Research on Application of Database Forensics Technology

Qin Liu

Department of Information Science and Technology, East China University of Political Science and Law, Shanghai 200042, China

Email: liuqin@ecupl.edu.cn

Keywords: Computer Forensics, Database Forensics, SQL Server

Abstract: Electronic evidence has become one of the new litigation evidence. Computer forensics as an effective means of obtaining electronic evidence has been widely studied. Database forensics is a new field of computer forensics. Based on the specific database products, this paper gives the common database attack behavior, analyzes the possible traces in the database, and summarizes the scope of the collection of evidence and commonly used analysis methods.

Introduction

With the wide use of computer and Internet, attack targeting at computer and Internet is also getting more and more severe. Computer has been using in more and more criminal activities, and cases in court relating to computer is also becoming commonly seen. The electronic evidence existing in computer and its related peripheral has become one of the new types of evidences[1]. Currently, among various computer applications, the way of storing data in database is the choice of most applications. Therefore, the scope of electronic evidence acquiring not only contains the traditional file system, memory data, and network data, but also shall include effective acquiring and analysis on data in database[2]. Based on this, the study on the evidence taking from database has become more and more urgent and important. Considering that different database products vary in memory structure and working principle, this paper would take mainstream database product SQL Server 2008 for instance to introduce the method and process of taking evidence from database.

Computer Forensics

There's no unified and correct definition of computer forensics. And currently, the most widely recognized concept is: computer forensics refers to the process of determining, collecting, protecting, analyzing, archiving and presenting on court of the electronic evidences that are acceptable by the court, reliable, and convictive, and existing in computer and its related peripheral devices. The purpose of forensics is to find out the invader (or invading machine) by the electronic evidence, and explain the invasion process. While cracking down computer-related criminals, the law-enforcing authorities haven't got a unified and standard procedure for computer forensics. Commonly, under the precondition of guaranteeing basic principles (timeliness, legality, standard, security, and completion, etc.), the computer forensics is divided into four steps: scene protection and investigation, evidence obtaining, evidence appraisal, evidence analysis, and evidence submission. The specific contents are as follow:

i) Scene protection and investigation: similar to other criminal forms, it needs to cordon off the scene and freeze the spot evidence. More specifically, it refers to a kind of detecting activity that the detectives adopt computer technical measures, investigating and visiting methods to inspect and search the places, articles, suspects, defendants, and persons that may possibly hide evidence of crimes, as well as detain and seal evidence materials relating to crimes according to *the Regulations on Processing of Criminal Cases of Public Security Authorities* and *the Regulations on Processing of Administrative Cases of Public Security Authorities.*

ii) Forensics: it refers to the process of identifying the electronic data in the physical devices that may contain electronic evidence and collecting the electronic data, or directly collecting the electronic data by technical measures. In essence, it is to find out the certain things from numerous unknown or uncertain things. Therefore, the task of this phase is to save all electronic data. Though there are lots of data sources, they can be approximately divided into two aspects: host and network. For example: system log, anti-virus log, database files, memory data, chat log, network packet, and firewall log, etc. As is known to all, database management system is a must to lots of application systems currently. The main core data of the application system is put in the database. Therefore, the evidence taking from database is of great importance. Specifically, it needs to collect the information about database system version, database data of memory, temporary database, event log, error log, database file, log file, audit and track file, etc.

iii) Evidence appraisal: the appraisal of computer evidence is mainly to realize appraisal of evidence completion. One of the hard points of computer forensics is to prove that the evidence collected by the forensics personnel hasn't been modified. However, the computer acquired evidence is easily to be changed or destroyed. If the acquired electronic data is not properly protected, it can be easily destroyed or even lost. Therefore, during the forensics process, necessary measures shall be taken to protect the evidence. The commonly used protective technologies include: chain of custody, digital time stamping technology, digital fingerprint technology, and data encryption technology, etc.

