INTRUSION DETECTION BASED ON ORGANIZATIONAL COEVOLUTIONARY FUZZY CLASSIFIERS

Liu Fang and Chen Zhen-guo
School of Computer Science and Technology, Xidian University, Xi'an, 710071, China

Abstract: To solve the intrusion detection question, we introduce the fuzzy logic into Organization CoEvolutionary algorithm[1] and present the algorithm of Organization CoEvolutionary Fuzzy Classification (OCEFC). In this paper, we give an intrusion detection models based on OCEFC. After illustrating our model and applying it to the real-world network datasets KDD Cup 1999, we obtain the better performance than other traditional methods.

Key words: Intrusion detection; Organization CoEvolutionary; Anomaly detection; Fuzzy Logic

1. INTRODUCTION

An intrusion detection system (IDS) is a component of the computer and information security framework. Many intrusion detection techniques have been used to IDS. Using data mining techniques over system audit data[2]. Using short sequences of system calls performed by running programs as discriminators between normal and abnormal [3]. In this paper, we present a new approach, based on OCEFC, to anomaly detection over network.

2. ORGANIZATIONAL COEVOLUTIONARY (OCE)

In OCE [1], the objects having some similarities in the value of attributes are gathered firstly. Then, the significance of attributes is used to guide the evolutionary process. The set of objects whose decision attribute have the same values form a target class: $DC_a = \{O \mid O \in U \wedge a \in V_D \wedge O_D = a\}$. A parameter

called attribute significance CI_c is introduced to weigh the influence of each attribute.

Details about the OCE can be seen in [1].

3. AN INTRUSION DETECTION BASED ON OCEFC

In 1965 Lotfi Zadeh[4] first published a description and analysis of Fuzzy Logic. This is a true superset of Boolean Logic and permits the description of functions and processes with a degree of vagueness or uncertainty.

3.1. Organizational CoEvolutionary Fuzzy Classification

We redefine the $SAME_{ORG}$ viewing from the fuzzy logic.

Definition 1: Same Attribute is a condition attribute whose value of all objects in organization ORG has the same logic value.

The algorithm of the evolution of attribute significance can be seen in [1].

Algorithm 1 Organizational CoEvolutionary Fuzzy Classification

Step1: Each object in every target class is added to the corresponding population $P_i(0)$ as a free organization. $i = 1,2,\cdots,|V_D|$; $t = 0$, $i = 1$

Step2: If $i > |V_D|$, go to Step8; else go to Step3;

Step3: If the number of organizations in $P_i(t)$ is larger than 1, go to Step4, otherwise go to Step7;

Step4: Two parent organizations ORG_{p1}, ORG_{p2} are randomly selected from $P_i(t)$; The **evolutionary operator**[1] will act on ORG_{p1}, ORG_{p2}, and we can get child organizations ORG_{c1}, ORG_{c2};

Step5: Compute the continuous attributes fuzzy logic value and Compute the fitness of child organizations ORG_{c1}, ORG_{c2};

Step6: Perform **selection mechanism**[1] on ORG_{p1}, ORG_{p2} and ORG_{c1}, ORG_{c2}, move the rest organization in $P_i(t)$ to $P_i(t+1)$,go to Step3;

Step7: $i = i+1$, go to Step2;

Step8: If meet the stopping criteria, go to Step9, otherwise, $t = t+1$, $i = 1$, go to Step2;

Step9: After the evolutionary process ends, we can extract the rules.

Definition 2: *weight*: the confidence of knowledge; CON: the condition set. M_O is the number of object when $x = a$, $a \in \{VeryLow, Low, Medium, High, VeryHigh\}$ and $class = c$, N_O is the number of object where $class = c$, SV_a^c is a statistic value and can be get by M_O / N_O. So we can get the confidence of knowledge by following equation: $weight = MIN\{SV_a^c(x)\}$, where x belongs to CON. $SV_a^c(x)$ is corresponding x.

Definition 3: Two organizations ORG_{p1}, ORG_{p2} are randomly selected from a population firstly, IF $USE_{ORG_1} \subseteq USE_{ORG_2}$ OR $USE_{ORG_2} \subseteq USE_{ORG_1}$, $THEN$ $ORG = ORG_1 \bigcup ORG_2$, $USE_{ORG} = USE_{ORG_1} \bigcap USE_{ORG_2}$.

Definition 4: The $SCALE_{ORG}$ is the scale of organization. The *weight* can determine the class that an object belongs to. If $weight_{object} > \lambda$, class=c, we can get $object_{decision} = c$ ($\lambda > 0.6$). It can get by following equation: $SCALE_{ORG} = |ORG| / |DC_a|$, where a is the value of decision attribute in ORG.

