Approach for Network Fault Diagnosis Based on Bayesian Model

Zhixian Ran[1] and Jianning Geng[2]
[1]School of Computer Science and Engineering, Zhengzhou University, Zhengzhou 450000, China
[2]School of Computer Science and Engineering, University of Electronic Science & Technology of China, Chengdu 610054, China

***Abstract*—User logon failure accounted for a relatively large proportion in network fault. Therefore, it is of great significance for service providers to diagnose and locate login failures quickly. In this paper, a login fault diagnosis method based on bayesian network is proposed. Firstly the login fault events are found by analyzing the data packets generated during PPPoE authentication, and the fault sample table is obtained based on expert knowledge and sample learning in order to further locate the root cause of login failure. And then, the fault Bayesian network model is established. Finally, the final fault reason is obtained through Bayesian reasoning process.**

Keywords—logon fault; fault diagnosis; Bayesian network; PPPoE

I. INTRODUCTION

The current operators can not take the initiative and to detect faults. They usually know failed event only when customer complaints. In the case of a network failure, it is not quickly and accurately to locate network faults, which can not bring a good user experience to the user. Addition, operators can quickly diagnose and locate the login fault for reducing the cost of maintenance, improve competitiveness also has great significance.

Bayesian network is a causal relationship between the uncertainty associated with the model, with a strong ability to deal with the problem of uncertainty. In this paper, the method process is that, analyzing the user to log failure data packet exchange process, extract fault symptoms, their causes and symptoms of the corresponding conditional probability of failure is then given by the experts based on experience, to establish the diagnosis Bayesian network model, and ultimately through Bayesian networks reasoning, locate the user logs on the root cause of the failure.

II. LOGIN FAULT DETECTION

A. Login Authentication Technology Principle

Currently user login authentication technology most commonly used is PPPoE protocol [1]. PPPoE (PPP over Ethernet), dissemination of PPP frames in Ethernet technology, compared with traditional access methods with a higher price, in the formation of a network is often used.

PPPoE protocol workflow is divided into two phases: Discovery (Discovery) phase and session (Session) phase [2]. The discovery stage and the stage of the session packet type field in the Ethernet frame values are 0x8863 and 0x8864, have been recognized by the IEEE. PPPoE workflow shown in Figure 1 [3].

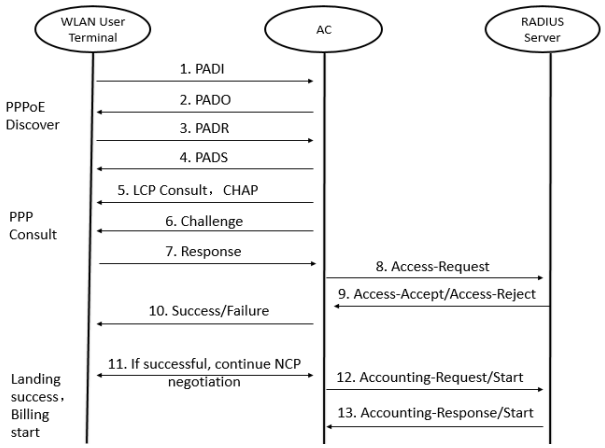

FIGURE I. PPPOE AUTHENTICATION FLOWCHART

Communication process discovery phase PPPoE protocol and session stages are as follows:

(1) Discovery phase

In the discovery phase, the host user to broadcast may find multiple access controller, and then will choose one of them. This stage is divided into four steps, after completion of the ends of the communication have been negotiated by the MAC address of PPPoE SESSION-ID and the communication ends, both of which defines a unique information on a PPPoE session. The four steps are as follows:

① Host launch PADI (PPPoE Active Discovery Initiation) packet, comprising at least one type of service tags, make the required service to the access controller.

② Access controller received PADI packet is transmitted to the user host PADO (PPPoE Active Discovery Offer) packet in response to the request. PADO packet contains the name of the access controller type label and service label, which indicates the service provided to the host server.

