Exploration and Research on Oil and Gas SCADA Security Defense Based on Dynamic Fuzzy Neural Network (DFNN)

Menghui Zhao[a], Xiedong Cao[b]
School of electrical and information, Southwest Petroleum University, Chengdu China
[a]smile_zmh@163.com, [b]cowyco@126.com

Abstract—This paper firstly analyses the security vulnerability of oil and gas SCADA systems. Combing the SCADA Security Defense Model (FSDM) based on factor neural network, the method to realize the function of the recognition module (the recognition module is a significant part of the IN neurons as the core of the FSDM) based on dynamic fuzzy neural network (DFNN) is proposed. The integration of dynamic character is to solve difficulty to make the SCADA system defensing rules which is hard for the domain experts. Fuzzy neural network can not only solve the fuzzy problems in malicious program judgment but also meet the real-time and rapid requirements in oil and gas SCADA system. As oil and gas SCADA security defense research is still in its infancy, dynamic fuzzy neural network which is proposed in this paper to solve oil and gas SCADA security defense problems provides a new solution for the future and lays the theoretical foundation.

Keywords-Oil and Gas SCADA; Security defense; Dynamic Fuzzy Neural Network; Factor space; Behavioral factors; Factors Neural Network

I. INTRODUCTION

In June 2010, "Stuxnet" virus Stuxnet began wantonly spread worldwide. "Stuxnet" virus Stuxnet computer can be said to be a revolutionary virus to industry, bringing new warning to industrial control systems security indicating “the Industrial virus age” has come. Oil SCADA system is a typical industrial control systems .So building the Oil and Gas SCADA system defense theory has important strategic significance.

II. SECURITY VULNERABILITY ANLYSIS OF OIL AND GAS SCADA SYSTEM

SCADA system security weaknesses highlight in the following four aspects:

(1) Operating system risk [1]. IPC in dispatching controlcenter operator workstation and the station control system mostly run operating system such as WINDOWS2000 and WINDOWS XP. In order to ensure the stablity of system ,IPC and industrial control servers which run WINDOWS operating system won’t install patches and any anti-virus software Microsoft releases on the original WINDOWS system and some even shut down firewalls after commissioning of system , which leads the system to a dangerous situation .

(2) Open industrial network protocols risk [2]. A large number of oil and gas pipeline‘s SCADA systems use open standard industrial protocol based on TCP/ IP, which may be taken advantage of by an attacker. By obtaining open *specification* documents and understanding the SCADA network protocols’ working mechanism and constructing control commands, an attacker can directly make threats to the normal operation of oil and gas pipeline SCADA system.

(3) System logging risk [1].Host computer system usually set different login user names and passwords depending on the different permissions, but the login password is set too simple and does not change for a long-term .,which leads system under attack by buried security risk.

(4) Illegally operational risk[1-2].During maintenance and commissioning work ,personal laptops access system for programming and debugging which appears frequently and workers may unconsciously use private removable storage devices like a U-disk with a Trojan or virus in the copping process of project backup .This kind of behavior can easily take the virus into the internal system.

From the four aspects analyzed above, obviously we can see the fact that security SCADA system is very fragile. Targeted attacks can be easily launched by hacker.

Successful attacking stories of domestic and foreign industrial control systems shows that unknown attacks (such as targeted attacks, APT attacks, new type vulnerability attacks) are the greatest threat industrial control system faces . The traditional defense system, based on signatures and rules matching technology, is a passive defense mode which can only deal with known attacks. Passive defense systems need frequent updating virus database that will adversely affect industrial systems but industrial systems for real-time and stability requirements are high . So the traditional passive defense system can not be applied to the oil and gas SCADA system. The SCADA security and defense based on fuzzy neural network this article to introduce is a proactive defense technology having the advantage against unknown attacks.