Algorithm 2 Rules extract

Step1: $i \leftarrow 1$, $RULES \leftarrow \varnothing$;

Step2: If $i > |V_D|$, go to Step4; Otherwise, Two organizations ORG_{p1}, ORG_{p2} are randomly selected from P_i, according to Definition 3: $P_i \leftarrow (P_i / \{ORG_1, ORG_2\}) \cup ORG$, until $\forall ORG_1, ORG_2 \in P_i$ not satisfy the Definition 3, go to Step3.

Step3: Extract rule from the same Attribute set of each ORG in P_i. Compute $SCALE_{ORG}$, then $RULES \leftarrow RULES \cup rule$; $i \leftarrow i+1$, go to Step2;

Step4: Sort by the $SCALE_{ORG}$ and get $RULES$.

Where P_i, i=1,2,…, $|V_D|$ is the evolutionary result; *RULES* is the rules set;

In this section, we provide a definition of the intrusion detection model.

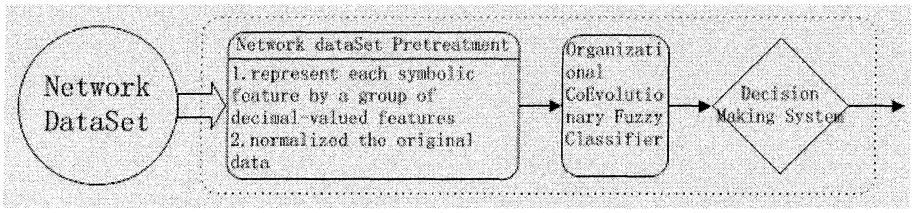

Figure 1: The IDS based on Organizational CoEvolutionary Fuzzy Classification

4. SIMULATION

In this paper, we use the dataset from original 10% KDD Cup 1999. To get a fuzzy dataset, where each numerical value in the dataset we normalized between 0.0 and 1.0 according to the equation: $x' = \ln(rx)/\ln(r\beta)$ $\alpha r \geq 1$, where x is the value of numerical attribute, β and α are separately the maximum and minimum value as the same attribute with x, r is a constant.

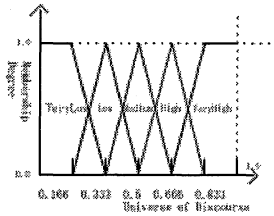

Figure 2: Fuzzy space for KDD-CUP99 dataset

For each continuous attribute we assign the fuzzy space shown in figure 2.

4.1. The Results of Simulation

We compared performance of OCEFC with the other methods. We use 20% randomly sampled as training data to evaluate the performance of a model. Another 40% randomly form the threshold determination set, which has no overlap with the training set. We use two performance measures in our simulation: FAR (false alarm rate) and TDR (true detection rate).

After running 100 times, we compare the average result based on OCEFC with other methods in *Table 1*. Ours method has better performance.

Table 1: Comparison of IDS performance based on OCEFC with others method

Method	OCEFC	GA	RIPPER-Artificial Anomalies [5]
FAR (%)	2.75	7.0	2.02
TDR (%)	99.23	93.14	94.26

5. CONCLUSION

This paper has investigated the use of OCEFC as one component of intrusion detection system. As the simulation shown, we achieve the better performance than others.

ACKNOWLEDGEMENTS

This work was Supported by the National Natural Science Foundation of China under Grant Nos. 60372045, 60133010; National High Technology Development 863 Program of China under Grant No. 2002AA135080.

REFERENCES

1. Liu Jing, Zhong Wei-Cai, Liu Fang, Jiao Li-Cheng. Classification based on organization coevolutionary algorithm. Chiense Journal of Computers, 2003, 26(4): 446-453 (in Chinese).
2. W. Lee, S. J. Stolfo, and K. W. Mok, "Mining audit data to build intrusion detection models", Proc. Int. Conf.Knowledge Discovery and Data Mining (KDD'98), pages 66-72, 1998.
3. S. A. Hofmeyr, A. Somayaji, and S. Forrest, "Intrusion detection using sequences of systems call", Journal of Computer Security, 6:151-180, 1998.
4. Zadeh, L..A.. Fuzzy Sets. Information and Control 8:338:353 1965
5. W. Fan, W. Lee, M. Miller, S. J. Stolfo, and P. K.Chan, "Using artificial anomalies to detect unknown and know network intrusions", Proceedings of the First IEEE International Conference on Data Mining, 2001.