③ Host review all received PADO, to send PADR(PPPoE Active Discovery Request) packet to the selected access controller.

④ When the access controller receives a PADR packet, it will begin to establish a PPP session, the client sends a PADS (PPPoE Active Discovery Session-confirmation) packets. When the client receives the PADS packet, the client and the access controller enters the PPP session stage.

(2) PPP session phase

The user host and the access controller start the PPP session according to the connection parameters negotiated between the two parties in the discovery phase. In this phase, the data of the two sides are encapsulated in the PPP data frame. If either end of the two sides want to end the session, you can terminate the PPPoE session by take the initiative to send PADT packets to each other [4].

B. Login Failure Detection

By analyzing the client and server-side data capture packets user login process, combined with the certification process when the user logs in, the user can know whether the normal login process. If login failures happened, it can also know which part of a problem. For example, in the PPPoE authentication process, if the client keeps sending PADI packet, but does not receive the server sent PADO, then the different physical links, possible reasons for this are: the line between cell appear to the server error, family WAN port cable connection exception, cats and central office server is not synchronized. If the client sends PADR packet but not receives the server side of PADS, the common cause of the error is the wrong account password or carriers bound computer network card MAC address.

III. BAYESIAN FAULT REASONING

By analyzing the user login message on the network, we can always know the reason of the network failure. However, considering the fact that the network topology is very complex, it is impossible to capture the data packets on the link, so it is impossible to accurately locate the fault by analyzing the data packets in some places. In order to accurately locate the network fault, this paper uses the Bayesian network to carry out fault reasoning, and finally locate the login fault.

A. Bayesian Network Foundation

Suppose the set of random variables U={$U_1, U_2, \ldots, U_k$}, $u1$ represents U_1 values, then P($U_1 = u1, U_2 = u2, \ldots, U_k = uk$) represents the joint probability, it must meet the following two properties [5]:

$$0 \le P(U_1, U_2, \ldots, U_k) \le 1 \tag{1}$$

$$\sum P(U_1, U_2, \ldots, U_k) = 1 \tag{2}$$

The formula (2) must contain all the random variables.

Given variable U_j, conditional probability U_t function with $P(U_t|U_j)$.

$$P(U_i|U_j) = \frac{P(U_i, U_j)}{P(U_j)} \tag{3}$$

In the formula, $P(U_t, U_j)$ is the joint probability of U_t and U_j, $P(U_j)$ is the edge probability of U_j. According to the formula (3), we can get:

$$P(U_i,U_j) = P(U_i|U_j)P(U_j) = P(U_j|U_i)P(U_i) = P(U_j,U_i) \tag{4}$$

So we can get very important Bias theorem:

$$P(U_i|U_j) = \frac{P(U_j|U_i)P(U_i)}{P(U_j)} \tag{5}$$

Bias theorem is mainly used to predict the probability of occurrence of events, this theorem shows that the event will happen depends on the number of times the event occurred in the event of the decision.

B. Bayesian Networks and Probabilistic Inference

Bayesian network is a directed acyclic graph, where nodes represent variables, there are between two nodes to the dependencies between variables representative of the arc, this dependency correlation between the size of the probability of this condition to the arc to indicate the size [6]. Network configuration diagram, although the relationship is not required for a node and its parent node must be a causal relationship, but in the Bayesian network diagram, this relationship can still show the dependencies between two nodes.

C. Diagnosis Bayesian Network

Bayesian network troubleshooting can be expressed as [7]:

First,after monitoring the system fails, according to the observations of the relevant information (according to the experts may also direct experience), the decision-making process of reasoning summarized by the maximum possible fault list of assumptions; then through Bayesian reasoning has been the most likely troubleshooting steps; after the previous step if the operating system function returned to normal, the diagnostic process is terminated, otherwise the new information obtained by the step to continue to calculate and to exclude other possible fault assumptions, the process is repeated until the system returns to normal function.

