A Discriminant Model of Network Anomaly Behavior Based on Fuzzy Temporal Inference

Ping He
Department of Information
Liaoning police Academy
Dalian 116036, China

***Abstract*—The aim of this paper is to provide an active inference algorithm for anomalous behavior. As a main concept we introduce fuzzy temporal consistency covering set, and put forward a fuzzy temporal selection model based on temporal inference and covering technology. Fuzzy set is used to describe network anomaly behavior omen and character, as well as the relations between behavior omen and character. We set up a basic monitoring framework of anomalous behaviors by using causality inference of network behaviors, and then provide a recognition method of network anomaly behavior character based on hypothesis graph search. As shown in the example, the monitoring algorithm has certain reliability and operability.**

Keywords- network anomaly behavior; anomalous omen; fuzzy temporal; set covering; hypothesis graph.

I. Introduction

Network anomaly behavior monitoring is a hotspot in reaches of network security. Up till now the basic idea of network anomaly discriminant lies in anomaly detection method, provided by Denning in 1987[1]. That is to say, according to abnormality situations of audit statement in monitoring system, the bad behaviors (i.e. events of violating safety norms) in network can be detected. The most study on network anomaly detection is based on the theory of data analysis. For example, probability and statistical method [2], data mining method [3], artificial immune algorithm [4], and corresponding artificial intelligence method [5] and so on. But the above methods have common constraint condition-data completeness.

To a certain extent, this condition is implementable. Meanwhile, a superior false alarm rate exists relatively. So it can not satisfy the requirement about reliability of network monitor. Literature [5] introduced an analysis method based on knowledge diagnosis. Literature [6] put forward a method for anomaly detection based on direct inference. Literature [7] brought forward an extrapolation inference diagnosis model based on set covering (GSC). It is one of knowledge diagnosis models with many advantages, such as intuition, parallel, leading into heuristic algorithm easily. Probability causality was led into the model by Peng in literature [7，8].

In order to solve problems about anomaly detection for fuzzy network behavior, literature [9] combined intrusion detection model with fuzzy theory. Among the above mentioned methods, it had no consideration of causality between anomalous behaviors and data in network, as well as temporal constraint relationship with each other. Because data features come into being with network anomaly behaviors, the interaction discriminant method based on anomalous behavior-data is an active network monitor and defensive strategy.

In this paper, we propose a fuzzy temporal inference method, and describe inaccuracy for network temporal knowledge by using of fuzzy set, and then constitute a model of network temporal generalized behaviors covering (NTGSBC). Finally, discriminant results are satisfactory.

II. Network Temporal Generalized Behaviors Covering (NTGSBC)

A. Fuzzy Temporal of Network Behaviors

Fuzzy temporal analysis method has reached satisfied result in the research of fault diagnosis [3, 4]. In this section, we will set up an inference method for network anomaly behavior by using the analysis method in literature [4]. As we know, the main character of complexity in network behaviors is time fuzziness of its behavior state. In fact, occurring temporal of network behavior is an uncertain time based on interval transition.

It is a fuzzy period of time, so definition is as follows:

Definition 2.1 [4]. Network behavior happened in a network fuzzy time interval (N.F.T.I), suppose I be a trapezoid fuzzy number (T.F.N) defined on the network behavior time axis T， $I=(\theta_{\mathrm{t}},t_s,t_f,\theta_h)$, the membership degree $\mu_I(t)$ is as follows：

$$\mu_I(t)=\begin{cases}0,t<t_s-\theta_l\\(t-t_s+\theta_l)/\theta_l,t_s-\theta_l\le t\le t_s\\1,t_s\le t\le t_f\\(t_f+\theta_h-t)/\theta_h,t_f\le t\le t_f+\theta_h\\0,t>t_f+\theta_h\end{cases}\tag{1}$$

The start time in I is expressed by start (I), and it is defined as:

$$\mu_{start(I)}=\begin{cases}0, t<t_s-\theta_l\\(t-t_s+\theta_l)/\theta_l, t_s-\theta_l\le t-t_s\\1, t=t_s\\0, t>t_s\end{cases}\qquad(2)$$

The end time in I is expressed by end (I), and it is defined as:

$$\mu_{end(I)}=\begin{cases}0, t<t_f\\1, t=t_f\\(t_f+\theta_h-t)/\theta_h, t_f<t\le t_f+\theta_h\\0, t>t_f+\theta_h\end{cases}\qquad(3)$$

When $t_s=t_f=t$, I can be reduced to a network fuzzy time point(F.T.P), $I=(\theta_l,t,t,\theta_h)$, so it turns into a triangular fuzzy numbers(T.F.N). When $\theta_l=\theta_h=0$, I can be reduced to an accurate time. F.T.I and F.T.P are shown as Figure 1.

