A New Security Model for Oil and Gas SCADA

Li Yang
School of Computer Science, Southwest Petroleum University, Chengdu, China
Ling Wang, Xinyu Geng, Xiedong Cao
Southwest Petroleum University, Chengdu, China
scxdyl@126.com

Abstract. To improve the ability of security defense in supervisor control and data acquisition (SCADA) for oil and gas field, a new security model for oil and gas SCADA based on factor neural network was presented. Firstly, definition of factor and factor Space were given. Then, formal description of knowledge factor expression, analysis factor neuron and analog factor neuron were described. Finally, experimental results indicated that the proposed model can effectively improve recognition rate of different attacks, which verify the effectiveness of the model and the algorithm, which lays the foundation for the research of the simulation method

Keywords: SCADA; Factor space; factor neuron; Knowledge expression

1 Introduction

Supervisor control and data acquisition (SCADA) systems are widely used in various fields of electric power, water, petroleum, chemical industry, metallurgy, transportation etc. Once the systems have security vulnerability, Industrial production and national economic security will be faced with serious threats[1-2]. In the past, SCADA systems were regarded as a relative isolate, safe system which has strong access control ability. Nowadays, Now a lot of SCADA system has been connected to the Internet enterprises, which become relatively open and transparent, which makes the SCADA system faced with many security problems, such as malicious virus, information leakage and tampering, and etc.. In recent years, attacks on industry control system are becoming more common[3-4]. For example: in 2010, "the network super weapon" Stuxnet virus went through the invasion of ICS system, and threatened seriously to the safe operation of Iran Bushehr nuclear power plant and nuclear reactor; In 2011, hackers entered supervisory control and data acquisition system SCADA, so that the water supply pumps of USA Illinois city water supply system were destroyed[5-6]. Therefore, strengthening the security of SCADA system has become a problem that can not be neglected in industrial control, security problem of SCADA system has been attached more and more attention.
2 Factor and Factor Space

Factors are basic elements which can describe objects, such as properties of objects and conditions in rules [7-8]. The meaning of the factor can be understood from the following three aspects:

(1) Attribution: it has two meanings. The first is that when looking for reasons from the results, factors are defined as objects cause some results. The second is that while the name is selected by state or feature, the factors will be symbols of a kind of state or a set of features [9-10].

(2) Analyticity: factors can be understood as a way to resolve the real world, an object can be described from different aspects in a different way, and the analysis process is the process of looking for factors. Factor is common character of an object, such as age, height, profession of a person.

(3) Descriptive: everything is the intersection of the various factors, a person can be identified from age, sex, height, weight, profession, educational history and etc., and a person is an intersection of the factors. And this means that it can build a broad cross-coordinate system, such objects can be described as a point of the generalized coordinates, and factor is the name of the dimension of the generalized coordinates.

3 Knowledge Factor Expression

[Definition 1] In the domain of U, the atomic model of knowledge factors is a triple,

\[ M(o) = < o, F, X > \]

Where \( o \) is a set of objects of the knowledge description about \( U \).
\( F \) is a factor set when \( U \) is used to describe \( o \).
\( X \) is a state set about \( F \) when \( F \) is used to describe \( o \), and

\[ X = \{ X_o(f) \mid f \in F, o \in O \} \]

[Definition 2] In the domain of U, the relation of knowledge mode is defined as

\[ R(O) = < RM, M(O), XM > \]

Where \( RM \) is a knowledge model.
\( M(O) \) is atomic model of knowledge representation in knowledge model.
\( XM \) is structure group state and state transformation relation of the atomic model \( M(O) \) in \( RM \).

