Journal of Theoretical and Applied Information Technology
31st December 2012. Vol. 46 No.2

ISSN: 1992-8645 www.jatit.org E-ISSN: 1817-3195

A KNOWLEDGE EXPRESSION METHOD OF SCADA NETWORK ATTACK AND DEFENCE BASED ON FACTOR STATE SPACE

[1, 2]LI YANG, [1]XIEDONG CAO, [1]XINYU GEN, [3]JUN ZHANG

[1]School of Computer Science, Southwest Petroleum University, Chengdu 610500, Sichuan, China

[2]College of Computer, Sichuan University, Chengdu 610065, Sichuan, China

[3]Department of Computer Science, University of Kentucky, Lexington 40506-0633, USA

ABSTRACT

To solve the problem of knowledge expression in SCADA network attack and defence system, a new knowledge expression method of SCADA network attack and defence based on factor state space is presented. Combined with factor space definition and the formal description of factor state space expression and analyzing attack factors, skill and attack aim, analysis and expression method of network attack and defense factors is developed. On the basis of analysis and expression of network attack and defense factors, equivalence class partition formal description is presented. For illustration, an attack simulation experiment is utilized to show the feasibility of the proposed method in solving network attack and defence knowledge expression. By introducing factor space canes, objects class can be represented. Empirical results show that our proposed method can effectively improve accuracy of attack classification. Knowledge expression method based on factor state space can effectively solve complexity of knowledge expression in network attack and defence system and provides a new method for solving similar application.

Keywords: *Factor; Knowledge Expression, Network Attack, Factor State Space, Factor Space Canes*

1. INTRODUCTION

SCADA systems are widely used in industries of petrochemistry, electric power, pipeline and etc. In early period of SCADA development, SCADA is a relative physical isolation system which is relative secure and have a strong capability of access control. In recent years, with the development of computer technology, network technology and communication technology, SCADA products are mainly used the open standard protocols, TCP/IP protocols and Ethernet technology are widely applied to SCADA systems so as to not only make systems achieve a better compatibility, but also make the integration and expansion of different systems become easier[1, 2]. Today, SCADA system has gradually developed into an open transparent standard system so that it is easy to maliciously attack against SCADA network. As a result, the key infrastructure, SCADA system, will be interfered or destructed which brings more security issues. Therefore, it is more important to study SCADA network security defence theory and build SCADA network security defence architecture to protect nation infrastructure, SCADA system, and it has important strategy significance [3, 4].

With the analysis of the composition and network topology structure of SCADA system, attack and defense of SCADA system can be taken as online digital intelligent antagonizing process and all reasoning judgment, thinking and expression in attack and defense, that is to say, it can be abstracted and established into a corresponding and equivalent network attack and defense knowledge system in practice. By the introduction of factor state space, a new knowledge expression method of SCADA network attack and defence based on factor state space is presented to provide base frame for building SCADA defence system [5, 6].

The remainder of the paper is organized as follows. We review the relevant basis theory on factor space in Section 2. Section 3 then describes analysis and expression model of network attack and defense factors. Equivalence class partition formal description of objects is given in Section 4. In Section 5, formal description of analog factor neuron is described. And empirical result is discussed in Section 6. Finally, we conclude our paper in Section 7, and provide suggestions for future work.

ISSN: 1992-8645 www.jatit.org E-ISSN: 1817-3195

2. THE BASIS THEORY OF KNOWLEDGE BASED FACTOR STATE SPACE

2.1 Factor

As a vocabulary of the factor space theory, factor has three meanings as follows. The first is that when looking for reasons from the results, factors are defined as the things which cause some results. While we understand factor concept from state or feature, the factors are symbols of a kind of state or a set of features [7, 8]. The second is analyticity, factors can be regarded as a way to resolve the real world, a thing can be described from different aspects in a different way, and the analysis process is the process of looking for factors. The third is descriptive; everything is the intersection of the various factors, which means that it can build a broad cross-coordinate system. Such system can be described as a point of the generalized coordinates, and factor is the name of the dimension of the generalized coordinates [9, 10].

2.2 Knowledge Factor Space Expression

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

$$M(o) =< o, F, X > \quad (3)$$

Where ○ is a set of objects of the knowledge description about U.

F is a factor set when U is used to describe ○.