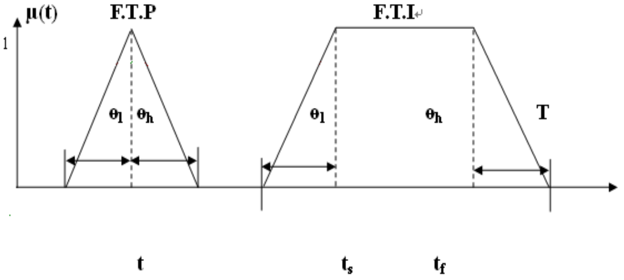

Figure 1. F.T.I and F.T.P

Fuzzy difference between two (T.F.N) reflect a kind of fuzzy temporal in the real network, and it exits

$$\mathrm{I}=\mathrm{I}_1\simeq I_2=(\theta_l^{(1)},t_s^{(2)},t_f^{(1)}-t_f^{(2)},t_f^{(1)}-t_s^{(2)},\theta_h^{(1)}+\theta_l^{(2)})\qquad(4)$$

B. Detection model for Anomalous Behavior

Definition 2.2 A discriminant model of network anomaly behavior is a relation mapping inversion space of data – behavior with fuzzy temporal, the formal representation can be shown as $P=<A,D,R,DEL,D^+,DOCT>$, where

(1) $A=\{a_1,a_2,...,a_m\}$, it is a non-empty finite set with anomalous behavior;

(2) $D=\{D_1,D_2,...,D_n\}$, it is a non-empty finite set with anomalous data;

(3) $R\subseteq A\times D$ is a relationship of network anomaly behavior defined on $A\times D$. $<a_i,D_j>\in R$, if and only if behavior a_i reflects data D_j in network, and $\forall D_j\in D$, there exists an element $a_i\in A$ at least, such that $<a_i,D_j>\in R$;

(4)DEL is a delay matrix on $|A|\times|D|$. For $<a_i,D_j>\in R, DEL(i,j)=(\theta_l^d, a^{(i,j)}, A^{(i,j)}, \theta_h^d)$ is a F.T.I. It expresses the delay time approximately “from $a^{(i,j)}$ to $A^{(i,j)}$” between “the beginning of anomalous behavior” and “the beginning of data D_j ”; for $<a_i,D_j>\notin R$, $DEL(i,j)$ has no definition.

(5) $D^+\subseteq D$ expresses a known anomalous data set of the discriminent target P. DOCT expresses |D| dimensional vector. For $D_j\in D^+$, DOCT $(j)=(\theta_l^m,t_s^{(j)},t_f^{(j)},\theta_h^m)$ is a F.T.I. It expresses the appearance time of the known anomalous data D_j approximately “from $t_D^{(j)}$ to $t_D^{(j)}$ ”. For $D_j\notin D^+, DOCT(j)$ has no definition. $cause(D_j)=\{a_i\mid a_i\in A,<a_i,D_j>\in R\}$ is an all possible anomalous behavior set caused D_j. $cause(D^+)=\bigcup_{D_j\in D^+}cause(D_j)$ is an all possible anomalous data set caused D^+.

$effect(a_i)=\{D_j\mid D_j\in D,<a_i,D_j>\in R\}$ is a data set caused by anomalous behavior a_i. $effect(A_I)=\bigcup_{a_i\in A_I}effect(a_i)$ is an all possible data set caused by $A_I\subseteq A$. For $A_I\subseteq A, D^+\subseteq D$, D_I is a covering of D^+, if and only if $D^+\subseteq effect(A_I)$. Network anomaly behavior $a_i\in A$ is relative to the known anomalous data $D_j\in D^+$, the start time of anomalous behavior $a_i\in cause(D_j)$ is as follows:

$$begin-OCT(a_i\mid D_j)=begin(DOCT(j))\simeq DEL(i,j)\qquad(5)$$

The end time of anomalous behavior a_i, relative to D_j, is as follows:

$$end-OCT(a_i\mid D_j)=end(DOCT(j))\qquad(6)$$

Definition 2.3 For $a_i\in A, D^{+(i)}\subseteq D^+$, $\{a_i\}$ is a temporal consistency covering on $D^{+(i)}$, if

(1) $\{a_i\}$ is a covering of $D^{+(i)}$.

(2) $\min(\max_{i\in T}\mu_{begin}(a_i \mid D^{+(i)})(t), \max_{i\in T}\mu_{end}(a_i \mid D^{+(i)})(t)) \geq \delta$ (7)

$$begin(a_i \mid D^{+(i)}) = \tilde{\bigcap_{s_j \in S^{+(i)}}} begin - OCT(a_i \mid D_j) \qquad (8)$$

$$end(a_i \mid S^{+(i)}) = \tilde{\bigcap_{D_j \in D^{+(i)}}} end - OCT(a_i \mid D_j) \qquad (9)$$

$0 \leq \delta \leq 1$is a threshold constant with temporal consistency, $\tilde{\bigcap}$ expresses intersection of fuzzy sets.

Different from literature [1,4], the solution of discriminant target in NTGSBC is not only pointed out anomalous behavior set A_I covered consistency data set D^+ , but also ensured $D_j \in D^+$ caused by $a_i \in A_I$, because of temporal consistency requirements.