Thus, the diagnostic information to be considered in the decision-making process, including fault symptom, the fault hypothesis, maintenance operations, observing the operation and the cost of diagnosis and the like. Among them, the symptom is a phenomenon that system running fault occurs, is the intuitive feeling; fault assumption is a failure mode set that when a failure occurs which may be caused by reasons. Repair work is based on the recommendations for service to repair or replace parts of a system, it may be to reinstall the software. Observing whether to exclude major operation by detecting

faults; diagnosis is a measure of the cost of a diagnostic method whether efficient, is to determine the merits of a standard diagnostic strategies.

D. Object Oriented Bayesian Network Class Definition

(1) node class

By means of the common nature of the node variables in the abstract Bayesian network, the node base class can be defined as follows:

```
class DBNnode
{
  public:
    DBNnode *father;
    double *pt;
    virtual double GetPt()=0;
    virtual int SetPt()=0;
    ……
  private:
    char name[32];
    NodeType nodetype;
    int stateCnt;
    int *state;
    double *  proTable;
}
```

NodeType is an enumeration type, which indicates the property of the current node, which is one of the five types of the fault symptom node, the fault hypothesis node, the maintenance operation node, the observation node and the other nodes.

Through the object oriented inheritance mechanism, it can be defined by the common Bayesian network node class definition of failure symptom node, fault hypothesis node, maintenance operation node, observation node and other nodes. For example, the failure of the hypothetical node, for example, the definition can be as follows:

```
class FaultDBNnode:public DBNnode
{
  public:
    FaultNodeType faultNodeType;
    double GetPt();
    int SetPt();
    ……
}
```

FaultNodeType is an enumeration type, said the fault hypothesis node attributes and values are final fault hypothesis nodes (denoting system the fundamental causes of malfunction), intermediate fault hypothesis node (node fault hierarchical process) and auxiliary fault hypothesis node (in order to facilitate the construction of Bayesian network and the introduction of auxiliary nodes). Because the conditional probability table and the conditional probability table of the nodes are two functions in the definition of the parent node, the two methods are implemented in the class of fault hypothesis.

For the maintenance operation node, its specific definition may be the following way:

```
class RepairDBNnode:public DBNnode
{
  public:
    DBNnode *depents;
    double *depCost;
    double GetPt();
    int SetPt();
    double GetCost();
    int SetCost();
    ……
  private:
    char describe[256];
    double cost;
    double    cn;
    bool isEXE;
}
```

Among them, cost variables are expressed in terms of the operation cost is consider dependencies node when the operation cost, with pointers to depents points rely on the manipulation of the relationships of nodes, depCost points to the corresponding dependence dependent weights correspond to the nodes. In addition to the cost of dependence, the operation of the node also costs a certain amount of CN, so the actual operation cost cost = depCost[i] * depents[i].cost + cn.

The remaining nodes are defined in accordance with the definition of the above two nodes.

Diagnosis Bayesian network(DBN) node base class and its sub class UML class diagram as shown in figure 2.

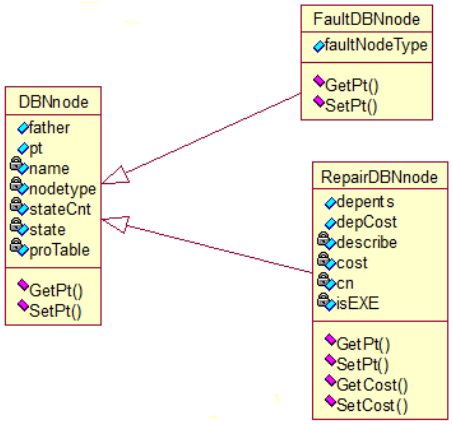

FIGURE II. DIAGNOSTIC BAYESIAN NETWORK NODE BASE CLASS AND ITS SUB CLASS UML CLASS DIAGRAM

(2) DBN class

According to the object oriented Bayesian network node attribute, the Diagnostic Bayesian Networks (DBN) can be defined as follows:

```
class DBN
{
  public:
    DBNnode *inputsNode;
    DBNnode *interNode;
    DBNnode *outNode;
    DBN *interDBN;
    DBNnode *GetDBN();
    void SetDBN();
  private:
    char name[32];
}
```

The UML class diagram of the diagnostic Bayesian network class is shown in Figure 3.

