyze the three main schemes for rule generation with genetics algorithms, namely the Genetic Cooperative-Competitive Learning (GCCL) (Greene & Smith, 1993), the Pittsburgh approach (Smith, 1980; Smith, 1983), and the Iterative Rule Learning (IRL) (Venturini, 1993). Additionally, in Victorie and Sakthivel (2012) authors

extend the previous work by defining a parallel environment for the execution of the population of rules.

Another topic of work is the integration of association rules and frequent episodes with fuzzy logic (Florez, Bridges, & Vaughn, 2002). In one of the latest publications (Tajbakhsh, Rahmati, & Mirzaei, 2009), authors use Apriori as baseline algorithm and fuzzify the obtained rules following the recommendations made in Kuok, Fu, and Wong (1998). Then, several implementation techniques were used to speed up the algorithm, i.e. to reduce items involved in rule induction without resulting into any considerable information loss.

The interest on the use of GFS have been also shown in the field of fuzzy association mining (Özyer et al., 2007). In this latter work, the procedure is divided into two stages: (1) authors generate a large number of candidate association fuzzy rules for each class; (2) with aims at reducing the fuzzy rule search space, a boosting GA based on the IRL approach is applied for each class for rule pre-screening using two evaluation criteria. However, it only optimizes classification accuracy and omits the necessity of interpretability optimization.

Finally, multi-objective GFS have been also analyzed in the context of IDS. In Tsang et al. (2007) the authors propose MOGFIDS (short for Multi-Objective Genetic Fuzzy Intrusion Detection System), which is based on the previous work of the authors related to an agents-based evolutionary approach for fuzzy rules (Wang, Kwong, Jin, Wei, & Man, 2005). This approach is based on the construction and evolution, in a Pittsburgh style, of an accurate and interpretable fuzzy knowledge base. Specifically, it is a genetic wrapper that searches for a near-optimal feature subset from network traffic data.

According to this brief review on the topic, we confirm that although GFS are effective approaches for solving classification problems, their use into IDS have been not address in depth. Therefore, we believe that our contribution will be of great interest for the research community, as it consolidates this line of work, and the lessons learned paved the way for future work on the topic.

3. Proposed methodology: Genetic Fuzzy Systems and pairwise learning

In this section we will present our approach for improving the behaviour in misuse detection for IDS using linguistic FRBCSs. Particularly, our scheme is based on the combination between GFS and the OVO learning scheme. According to the former, in Section 3.1 we will first describe the features of the FARC-HD algorithm, which has been selected as baseline technique. Then, Section 3.2 presents the details for the OVO binarization technique and the procedure for the decision step. Finally, we will develop a brief discussion on the benefits of the combination of both approaches for IDS, which will be carried out in Section 3.3.

3.1. Genetic Fuzzy Systems: FARC-HD algorithm

In this paper we have make use of a novel and robust linguistic FRBCS, i.e. the FARC-HD approach (Alcalá-Fdez et al., 2011). This algorithm is based on association discovery, a commonly used technique in Data Mining for extract interesting knowledge from large datasets (Han & Kamber, 2006) by means of finding relationships between the different items in a database (Zhang & Zhang, 2002). The integration between association discovery and classification leads to precise and interpretable models.

FARC-HD is aimed at obtaining an accurate and compact fuzzy rule-based classifier with a low computational cost. In short, this method is based on the following three stages (as depicted in Fig. 1):

Stage 1 *Fuzzy association rule extraction for classification*: A search tree is employed to list all possible frequent fuzzy item sets and to generate fuzzy association rules for classification, limiting the depth of the branches in order to find a small number of short (i.e., simple) fuzzy rules.

Stage 2 *Candidate rule pre-screening*: Afterwards the rule generation, the size of the rule set can be too large to be interpretable by the end user. Therefore, a pre-selection of the most interesting rules is carried out by means of a "subgroup discovery" mechanism based on an improved weighted relative accuracy measure (wWRAcc') (Kavsek & Lavrac, 2006).

Stage 3 *Genetic rule selection and lateral tuning*: Finally, in order to obtain a compact and accurate set of rules within the context of each problem, an evolutionary process will be carried out in a combination for the selection of the rules with a tuning of membership function, as its positive synergy has been shown in previous work on the topic (Casillas, Cordón, del Jesús, & Herrera, 2005; Alcala, Alcalá-Fdez, & Herrera, 2007).

3.2. Classification via binarization techniques: One vs One

Multiple classes imply an additional difficulty for Data Mining algorithms, as the boundaries among the classes may overlap, causing a decrease in the performance level. In this context, decomposition strategies have been widely used in the literature to address this problem (see Lorena et al., 2008 for an extensive review).

The main idea behind this procedure is to transform the original multiple-class problem into binary subsets, which are easier to discriminate, via a class binarization technique (Allwein, Schapire, & Singer, 2000; Dietterich, 2000; Galar et al., 2011). Among them, the OVO approach (Hastie & Tibshirani, 1998) is one of the most extended schemes, being established by default in several widely used software tools (Alcalá;-Fdez et al., 2009; Chang & Lin, 2011; Hall et al., 2009) to handle multi-class problems using SVMs.

In OVO, also known as Pairwise classification, the original m-class problem is divided into $m \cdot (m-1)/2$ two-class problems, one for each possible pair of examples. Then, a binary classifier is trained ignoring those examples that do not belong to its related classes. An example of this binarization technique is depicted in Fig. 2.

When classifying instances, a query is submitted to all binary models, and the predictions of these models are combined into an overall classification (Hüllermeier & Brinker, 2008; Hüllermeier & Vanderlooy, 2010). In order to do so, it is usual to construct a score-matrix R containing these outputs, which are used to decide the final class:

$$R = \begin{pmatrix} - & r_{12} & \dots & r_{1m} \\ r_{21} & - & \dots & r_{2m} \\ \vdots & & & \vdots \\ r_{m1} & r_{m2} & \dots & - \end{pmatrix} \tag{6}$$

where $r_{ij} \in [0,1]$ is the confidence of the classifier discriminating classes i and j in favor of the former; whereas the confidence for the latter is computed by $r_{ji} = 1 - r_{ij}$ (if it is not provided by the classifier). Once the score-matrix is constructed, any combination can be used to infer the class.

