# Computer Network Information Discovery Based on Information Fusion

TAO ZHANG, MING-ZENG HU, XIAO-CHUN YUN, YONG-ZHENG ZHANG
Research Center of Computer Network and Information Security Technology
Harbin Institute of Technology
No.92, West Da-Zhi Street, Harbin, 150001
CHINA

*Abstract:* - The existing tools detecting network information can not satisfy the researcher's requirement in both range and precision, therefore the information fusion technology is applied to develop a new tool. Data from Multi-sensors can completely reflect the information of computer network, in this paper we implement the network data collection based on TCP/IP protocol utilizing multi-sensors, fuse these data in data layer, identify system type based on fuzzy logical statistic method, obtain port and network service information with the support of system knowledge database, and get the network topology information utilizing the known information.

*Key-Words:* - Information Discovery; Information Fusion; TCP/IP; Sensor; Active Probe; Passive Detect

## 1 Introduction

As the high-speed network grows rapidly, scale of network extends quickly, network devices and services increase and the attack methods to network update ceaselessly, the demand for real-time, security, scalability and usability of network security evaluation system become higher and higher. Especially, when we analyze the network security in a large scale, high speed and multiple management domains network environment, the high capability for data analysis and processing to detect the system information and various network devices and hosts accurately is required.

Information fusion comprehensively processes data from multiple sensors and multiple resources, thereby making more accurate and more generally believable conclusion on object system. Based on this theory, this article presents a computer network information discovery method and we designs and implements a prototype system for system security evaluation.

## 2 Related Works

The method of actively sending packet to explore system information is to extract the information feature of known system, conclude a principle to express, build a "system feature library" in which each feature correspond with a principle, and match the collected object system information with existed principles one by one. For example ISS [1], Nmap [2], Xprobe [3], etc.. But there are some problems existing in accuracy when these tools are used to explore, and it's easy to be influenced by network filtering devices. On the other hand, a mass of data packet will influence the common work of network, and make this kind of tools function discontinuously, which is limited by time deeply.

The passive detection can analyze the system information and existed vulnerabilities too [4-5], its basic working principle is similar with an intrusion detection system based on rule matching, but there is big difference between rules and matching. This kind of tools can run at the gate of network incessantly, monitor the communication between outside and inside of network, get the information of system topology, services and ports by packets capturing, protocol recuperation and analysis which cover the shortage of discontinuously working of active scan tools. But the limitation of this kind of tools is space, it can't capture the internal communication of the system, if the monitor point is placed inside of the network, it just can capture a part of the data because of the using of switch network.

The researchers already started using information fusion which is based on multiple resources data capture, analysis, classification and decision on network security analysis. University of Massachusetts excogitated a decision tree algorithm to analyses UNIX system logs, which uses various signal detection technologies to check if intrusions occur. In order to increase the accuracy of detection, Nong Ye uses decision level fusion to acquired result weights of various intrusion detection tools [6]. Based on EMERAL, Alfonso Valdes raised an

association decision for intrusion detection tools, which include three levels: event aggregation, sensor coupling and alarm fusion, and implement a prototype system [7]. Tim Bass proposed to integrate various attack detection tools to implement the prototype of distributed network intrusion detection, and he analyzed the impossible network intrusion events comprehensively for data, information and knowledge.

# 3 System Design

As the informationalization progresses continuously, the network scale in management domain extends ceaselessly, network devices and services increase, the data speed increases continuously, the structure of information system has some stability, and at the same time the situation that has dynamic change appears. Face these characteristics, the structure of computer network information discovery system based on information fusion which is proposed in this article is shown in Fig.1, which uses a structure of distributed detection, statistical analysis and centralized management.