iv) Evidence analysis: evidence analysis is the core of the computer forensics. It mainly analyzes the acquired data and determines the types of the evidences, including inspecting files and directory contents and resuming the deleted contents, analyzing the computer type and adopted operating system to see whether it is multi-OS or not, or having or not hidden division; having or not suspected peripheral devices; having or not remote control, Trojan program; and the status of network environment; and it also deducts conclusions according to the found evidence by scientific method. For database forensics, the core is to analyze and find out the suspected content data, derived data, environment data, communication data from the acquired data according to the crime clue, thus to form effective evidence chain.

v) Evidence presentation: it is mainly to print the overall analysis and tracking results of the targeted computer system, all possibly useful files, and the list of the mined file data, thus to give out the analysis conclusion, mainly concerning computer crime date and time, hardware disc partition, OS version, completion of data and OS when running the forensics tool, virus evaluation, found file structure, data, and author information, any attempt to hide, delete, protect, encrypt to information, and other relative information found during the investigation process; mark the extraction time, site, machine, extractor and witness; and give out necessary expert certificate or testimony on court; and finally present to the judicial authority according to legal procedure in form of evidence.

In a word, the process of computer forensics concerns lots of aspects. This paper only targets at the database forensics process. Since the database products are different, for each specific forensics step, the concerned methods are also different. This paper takes Server 2008 for instance to introduce and study the process of forensics from database.

Analysis on Database Attacks

Brute-force Breaking Password: SQL Server database provides two modes of ID authentication, an integrated mode: user of operating system is also the user of database, that is, as long as the user logs in to the operating system, it could directly connect to SQL server via trusted links[3]. Because the Windows authentication uses Kerberos security protocol, so it is the default mode to provide the password policy enforcement according to the complexity of strong password authentication, provide support of account lock, and support password expiration. The other is the mixing mode: it allows users to use Windows authentication and SQL Server authentication to connect. When choosing a

mixing mode authentication, it needs to enter the system administrator password (Sa). SQL Server is a Winsock application. Once the SQL Server service is started, monitor connection request would be proposed on a specific port: the default port is 1433. Safety audit mechanism of user login failure times limit is adopted in SQL Server, which makes malicious attackers possible perform brute-force breaking of Sa user password by generating a huge number of password dictionary to break the known account as Sa of SQL Server. This kind of attacks usually adopts tools as Xscan and 1433 chicken tools etc.

In order to prevent brute-force password breaking, it can add certain login restriction or authentication code to web application.

SQL Injection: SQL injection means that the attacker constructs special SQL commands and submit them to Web applications to deceit the server to execute malicious SQL command, thus to reach the purpose of controlling database administration privilege and modifying data.

The main SQL injection mode is to insert the malicious code into the user input variables, thus to make the application have no security check and consider it as legal SQL command to execute. Please see the following code:

Dim name

name=Request.form(“name”)

Sql=”select * from goods where name=’”&name&”’”

Attacker inputs: Wonderful World’;drop table goods—

The script would generate such code: select * from goods where name=’Wonderful World’;drop table goods--‘

The program was to acquire the commodity names that the user wants to inquire according to the web table thus to display the commodity information. But the attacker destroys the data and deletes the commodity information table by drop command.

Various SQL sentences could be constructed manually for SQL injection. Or lots of modularized and automatic tools can be adopted. For example: SSQLInjection, NBSI, and HDSID, etc.

The most easy and simple way to prevent SQL injection is to restrict user input, for example, not allow user to input single quote mark“ ’ ” and comment symbol“--”; secondly, the user may also program to not use the above mentioned Mosaic SQL, but parameterized SQL sentences; besides, it is better to use the safe parameters that the SQL Server database is with.