In our case, we will make use of the preference relations solved by Non-Dominance Criterion (ND) (Fernández, Calderón, Barrenechea, Bustince, & Herrera, 2010). ND was originally defined for decision making with fuzzy preference relations (Orlovsky, 1978). In this case, the score matrix is considered as a fuzzy preference relation, which has to be normalized. Then the degree

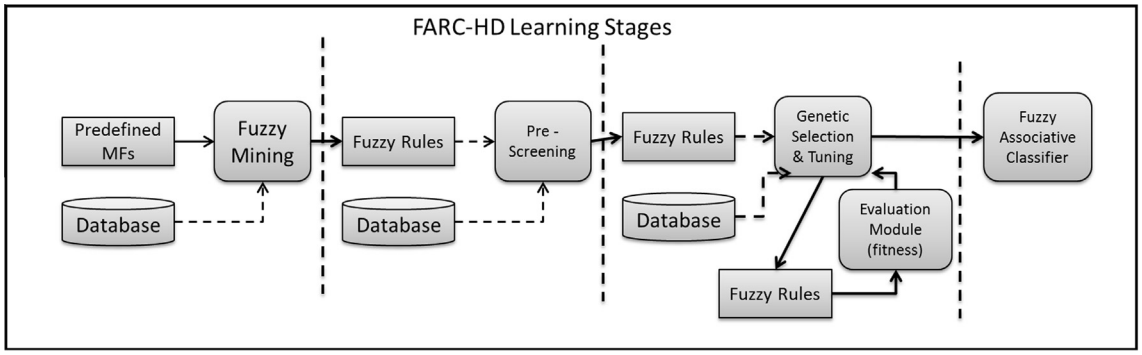

Fig. 1. Learning stages for the FARC-HD algorithm.

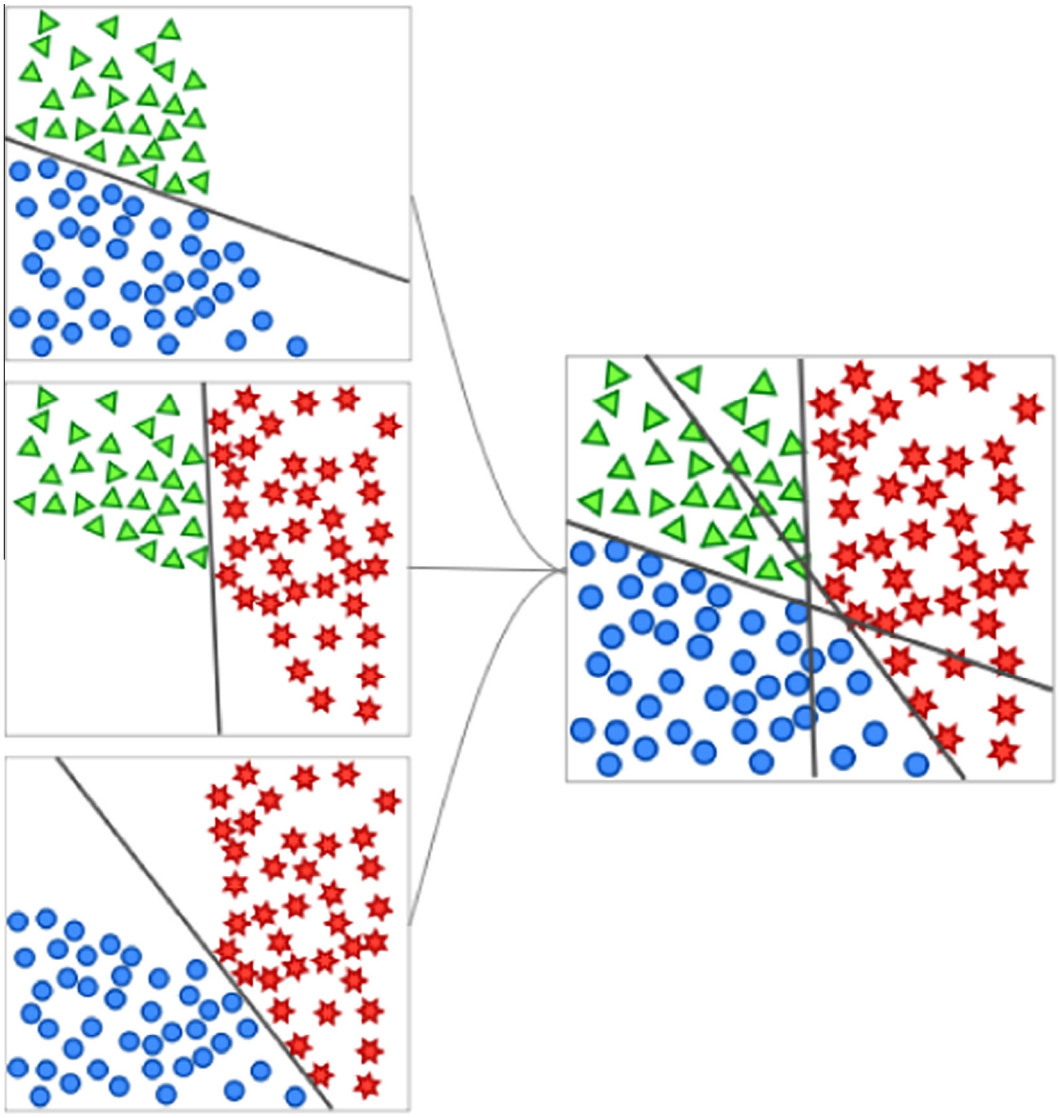

Fig. 2. Example of the OVO binarization technique for a 3-class problem.

of non-dominance is computed (the degree to which the class i is dominated by none of the remaining classes) and the class with the largest degree is predicted:

$$Class = \arg\max_{i=1,\ldots,m}\left\{1 - \sup_{1\leqslant j\neq i\leqslant m} r'_{ji}\right\} \tag{7}$$

where r'_{ji} corresponds to the normalized and strict score-matrix. In cases where we have a pattern of output for which more than two classes obtain the same vote, the instance will be classified according to the maximum a priori probability, i.e., the majority class. If a vote remains tied, the class is assigned randomly from the previous possibilities.