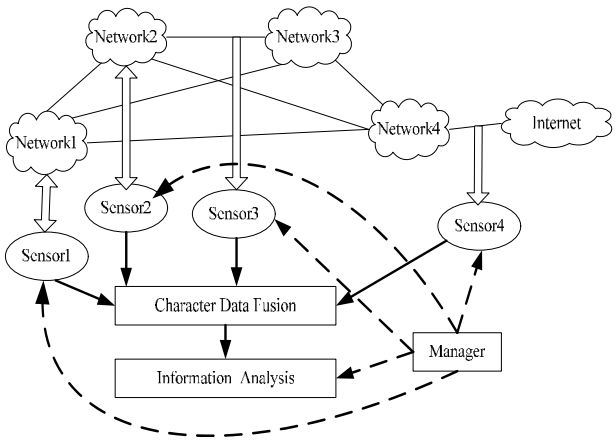


Fig.1 System Design Frame

According to the size of management domain and detection demand, management system sets multiple and multiple kinds sensors to capture network data, in order to fulfill the time and space requirement of network data collection and define, configure and modify the sensor features of working time and working scope. Additionally, it manages the communication, data exchange and dispatch of implementation order between modules.

The sensors' responsibility is collecting and filtering the network packets and do initial analysis, and produce the initial network future data according to the uniform data expressive format.

Character data fuser implements information intelligent identification and keywords match on network character data according to characters of protocol stack and application, judges the system type and port services information of the object, stores them according to the defined format and get the network topology information according to IP, routers and subnet mask.

## 3.1 Sensor Types

In order to implement the full-scale information discovery on information system, we must avoid the limitation on time and space when use single detect tool to capture data. In this system we designed two kinds of network data sensors which are described above, passive sensors and active sensors. We set up monitor point at network gate or some position, capture the data flowed by, analysis the protocols and provide necessary information, which is shown as sensor $S_1$ in Fig.2. Another kind of active detect sensor is shown as sensor $S_2$ in Fig.2, which provide necessary information based on answer messages, and is used to scan the network in fine granularity.

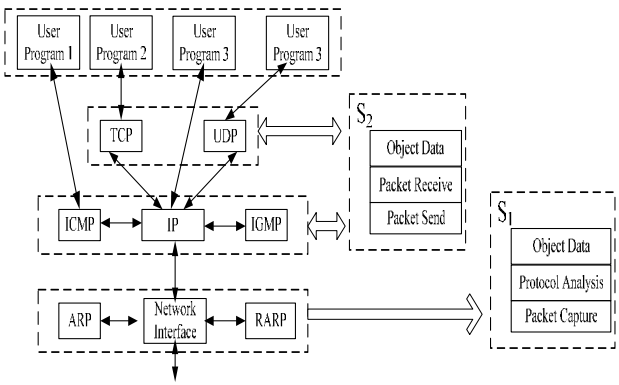


Fig.2 Sensor Classes

As shown in Fig.1, n sensors nodes, $S_1, S_2, \ldots, S_n$ represent n different type data capture tools distributed in network which collect initial data, $Y_1, Y_2, \ldots, Y_n$, then process at partial node, send the captured information $u_1, u_2, \ldots, u_n$ to character data fusion center, character data fusion center make decision according to the experimenting methods. At the same time, considering the change of information system scale, the suitable coverage scope can be provided by adding or removing sensors.

## 3.2 Data Fusion Method

The data level fusion provides the detailed information which other fusion levels can't provide, and fusion methods of data level can analyze and

process data from sensors directly. The advantage of the fusion is that it can keep the native data as much as possible. On the other hand, the load of data communication is big.

### 3.2.1 System Type Identify

Because the implementation of TCP/IP protocol stack is different in various systems [8-9], we can use some character packets to identify the different system types. The character packet type include: FIN Probing, TCP ISN Sampling, TCP Option, no-section mark, TCP initialization window, ACK value, ICMP Error Quoting, ICMP Error Message Echo Integrity, ICMP Error Message Type of Service, TCP Timestamp, SYN flood limit, etc.

As the setting of network filter devices and the difference of topology structure, there is a problem of accuracy existed in the character packets captured by the sensor. And at the same time, character packets have some fuzzy degree which is influenced by the system configurations and the installed software. The recognisation with high accuracy method described in this article uses a statistic method based on the fuzzy logic by multi-dynamic character packet types.