Definition 2.4 Suppose a complete explanation of discriminant target P in network behaviors *be*

$Co - \exp(P) = \{(a_i, D^+(a_i)) \mid a_i \in A,\ D^+(d_i) \subseteq D^+\}$,

and satisfying the following conditions:

(1) For $(a_i, D^+(a_i)) \in Co - \exp(P)$, $\{a_i\}$ is a temporal consistency covering on $D^+(a_i)$.

(2) $\bigcup_{(a_i, D^+(a_i)) \in Pa - \exp(P)} D^+(a_i) = D_1^+$ expresses a known anomalous omen set.

$A_I = \{a_i \mid ((a_i, D^+(a_i)) \in Co - \exp(P)\}$ is called an explanation omen set of complete explaining $Co - \exp(P)$. Obviously A_I covers D^+.

Definition 2.5 Suppose a partial explanation of discriminant target *P* in network behaviors be

$Pa - \exp(P) = \{(a_i, D^+(a_i)) \mid a_i \in A, D^+(a_i) \subseteq D^+\}$,

and satisfying the following conditions:

(1) For $(a_i, D^+(a_i)) \in Pa - \exp(P)$, $\{a_i\}$ is a temporal consistency covering on $D^+(a_i)$.

(2) $D_1^+ = \bigcup_{(a_i, D^+(a_i)) \in Pa - \exp(P)} M^+(a_i) \subset D^+$,

and for $D^+ - D_1^+$,

$cause(D_2^+) \subseteq A_I = \{a_i \mid (a_i, S^+(a_i)) \in Pa - \exp(P)\}$.

D_2^+ is called a non-covering anomaly data set of $Pa - \exp(P)$, A_I is called an explanation anomaly behavior set of $Pa - \exp(P)$.

In the definition 2.5, due to $\text{cause}(D_2^+) \subseteq A_I$, it has $D_2^+ \subseteq effect(A_I)$, namely A_I is a covering of D_2^+ , so A_I is a covering of the whole known anomalous data set $D^+ = D_1^+ \cup D_2^+$. But there exists a constraint condition of temporal consistency, all the a_i in A_I only can explain (or cover) D_1^+ , a part of D^+.

A complete solution $Co - \exp(P)$ of discriminant target *P* is a complete explanation of *P*. and the cardinality of anomalous data set D_I , explained $Co - \exp(P)$, is minimum. A partial solution of *P* is a partial explanation $Pa - \exp(P)$, and the cardinality of non-covering anomaly omen set D_2^+ is minimum.

III. SOLVING DISCRIMINANT TARGET IN NETWORK BEHAVIORS

A. Hypothesis Graph

Solving problems of a discriminant target in network behaviors is based on a search method of hypothesis graph. Hypothesis graph is defined by network nodes and successor.

Definition 3.1 Suppose a node n_i be

$n_i = (A_I^{(i)}, D_1^{+(i)}, D_2^{+(i)})$,

in the hypothesis graph G(P) of discriminant target *P*, where

$$A_I^{(i)} \subseteq A, D_1^{+(i)} \subseteq D^+ \cap\ effect(A_I^{(i)}),$$
$$D_2^{+(i)} = D^+ - D_1^{+(i)}$$
$$n_i = (A_I^{(i)}, D_1^{+(i)}, D_2^{+(i)})$$
.

in G(P) is divided into three types:

1) Complete end node: for $n_i, D_2^{+(i)} = \phi$ 。

2) Partial end node: $n_i, D_2^{+(i)} \neq \phi$ *and* $cause(D_2^{+(i)}) \subseteq A_I^{(i)}$.

3) Non-end node: the other nodes except the above two types nodes.

Definition 3.2 *For the node* $n_k = (A_I^{(k)}, D_1^{+(k)}, D_2^{+(k)})$ in the hypothesis graph G(P), if it has $D_j \in D_2^{+(k)}, d_i \in cause$ $(D_j) - A_I^{(k)}$, then a successor node of n_k is as follows:

$$succ(n_k, a_i) = (A_I^{(k)} \cup \{a_i\}, D_1^{+(k)} \cup some(D_2^{+(k)} \cap$$
$$effect(a_i)), D_2^{+(k)} - some(D_2^{+(k)} \cap\ effect(a_i)))$$

Where $\{a_i\}$ is a temporal consistency covering on $some(D_2^{+(k)} \cap\ effect(a_i)) \subseteq D_2^{+(k)} \cap\ effect(a_i)$.

When constructed G (P), it began from an initial node $n_0 = (\phi, \phi, D^+)$, then continually expanded nodes and generated its successor node, until to the expansion node translated into non-expansion complete nodes or partial nodes. The following two theorems point out corresponding relationship in G (P) between a path and discriminant targets.