3.3. Benefits on the synergy of GFS and OVO for IDS

The main aim behind our proposal is to develop a good misuse detection mechanism. Therefore, we must not just focus on a global attack detection rate, but improving individual accuracy rates so that different actions must be carried out depending on the type of attack discovered. As natural, our methodology should also maintain a low false alarm rate. These constraints can be summarized as obtaining a high and homogeneous precision for all classes of the IDS problem.

The hypothesis for the success of our proposed methodology, regarding the previous fact, is basically based on two pillars:

1. The necessity of a robust and accurate learning classifier which could reach the desirable quality regarding the performance metric. GFSs satisfy the former premise in a twofold way: on the one hand, the use of fuzzy sets, and especially linguistic labels, allows at smoothing the borderline areas of the classes in the inference process for the rule set. On the other hand, the application of a genetic optimization supposes a leap of quality as the fuzzy system is adapted to the context of the problem.
2. For IDSs, we must overcome which is known as the Base-rate Fallacy (Axelsson, 2000). This phenomenon implies that when having a higher proportion of benign network activities in network data, the factor governing the false alarm rate dominates the factor governing the detection rate. The ultimate consequence of this, is that the detection rate of some intrusive events, i.e. the rare categories, to be less than the false alarm rate, so that they are usually ignored in a standard analysis.
   In our case, all types of attacks are worth to process. The pairwise learning procedure, working in a "divide-and-conquer" scheme, the learning system can better focus on the different relations among all the classes of the problem, avoiding the bias for the majority cases. Additionally, we must aware that the decision boundaries of these binary problems are much simpler than in the original multi-class dataset.

Regarding the former facts, if we contrast our new approach versus the algorithms from the state-of-the-art presented in Section 2.3, we may stress two clear differences: (1) the FARC-HD baseline fuzzy classifier has shown a robust behaviour for different classification scenarios (López, Fernández, & Herrera, 2013; Sanz, Fernández, Bustince, & Herrera, 2013) and it is especially well-suited for high-dimensional problems (Alcalá-Fdez et al., 2011), such as the one we are addressing in IDS. (2) To the best of our knowledge, our approach is the only GFS scheme that aims at providing an average classification for all concepts of the IDS problem, and not a single metric of performance.

4. Experimental framework

In this section we first provide details of the real-world binary-class imbalanced problems chosen for the experiments (subSection 4.1). Then, we will describe the learning algorithms selected for this study and their configuration parameters (subSection 4.2). Finally, we present the metrics of performance applied to compare the results obtained with the different classifiers (subSection 4.3).

4.1. Benchmark data: KDDCUP'99 problem

For the evaluation of our proposed methodology, we will make use of the KDDCUP'99 problem, a dataset prepared by Lee and Stolfo (2000) from the DARPA'98 intrusion detection evaluation program. Behind this decision lies the widest use of this problem in the network intrusion detection domain, which makes it a standard until today (Benferhat, Boudjelida, Tabia, & Drias, 2013; Khor et al., 2012; Chung & Wahid, 2012).

The dataset was managed by Lincoln Labs and consists of an environment of a local area network (LAN) that simulates a typical U.S. Air Force LAN, including several weeks of raw TCP dump data with normal activities and various types of attacks. Each connection is described by 41 discrete and continuous features that can be basically grouped into three categories: basic features of individual connection, content features within a connection, and traffic features which are computed using 2 s time windows (Lee, Stolfo, & Mok, 1999) (See Table 1).

The original size of this problem was too large (approximately 5 million examples) so that it truly affects the time spent in the training stage. For this reason, usually a small subset containing the 10% of the instances is used for this proposal. In particular, it contains a number of 494,021 records, which can be labeled to be normal, or an attack. 24 different attack types can be found in this problem, but they can be grouped into four major categories, namely:

- Denial of Service (DOS): make some machine resources unavailable or too busy to answer to legitimate users requests (SYN flooding).
- Probing (PRB): Surveillance for information gathering or known vulnerabilities about a network or a system (port scanning).
- Remote To Local (R2L): use vulnerability in order to obtain unauthorized access from a remote machine (password guessing).
- User To Root (U2R): exploit vulnerabilities on a system to gain local super-user (root) privileges (buffer overflow attack).

This attack clustering helps to enhance the detection rates, but the dataset still show an imbalanced distribution, as shown in Table 2. This issue could still hinder the learning and detection of the minority categories (R2L and U2R), which are affected by the dominant ones (Normal and DOS).

In order to overcome this difficulty, some research works used various sizes of datasets prepared by making random selection, sampling or taking a subset from the KDDCUP'99 datasets (Chen, Abraham, & Bo, 2007; Muda, Yassin, Sulaiman, & Udzir, 2011; Yi, Wu, & Xu, 2011). In our case, we have found a huge quantity of

Table 1
Information of the classes for the KDDCUP'99 problem: subclasses and distribution of instances.

Class	SubClasses	Size (distribution %)
Normal	Normal	97,278 (19.6911)
DOS	back, land, neptune, pod, smurf, teardrop	391,458 (79.2391)
PRB	ipsweep, nmap, portsweep, satan	4,107 (0.8313)
R2L	ftp_write, guess_passwd, imap, multihop, phf, spy, warezclient, warezmaster	1,126 (0.2279)
U2R	buffer_overflow, loadmodule, perl, rootkit	52 (0.0105)

Table 2
Characteristics of the pre-processed (no redundant examples) KDDCUP'99 problem used in the experimental study. New class distribution and percentage of reduction from the original problem.