X is a state set about F when F is used to describe ○, and

$$X = \{X_O(f) \mid f \in F, o \in O\} \quad (4)$$

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

$$R(O) =< RM, M(O), XM > \quad (5)$$

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.

3. EXPRESSION OF NETWORK ATTACK AND DEFENCE FACTORS

An object is described as a network attack and defence system, and the main purpose of this paper is to build a knowledge network. We want to use this knowledge to constitute an integrated analysis model, which builds an organizational structure of the network attack and defence systems, behavior rules information together to make itself become knowledge representation model and knowledge application model. A feasible way is as follows: factors are used to build the structure of knowledge networks, to organize and package description, declarative and procedural knowledge, relationship factors in the structure relationship slot are used to construct various relationship chains in knowledge networks, so that the whole knowledge system becomes a network system in which a framework is a node, and a relationship chain is a boundary relationship factor.

Let U be the considered domain, SA is considered as an offensive and defensive SCADA system. A real-world SCADA network system consists of a number of different types of subsystems, attributes and their relationships. According to different perspectives, an offensive and defensive SCADA system can be described as follows:

$$(SA = \{ds, as, fs, ys \mid ds \in D, as \in A, fs \in F, ys \in Y, s \in S\} < s >) \quad (6)$$

SA is an offensive and defensive SCADA system, $S = \{s\}$ is various cognition and description set, where s is described as a viewpoint of cognition and description.

$A = \{as\}$ is a describable explicit attack factor set according to the viewpoint of s. {as} includes various representable existence condition and attack function.

$D = \{ds\}$ sets the structure of the system, it expresses various behavior pattern of relationships, includes relationship patterns between the state space, behavior patterns, state transition pattern of relationships and constraints, and so on.

It expresses inference relations of condition-function-result between a variety of explicit knowledge attack factors, such as partial order relationship, the same source relationship and the same result relationship.

$F = \{fs\}$ is a set of factor state space according to the viewpoint of s.

$fs = (f+, aer, f-)$ are used to represent attack promotion state, attack inhibition state, state of the

ISSN: 1992-8645 www.jatit.org E-ISSN: 1817-3195

attacker. In $f=(fl,uspw,pc,sev,io,nc,sc)$,fl represents a file variable, uspw state is user state, pc is a process parameter, sev is system service, io represents input/output parameter, nc is the network connection parameter, sc is the system environment variable.

$Y=\{ys\}$ is function state space.

For example, the state space of fl contains search, upload, download, modification, deletion, and so on. $fl=\cup fli(i=1,...,n)$ represents file type. Fl is a factor state space. In order to express offensive and defensive SCADA network (SA). Need to summarize and abstract SA. For example, things has the same characteristics, which obeys and abides by the rules of the specific acts, the same things are abstracted as a cognitive "object", similar entities that have specific traits are abstracted as common factors. Offensive and defensive behavior in the specific description of the network performance factors are described as a state factor.

4. EQUIVALENCE CLASS PARTITION OF OBJECTS

In offensive and defensive SCADA network systems, we need to classify things on certain viewpoint. Within the domain U, S is selected as a cognitive viewpoint, and according to the viewpoint of S, {as} will be classified in accordance with an "equivalence relation".

[Definition 1] Let A be a set, R is a relationship on A, and R is the view of S under an equivalence relationship on A. According to the viewpoint of S, for x, y, z ∈ A, R (s) satisfies

(1)Reflexive $xR(s)x$

(2)Symmetry If $xR(s)y$,then $yR(s)x$

(3)Transitivity If $xR(s)y$, $yR(s)z$ then $xR(s)z$

Under the viewpoint of S, an equivalence relation on R (s) is usually denoted by

$$(x)R(s)(y) or (x)R(s)(y)(x,y\in A) \quad (7)$$

[Definition 2]If $a\in A, o=\{y\mid (y)R(s)(a)\}$,then equivalence relationship R (S) under A and O is called an equivalence class.

[Definition 3]If o is equivalence relation on R (s), o is an abstract object for system SA in the viewpoint of s, set O＝ {o|o is an abstract object for system SA in s viewpoint}, where O is an object set for system SA in the viewpoint of s.

[Definition 4] Let Ai be subset of A, if A = ∪ Ai, {Ai} is called as a division level of A; If Aα∩Aβ=φ (α≠β, A α ,Aβ∈A),{Ai} is called as a deterministic division of A.