Theorem 3.1 *Suppose a path in hypothesis graph* G (P), from an initial node $n_0 = (\phi, \phi, D^+)$ to some complete end node $n_l = (A_I^{(l)},\ D_1^{+(l)}, D_2^{+(l)})$, be (without loss of generality) $n_0, n_1, ..., n_i, n_{i+1}, ..., n_l$, where n_{i+1} is a successor node of n_i, $0 \le i \le l-1, n_i = (A_I^{(i)}, D_1^{+(i)}, D_2^{+(i)}),$ then

$$D^+(a_k)) \mid a_k \in A_I^{(i+1)} - A_I^{(i)}, D^+(a_k) = D_1^{+(i+1)} - D_1^{+(i)},$$
$$0 \le i \le l-1\}$$

is a complete explanation $Co - \exp(P)$ of the target *P*, and the set of anomalous behavior explanation $Co - \exp(P)$ is $A_I = A_I^{(l)}$.

Theorem 3.2 Suppose a path in hypothesis graph *G(P),* from an initial node $n_0 = (\phi, \phi, D^+)$ to some partial end node $n_l = (A_I^{(l)},\ D_1^{+(l)}, D_2^{+(l)})$, be (without loss of generality) $n_0, n_1, ..., n_i, n_{i+1}, ..., n_l$, where n_{i+1} is a successor node of n_i, $0 \le i \le l-1, n_i = (A_I^{(i)}, D_1^{+(i)}, D_2^{+(i)}),$ then

$$\{S^+(a_k)) \mid a_k \in A_I^{(i+1)} - A_I^{(i)}, D^+(a_k)$$
$$= D_1^{+(i+1)} - D_1^{+(i)}, 0 \le i \le l-1\}$$

is a partial explanation $Pa - \exp(P)$ of the target *P*, and the set of non-covering anomaly omen $Pa - \exp(P)$ is $D_2^+ = D_2^{+(l)}$.

The proof of above two theorems is shown in appendix.

B. *Operating process of the discriminant system*

According to theorem 3.1and 3.2, the operating of monitor system is based on the search of hypothesis graph. If it exists complete end nodes in the final hypothesis graph G(P), then the complete explanation $Co - \exp(P)$ of the target *P* can be obtained. And then a minimum in $|A_I|$ is taken for a complete solution $Co - \exp(P)$ of the target *P* , where A_I is a set of explanation anomaly behavior $Co - \exp(P)$. If it only exits partial end nodes in G(P), then the partial explanation $Pa - \exp(P)$ of the target *P* can be obtained. And then a minimum in $|S_2^+|$ is taken for a partial solution $Pa - \exp(P)$ of the target *P*, where S_2^+ is a set of non-covering anomaly omen $Pa - \exp(P)$.

Solution procedure is as follows:

```
(1) Algorithm Solve-TGSC(A, D, R, DEL, $D^+$ , MOCT)
(2) Variable $n_i$ : node, $D_j$ : data , $a_k$ : anomalous behavior, table OPEN, table CLOSE
(3)Begin OPEN: = $\{n_0 = (\phi, \phi, D^+)\}$
(4) CLOSE: = $\phi$ : {Initializing table OPEN, CLOSE}
(5) While there are non-terminal nodes in OPEN do
(6) Begin $n_i$ : = POP(OPEN)
    {Removing non-end node $n_i = (A_I^{(i)}, D_1^{+(i)}, D_2^{+(i)})$ from table OPEN, and add to table CLOSE}
(7) $D_j := select(D_2^{+(i)});$ {Selecting $D_j$ from $D_2^{+(i)}$ }
(8) For each $a_k \in cause(s_j) - A_I^{(i)}$ do
(9) Begin SUB: = choose $(D_2^{+(i)} \cap effect(a_k))$ ; {Constructing set SUB}
(10) For each $sub_i \in SUB$ do
(11) Begin such（$n_i, a_k$）：=
        $(A_I^{(i)} \cup \{a_k\}, D_1^{+(i)} \cup sub_i, D_2^{+(i)} - sub_i);$
     {Constituting successor node of $n_i$ }
(12) INSERT（such（$n_i, d_k$），OPEN）;{ renewing table OPEN}
(13)          End for
           End for
       End for
(14) If there are complete terminal nodes in OPEN
(15) Then solving the complete solution of problem P, return $Co - sol(P)S;$ {theorem 3.1}
(16) Else solving the partial solution of problem P, return $Pa - sol(P)S;$ {theorem 3.2}
(17) End.
```

In the step (9), it constructs a set by using function choose, satisfying the following conditions

$$SUB = \{sub_i \mid sub_i \subseteq D_2^{+(i)} \cap \ effect(a_k), sub_i \ne \phi\}$$

(a) For $sub_i \in SUB, \{a_k\}$ is a temporal consistency covering.

(b) For $sub_i \in SUB$, It does not exist

$$sub_i \subseteq D_2^{+(i)} \cap \ effect(a_k),$$

such that $sub_j \supset sub_i$, and $\{d_k\}$ is a temporal consistency covering on sub_i.