Class	New examples (distribution %)	Original size (reduction %)
Normal	87,832 (60.3303)	97,278 (9.7103)
DOS	54,572 (37.4876)	391,458 (86.0593)
PRB	2,130 (1.4630)	4,107 (48.1373)
R2L	999 (0.6861)	1,126 (11.2788)
U2R	52 (0.0357)	52 (0.0000)
Total	145,585(100.00)	494,021 (70.5306)

repeated registers practically in all classes. Hence, we have first proceeded to the removal of these duplicated instances, obtaining a total of 145,585 examples which follows a new class distribution which is shown in Table 2.

When conducting the experiments, we will proceed with a hold-out based validation methodology. In this way, we will use a 10% of the total examples for training and a 90% for test. We must point out that, instead of carrying out a standard hold-out, we have taken into account the initial distribution of classes, so that a 10% of the examples for each class will be used for training except for U2R for whom we will include a 50%. Table 3 shows the final distribution of examples for each partition/class.

4.2. Algorithms and parameters

In this paper, we have considered several algorithms for a fair analysis of the behavior of our proposal. In particular, we will first contrast our results versus the standard FARCHD. Regarding GFS developed for IDS, as stated in Section 2.3 there are, to the best of our knowledge, few works of interest which cover this category. In this way, we must stress the MOGFIDS algorithm (Tsang et al., 2007), the GFS approaches from Abadeh et al. (2011), and a genetic approach for boosting fuzzy association rules (Boost-FAR) (Özyer et al., 2007) as the most suited algorithms of the same paradigm. Additionally, we will include C4.5 (Quinlan, 1993) in the experimental study as a state-of-the-art rule induction algorithm.

1. **FARC-HD** (Alcalá-Fdez et al., 2011): First, we have selected 5 labels per variables for the fuzzy sets, product t-norm as conjunction operator and additive combination for the inference procedure. As specific parameters of the learning stage, we have set up the minimum support to 0.05 and the minimum confidence to 0.8. Finally, we have fixed the maximum depth of the tree to a value of 3, and the k parameter for the pre-screening to 2. For more details about these parameters, please refer to Alcalá-Fdez et al. (2011).
2. **MOGFIDS** (Tsang et al., 2007): We apply 12 different chromosomes (rule sets) each of which has 10 fuzzy rule sets solutions. Therefore there are 120 fuzzy systems generated in total for initialization. MOGFIDS is then trained using 80 iterations for the evolution of the individuals.
3. **GFS from Abadeh et al.** (Abadeh et al., 2011): In this case, there are three different schemes for the GFS:
   - The **"GFS-GCCL"** approach will content 20 chromosomes as population, with a crossover and mutation probability of 0.9 and 0.1 respectively. The number of mutation attempts will be set up to 20, and the replacement percentage to the 20%. Finally, the number of generations will be just 100.
   - The Pittsburgh approach (**"GFS-Pitts"**) will share the same parameters than the GCCL technique, but in this case the population size is incremented to 50 individuals.
   - The **"GFS-IRL** scheme also shares the parameters of the previous approaches. However, the population size is increased to 100 individuals. Finally, as specific parameters we must take into account the fraction of instances that should be covered by a rule which has been defined at 0.5, the maximal tolerance for the error made by an individual rule is 0.2, and the maximum number of generations for weight adjustment has been set up to 20.
4. GFS for Association Rules (**Boost-FAR**) (Özyer et al., 2007): Initial rules are obtained by the standard fuzzy Apriori (Kuok et al., 1998). Pre-screening is made for the first 1000 rules per class (5000 in total). Then, the GFS procedure considers the following parameters: 100 chromosomes as population size, consider 20 to be the number of elite solutions. The replacement ratio is set to 0.8 and the crossover and mutation probabilities are set to 0.9 and 0.09 respectively. 30 iterations are carried out when extracting each rule. Finally, the parameter value of the fitness function $k_{max} = 0.2$.
5. **C4.5** (Quinlan, 1993): For C4.5 we have set a confidence level of 0.25, the minimum number of item-sets per leaf was set to 2 and the application of pruning was used to obtain the final tree.

Table 3
Number of examples per class in each dataset partition.

Class	#Ex. Training	#Ex. Test
Normal	8,783	79,049
DOS	5,457	49,115
PRB	213	1,917
R2L	100	899
U2R	26	26
Total	14579	131,006

4.3. Performance metrics for IDS

In the specialized literature for IDS in general, and for misuse detection in particular, authors have made use of several metrics of performance for the evaluation of their results in comparison with the state-of-the-art. In this paper, we have selected different measures which will allow us to analyze the behaviour of our approach under several perspectives:

1. *Accuracy*: It stands for the global percentage of hits. In our case (IDS), its contribution is low as it does not take into account the individual accuracies of each class, but it has been selected as a classical measure.

$$Acc = \frac{\sum_{i=1}^{C} TP_i}{N} \tag{8}$$

   where C stands for the number of classes, N stands for the number of examples and TP_i is the number of True Positives of the ith class.
2. *Mean F-Measure*. In the binary case, the standard f-measure computes a tradeoff between precision and recall of both classes. In this case, we compute the average for the F-measure achieved for each class (taken as positive) and the remaining ones (taken as a whole as negative):

$$MFM = \frac{\sum_{i=1}^{C} FM_i}{C} \tag{9}$$

$$FM_i = \frac{2 \cdot Recall_i \cdot Precision_i}{Recall_i + Precision_i} \tag{10}$$

$$Precision_i = \frac{TP_i}{TP_i + FP_i} \tag{11}$$

$$Recall_i = \frac{TP_i}{TP_i + FN_i} \tag{12}$$

   where TP_i, FP_i and FN_i are the number of true positives, false positives and false negatives of the ith class respectively.
3. *Average accuracy*. It is computed as the average of the individual hits for each class. For this reason, it is also known as the average recall:

$$AvgAcc = \frac{1}{C}\sum_{i=1}^{C} Recall_i \tag{13}$$

4. *Attack Accuracy*. In this case we omit the "Normal" instances and we focus in checking whether we guess correctly the different "Attack" types individually.

$$AttAcc = \frac{1}{C-1}\sum_{i=2}^{C} Recall_i \quad (14)$$

In this case, the first class ($i = 1$) is considered to be the "Normal" class.