CC Attack: CC refers to the way that the attacker takes use of agent server to generate legal request targeting at the victim machine in order to realize DOS (denial of service) and masquerading. CC attacks mainly send requests crazily to places of Web application consuming lots of resources, which can result in waste of server resource and 100% occupation of CPU for long time. It leaves numerous data connection requests till appearing net jam and normal access terminates. Compared with other DOS attacks, CC attack makes you hardly see the IP of the real source or abnormal flux, but it could use up the server resource to make it unable to normally connect. Though CC attack can not break down the database, it could result in slow response of database. It is a kind of key database attack measure. The commonly used CC attack tools include CC Attacker and Variable CC Attacker, etc.

CC attack can be prevented by multiple methods, including forbidding website agent access, restrict connection quantity, and modify maximum timeout time, etc[4].

Though it could prevent attacks to some extent, it is impossible to completely eradicate illegal access and damage. Once the database is infracted, judicial personnel shall collect and analyze evidences according to the computer forensics steps and principles.

Evidence Collection and Analysis

Database is the system software running on the operating system. In order to guarantee the completion and coherence, during the forensics process, it needs not only to obtain the evidence in the database, but also obtain the evidence in the operating system. However, when the crime happens, as time goes by, some evidences may be covered legally or by malicious data. Therefore, during the process, it not only needs to determine the scope of forensics, but also needs to determine the sequence of data extracting, namely the priority (evidences of lower priority shall be collected earlier). Commonly, sequence shown in Table 1 can be taken for evidence collecting[5]:

Table 1 Scope of Evidence Collecting

Level	Category	Description	Priority
OS	Basic Data	Version, Active Users, Active Request, etc.	2
	Memory Data	Data stored in memory and cache memory	1
	System Event Logs	System logs, application logs, security logs, etc.	4
Database Management System	Basic Data	Version, active users, etc.	2
	Database Files	Contain all data objects in database	5
	Event Logs	Record all changes on database	4
	Error Logs	Record all current or potential problems of database	6
	Track Files	Record the recorded track data under database audit tracking system	7
	Temporary Table Space	Temporary storage table, temporary storage process	3

i) Basic data: it includes data of operating system and database versions and is used to get to know the basic architecture of the whole application service. Usually, on the server, multiple database systems are run at the same time. So the evidence taking personnel shall find out the value and name of the investigated database according to the configuration of the application server.

ii) Memory data: it mainly refers to the data in memory and high speed cache memory. When the personnel is taking evidence, if the system is not shutdown and allows online investigation, then the memory data shall be the one that needs to be collected firstly.

iii) System event logs: it includes the logs of operating system, logs of application, and security logs. If the web service is started, service logs shall also be collected. Taking Windows10 for instance, it allows viewing the event list in well classified types via “Event Viewer”, or there’re also log files of relative evt type in system file folder.

iv) Database files and event logs: all objects of database are stored in the data files. The log files keep all update operations to database. They are the main collecting contents for database forensics. Different database products vary in data organizing modes. Taking SQL Server 2008 for instance, the database files mainly include data files (.mdf files) and several less important database files (.ndf files). And the log files are 1df files.

v) Error logs: it records all current or potential problems existing in the database system, including inner error message, automatic message resuming of database reboot or other server level error messages. The 7 ErrorLog files stored in SQL Server would be helpful for forensics.

vi) Track files: it records various operating information of the monitored database. The contents the contents of the track files have relation with the database system and the user set audit tracking policy. In SQL Server, the track file is defaulted as being stored under the installation directory with

".trc" suffix. And the system is defaulted to provide 5 track files. Each track file is defaulted as 20MB. The system would maintain the 5 files itself. When it is rebooted or reaches the maximum value, it would re-generate new files and delete the earliest files, thus to update in proper order.

vii) Temporary table space: the temporary database is used to provide space for all temporary table, temporary storage process, and other temporary operating. The temporary database of SQL Server is tempDB. The temporary database is also stored on hardware disc, but it is temporary. When the SQL Server is rebooted, the system would recreate new and empty tempdb database.