In the step (12), for successor node of n_i,

$$n_j = succ(n_i, a_k) = (A_I^{(j)}, D_1^{+(j)}, D_2^{+(j)}),$$

INSERT revises table OPEN as follows:

(i) If n_j is a complete or partial end node, n_j will be put into the table OPEN directly. Otherwise n_j is a non-end node, and taken by the following steps:

(ii) If it has $n_t=(A_I^{(t)},D_1^{+(t)},D_2^{+(t)})$ in the table OPEN and *CLOSE*, such that $D_2^{+(j)}=D_2^{+(t)}$, $A_I^{(t)}\subseteq A_I^{(j)}$, then the node n_j is abandoned, the table *OPEN* is not change.

(iii) If it has $n_t=(A_I^{(t)},D_1^{+(t)},D_2^{+(t)})$ in the table OPEN, such that $D_2^{+(j)}=D_2^{+(t)}$, $A_I^{(t)}\supset A_I^{(j)}$, then the node n_t is deleted from the table OPEN, and put into the table CLOSE, meanwhile n_j is put into the table OPEN.

(iv) Otherwise, n_j is put into the table OPEN.

In afore-mentioned algorithm, the table OPEN is using to deposit expanding nodes, and the table CLOSE is using to deposit expanded nodes. In the process of constructing G(P), successor nodes is generated in basis of causality and temporal constraint(temporal consistency) between anomalous behavior and anomalous omen. Suppose the graph is acyclic graph, and the number of nodes is limited. The complete or partial nodes can be obtained through successor expanded $succ(n_i,a_k)$ after undergoing limited steps (no more than |D|). Therefore termination of algorithm is quite obvious.

IV. Instance Analysis

A discriminant target of network behaviors

$P=<A,D,R,DEL,D^+,DOCT>$ has the following definition :

$$D^+=\{D_1,D_2,D_3,D_4,D_5\},\quad A=\{a_1,a_2,a_3\},$$

then the relation matrix of behavior-data in network is as follows :

$$R=\begin{array}{c} \\ D_1\\ D_2\\ D_3\\ D_4\\ D_5\end{array}\begin{array}{c}\begin{array}{ccc}a_1 & a_2 & a_3\end{array}\\ \begin{bmatrix}1 & & 1\\ 1 & & \\ 1 & 1 & \\ & 1 & 1\\ & 1 & 1\end{bmatrix}\end{array}$$

Based on the theorem 3.1 and 3.2, it can be obtained the following values:

DOCT(1)=(1,8,10,1), DOCT(2)=(1,9.8,9.8,1),
DOCT(3)=(1,12,13,1),DOCT(5)=(1,10,11,1)

$$DEL(1,1)=(1,3,4,1),DEL(1,3)=(1,3,3,1),$$
$$DEL(2,4)=(1,7,7,1),DEL(3,1)=(1,8,8,1),$$
$$DEL(1,2)=(1,4.6,5.7,1),DEL(2,3)=(1,4,5,1),$$
$$DEL(2,5)=(1,5,6,1),DEL(3,4)=(1,9,10,1)$$
$$DEL(3,5)=(1,6,$$

It takes a temporal consistency threshold δ =0.6, it is taken the first in first out strategy the table OPEN, and generated a hypothesis graph, it is shown as Figure 2.

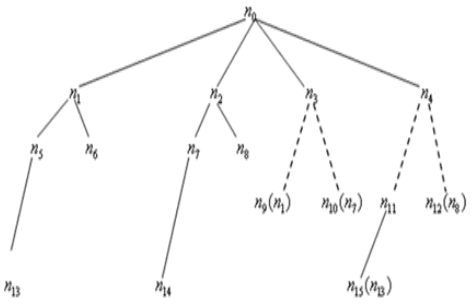

Figure 2. Hypothesis graph of discriminant target *P*

Where

$$n_0=(\phi,\phi,D^+),n_1=(\{a_1\},\{D_1,D_2\}),$$
$$n_2=(\{a_1\},\{D_3\}),n_3=(\{a_3\},\{D_1\}),$$
$$n_4=(\{a_3\},\{D_5\}),n_5=(\{a_1,a_2\},\{D_1,D_2,D_3\}),$$
$$n_6=(\{a_1,a_2\},\{D_1,D_2,D_5\}),n_7=(\{a_1,a_3\},\{D_3,D_1\}),$$
$$n_8=(\{a_1,a_3\},\{D_3,D_5\}),n_9=(\{a_3,a_1\},\{D_1,D_2\}),$$
$$n_{10}=(\{a_3,a_1\},\{D_1,D_3\}),n_{11}=(\{a_3,a_1\},\{D_5,D_1,D_2\}),$$
$$n_{12}=(\{a_3,a_1\},\{D_5,D_3\}),n_{13}=(\{a_1,a_2,$$
$$\{D_1,D_2,D_3,D_5\}),$$
$$n_{14}=(\{a_1,a_3,a_2\},\{D_3,D_1,D_5\}),$$
$$n_{15}=(\{a_1,a_3,a_2\}\{D_5,D_1,D_2,D_3\})$$

Because of $D_1^{+(i)}\cup D_2^{+(i)}=D^+$, n_i is written n_i = $(A_I^{(i)},D_1^{+(i)})$ for short. In the beginning of algorithm, it is taken $D_1\in D_2^{+(0)}=D^+$ from a initial node $n_0=(\phi,\phi,D^+)$, then $a_1,a_3\in cause(D_1)-A_I^{(0)}$. Based on a behavior a_1, it can construct sub-set $SUB=\{\{D_1,D_2\},\{D_3\}\}$, and generate successor nodes n_1,n_2 in the basis of n_0, then put it into table *OPEN*. In the same way, based on a_3, it can construct sub-set $SUB=\{\{D_1\},\{D_3\}\},$ and then obtain the successor nodes n_3,n_4 in the basis of n_0.