5. *Attack Detection Rate*. It stands for the accuracy rate for the attack classes. Therefore, it is computed as:

$$ADR = \frac{\sum_{i=2}^{C} TP_i}{\sum_{i=2}^{C} TP_i + FN_i} \quad (15)$$

Reader must take into account that also in this case, the first class ($i = 1$) is considered to be the "Normal" class.

6. *False Alarm Rate*. In this case, we focus on the "Normal" examples, and we check which is the percentage of "false negatives" found, i.e. those instances identified as "alarms" but which are actually normal behavior.

$$FAR = \frac{FP_1}{TP_1 + FP_1} \quad (16)$$

As in the former metric (*ADR*), the "Normal" class has the first index ($i = 1$).

5. Experimental study

This section is aimed at analyzing the goodness of our pairwise learning proposal over GFS, in order to develop an accurate IDS regarding different metrics of performance. Therefore, the goal is being able to achieve a global accuracy for all concepts of the problem, so that we must be careful in not focusing on a single measure, but studying all of them in a whole.

Table 4 shows the experimental results for the test partition of KDDCUP'99 (90% of the dataset) over the selected accuracy measures. From this table, we must stress the good performance shown by our approach with respect to the remaining methodologies. Among all algorithms of study, FARCHD-OVO excels as the one which reaches a high value for each and every one of the metrics of performance, being the most robust approach on average. We must point out that in the case of Boost-FAR results for the F-measure are not included as the confusion matrix is not available in the paper of this proposal.

When contrasting the results of our OVO technique versus the standard FARCHD, we observe that there is a clear improvement in all cases, especially regarding the average accuracy, attack accuracy, and the false alarm rate. Specifically, a higher average accuracy implies that the individual classes of the problem are better identified, both the normal class, and the four types of attack independently, as observed by the high value for *AttAcc*. This is due to the fact that our proposed technique carries out binary comparisons within the decision step, thus considering a higher degree of confidence such in a decision making problem. The low false alarm rate is a direct consequence of the former behavior.

Regarding the individual comparison versus the remaining GFSs algorithms of this study, the superior quality of FARCHD-OVO is undoubted. The only model that is somehow near the performance of our approach is GFS-IRL, as it shows a high value for the mean F-measure and a low false alarm rate. In the latter case, although good results for mean F-measure implies a high average precision, the detection of the different attack classes is poor, which results on a undesirable behaviour. As a final remark, we may observe than the quality of the GFS algorithms for comparison is even under the average performance of the standard FARCHD approach.

Finally, when contrasting our results with the C4.5 decision tree, we focus basically on the average metrics such as the mean F-measure and the average accuracy, whose results excel in contraposition with that of C4.5. We must also point out that the main advantage of FARCHD-OVO in contrast to C4.5 is, as in the case of the standard FARCHD algorithm, its ability to achieve a robust behavior for all concepts of the problem, i.e. a good tradeoff between recall (average accuracy) and precision (mean F-measure).

In order to complement our experimental study, we also show in Table 5 the individual accuracy results, i.e. the recall measure for all classes of the KDDCUP'99 dataset.

This table of results also supports the findings extracted throughout the experimental study, from which we stated the superior behaviour of FARCHD-OVO in terms of average performance. Our initial aim was to develop a methodology that was able to identify correctly all classes of the problem, and not just focus on a good detection rate, or a simple high accuracy over all examples, disregard the class distribution.

According to this last fact, we observe that FARCHD-OVO is capable at achieving a high recall for the most minority classes (R2L and U2R), thus enhancing the behaviour of C4.5 which only focuses on the majority ones (Normal and DOS). We acknowledge that this is also the case of GFS-GCCL (for R2L) and GFS-Pitts (for U2R), but the hitch here is that these techniques reach this good performance under the premise of raising a higher ratio of false negatives (normal cases identified as errors), which is completely inappropriate for a comfortable use of the system. Additionally, the results for PRB in FARCHD-OVO are clearly superior than the case of the remaining FRBCS.

From this experimental analysis, we must emphasize that a potential scenario for further research is the addressing of the lesser represented types of attacks, according to the existing class imbalance. We must study the intrinsic data characteristics of this problem (López, Fernández, García, Palade, & Herrera, 2013) in order to propose more specific solutions that are able to learn correctly the boundaries for all classes.

Finally, we include the confusion matrices obtained in the training and test partitions for FARCHD-OVO algorithm. This is done with aims at complementing the experimental results, so that any interested research could reproduce and extend the current

Table 4
Experimental results in test over the reduced KDDCUP'99 dataset for different metrics of performance: Accuracy (Acc), Mean F-measure (MFM), Average Accuracy (AvgAcc), Attack average accuracy (AttAcc), Attack Detection Rate (ADR), and False Alarm Rate (FAR).

Classifier	Acc	MFM	AvgAcc	AttAcc	ADR	FAR
FARCHD-OVO	99.00	84.12	**89.32**	**86.70**	97.77	.1910
FARCHD	98.30	84.26	87.76	84.77	96.17	.2947
MOGFIDS	92.77	61.68	62.19	53.15	91.41	1.6599
GFS-GCCL	98.68	77.87	85.59	82.12	97.49	.5214
GFS-Pitts	98.64	75.55	86.07	82.69	97.26	.4016
GFS-IRL	98.64	**85.42**	85.18	81.57	97.16	.3777
Boost-FAR	97.61	74.26	67.47	59.36	94.13	**.0845**
C4.5	**99.59**	81.81	87.79	84.79	**99.29**	.2062

The bold values indicate the best result for each performance measure.

Table 5
Individual accuracy results in test for every class of the reduced KDDCUP'99 dataset.