III. In Neurons Of SCADA Security Defense Model (FSDM) Perform Implementation Based On Factors Neural Network

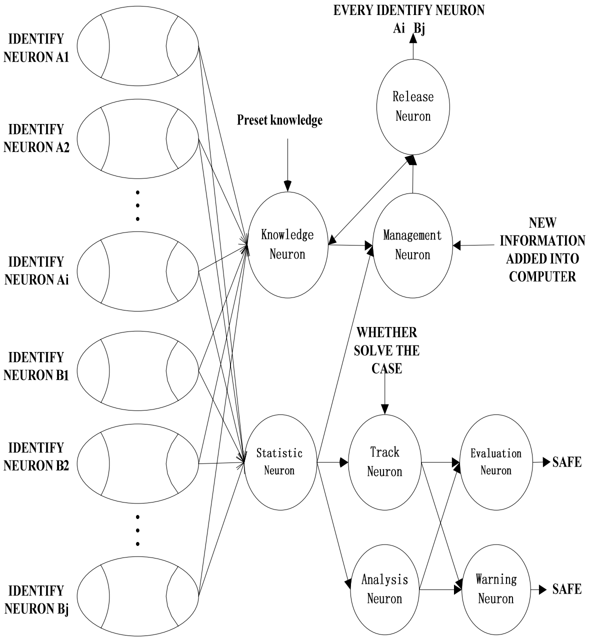

Figure 1. Oil and Gas SCADA Security Defense Mode based on FNN(FSDM)

Depending on the different functions of factor analytic neuron [3] in SCADA transport of oil and gas gathering security defense mode, the model neurons can be divided into nine categories. Wherein IN (Identify Neuron) is the core and the most basic class neuron of the entire mode and plays the most important role in the completion of security defense function. IN is configured in IPC and in host and console computers of engineer station and operator stations of all levels. IN is responsible for capturing program behavioral and extracting behavioral factors. By comparing with the factor knowledge base in IN, IN can judge whether the program is a malicious program and take appropriate action.

The internal structure of IN Neuron in SCADA Security Defense Model (FSDM) based on factor neuron network is as shown in Figure 2.

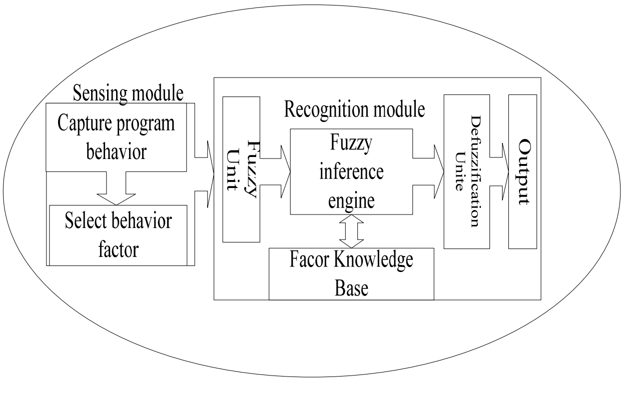

Figure 2. The internal structure of IN Neuron

IN neurons is composed of the sensing module and the recognition module .The sensing module is responsible for capturing program behavioral and extract the malicious program behavior factors as input of the recognition module. The recognition module based on Dynamic Fuzzy Neural Network method gives out the result of malicious judgment after the process of fuzzification, fuzzy reasoning and defuzzification. In this solution it is firstly necessary to select the appropriate characteristic behaviors of malicious program as behavioral factors which are inputs of the Identify Neurons. This article focuses on the recognition module of Identify neurons using the Dynamic Fuzzy Neural Network solution. The following analyze why using such a solution.

The process of analyzing whether a program is normal or malicious involves in a lot of fuzzy and uncertain concepts and features [4]. Malicious program judgment is not a simple two-phase problem (the results are only yes or no) but largely vague uncertainties. Usually when a program consists of multiple obvious malicious sequences of behaviors (we call a program behavior in this paper), then we can determine the program is malicious. A program behavior itself is normal or malicious is an ambiguity problem rather than certainty problems. Only in terms of the program behavior - a file is deleted- although relying on experience we know it has a high degree of maliciousness, but we can not conclude that it was a malicious behavior (unless some conditions exist, such as whether it is allowed by the administrator or to cause serious consequences, etc.). For another example a program behavior directory traversal of the files on the disk , for this behavior it will not cause direct harm to the system security and only take up a certain amount of system resources. But it is often used in malicious program code while the normal program code rarely uses, from this perspective, we can say that this program behavior has slight degree of maliciousness. While deleting the file this behavior may cause direct harm to system security, such as deleting important files in system folder leads the system can not operate normally. So from this point of view we can say deleting files and traversing the disk file both have a certain degree of maliciousness but deleting the file level has much higher degree of maliciousness than traversing the disk. So this is a vague question.