After generated $n_9=(\{a_3,a_1\},\{D_1,D_2\},\{D_3,D_5\})$ in the step (4), due to $A_I^{(1)}=\{a_1\}\subset A_I^{(9)},D_2^{+(1)}=D_2^{+(9)}$ in n_1, according to principle of INSERT in the step (ii), the node n_9 could be abandoned, namely it may be pruning correspondingly. Based on INSERT principle, n_{10} and n_{12} are treated in the same way. The corresponding complete solution of paths $n_0\to n_1\to n_5\to n_{13}$ and $n_0\to n_4\to n_{11}\to n_{15}$ is

$$Co-sol(P)=\{(a_1,\{D_1,D_2\}),(a_2,\{D_3\}),(a_3,\{D_5\})\}$$

. Partial end nodes n_6 and n_{14} are corresponding to partial solutions respectively. For example n_6 corresponds to $Pa-sol(P)=\{(a_1,\{D_1,D_2\}),(a_2,\{D_5\})\}$, it is known that the omen m_3 does not explain. The occurrence time of each known omen and corresponding anomalous behavior is shown in Figure1.

V. Conclusion

The solution for network anomaly detection in this paper is a further development based on literature [5，6]. Actually, it is breadth- first search method. So the search cost is still greater, especially for a large amount of data, though the method presented in this paper makes pruning, in order to decrease the number of network nodes, by using of temporal consistency in the step of SUB, INSERT etc. Trying to resolve this conflict, a possible method is to convert the original method into depth-first search method by introducing node evaluation function, such as literature [7，8，9]. But in the model of NTGSBC, node evaluation function must reflect causality and temporal constraint between anomalous behavior and data at the same time. It is more complex than pure probability causality in literature [2, 3]. It is yet to be further studied about how to seek appropriate node evaluation function in the model of NTGSBC. The other possible solution is problem decomposition. For example, in total behavior detection system modeling of a large website, we will divide the total detection process into many subsystems according to structure and function of website system. And define the causality among subsystems. Moreover the subsystem itself is defined by the model of NTGSBC. Anomaly behavior detection process is separated into inner inference for NTGSBC monitor system and anomaly causality diffusion among subsystems. The advantage of this method is as follows: (1) the detection scale of subsystems, obtained after decomposition, is smaller. It is fit for NTGSBC modeling and problem solving. (2) Through problem decomposition and defining the diffusion causality among subsystems, multi-layered causality model about total monitor targets will be built up, based on two layer causality from anomalous behavior to data. Multi-layered causality model is more suitable for detection target describing and problem solving.

Temporal consistency set covering is defined in this paper. And based on this definition, we described the basic framework of NTGSBC and the method of problem solving. Fuzzy temporal information is introduced in the model of NTGSBC, it makes generalized inference detection model and method more fitting for practical problems in other fields. Certainly, more detailed studies should be continued in the further.

References

[1] Bykova M, Ostermann S, Tjaden B. Detecting network intrusions via a statistical analysis of network packet characteristics. In: Proc. of the 33rd Southeastern Symp. on System Theory. 2001. 309314. http://masaka.cs.ohiou.edu/papers/ssst2001.pdf

[2] Denning DE. An intrusion-detection model. IEEE Trans. on Software Engineering, 1987,13(2):222-232.

[3] Lee W, Stolfo SJ. A framework for constructing features and models for intrusion detection systems. ACM Trans. on Information and System Security, 2000,3(4):227-261.

[4] Valdes A, Skinner K. Adaptive, model-based monitoring for cyber attack detection. In: Debar H, M¨ L, Wu SF, eds. Proc. of the 3rd Int!¯lWorkshop on the Recent Advances in Intrusion Detection (RAID 2000). LNCS 1907, Heidelberg: Springer-Verlag, 2000. 80-93.

[5] Aickelin U, Greensmith J, Twycross J. Immune system approaches to intrusion detectionA review. In: Nicosia G, et al., eds. Proc. of the 3rd Int!¯l onf. on Artificial Immune Systems. LNCS 3239, Heidelberg: Springer-Verlag, 2004. 316-329.

[6] Lee W, Stolfo SJ. A Data mining framework for building intrusion detection models. In: Gong L, Reiter MK, eds. Proc. of the !99 IEEE Symp. on Security and Privacy. Oakland: IEEE Computer Society Press, 1999. 120-132.