Classifier	Normal	DOS	PRB	R2L	U2R
FARCHD-OVO	**99.81**	98.05	95.83	87.54	65.38
FARCHD	99.71	96.58	93.84	79.42	**69.23**
MOGFIDS	98.36	97.20	88.60	11.01	15.79
GFS-GCCL	99.48	98.16	81.72	**94.05**	54.55
GFS-Pitts	99.60	98.00	72.57	93.50	66.67
GFS-IRL	99.62	98.03	78.17	88.16	61.90
Boost-FAR	99.92	97.07	46.50	31.35	62.50
C4.5	99.79	**99.68**	**96.14**	85.65	57.69

The bold values indicate the best result for each performance measure.

Table 6
Confusion Matrix in the training partition for the FARCHD-OVO approach.

	Normal	DOS	PRB	R2L	U2R	Recall
Normal	8776	4	1	2	0	99.92
DOS	96	5361	0	0	0	98.24
PRB	5	1	207	0	0	97.18
R2L	9	0	0	91	0	91.00
U2R	0	0	0	1	25	96.15
Precision	98.76	99.91	99.52	96.81	100.00	

Table 7
Confusion Matrix in the test partition for the FARCHD-OVO approach.

	Normal	DOS	PRB	R2L	U2R	Recall
Normal	78898	51	37	35	28	99.81
DOS	954	48158	3	0	0	98.05
PRB	54	25	1837	0	1	95.83
R2L	83	2	0	787	27	87.54
UR2	4	1	0	4	17	65.38
Precision	98.63	99.84	97.87	95.28	23.29	

study for additional future work. The aforementioned information is shown in Tables 6 and 7.

In accordance with the all the experimental results, one of the main advantages of our new our methodology are the achievement of a system that is able to improve the attack detection rate among all algorithms while maintaining a low false alarm rate. Another goodness of our approach is the homogenous accuracy for all classes of the problem. Finally, we must highlight the interpretability of the rule set, as it is composed by a low number of rules including few attributes in the antecedents of the rules. By contrast, the remaining GFS approaches consider from 100 up to 5,000 rules, which degrades the interpretability of the fuzzy approach.

We must take into account that the efficiency of our proposal is affected by the pairwise learning approach, as the number of classifiers to be learnt increases with the number of classes. Nevertheless, the significance of this problem is lowered according to the following issues: (1) the number of training instances considered by each binary classifier is lower than the general case; (2) misuse detection algorithms are developed as offline systems, so that the training time cannot be considered as a key factor.

6. Concluding remarks

In this work we have proposed a new methodology based on GFS and pairwise learning for the development of a robust and interpretable IDS. Concretely, this approach is based on the FARCHD algorithm, which is a linguistic fuzzy association rule mining classifier, and the OVO binarization that confronts all pairs of classes in order to learn a single model for each couple.

The quality of the results for this proposal has been tested by considered an appropriate experimental framework. The algorithms for comparison have been selected from the state-of-the-art in GFS for IDS. Specifically, we have make use of a multi-objective fuzzy model (MOGFIDS), three different GFS schemes developed by Abadeh et al., and a genetic approach for boosting fuzzy association rules. Additionally, we have included C4.5 as a baseline rule induction algorithm for comparison. The KDDCUP'99 has been selected as benchmark dataset following the standards for works on this topic. Finally, several metrics of performance have been employed for determining the robustness of our proposal under different perspectives.

Experimental results show that our FARCHD-OVO approach has the best tradeoff among all performance measures, especially in the mean F-measure, the average accuracy and the false alarm rate. The good behavior shown by our methodology is supported by the advantages derived from the use of fuzzy logic and linguistic labels. First, this paradigm allows at better managing the numerical variables associated with intrusion detection. Furthermore, with this tool we may address properly the vague division that exists between normal and anomalous accesses. Regarding the decomposition techniques for the learning stage, it has been carried out with aims at achieving a more precise identification for all types of attacks, including the minority ones. The application of this "divide-and-conquer" strategy improves the individual accuracy for the different classes of the problem, which is reflected on the high value for the average accuracy metric.

The promising performance achieved by this model allows us to consider several ways for future work. On the one hand, analyzing in depth the intrinsic properties of the IDS problem for the design of an "ad hoc" GFS algorithm to solve this problem. On the other hand, make the most of the possibilities that the decomposition-based learning offers, and focusing on different schemes for both carrying out both the binarization and the final decision process from the score-matrix. Finally, we can consider a multi-objective evolutionary algorithm for maximizing several metrics of performance and being able to select the most appropriate solution depending on the context.

Acknowledgment

This paper is funded by King Abdulaziz University, under Grant No. (3-611-1434-HiCi). The authors therefore, acknowledge technical and financial support of KAU.

References

Abadeh, M. S., Habibi, J., & Lucas, C. (2007). Intrusion detection using a fuzzy genetics-based learning algorithm. *Journal of Network and Computer Applications, 30*(1), 414–428.

Abadeh, M. S., Mohamadi, H., & Habibi, J. (2011). Design and analysis of genetic fuzzy systems for intrusion detection in computer networks. *Expert Systems with Applications, 38*(6), 7067–7075.

Alcalá;-Fdez, J., Sánchez, L., García, S., del Jesus, M. J., Ventura, S., Garrell, J. M., et al. (2009). KEEL: A software tool to assess evolutionary algorithms for data mining problems. *Soft Computing, 13*, 307–318.

Alcala, R., Alcalá-Fdez, J., & Herrera, F. (2007). A proposal for the genetic lateral tuning of linguistic fuzzy systems and its interaction with rule selection. *IEEE Transactions on Fuzzy Systems, 15*(4), 616–635.

Alcalá-Fdez, J., Alcalá, R., & Herrera, F. (2011). A fuzzy association rule-based classification model for high-dimensional problems with genetic rule selection and lateral tuning. *IEEE Transactions on Fuzzy Systems, 19*(5), 857–872.

Alcalá, R., Nojima, Y., Ishibuchi, H., & Francisco, H. (2012). Special issue on evolutionary fuzzy systems. *International Journal Of Computational Intelligence Systems, 5*(2), 209–211.

Allwein, E. L., Schapire, R. E., & Singer, Y. (2000). Reducing multiclass to binary: A unifying approach for margin classifiers. *Journal of Machine Learning Research, 1*, 113–141.