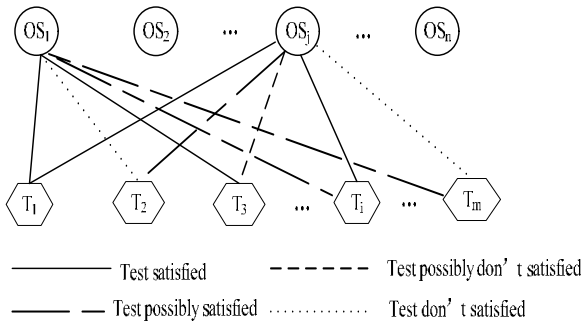


Fig.3 Character Packets and Test Result

Assume Ti as a method example using one kind of character packet, there are m kinds of sample, all of the test method examples can be represented by the test method set $(T_1, T_2, \ldots, T_i, \ldots, T_m)$. $OS_j$ is a operation system type, there are n kinds of operation systems $(OS_1, OS_2, \ldots, OS_j, \ldots, OS_n)$. The match result of system type to a testing sample can be divided into: test satisfied, test possible satisfied, test possible don't satisfied and test unsatisfied, as shown in Fig.3.

Corresponding to the return result of test method example $T_i$, the probability of the object system type belong to operation system $OS_j$ can be represented by $\mu_{(i,j)}$, then we can use $(\mu_{(i,1)}, \mu_{(i,2)}, \ldots, \mu_{(i,j)}, \ldots \mu_{(i,n)})$ to represent the degree of this test sample belong to all operation system type. The correspond of object system type to testing sample is shown in Table 1, the known apriori knowledge $\mu_{(i,j)}$ value can be got according to Table 2.

Table 1Correspond of object system type to character packet example

| | OS1 | OS2 | ... | OSn |
|---|---|---|---|---|
| T1 | μ (1,1) | μ (1,2) | ... | μ (1,n) |
| T2 | μ (2,1) | μ (2,2) | ... | μ (2,n) |
| ... | ... | ... | ... | ... |
| Tm | μ (m,1) | μ (m,2) | ... | μ (m,n) |

Table 2 Apriori knowledge of μ(i,j) value

| μ(i,j) | Description |
|---|---|
| 0 | Ti don't satisfy the character of OSj |
| 0.3—0.4 | Ti maybe don't satisfy the character of OSj |
| 0.7—0.8 | Ti maybe satisfy the character of OSj |
| 1 | Ti satisfy the character of OSj |

For one object detection, to assign value for $\mu_{(i,j)}$ according to the corresponding response in Table 3, object system type $OS_x$ will be decided if the condition can be fulfilled:

$$x = \{j \mid \sum_{i=1}^{m} m_{(i,j)} == (\mathrm{MAX}(\sum_{i=1}^{m} m_{(i,1)}, \sum_{i=1}^{m} m_{(i,2)}, \ldots, \sum_{i=1}^{m} m_{(i,j)}, \ldots, \sum_{i=1}^{m} m_{(i,n)})\} \quad (1)$$

This method can increase the detection precision effectively. But its shortcoming is that it needs too many character packets. An effective improved method is to combine apriori knowledge and experiment results, dynamically delete the testing samples which influence the $\sum\mu_{(i,j)}$ （j=1,2,3,…,n） lightly and add the testing samples which influence it weightily, and increase or reduce the amount of $\mu_{(i,j)}$ according to the great amount of statistic data. Compared with traditional strict match method, this method has a stronger anti-jamming ability.

### 3.2.2 Port and Service Identify

According to the definition of RFC and default rule, the ports and open services are regulated in detail, for example, port 80 for HTTP service, port 21 for FTP service. After the open port information gotten, the corresponding service can be mapped using port service mapping table. When the server and the client make the initial connection, generally, the interacted server and client name, version, developer and etc. can be provided. In order to get the service

information detail, sensors capture the corresponding packets which can implement TCP complete connection or UDP connection and protocol analysis. The fuser implements keyword match between the captured data and the server keyword list stored in system knowledge library, so recognizes corresponding service information. This system can complete most recognisation works of services.