For the fuzzy and uncertain characteristics mentioned as above, we making the use of fuzzy logic judgment has obvious advantages. Advantages of fuzzy logic is that it is more suitable for expressing vague or uncertain knowledge, and the reasoning is close to human thinking pattern. On the other hand, Oil and SCADA System has high demand for real-time, so security defense software deployed in the system can not take up a lot of system resources and calls for the speedy handling of security threats. The neural network has a parallel structure, parallel structure determines it has a very strong advantage of rapidity which is benefit for increasing hardware utilization and conserving system resources. Considering the advantages above, SCADA malicious program determining based on fuzzy neural network can be a good solution to solve the ambiguity and uncertain problem but also to meet industrial real-time requirements. So we build execution neurons based on Fuzzy Neural Network. Taking into the complexity of the oil and gas SCADA system account, we use the method of dynamic fuzzy neural network to acquire knowledge the rules.

IV. Fuzzy Dynamic Neural Network Recognition Model

A. The Mathematical Description of Recognition Model

On the domain U define malicious behavior factor sets for representing the maliciousness of a executable program. If there are n types of malicious behavioral factors (n is a finite number) then we get the expression U={u_1, u_2 …… u_n };On the domain V define a malicious classification fuzzy sets based on the degree of maliciousness of a program. Assuming the number of the fuzzy subsets is p, then V={v_1, v_2, ⋯⋯ , v_p } .

As for all programs given to be analyzed the complete works is A. Program a (a ∈A) is to be analyzed. After analysis we obtain malicious behavioral factors sets M of program a then M={ u_{z1}, u_{z2}, ⋯⋯ u_{zk} ⋯⋯ , u_{zm} }□U (k<m≦n). Assuming any fuzzy subset v_i ∈V are on domain M ,define expression μ_{v_i} (μ_{v_i} ∈ [0 1]) is a membership function for subset v_i.For any behavioral factor u_{zk} (u_{zk} ∈ M), expression $\mu_{v_i}(u_{zk})$ represents the membership a behavior factor u_{zk} (u_{zk} ∈ M) belonging to a fuzzy subset v_i (v_i ∈ V). According Zadeh notation, subset v_i can be expressed as the following formula (1):

$$v_i = \frac{\mu_{v_i}(u_{z1})}{u_{z1}} + \frac{\mu_{v_i}(u_{z2})}{u_{z2}} + \cdots\cdots + \frac{\mu_{v_i}(u_{zm})}{u_{zm}} + \cdots\cdots + \frac{\mu_{v_i}(u_{zm})}{u_{zm}} \quad (1)$$

On domain V define fuzzy subset N is the result of malicious program classification of programs to be analyzed. And expression $\mu_N(v_i)$ represents the membership any subset v_i (v_i ∈ V) belonging to a fuzzy subset N. According Zadeh notation, subset N can be expressed as the following formula (2) [5-6]:

$$N = \frac{\mu_N(v_1)}{v_1} + \frac{\mu_N(v_2)}{v_2} + \cdots\cdots + \frac{\mu_N(v_p)}{v_p} \quad (2)$$

With the definitions above, a rule judging the degree of the maliciousness of a program can be expressed as IF-THEN implication relationship (3) as below:

$$R_s : M_{s1}, M_{s2}, \ldots\ldots, M_{sp} \rightarrow N_{s1}, N_{s2}, \ldots\ldots, N_{sp} \quad \omega_s \quad (3)$$

Wherein, the expression R_s and s all represent the r-th rule in the relationship (3). The expression $M_{si} = \sum_{k=1}^{m} \frac{\mu_{v_i}(u_{zk})}{u_{zk}}$ (1≦i≦p) represents the membership u_{z1}, u_{z2} ⋯⋯ u_{zm} (∈ M) in turn belonging to the fuzzy subset v_i (v_i ∈ V). $N_{sr} = \mu_N(v_r)$ (1≦r≦p) represents the membership v_r (v_r ∈ V)belonging to fuzzy subset N. The expression $M_{s1}, M_{s2} \ldots\ldots M_{sm}$ represents the rule antecedent, and the expression $N_{s1}, N_{s2} \ldots\ldots N_{sp}$ represents the rule post, the expression ω_s represents the weight of rule R_s .Therefore, the judgment of maliciousness of a program can be expressed as a mathematical description :

Rules:

$$R_1 : M_{11}, M_{12}, \ldots\ldots, M_{1m} \rightarrow N_{11}, N_{12}, \ldots\ldots, N_{1p} \quad \omega_1$$
$$R_2 : M_{21}, M_{22}, \ldots\ldots, M_{2m} \rightarrow N_{21}, N_{22}, \ldots\ldots, N_{2p} \quad \omega_2$$

……

$$R_t : M_{t1}, M_{t2}, \ldots\ldots, M_{tm} \rightarrow N_{t1}, N_{t2}, \ldots\ldots, N_{tp} \quad \omega_t$$