[7] Eskin E, Arnold A, Prerau M, Portnoy L, Stolfo SJ. A geometric framework for unsupervised anomaly detection: detecting intrusions in unlabeled data. In: Barbar¨ D, Jajodia S, eds. Applications of Data Mining in Computer Security. Boston: Kluwer Academic Publishers, 2002. 78-99.

[8] Proedru K, Nouretdinov I, Vovk V, Gammerman A. Transductive confidence machine for pattern recognition. In: Elomaa T, et al., eds. Proc. of the 13th European Conf. on Machine Learning. LNAI 2430, Heidelberg: Springer-Verlag, 2002. 381-390.

[9] Barbar¨ D, Domeniconi C, Rogers JP. Detecting outliers using transduction and statistical testing. In: Ungar L, Craven M, Gunopulos D, Eliassi-Rad T, eds. Proc. of the 12th ACM SIGKDD Int!¯l onf. on Knowledge Discovery and Data Mining. New York: ACM Press, 2006. 55-64.

[10] Angiulli F, Pizzuti C. Outlier mining in large high-dimensional data sets. IEEE Trans. on Knowledge and Data Engineering, 2005, 17(2):203-215.

[11] Ghosh AK, Schwartzbard A. A study in using neural networks for anomaly and misuse detection. In: Proc. of the 8th USENIX Security Symp. 1999. 141-151. http://www.usenix.org/events/sec99/full_papers/ghosh/ghosh.ps

[12] Manikopoulos C, Papavassiliou S. Network intrusion and fault detection: A statistical anomaly approach. IEEE Communications Magazine, 2002,40(10):76-82.

[13] Laskov P, Schafer C, Kotenko I. Intrusion detection in unlabeled data with quarter-sphere support vector machines. In: Proc. of the Detection of Intrusions and Malware & Vulnerability Assessment (DIMVA 2004). 2004. 71-82. http://www2.informatik.hu-berlin. de/wm/journalclub/dimva2004.pdf

Authors Profile

He Ping is a professor of the Department of Information at Liaoning Police Academy, P.R. China. He is currently Deputy Chairman of the Centre of Information Development at Management Science Academy of China. In 1986 He advance system non-optimum analysis and founded research institute. He has researched analysis of information system for more than 20 years. Since 1990 his work is optimization research on management information system. He has published more than 200 papers and ten books, and is editor of several scientific journals. In 1992 awards Prize for the Outstanding Contribution Recipients of Special Government Allowances P. R. China.

Appendix

The proof of theorem 3.1is as follows:

(1) Note

$$Co-ps=\{(a_k,D^+(a_k))\,|\,a_k\in A_I^{i+1}-A_l^i,D^+(a_k)$$
$$=D_1^{+(i+1)}-D_1^{+(i)},0\le i\le l-1\}$$
,

as n_{i+1} is a successor node of n_i , according to the definition 3.2， it exists $D_j\in D_2^{+(i)},a_k\in cause(D_j)-A_I^{(i)}$ ， such that

$$n_{i+1} = succ(n_i, a_k)$$
$$= (A_I^{(i)} \cup \{a_k\}, D_1^{+(i)} \cup some(D_2^{+(i)} \cap effect(a_k)), D_2^{+(i)}$$
$$-some(D_2^{+(i)} \cap effect(a_k))),$$

a_k is a temporal consistency covering on

$$some(D_2^{+(i)} \cap effect(a_k)) \subseteq D_2^{+(i)} \cap effect(a_k).$$

Due to $a_k \in cause(D_j) - A_I^{(i)}$, it exists

$$\{a_k\} \cap A_I^{(i)} = \phi, A_I^{(i+1)} - A_I^{(i)} = A_I^{(i)} \cup \{a_k\} - A_I^{(i)} = \{a_k\},$$
$$a_k \in A_I^{(i+1)} - A_I^{i(}.$$

Based on $D_2^{+(i)} = D^+ - D_1^{+(i)}$, it exists

$$D_1^{+(i)} \cap D_2^{+(i)} = \phi, D_1^{+(i)} \cap some(D_2^{+(i)} \cap effect(a_k)) = \phi,$$

Then it obtains

$$D^+(a_k) = D_1^{+(i+1)} - D_1^{+(i)}$$
$$= D_1^{+(i)} \cup some(D_2^{+(i)} \cap effect(a_k)) - D_1^{+(i)}$$
$$= some(D_2^{+(i)} \cap effect(a_k))$$

So for $(a_k, D^+(a_k)) \in Co-ps, \{a_k\} = A_I^{i+1} - A_l^i$ is a temporal consistency covering on $D^+(a_k)$.