Axelsson, S. (1998). Research in intrusion-detection systems: A survey. Technical Report 98–17, Department of Computer Engineering, Chalmers University of Technology, Goteborg, Sweden.

Axelsson, S. (2000). The base-rate fallacy and the difficulty of intrusion detection. *ACM Transaction on Information Systems and Security, 3*(3), 186–205.

Benferhat, S., Boudjelida, A., Tabia, K., & Drias, H. (2013). An intrusion detection and alert correlation approach based on revising probabilistic classifiers using expert knowledge. *Applied Intelligence, 38*(4), 520–540.

Casillas, J., Cordón, O., del Jesús, M. J., & Herrera, F. (2005). Genetic tuning of fuzzy rule deep structures preserving interpretability and its interaction with fuzzy rule set reduction. *IEEE Transactions on Fuzzy Systems, 13*(1), 13–29.

Chang, C.-C., & Lin, C.-J. (2011). LIBSVM: A library for support vector machines. *ACM Transactions on Intelligent Systems and Technology, 2*, 27:1–27:27.

Chebrolu, S., Abraham, A., & Thomas, J. P. (2005). Feature deduction and ensemble design of intrusion detection systems. *Computers and Security, 24*(4), 295–307.

Chen, Y., Abraham, A., & Bo, Y. (2007). Hybrid flexible neural-tree-based intrusion detection systems. *International Journal Intelligent Systems, 22*(4), 337–352.

Chung, Y. Y., & Wahid, N. (2012). A hybrid network intrusion detection system using simplified swarm optimization (SSO). *Applied Soft Computing, 12*(9), 3014–3022.

Cordón, O., del Jesus, M. J., & Herrera, F. (1999). A proposal on reasoning methods in fuzzy rule-based classification systems. *International Journal of Approximate Reasoning, 20*(1), 21–45.

Cordón, O., Gomide, F., Herrera, F., Hoffmann, F., & Magdalena, L. (2004). Ten years of genetic fuzzy systems: Current framework and new trends. *Fuzzy Sets and Systems, 141*(1), 5–31.

Debar, H., Dacier, M., & Wespi, A. (1999). Towards a taxonomy of intrusion-detection systems. *Computer Networks, 31*(8), 805–822.

Denning, D. E. (1987). An intrusion detection model. *IEEE Transactions on Software Engineering, 13*(2), 222–232.

Dickerson, J., & Dickerson, J. (2000). Fuzzy network profiling for intrusion detection. In Proceedings of the 19th International Conference of the North American Fuzzy Information Society (NAFIPS'00), (pp. 301–306). Atlanta, GA, USA. IEEE Press.

Dickerson, J., Juslin, J., Koukousoula, O., & Dickerson, J. (2001). Fuzzy intrusion detection. In Proceedings of the 20th International Conference of the North American Fuzzy Information Society (NAFIPS'01) and Joint the 9th IFSA World Congress, vol. 3, (pp. 1506–1510). Vancouver, Canada. IEEE Press.

Dietterich, T. G. (2000). An experimental comparison of three methods for constructing ensembles of decision trees: Bagging, boosting, and randomization. *Machine Learning, 40*, 139–157.

Eiben, A., & Smith, J. (2003). Introduction to evolutionary computation. In *Natural Computing Series*. Springer.

Fernández, A., Calderón, M., Barrenechea, E., Bustince, H., & Herrera, F. (2010). Solving multi-class problems with linguistic fuzzy rule based classification systems based on pairwise learning and preference relations. *Fuzzy Sets and Systems, 161*(23), 3064–3080.

Florez, G., Bridges, S., & Vaughn, R. (2002). An improved algorithm for fuzzy data mining for intrusion detection. In Proceedings of the 21st North American Fuzzy Information Processing Society Conference (NAFIPS'02), (pp. 457–462). New Orleans, LA.

Fürnkranz, J. (2002). Round robin classification. *Journal of Machine Learning Research, 2*, 721–747.

Gacto, M., Alcalá, R., & Herrera, F. (2011). Interpretability of linguistic fuzzy rule-based systems: An overview of interpretability measures. *Information Sciences, 181*(20), 4340–4360.

Galar, M., Fernández, A., Barrenechea, E., Bustince, H., & Herrera, F. (2011). An overview of ensemble methods for binary classifiers in multi-class problems: Experimental study on one-vs-one and one-vs-all schemes. *Pattern Recognition, 44*(8), 1761–1776.

Gomez, J. % Dasgupta, D. (2001). Evolving fuzzy classifiers for intrusion detection. In Proceedings of IEEE Workshop on Information Assurance, (pp. 68–75). United States Military Academy, West Point, New York.

Greene, D. P., & Smith, S. F. (1993). Competition-based induction of decision models from examples. *Machine Learning, 13*(2-3), 229–257.

Guo, C., Zhou, Y., Ping, Y., Zhang, Z., Liu, G., & Yang, Y. (2014). A distance sum-based hybrid method for intrusion detection. *Applied Intelligence, 40*(1), 178–188.

Hall, M., Frank, E., Holmes, G., Pfahringer, B., Reutemann, P., & Witten, I. H. (2009). The WEKA data mining software: An update. *SIGKDD Explorations, 11*(1), 10–18.

Han, J., & Kamber, M. (2006). Data mining. In *Concepts and Techniques*. Morgan Kaufmann.

Hastie, T., & Tibshirani, R. (1998). Classification by pairwise coupling. *The Annals of Statistics, 26*(2), 451–471.

Herrera, F. (2008). Genetic fuzzy systems: Taxonomy, current research trends and prospects. *Evolutionary Intelligence, 1*, 27–46.

Hüllermeier, E., & Brinker, K. (2008). Learning valued preference structures for solving classification problems. *Fuzzy Sets and Systems, 159*(18), 2337–2352.

Hüllermeier, E., & Vanderlooy, S. (2010). Combining predictions in pairwise classification: An optimal adaptive voting strategy and its relation to weighted voting. *Pattern Recognition, 43*(1), 128–142.