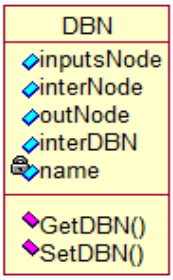

FIGURE III. UML CLASS DIAGRAM OF DIAGNOSTIC BAYESIAN NETWORK CLASS

IV. LOGIN FAULT LOCATION

A. Establish Diagnosis Bayesian Network of Login Fault

For specific systems, diagnostic Bayesian model constructed in accordance with the following basic steps: ① establish directed acyclic graph through the fault symptom and cause of the fault table; ② establish conditional probability table; ③ establish the cost of operating table; ④ Creating and editing models. Here follow these steps to build the user's logon failure Bayesian network model.

In the packet analysis by client login process, the observed failure symptoms such as shown in Table 1:

TABLE I. USER LOGIN FAILURE SYMPTOMS

fault symptom ordinal	fault symptom description	fault occurrence probability
S1	Client initiated PADI, less than PADO	0.4
S2	Client initiated PADR, less than PADS	0.4
S3	In the process of link negotiation, the client receives PADT	0.1
S4	The server sent PADO, received less than PADR	0.05
S5	The server receives PADR, does not send PADS	0.05

Failure reasons: F1: user hosts and routers connect impassability; F2: District to the server link connection exception; F3: PPPoE server is not working properly; F4: MAC address binding with the user account; F5: account password error or is in arrears.

The corresponding relationship between the fault symptoms and the fault causes is shown in table 2.

TABLE II. FAULT SYMPTOMS AND CAUSES OF FAILURE

Fault symptom	Cause of failure
S1	F1, F2, F3
S2	F1, F2, F3, F4, F5
S3	F4, F5
S4	F1, F2
S5	F3, F4, F5

Fault causes, F1, F2, F3 and F5 is the cause of the middle of the fault point of reason, F4 is the ultimate failure of the node. F1, F2, F3 and F5 corresponding to the final cause of the failure of the point as shown in table 3.

TABLE III. FINAL FAILURE REASON HYPOTHESIS POINT

Intermediate fault hypothesis point number	Final failure point
F1	F11: Home network cable is not connected; F12: Local connection disabled; F13: Network card driver damage; F14: Router does not work properly.
F2	F21: Cell-to-server link open circuit; F22: Intermediate device CPU 100% load operation; F23: Intermediate device caches enough lead to loss.
F3	F31: Server CPU100% work load; F32: The server cache is not enough to cause the packet loss.
F5	F51: Account in arrears: F52: Account password input error.

Using the reason for the failure symptoms and the corresponding fault, it can establish the diagnosis Bayesian network directed acyclic graph. Each corresponds to a single source of the failure symptoms DBN model. For example, S1 fault, the "client-initiated PADI, can not receive PADO", it's Bayesian network model is shown in Figure 4.

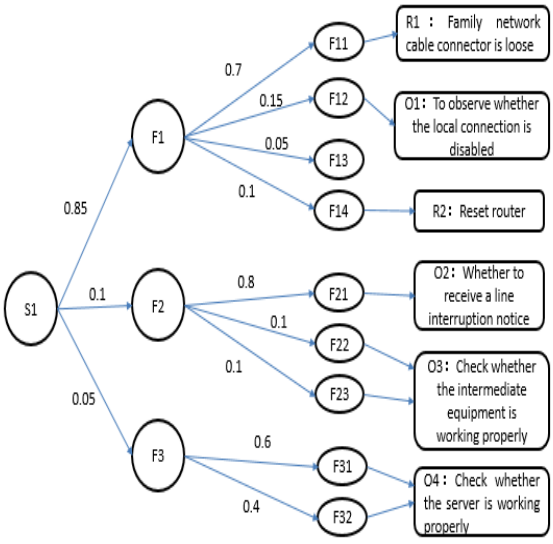

FIGURE IV. FAULT S1 CORRESPONDING DBN MODEL

Through the sample and case analysis, the conditional probability of failure assumptions given node in Figure 4 in the Bayesian model. In the final breakdown of the reasons given in the appropriate manner, such as R1 and R2 are maintenance operations, O1, O2, O3, O4 is observation operations, which corresponds to the cost of operation, as shown in Table 4.