The evidence takers collect lots of basic data. However, it is a kind of very fuzzy work to find out data files relating to the case from numerous data. The preliminary filter can be done from two aspects: time and relevance of elements. For example: determine the time period of illegal attack happens according to the case clues, thus to lock routine, table, and operation, minimizing the scope of search.

Case Analysis

The database server of a certain company is attacked by hacker. And part of the data in the business database (database name: business) is deleted. The detectives firstly get to know the architecture, the started service, developed port, operating system, and various basic server data of the system:

i) Database server: SQL Server2008[6], open TCP1433 port (SQL Server default port), Windows Server2008 OS.

ii) Web Server: IIS6.0, open TCP80 port

The rough attack process is: the attacker takes use of the port scanning program to find that the host opens 1433 port, and perform brute-force password decoding to account sa. When getting the password, it destroyes business database. Then the detectives collect relative evidence materials according to the contents described in Section 4.

i)Memory data: SQL Server adopts resuming technology with Checkpoint, so that the record of logs after the checkpoint is stored in the cache memory temporarily. The active logs in the memory can be read by DBCC LOG command.

ii) Fixing of hardware disc evidence: use the disc mirroring tools to mirror relative files according to byte (for example: database files, logs, error logs, and track files, etc.). And meanwhile, perform MD5 HASH check to mirroring files to guarantee that the data hasn't been changed during the copy process.

After acquiring, it needs to analyze the data from two aspects:

i) Analyze from aspect of time: firstly, error log records the validation data of failure of SQL Server login. It can be analyzed from the error log that, during a certain time period, a series of IPs trigger sa account login failure events, and a successive timepoint sa login succeeds, then it can perform analysis on the client operation on database after this time period. Use command "select * from sys.traces" to determine the Trace file that is using by the database currently; use Trace file analysis tool provided by SQL Server to extract this Trace file to verify the brute-force decoding of the sa account password. And meanwhile, it can analyze and obtain the happening time period, SPID and certain operation made to a certain dataset (since the database uses default trace policy and hasn't performed any trace to specific modification, the trace file analyzing would only analyze a certain operation made to database).

ii) Log content analysis: the key point shall be the operating of the above mentioned SPID routine to database. Log content analysis adopts build-in function "DBCC LOG" command to view specific database event log; it uses BDCC LOG (business,3) command to view the illegal operation of specified SPID attack event to database. There're lots of fields in log file. Those relating to the case

are roughly: AllocUnitName, which is the modified table name; Page ID, which is the number of storing page; Slot ID, which is number of record item of data in the page; and Partition ID, which is the ID of partition of data in the page. After positioning which table, page, and partition the SPID modifies, it shall use DBCC PAGE command to analyze the data record page to obtain the values before and after the illegal modification made of routine to database.

Conclusions

This paper takes SQL Server database management system for instance to introduce the application of computer forensics technology in aspect of database. Though different database products are different in storage structure and working principle, the basic principles and steps are of correlates. Huge amount of data is stored in database. And during the forensics process, it usually collects numerous data, which would cost lots the detectives' lot of time to find out clues. Moreover, it mainly relies on the experience of personnel handling the case. With the optimization of data analysis, especially the big data analysis technology, such status could be improved. This is also the key working point of the next step work of the project team.

References

[1] Ayers R, Brothers S, Jansen W. Guidelines on mobile device forensics. NIST Special Publication. 80(1)(2013) 101-104.

[2] Ding Li-ping, Wang Yong-ji. Study on Relevant Law and Technology Issues about Computer Forensics. Journal of Software. 16 (2)(2005) 260-275.

[3] Dong Xian-hui. Evidence Collection and Analysis of SQL Server Database. Chongqing University. 2013

[4] Brian Carrier. File System Forensic Analysis. Boston: Addison Wesley Professional, 2005:

16-27

[5] Song Lei, Zhang Peijing. Study on database system forensic. Police Technolgy. 2011.1 37-39.

[6] SQL Server 2008 Books Online on https://msdn.microsoft.com/en-us/sqlserver/cc514207.aspx