### 3.2.3 Topology Build

For building topology structure, the host IP, the router and the subnet mask are needed. The IP is assigned by user. For router information, the tradition Traceroute program use ICMP or UDP packet to implement route tracking. In order to reduce the influence of packet filter device to data collection, this system uses ICMP, UDP and TCP SYN packet to detect route information and subnet mask of object system, sensors of passive detection type provide the supplementary information of subnet mask. Combining the completed recognisation of computer network information above, we can connect network and host, and build the topology of information system.

## 4 System Implement

The implement of prototype system based on information fusion described in this article is shown as Fig.4.

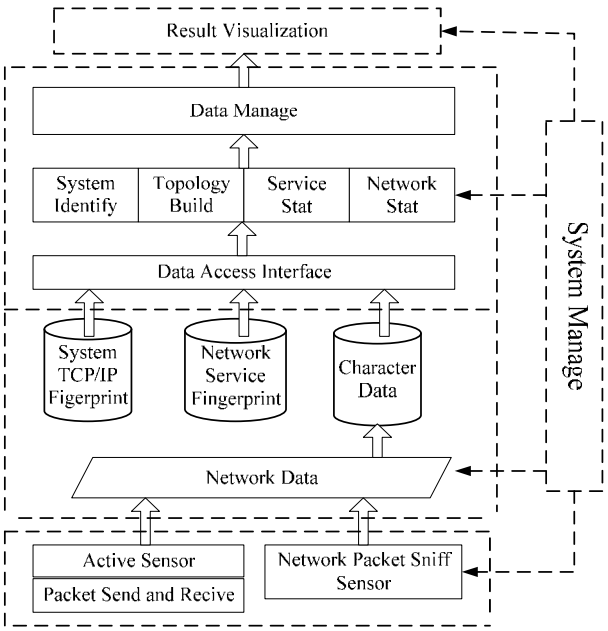


Fig.4 System Implement

The system is constructed by four parts:

(1) Sensor module: According to detection requirements, sensors are arranged in some data collect points on network to fulfill the spatio-temporal demand for the data discovery, and provide and arrange the data by the defined formats.

(2) Data fusion module: It implements information intelligent recognisation and keyword match based on the protocol stack and the feature of application program, judge the type of the object system, the port service information and the topology information. It provides the information to the model of network security analysis according to the defined format.

(3) System knowledge library: This module stores system vulnerability, attack knowledge, operation system protocol stack fingerprint features, common port service mapping information, network service keywords, common backdoor program features.

(4) System management and control module: a) Managing the communication, data exchange and dispatch of implementation order. b) Defining, configuring and modifying sensor work time, work scope according to features of detecting object and demand of detecting information. c） Displaying result via visual interface.

## 5 Conclusions

In this article, we combine the active detection and passive monitor to capture network data, using the advantages of them to increase the scalability, fault-tolerant and usablity. In the system, using high precision recognisation technology and detect algorithm to analyze the network data, increase the detection precision of network information, and combining multiple resource provide a synthetical and full-scale computer network information.

*References:*


[1] ISS. http://www.iss.com/, 2005.

[2] NMAP. http://www.insecure.org/nmap/index.html, 2005.

[3] Ofir Arkin. ICMP Usage in Scanning. Version 3.0, June 2002.

[4] Libpcap. http://www.tcpdump.org/, 2005.

[5] Passive Vulnerability Scanning. http://www.tenablesecurity.com/nevo.html,2005.

[6] Nong Ye. Giordano, J. Feldman, J. Qiu Zhong. Information fusion techniques for

network intrusion detection. Information Technology Conference, 1998. IEEE.
[7] Alfonso Valdes, Keith Skinner. An Approach to Sensor Correlation. 2000.
[8] W.R. Stevens. TCP/IP Illustrated, Vol. 1 Addison-Wesley, 1994.
[9] Marco de Vivo, Eddy Carrasco. A review of port scanning techniques. ACM SIGCOMM Computer Communication Review, Volume 29, Issue 2, April 1999.