Given: $M_1^*, M_2^*, \cdots\cdots,\ M_m^*$

Solving: $N_1^*, N_2^*, \cdots\cdots,\ N_p^*$

B. Dynamic Fuzzy Neural Network Model (DFNN) and the presentation of dynamic

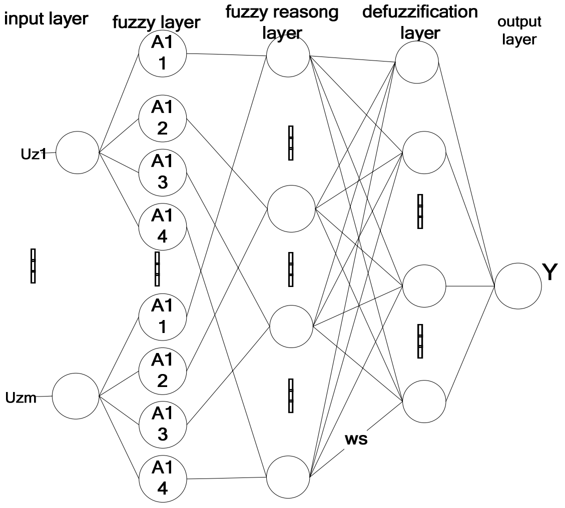

Figure 3. Network topology of DFNN

Network topology of DFNN is as shown in figure 3, which is based on expansion of RBF neural network. There are five layer structure: input layer, fuzzy layer, fuzzy reasoning layer, defuzzification layer and an output layer [7-8].

As shown in figure 3, Y is output variable, and A11 A12 A13 A14 represents a membership function. The third layer is fuzzy reasoning layer and each neuron represents one fuzzy rule, the fourth layer is defuzzification layer and each neuron represents a normalized node connecting to the right outputs. ω_s represents the weight of rule s .

Here is a brief introduction of the dynamic character of DFNN. Its dynamic property mainly reflects in the dynamically changing of network structure caused by the number of neurons in reasoning layer to increase. For this layer each neuron represents a fuzzy rule. Before the neural network trained, the number of neurons in the layer is 0. So the number of rules stored is 0. With fuzzy neural network being trained, the number of fuzzy rules increase one by one, followed the number of neurons in reasoning layer dynamically increase (dynamic performance). Since the SCADA system is a complex system, SCADA Security Defense research is in its infancy, experts can not make good defense rules. The integration of dynamic is mainly to solve difficulty to make the SCADA system defensing rules which is hard for the domain experts.

C. Map Malicious Programs Judgment Process to Dynamic Fuzzy Neural Networks

1) The first layer: the input layer:

It is necessary to select the appropriate program behaviors as behavioral factors as input of Indentify Neurons. Generally after the process of capturing program behavioral we will get a lot of program behaviors. Obviously, if all these behaviors as input of dynamic fuzzy neural network, the number of input nodes will be very large, it is not conducive to the training of the network. Therefore, we must reduce the dimensions. Malicious behavior factors selection process is reducing dimension process.

That we extract behavioral factors is to establish a generalized coordinate system, every behavior factor is a dimensional coordinate system name[9], any program to be analyzed after capturing behavioral factors can be described as a point in generalized coordinate system. According to the factor space theory, factors can be divided into four types: 1 variable type 2 Symbolic type 3 Switching type 4 extent type[9] .According to defense requirements, different defense rules and so on , malicious behavior factors can have different selection methods. But no matter what kind of behavioral factors selection method we choose, behavioral factors have to fully reflect the maliciousness of the program. If traditional IT systems and Oil and Gas Gathering SCADA Systems as a defense object, there is a big difference in choosing factors. Therefore, Oil and Gas SCADA System should choose targeted behavior factors. The behavioral factors proposed by this paper as inputs are the switching type and extent type behavior factors. Switching behavior factors is composed of two states Yes or No. For example, "appearance of modifying the SCADA system privileges" this behavior factor are composed of two states, appear and does not appear, you can use the values 1 and 0 respectively represent them [10]. Behavioral factors such as "the danger and important degree of files deleted and modified" as behavioral factors, ranging between state space [0 1]. It can be determined by expertise. The number of neurons of input layer is the number of malicious behavior factors dimensions.