TABLE IV. OPERATING EXPENSE SPENT NODE

Description of Operation	operation cost (min)
R1: Family network cable connector is loose	5
R2: Reset router	3
O1: To observe whether the local connection is disabled	3
O2: Whether to receive a line interruption notice	1
O3: Check whether the intermediate equipment is working properly	40
O4: Check whether the server is working properly	30

B. Logon Failure Bayesian Network Positioning

As S1 fault fault example, Reasoning as follows:

(1) When the S1 failure symptoms, select the maximum conditional probability of failure assumptions node, select the F1 (hosts and routers connected to nowhere).

(2) F1 corresponding to the four ultimate failure assumptions node, which has the greatest probability of three nodes also corresponds to an operation node. Calculate the next optimal choice,

$$W = P / C \quad (6)$$

W represents an optimal weight, P represents the conditional probability, and C represents the cost of the operation node. The greater the W indicates the node should first be selected. In Figure 4 and Table 4, calculate the optimal weight of three nodes in order as 0.14, 0.05, 0.03. So, the first fault identified as F11, that is cause of the malfunction is home to the network cable is not connected;

(3) R1 maintenance operations, troubleshooting home network cable connector is loose, and re-plug the network cable connector.

(4) Check whether the fault is rectified. If you normal landing, the diagnostic process ends, and then increase the success rate of this approach in the case of failure to resolve a library, Figure 4 Bayesian network model conditional probability F11 will increase accordingly. The entire process ends. If the fault persists, then transferred to (5) steps.

(5) Reduce the success rate of (3) step troubleshooting methods in the case base. Select the optimal weight second largest F12 fault hypothesis node that finds cause of the malfunction is a local connection is disabled.

(6) Observing the local connection is disabled, if you disable it open.

(7) Check whether the fault is rectified. If you normal landing, the diagnosis process ends. If you do not work, then repeat the above steps until troubleshooting. If the cause of the failure is not the final diagnosis of any fault hypothesis node Figure 4 (this situation is inevitable, incomplete samples and expertise will always generate this case), you should adjust the dynamic Bayesian network structure.

C. System Testing and Verification

The test is done in the network lab environment. Topology experimental environment shown in Figure 5.

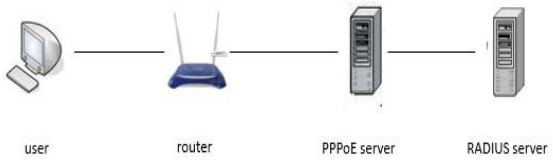

FIGURE V. SYSTEM TEST LAB ENVIRONMENT

The PPPoE server and the RADIUS server are built on the computer of the two Win7 system. The user name and password are stored on the RADIUS server.

Test one: the user side of the network cable pulled out from the router.

With Wireshark software intercepted a key part of the software package screenshot shown in figure 6.