In the network attack and defense system, the division of the object reflects a kind of cognition and description viewpoint of things. This viewpoint is relate to the levels and considering problems angle, different division degree is sometimes called the particle size of the classification .

[Definition 5] $R(s1),R(s2)$ is equivalence relation of two different object divisions of the classification A. If $(x)R(s1)(y)\rightarrow(x)R(s2)(y)$, $R(s1)$ is fine than $R(s2)$, then $R(s1)<<R(s2)$.

A network attack and defence system SA, when one of the behaviors is divided into different size parts, the relationship between objects, the factors of the system and the system state will also change. The system can have a variety of different partition,produce all kinds of different particle size of the objects. By the introduction of different kinds of system structure and factors description, we can create all kinds of different cognition and description model. Some common characteristics in the model are as follows:

(1)The model should have a hierarchical relationship.

(2)In the same level, different aspects of each model can be merged into a comprehensive model.

(3)The nature of the model can not change between different levels. For instance, if the original system is topology structure, the topological properties in different levels of the model shall remain unchanged, if the original system is partial order structure, it also should have partial order in the various models.

[Definition 6] $M=<<O,G>,F,X>SA$ is a cognition or description model in the viewpoint of S.

If $O=\{O\}$ < object set in M >, where

O is equivalent clustering of objects in SA in the viewpoint of S.

$G=\{G\}$ < structure- called of M >

G is equivalent transformation of SA in relationship d in the viewpoint of S.

$F=\{f\}$ < cognition or description factor sets of M >

F is cognition and description factor of SA in the viewpoint of S

$X=\{X\}$ < factor expression state set of M >, and X is a factors state set of SA in the viewpoint of S.

[Definition 7] $SA=(<A,D>,Fo,Y)$ is a practical system, A ', A have H property. In the viewpoint of S, we can obtain the following formula from SA model system.

Journal of Theoretical and Applied Information Technology
31st December 2012. Vol. 46 No.2

ISSN: 1992-8645 www.jatit.org E-ISSN: 1817-3195

$$M =<< O, G >, F, X > \quad (8)$$

Make s: $s : A' \to s(A') \in M$,

S (A ') also has H property; M has H property of SA.

[Definition 8] S1 and s2 set are two different cognition and description viewpoint of SA,

Abstract cognition and description models are as follows:

$$M1 =<< O1, G1 >, F1, X1 > \text{ and} \quad (9)$$

$$M2 =<< O2, G2 >, F2, X2 > \quad (10)$$

If $s3 = s1 \quad s2$, then

$$M3 = M1 \quad M3 =<< O3, G3 >, F3, X3 >$$

$$=<< O1, O2 \quad G1 \quad G2 > F1 \quad F2, \quad X1 X2 > \quad (11)$$

If R is equivalence relationship, then

$$\forall x, y \in A(x) R(s1 \quad s2)(y) \leftrightarrow (x) R(s3)(y) \quad (12)$$

5. FORMAL DESCRIPTION OF FACTOR NEURON

5.1 Formal Description of Analysis Factor Neuron

An analysis factor neuron model can be described as follows:

$$M = \{<< O, G >, F, X >, < P, Q, R >, < A, B >\} \quad (13)$$

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;

X 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, judgement and control rule set. They together complete main independent operations and control functions;

A is input information 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:

$$Mi = \{< Gi, Fi, Xi >, < p, q, r >, < a, b >\} \quad (14)$$

Where $< Gi, Fi, Xi >$ together describes the structure, factor and states of factor neuron;

P, q, r respectively implements the reasoning, judgement and inner control function of factor neuron;

A is input information; b is the target or response of factor neuron reasoning.

5.2 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. F1,…,fm are input factors relate to o, each factor is called a perceptible channel of analog factor neuron. G1,…,gn are output factors relate to o, they represent different output response.