2) The second layer: fuzzy layer

The number of neurons in fuzzy layer is determined by the number of the fuzzy sets. For example, if domain A is sets of all programs, the fuzzy subset F (A) can be selected as follows: high malicious program A11, low malicious programs A12, suspicious programs A13, normal procedure A14 .Then in fuzzy layer there are 4 nodes corresponding to each node in input layer. This paper select Gaussian membership (4) function to fuzzy the input.

$$\mu_{ij}(u_{mi}) = \exp[-\frac{(u_{mi}-c_{ij})^2}{\sigma_j^2}] \qquad (4)$$

Wherein $i = 1,2,\cdots\cdots$ m $j = 1,2,\cdots\cdots$ t "μ_{ij}"is the j-th membership function of the i-th input variable. "c_{ij}"is the center of the membership function , "σ_j" is the width of the membership function, m is the number of input variables, "t" is the number of membership functions, also indicates the total number of fuzzy rules [11].

3) The third layer: fuzzy inference layer

Each node of fuzzy inference layer represents a fuzzy rule, the equivalent of a fuzzy rule's IF- part . The output of this layer represents the trigger right of a rule. The output the j-th rule is as the membership expression (5) below:

$$\Phi_j = \exp[-\frac{\sum_{i=1}^{m}(u_{zi} - c_{ij})^2}{\sigma_j^2}] \qquad (5)$$

4) The fourth layer: defuzzification layer

Defuzzification layer also known as normalization layer, achieve normalized calculation as membership expression (6) below:

Input: φ_j

Output:

$$\overline{\Phi_j} = \frac{\Phi_j}{\sum_{k=1}^{t}\Phi_k} \qquad (6)$$

5) Fifth layer: output layer

Each node represents a variable output, the output is the superposition of each input, and there is only one output as the membership expression (7).

$$Y = \sum_{j=1}^{t} \omega_j \bullet \overline{\varphi_j} \qquad (7)$$

Where in Y is the output variable, ω_j is the weight of the j-th rule.

Dynamic fuzzy neural network's learning algorithm mainly including generating the fuzzy rules, pruning techniques of fuzzy rules, determining the premise parameters and the weight and so on [11]. On the above theory we can design and train a fuzzy neural network and then we get the fuzzy rules and put them into neural factors repository. Deploy Fuzzy neural network trained to the Indentify Neurons in connection with the perception module can complete the malicious judgment of unknown programs.

V. Conclusion

This paper analyzes the security status of the modern Oil and Gas SCADA system. Based on the Factor SCADA Security Defense Model (FSDM), dynamic neural network model of Identify Neuron was constructed to judge malicious program. The idea of Dynamic Fuzzy Neural Network (DFNN) applied to industrial SCADA systems is to meet the real-time and anti -unknown attacks requirements. Furthermore it provides a new method for SCADA security and defense system.

Acknowledgment

The paper is supported by National Natural Science Foundation Project (Grant No. 61175122).

References

[1] Guofu Wei.Discussion of long gas pipeline SCADA system information security issues[J]. Technology innovation and application, 2014, 36:12-13.

[2] Qi Xiong, Zhonghua Dai, Yong Peng, Weisheng Yi, Ting Wang. The exploration of Security Risk Analysis and protective Framework of oil and gas pipeline SCADA system network[J]. communication technology, 2014, 08:919-924.

[3] Zengliang Liu.Study on factor neural network theory and application[M]. Guiyang:Guizhou Technology Press, 1994:17-36.

[4] Feng Yue. Key technology research on malicious program behavior judgment based on dynamic fuzzy neural network[D]. The PLA Information Engineering University, 2010.

[5] Xiedong Cao. Fuzzy Information Processing and Application[M]. Beijing: Science Press, 2003.

[6] L.A.Zadeh..Fuzzy Sets[J]. Information and Control, 1965, 8:338-365.

[7] ShiQian Wu, Jun Xu. Dynamic fuzzy neural network design and application[M]. Beijing: Tsinghua University Press, 2008. 1-36, 55-66, 107-121.

[8] Wu S Q, Er M J. Dynamic Fuzzy Neural Networks: A Novel Approach to Function. Approximation. IEEE Trans Syst, Man, Cybern. Part B. 2000. 30:358-364.

[9] Chunxia Guo, Zengliang Liu, Zhinan Zhang, Yuan Tao.Knowledge factor space model of cyberattacks [J]. Telecommunication Engineering, 2009,10: 11-14.

[10] Hongxing Li.Mathematical framework of factor space theory and knowledge representation (Ⅰ) —— axiomatic definition and description of factor space [J]. Beijing Normal University (Natural Science), 1996, 04: 470-475.

[11] Rongrong Mei. Application and Research on Dynamic Fuzzy Neural Network [D]. Jiangnan University, 2011.