(2) In order to express itself clearly note $a_k^{i+1} \in A_I^{(i+1)} - A_I^{(i)}, D^{+(i+1)}(a_k^{i+1}) = D_1^{+(i+1)} - D_1^{+(i)}$, then

$$\bigcup_{(a_k, D^+(a_k)) \in co-ps} D^+(a_k) = \bigcup_{0 \le i \le l-1} D^{+(i+1)}(a_k^{+(i+1)}) \quad \text{（A.1）}$$

A proof by mathematical induction is adopted firstly. For $p \le l-1$, the following expression holds

$$\bigcup_{0 \le i \le p} D^{+(i+1)}(a_k^{+(i+1)}) = D_1^{+(p+1)} \qquad \text{（A.2）}$$

When p=0, noticed that $S_1^{+(0)} = \varphi$ for initial node n_0, it exits

$$\bigcup_{0 \le i \le 0} D^{+(i+1)}(a_k^{(i+1)}) = D^{+(1)}(a_k^{(1)}) \quad .$$
$$= D_1^{+(1)} - D_1^{+(0)} = D_1^{+(0+1)}$$

Suppose the expression (A.2) is set up when $p \le t, t \le l-2$, then

$$\bigcup_{0 \le i \le t+1} D^{+(i+1)}(a_k^{(i+1)})$$
$$= (\bigcup_{0 \le i \le t} D^{+(i+1)}(a_k^{(i+1)})) \cup D^{+(t+2)}(a_k^{(t+2)})$$
$$= (D_1^{+(t+1)} \cup (D_1^{+(t+2)} - D_1^{+(t+1)}))$$
$$= D_1^{+(t+1)} \cup some(D_2^{+(t+1)} \cap effect(a_k^{t+2}))$$
$$= D_1^{+(t+2)}$$

So the expression (A.2) is set up. As well as $n_l = (A_I^{(l)}, D_1^{+(l)}, D_2^{+(l)})$ is a complete end node, $D_1^{+(l)} = D^+, D_2^{+(l)} = \varphi$, hence

$$\bigcup_{0 \le i \le l-1} D^{+(i+1)}(a_k^{(i+1)}) = D_1^{+(l)} = D^+ .$$

According to the results of expression (1) and (2), and definition 2.4, the set Co-ps is a complete explanation Co-exp（P）for the problem *P*.

(3) As Co-ps is a Co-exp（P）of the problem *P*, according to definition 2.4, a fault set is as follows:

$$A_I = \{a_k \mid (a_k, S^+(a_k)) \in Co-ps\}$$
$$= \bigcup_{0 \le i \le l-1} \{a_k^{(i+1)}\} \quad .$$
$$= \bigcup_{0 \le i \le l-1} A_I^{(i+1)}(A_I^{(i+1)}) - A_I^{(i)})$$

Resembling the proof of the expression (A.2) in the step (2), a proof by mathematical induction is adopted. So the set of anomalous behavior explanation $Co-\exp(P)$ is as follow:

$$A_I = \bigcup_{0 \le i \le l-1} (A_I^{(i+1)}) - A_I^{(i)}) = A_I^{(i)}$$

Theorem 3.2 is proved as follows:

$$Pa-ps = \{(a_k, D^+(a_k)) \big| a_k \in (A_I^{(i+1)} - A_I^i),$$
$$D^+(a_k) = D_1^{+(i+1)} - D_1^{+(i)}, 0 \le i \le l-1\}$$

(1) it is similar to the step (1) in theorem 3.1, it exists $(a_k, D^+(a_k)) \in Pa-ps$. $\{a_k\}$ is a temporal consistency covering of $D^+(a_k)$.

(2) For the set Pa-ps, resembling the proof of step (2) in theorem 3.1, it can be obtained

$$D_1^+ = \bigcup_{(a_k, S^+(a_k)) \in Pa-ps} D^+(a_k) = \bigcup_{0 \le i \le l-1} D^{+(i+1)}(a_k^{+(i+1)}) = D_i^{+(i)} \quad \text{（B.1）}$$

Resembling the proof of step (3) in theorem 3.1, for Pa-ps, it exists

$$A_I = \{a_k \Big| (a_k, D^+(a_k)) \in Pa-ps\} = \bigcup_{0 \le i \le l-1} \{a_k^{+(i+1)}\} \qquad \text{(B.2)}$$
$$= \bigcap_{0 \le i \le l-1} (A_I^{(i+1)} - A_I^i) = A_I^i$$

As $n_t = (A_I^i, D_1^{+(i)}, D_I^{+(i)})$ is partial end nodes, so

Cause $\{D_2^{+(i)} \subseteq A_I^i, D_2^{+(i)} \ne \varnothing, D_1^{+(i)} = D^+ - D_2^{+(i)} \subset D^+$

Then combining with the expression (B.1) and (B.2), it exists

$$D_1^+ = \bigcup_{(a_k, M^+(a_k) \in Pa-po} D^+(a_k) = D_I^{+(l)} \subset D^+ .$$

For

$$D_2^+ = D^+ - D_1^+ = D^+ - D_1^{+(l)} = D_2^{+(l)},$$
$$cause(D_2^+) = cause(D_2^+) \subseteq A_I^{(l)} = A_I,$$
$$cause(D_2^+)k \subseteq A_I$$
$$= \{a_k \mid (a_k, D^+(a_k) \in Pa-ps\}.$$

According to the conclusions in step (1) and (2), and definition 2.5, Pa-ps is a partial explanation of the problem *P*, namely theorem 3.2 can be established.