Formal description of analog factor neuron is as follows:

$$F_o = \{ f_1, f_2, ..., f_m \} \quad (15)$$

$$G_o = \{ g_1, g_2, ..., g_m \} \quad (16)$$

$$X_o(F_o) = \{ X_o(F_i) \mid i = 1, 2, ..., m \} \quad (17)$$

$$Y_o(G_o) = \{ X_o(g_j) \mid j = 1, 2, ..., m \} \quad (18)$$

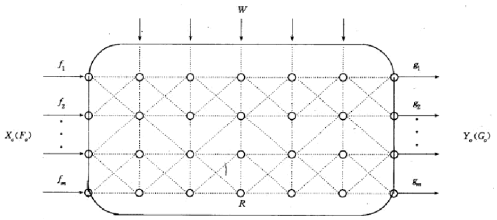

Figure 1: Analog Factor Neural Network Structure

For analog factors neuron, its external function can be expressed as:

$$Y_o(G_o) = R(X_o(F_o)) \quad (19)$$

5.3 Formal Description of Combination Factor Neurons

If combination neural network system M is composed of N independent units or subsystems, then

M is denoted as

$$M = \{M_i\} (i = 1, 2, ..., N) \quad (20)$$

Let the state of subsystem M_i be X_i, function state X can be made up of the various subsystems function state set Xi vector, that is to say:

$$X = \{X_1, X_2, X_3, ..., X_N\} \quad (21)$$

A general order of factor neural network system can be realized by the weighted and controllable connections of various subsystems, Let R = $\{r_{ij}\}$ be the relationship between the various subsystems, then:

$$r_{ij} = W_{ij} \bullet e_{ij} \quad (22)$$

Where e_{ij} is the connection relationship of function and structure between subsystems. W_{ij} is a transformation parameters of controllable function connection in the system.

Combination factor neural network subsystems set $\{M_{ij}\}$, subsystem state set $\{X_{ij}\}$ and controllable connection $R = \{r_{ij}\}$ together make up system structure.

The system function realization is reflected by system dynamic equation.

$$U = F(t, X(t), I(t)) \quad (23)$$

ISSN: 1992-8645 www.jatit.org E-ISSN: 1817-3195

Where $X(t)$ is the system's overall state; $I(t)$ is input states form outside; $F()$ is state mapping function of the system.

When N systems make up a factor neural network system, state function dynamic equation of each subsystem is expressed as:

$$U_i = h_i(t, X_i, I_i + \sum g_{ij}(t, X_i, I_j))(i = 1,2,...,N) \quad (24)$$

Where X_i, X_j is respectively current state of subsystem i and j;

I_i, I_j is respectively current input information of subsystem i and j;

H_i is feedback state mapping function of subsystem i;

G_{ij} is state-change effect mapping function from subsystem j to subsystem i.

6. EXPERIMENTAL VERIFICATION

The experiment uses KDD CUP99 to reason and verify knowledge expression model based on factor state space. With the knowledge representation theory of factor space, we can get O={all attacking behaviors}, F={data link feature set, the attacking type}={F1,F2}, F2=abnormal behavior type, for F2 is able to generate G={PROBE, DOS, U2R, R2L}, so we can build the following factors space canes according to factor space canes theory, the factors space canes is shown in Figure 2.

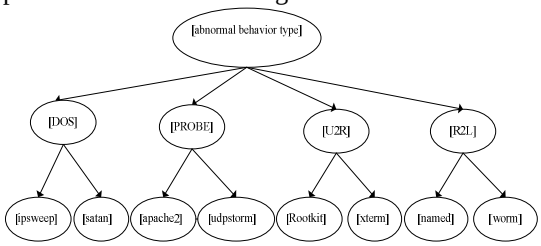

Figure 2 Factors Space Canes

If F2 V (abnormal type) (V means factors set), concerning "abnormal type" factors spaces, [abnormal type] {X(f)f ∈ F2}.F21 V(DOS)\V(abnormal type)(V means DOS factor set), F22 V(PROBE)\V(abnormal type), F23 V(U2R)\V(abnormal type), F24 V(R2L)\V(abnormal type), then [DOS] {X(f)}f ∈ F21, [PROBE] {X(f)}f ∈ F22,[U2R] {X(f)}f ∈ F23, [R2L] {X(f)}f ∈ F24.

F1 is a link factor set, and F1={TCP link basic features, TCP link content features , time-based network traffic statistical features, host-based network traffic statistical features}={F11, F12, F13, F14}, F11can generate 9 seed factors, such as G11={duration, protocol type, service,…}etc; F12 can generate 13 seed factors,F13 can produce 9 seed factors, F14 can generate 10 seed factors. Similarly, it can produce melon factor space canes.