Ishibuchi, H., Nakashima, T., & Nii, M. (2004). *Classification and modeling with linguistic information granules: Advanced approaches to linguistic Data Mining*. Springer–Verlag.

Ishibuchi, H., & Yamamoto, T. (2005). Rule weight specification in fuzzy rule–based classification systems. *IEEE Transactions on Fuzzy Systems, 13*, 428–435.

Kavsek, B., & Lavrac, N. (2006). Apriori-sd: Adapting association rule learning to subgroup discovery. *Applied Artificial Intelligence, 20*(7), 543–583.

Khor, K.-C., Ting, C.-Y., & Phon-Amnuaisuk, S. (2012). A cascaded classifier approach for improving detection rates on rare attack categories in network intrusion detection. *Applied Intelligence, 36*(2), 320–329.

Kuok, C., Fu, A., & Wong, M. (1998). Mining fuzzy association rules in databases. *SIGMOD Record, 27*(1), 41–46.

Lee, W., & Stolfo, S. (2000). A framework for constructing features and models for intrusion detection systems. *ACM Transactions on Information Systems and Security, 3*(4), 227–261.

Lee, W., Stolfo, S. J., & Mok, K. W. (1999). A data mining framework for building intrusion detection models. In *IEEE Symposium on Security and Privacy* (pp. 120–132). IEEE Computer Society.

Lee, W., Stolfo, S., & Mok, K. (2000). Adaptive intrusion detection: A data mining approach. *Artificial Intelligence Review, 16*, 533–567.

López, V., Fernández, A., García, S., Palade, V., & Herrera, F. (2013). An insight into classification with imbalanced data: Empirical results and current trends on using data intrinsic characteristics. *Information Sciences, 250*(20), 113–141.

López, V., Fernández, A., & Herrera, F. (2013). Addressing covariate shift for genetic fuzzy systems classifiers: A case of study with farc-hd for imbalanced datasets. In *FUZZ-IEEE'2013* (pp. 1–8). IEEE.

Lorena, A. C., Carvalho, A. C., & Gama, J. M. (2008). A review on the combination of binary classifiers in multiclass problems. *Artificial Intelligence Review, 30*(1–4), 19–37.

Muda, Z., Yassin, W., Sulaiman, M., & Udzir, N. (2011). A k-means and naive bayes learning approach for better intrusion detection. *Information Technology Journal, 10*(3), 648–655.

Orlovsky, S. A. (1978). Decision-making with a fuzzy preference relation. *Fuzzy Sets and Systems, 1*(3), 155–167.

Özyer, T., Alhajj, R., & Barker, K. (2007). Intrusion detection by integrating boosting genetic fuzzy classifier and data mining criteria for rule pre-screening. *Journal of Network and Computer Applications, 30*(1), 99–113.

Patcha, A., & Park, J.-M. (2007). An overview of anomaly detection techniques: Existing solutions and latest technological trends. *Computer Networks, 51*(12), 3448–3470.

Peddabachigari, S., Abraham, A., Grosan, C., & Thomas, J. P. (2007). Modeling intrusion detection system using hybrid intelligent systems. *Journal of Network and Computer Applications, 30*(1), 114–132.

Quinlan, J. (1993). *C45: Programs for Machine Learning*. Morgan Kaufmann.

Sáez, J. A., Galar, M., Luengo, J., & Herrera, F. (2014). Analyzing the presence of noise in multi-class problems: Alleviating its influence with the one-vs-one decomposition. *Knowledge and Information Systems, 38*(1), 179–206.

Sanz, J. A., Fernández, A., Bustince, H., & Herrera, F. (2013). Ivturs: A linguistic fuzzy rule-based classification system based on a new interval-valued fuzzy reasoning method with tuning and rule selection. *IEEE Transactions on Fuzzy Systems, 21*(3), 399–411.

Smith, S. (1980). A learning system based on genetic algorithms. (Ph.D. thesis), University of Pittsburgh, Pittsburgh, PA.

Smith, S. (1983). Flexible learning of problem solving heuristics through adaptive search. In 8th International Joint Conference on Artificial Intelligence, (pp. 422–425).

Tajbakhsh, A., Rahmati, M., & Mirzaei, A. (2009). Intrusion detection using fuzzy association rules. *Applied Soft Computing, 9*(2), 462–469.

Tsang, C.-H., Kwong, S., & Wang, H. (2007). Genetic-fuzzy rule mining approach and evaluation of feature selection techniques for anomaly intrusion detection. *Pattern Recognition, 40*(9), 2373–2391.

Venturini, G. (1993). SIA: A supervised inductive algorithm with genetic search for learning attributes based concepts. In P. Brazdil (Ed.), *Machine learning ECML–93. LNAI* (vol. 667, pp. 280–296). Springer.

Victorie, T. A., & Sakthivel, M. (2012). A local search guided differential evolution algorithm based fuzzy classifier for intrusion detection in computer networks. *International Journal of Soft Computing, 6*(5-6), 158–167.

Wang, H., Kwong, S., Jin, Y., Wei, W., & Man, K.-F. (2005). Agent-based evolutionary approach for interpretable rule-based knowledge extraction. *IEEE Transactions On Systems, Man, And Cybernetics - Part C: Applications And Reviews, 35*(2), 143–155.

Wu, S. X., & Banzhaf, W. (2010). The use of computational intelligence in intrusion detection systems: A review. *Applied Soft Computing, 10*(1), 1–35.

Yi, Y., Wu, J., & Xu, W. (2011). Incremental SVM based on reserved set for network intrusion detection. *Expert Systems with Applications, 38*(6), 7698–7707.

Zhang, C., & Zhang, S. (2002). Association rule mining, models and algorithms. *Lecture Notes in Computer Science* (vol. 2307). Springer.

Zhu, D., Premkumar, G., Zhang, X., & Chu, C.-H. (2001). Data mining for network intrusion detection: A comparison of alternative methods. *Decision Sciences, 32*(4), 635–660.