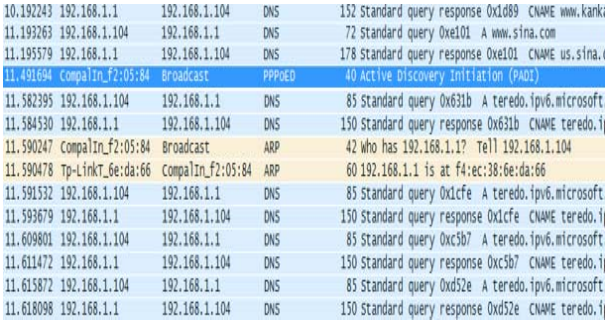

FIGURE VI. TEST A FAULT DATA PACKAGE SCREENSHOT

As can be seen from the figure 6 (blue selected data packets), when the user logs on to launch the PPPoE certification, in the form of radio issued a PADI packet. Known from PPPoE authentication process described in Chapter 2.1, normal landing process interaction server received PADI packet after sending PADO data to the client package, then the client sends PADR packet. Finally, the server sends pads data packet to complete the discovery process of PPPoE authentication. And figure 6 shows that the user side after receiving the PADI packet and did not get the follow-up of the certification data package, which shows the user login failure. Further analysis of the PADI data packet, open the data packet after the detailed information as shown in figure 7.

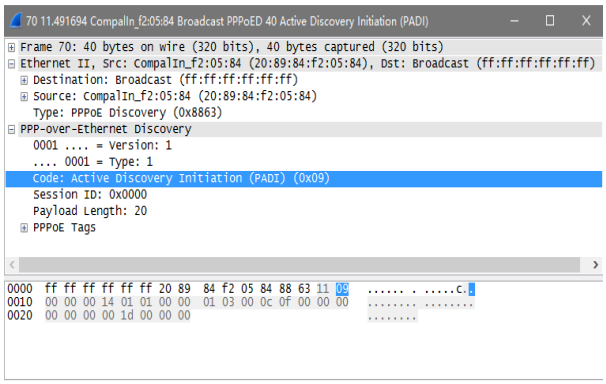

FIGURE VII. PADI PACKAGE DETAILS SCREENSHOT

Shown in Figure 7, MAC address 20:89:84:f2:05:84 users launched a broadcast packet (known from Ethernet frame destination address ff:ff:ff:ff:ff:ff), Ethernet frame protocol field 0x8863, that is PPPoE discovery stage of data packets, and code fields 0x09 shows the PADI packet (active discovery initiation).

In the process of testing and verification, the system successfully detects the user login failure and the final results.

In the experiment, five times experiments are carried out for each kind of final failure listed in table 3, a total of Fifty-five times experiments. The successful positioning of the number and success ratio of the successful registration as shown in table 5.

TABLE V. EXPERIMENTAL RECORD

Experimental fault	Successful times	Success ratio
F11	5	100%
F12	5	100%
F13	4	80%
F14	5	100%
F21	4	80%
F22	4	80%
F23	4	80%
F31	5	100%
F32	4	80%
F51	5	100%
F52	5	100%

As can be seen from the experiment record sheet, log on fault diagnosis method based on Bayesian proposed experiment achieved good results.

V. CONCLUSIONS

To be able to proactively identify login failures, this thesis detect faults By analyzing data packets, and then through the establishment of a diagnostic Bayesian network model to inference network failure, and finally achieved good results.

ACKNOWLEDGEMENT

This paper was partially supported by the National Key Research and Development Program of China, No. 2018YFB08040505; the Science and Technology Project of State Grid Corporation of China, No. 522722180007.

REFERENCES

[1] Tan Liang, Zhou Mingtian. The design and implementation of a new user login trusted authentication scheme [J]. computer application, 2007.

[2] Yu Lu. research of wireless access user login security[D]. Beijing: Beijing University of Posts and Telecommunications, 2014.

[3] Yang Qiru. Realization and optimization of PPPoE [D]. Chengdu: University of Electronic Science and technology, 2006.

[4] Lin Meijun. Research and improvement of PPPoE protocol [D]. Chengdu: University of Electronic Science and technology, 2014.

[5] Li Jianchuan. Fault diagnosis and maintenance decision method and its application in Bayesian networks [D]. Changsha: national defence science and Technology University, 2002.

[6] Li Yuchun. Research on remote intelligent fault diagnosis of CNC machine tools based on Bayesian network [D]. Hangzhou: Zhejiang University, 2010.

[7] Wu Xin. Study on power system fault diagnosis based on Improved Bayesian network method [D]. Hangzhou: Zhejiang University, 2005.