Therefore, we can establish reasoning rules from F1 link factors to F2 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 generate

$$X_{F1} \bullet R \to X_{F2} \quad (25)$$

When the system inputted a set of connection state X'F1, according to R, the computer can infer as follows:

$$X'_{F1} \bullet R \to X'_{F2} \quad (26)$$

When analysis factor neural network can not definitely infer the results, then the system starts analog factor neural network to detect. Analog factor neural network has 9 input nodes, and 9 connection factors. Hidden layer uses 40 nodes, output layer uses 1 nodes. Initial learning rate is set to 0.3, by training 200 times. As shown in table 1, since percentage of PROBE and DOS in the dataset is high, the recognition rate of PROBE and DOS is high and its false rate and missing rate is relatively low, there is little different in recognition rate between combination factor neural network and analog factor neural network. As percentage of R2L in the dataset is low, recognition rate of combination factor neural network is higher than analog factor neural network. As a result, in general, our proposed model has improved detecting efficiency , reduced miss rate and false rate.

Table 1 Detection Rate Comparison Of Analog Factor Neural Network With Combined Factor Neural Network Under Different Type Attacks

Type		Accurate rate	False rate	Missing rate
Normal	Analog	98.30	1.70	0.00
	combination	98.33	1.67	0.00
Probe	Analog	99.82	0.18	0.00
	combination	99.83	0.17	0.00
Dos	Analog	99.88	0.12	0.00
	combination	99.90	0.10	0.00
U2R	Analog	96.02	3.51	0.47
	combination	98.30	1.10	0.60
R2L	Analog	21.14	3.43	75.43
	combination	80.40	6.83	12.77

ISSN: 1992-8645 www.jatit.org E-ISSN: 1817-3195

7. EXPERIMENTAL VERIFICATION

Combined with factor space theory and the formal description of factor state space expression and analyzing attack factors, skill and attack aim, this paper presents a new knowledge expression model of SCADA network attack and defence system based on factor state space, gives an analysis and expression method of network attack and defense factors and proposes formal description of network attack and defence factors. Empirical results further verify the valid of our proposed method. Knowledge expression method 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.

ACKNOWLEDGEMENTS

This work was supported National Natural Science Foundation Project under grants 61175122, 61173182 and 61179071, as well as by Applied Basic Research Project(2011JY0124) and International Cooperation and Exchange Project (2012HH0004) of Sichuan Province.

REFERENCES:

[1] D.Vivo M, D.Vivo G O, I. G, "Internet security attacks at the basic levels", *Operating Systems Review*, Vol.32, No.2, 2002, pp.40-48.

[2] T. L,Z. Y,A. G, "Intrusion detection force:an infrastructure for internet-scale intrusion detection", *Proceedings of the First IEEE International Workshop on Information Assurance(IWIA'03)*, IEEE Conference Publishing Services,March 24,2003,pp.73-91.

[3] S.W, L.S T, C. K J, "A frame knowledge system for managing financial decision knowledge", *Expert Systems with Applications*, Vol.35, No.3, 2008, pp.1068-1079.

[4] X. Li, J. Chen, "Research on Knowledge Acquisition and Reasoning Mechanism of Pathologic Diagnosis Expert System", *Journal of Convergence Information Technology*, Vol. 7, No. 20, 2012, pp. 550-556.

[5] C. Chen, "A Usability Inspection Expert System based on HE, GRY and GST", *International Journal of Intelligent Information Processing*, Vol. 3, No. 1, 2012 ,pp. 1-15.

[6] C. Xiedong, et al., "The Geological Disasters Defense Expert System of the Massive Pipeline Network SCADA System Based on FNN", *Lecture Notes in Computer Science*, Vol.1, No.7234, 2012, pp.19-26.

[7] L. Yang,etal, "A New Formal Description Model of Network Attacking and Defence Knowledge of Oil and Gas Field SCADA System", *Lecture Notes in Computer Science*, Vol.1, No.7234, 2012, pp.2-10.

[8] G.Chunxia, L. Zeng-liang, M. Qing, "Network attack planning model and its generating algorithm", *Computer Engineering and Applications*, Vol.46, No.31, 2010, pp.121-123.

[9] S. Zhong, G. Xu, Y. Yang, W. Yao, Y. Yang, "Algorithm of Generating Host-based Attack Graph for Overall Network", *Advances in Information Sciences and Service Sciences*, Vol. 3, No. 8, 2011, pp. 104-110.

[10] L. Rui, "Computer network attack evaluation based on incremental relevance vector machine algorithm", *Journal of Convergence Information Technology*, Vol. 7, No. 1, 2012, pp. 245-252.