The atomic model of the knowledge factor representation gives a discrete set that describes objects; this is the basis of knowledge representation with factors. The relation mode of knowledge factor representation can associate with various related knowledge or different knowledge representation; this can realize the transformation of the different ways of knowledge and knowledge reasoning. They provide the basis of representation and processing of knowledge in using factors neural network.
4 Formal Description of Analysis Factor neuron

An analysis factor neuron model can be described as follows:

\[ M = \{ \langle O,G \rangle, F, X \rangle, \langle P,Q,R \rangle, \langle A,B \rangle \} \]

Where O is a set of objects in the network system; G is the structure relation in the network; F represents cognition and description factor sets is state space of factor set in the network; O,G,F,X determine the state and structure of the system together; P,Q,R is respectively reasoning, judgment and control rule set. They together complete main independent operations and control functions;

A is input from outside and B is the target or response of information processing.

As a network consists of many neurons, an analysis factor neuron with reasoning function can be rewritten as:

\[ M_i = \{ \langle G_i,F_i,X_i \rangle, p,q,r \rangle, \langle a,b \rangle \} \]

Where \( G_i,F_i,X_i \) together describes the structure, factor and states of factor neuron.

\( p,q,r \) respectively implements the reasoning, judgment and inner control function of factor neuron; \( a \) is input information; \( b \) is the target or response of factor neuron reasoning.

5 Formal Description of Analog Factor Neuron

As shown in Figure 1, in the network, there is a controllable series-parallel connection network which consists of many mini-cells. \( f_1, ..., f_m \) are input factors relate to \( o \), each factor is called a perceptible channel of analog factor neuron. \( g_1, ..., g_n \) are output factors relate to \( o \), they represent different output response.

Let 

\[ F_a = \{ f_1, f_2, ..., f_m \} \]

\[ G_a = \{ g_1, g_2, ..., g_n \} \]

\[ X_a(F_a) = \{ X_a(f_i) \mid i = 1, ..., m \} \]

\[ Y_a(G_a) = \{ X_a(g_j) \mid j = 1, ..., m \} \]

\[ \text{Fig. 1. Analog factor neural network structure} \]

For analog factors neuron, its external function can be expressed as:
Simulation experiment on SCADA network attack and defence

We collected 494,021 records by tcpdump tool. Each record contains 41 feature values and 1 attack type description. With the knowledge representation theory of factor space, we got the following attack type factor space set. \( O = \{ \text{all attacking behaviors} \}, F = \{ \text{data link feature set, the attacking type} \} = \{ F_1, F_2 \}, F_2 = \text{abnormal behavior type} \), for \( F_2 \) is able to generate \( G = \{ \text{PROBE, DOS, U2R, R2L} \} \), so you can build the following factors space canes according to factor space canes theory:

Therefore, we can establish reasoning rules from \( F_1 \) link factors to \( F_2 \) abnormal type factors, these rules have shown the corresponding relationship between connection factor state vector sets and abnormal type state vector sets, in addition, on the basis of this, we can establish deduction matrix to make

\[
Y \cdot \left( G_n \right) = R \left( X \cdot \left( F_n \right) \right)
\]

When the system inputted a set of connection state \( X'_{F1} \), according to \( R \), the computer can infer as follows: \( X'_{F1} \cdot R \rightarrow X'_{F2} \).

when analysis factor neural network can not definitely infer the results, then system starts analog factor neural network to detect. Analog factor neural network has 41 input nodes, and 41 connection factors. Hidden layer uses 80 nodes, output layer uses 3 nodes. Initial learning rate is set to 0.5, by training 6000 times, as a result, the accuracy rate of using analog factor neural network is higher than not. The proposed model has improved detecting efficiency, reduced miss rate and false rate.

Conclusions

Analyzing factor, factor space, attack factors, skill and attack aim, this paper gives a new knowledge factor expression model of SCADA network attack and defence system, presents factor neuron model unrelated to time based on factor space of network attack and defense and proposes factor neuron model based on variable weight. The experimental results further verify the valid of our proposed method, which has a higher true recognition rate than other methods. Factor neuron model based on factor state space is a new modeling approach which can be used in factor analysis and expression of network attack and other fields. We will make research on a new factor space reasoning model and discuss equivalence partitioning and associating mechanisms in an analysis factor neuron in the next work.
References