Send Orders for Reprints to reprints@benthamscience.ae

The Open Automation and Control Systems Journal, 2014, *6*, 919-926

Research on Data Leak Protection Technology Based on Trusted Platform

Wang Xingkui* and Peng Xinguang

College of Computer Science and Technology, Taiyuan University of Technology, Taiyuan, China

Abstract: In order to guard against the leakage of important data in the system, a new model of data protection is proposed in the paper, which is also targeted at the shortcomings in the existing data protection technology in the field of data leakage prevention. By taking advantage of TPM data protection technique as well as the combination of symmetric encryption and asymmetric encryption, a new data protection method that is based on Trusted Platform has been also established. Once the encrypted data and platform configuration value is sealed, as long as the system state value is subject to change, the encrypted data will never be revealed and thus the data safety will be preserved, whether it is data stealing or active leakage. According to the test results, this method of data protection will be of great use to prevent the active leakage through the network or mobile storage device.

Keywords: Data leak Prevention (DLP), data seal, trusted computing, trusted platform module, platform configuration register(PCR).

1. INTRODUCTION

As the information resource becomes more and more digital and network oriented, it has also given rise to an increasing possibility of the data stealing, leakage and damage. According to the 27th statistical report of the state of China's Internet development issued by China Internet Network Information Center(CNNIC) on January 19, 2011, the number of Chinese citizens who have suffered from the attack of computer virus or Trojan totaled up to 209 million in 2010 alone, accounting for 45.8 percent of its online population. For those people that once had their account or password stolen, the total number also came to 99.69 million, occupying 21.8 percent of its online population. In the year of 2008 alone, there were about 250 million pieces of records concerning the data safety. What is more, the Data Break Investigation Report (DBIR) made by Versizon in 2012 has also sounded the alarm for the great number of large-and-medium companies in China to pay more attention to the safety of their database. In November 2011, some huge Chinese E-commerce websites including CSDN, Tianya, Paypal and Dangdang were left in a dilemma to find that a large amount of their user's information were exposed in the network, which also marks the largest event of information leakage in China's internet history. From the perspective of leakage channel, the data leakage is in many cases initiated by the internal staff, hacker or third-party developer to steal user's information from the database and sell it in the network with the purpose of gaining profits. According to relevant data, the total number of Chinese citizens that once had their personal information disclosed has almost come to 120 million, which also makes the effort of enhancing system safety, implementing safety standard and improving the capabilities of preventing invasion, attack and theft quite an urgent task. On December 29, 2012, the Ministry of Industry and Information Technology of China called on the relevant websites in its notice the about recent events of user's information leakage to put great emphasis on the issue of data safety and also required them to carry out the data encryption as soon as possible.

2. DATA LEAK PREVENTION

As an important part of information safety, Data Leakage Prevention (DLP) has also emerged as a hot issue. According to the research report of the safety issue among global financial institutions in 2010 issued by Deloitte, which is targeted at the safety protection technology developed, implemented or planned by more than 350 large-and-medium financial institutions distributed in 45 countries around the world, the technology of DLP has stood out as the most promising and feasible solution, besides the security log, event management system and data encryption.

The DLP is referred to as the technology that aims to guard against the leakage of user's data or information, whether it is intentional or accidental, that goes against the safety strategy. So far, many scholars and research centers both at home and abroad have carried out quite a large number of investigations into DLP and also come up with some valuable solutions. In terms of the technology adopted, these solutions can be classified as control technology, encryption technology, content supervision, filtering technology and combination strategy. Considering that most of the technologies are still centered on the software at the present stage, they are inevitably accompanied by some shortcomings. For example, some problems featured by passive leakage can not be solved effectively, usually seen in the loss of mobile

1874-4443/14

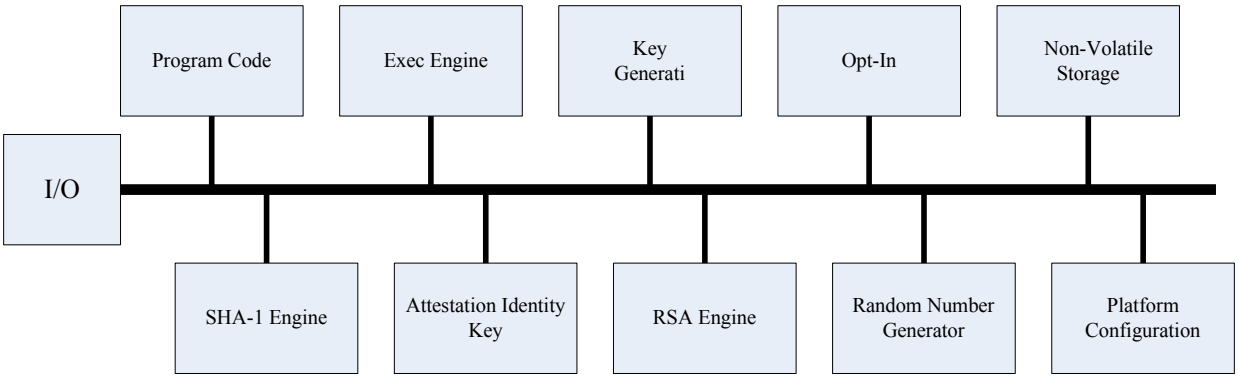

Fig. (1). Structure of TPM.

storage device, mobile end or laptop. What is more, some technologies also involve the bottom layer of the operating system, which means that it will be complicated to carry out and even conflict with other driver programs. As to the active or intentional leakage, it can not be solved by means of network. And the data transmitted through the external storage medium can not be protected as well. Therefore, it has been quite necessary for us to work out a reliable solution to provide protection for the sensitive data stored in the computer.

3. TRUSRED COMPUTING

According to Trusted Computing Group (TCG), the trusted behavior is referred to as expected behavior that is conducted in certain mode so as to achieve a specific goal [1]. The trusted computing platform is defined as the computing platform that is able to report the system status in an accurate manner. Judging by two basic concepts of TCG, the trusted behavior of the program is more of its predictability. In order to determine whether the program is running in the expected mode, we should first set up a mode of expected behavior as a reference so as to compare it with the actual running mode. As a trusted computing platform, the TPM will also provide a root of trust formeasurement (RTM) that is in charge of measuring the integrity of each unit in the trusted computing platform. What is more, the actual running mode of the program will be also used as the value of RTM so as to build up the trusted link of the trusted computing platform. On the whole, the measurement of integrity has been considered as the prerequisite to determine whether the computing platform is trusted and report the system status accurately or not.

At the present stage, the technology of trusted computing (TC) has evident advantage over many others in a variety of fields such as the safe startup in the end platform, the building of trusted domain, safe storage and key management. Based on a range of technologies such as the hardware-level key management center, trusted authentication, trusted measurement, trusted storage and trusted network, TC has been able to provide a solid technical basis for some key issues in the filed of DLP. Therefore, this paper will make a study of the protection over sensitive data in the computer.

3.1. Trusted Platform Module

The core part of trusted computing is known as the trusted platform module (TPM), which is also referred to as the trusted origin and aims to turn personal computer into a safe and reliable platform. TPM is in essence a device capable of key generation, encrypt and decrypt algorithms. What is more, it is also equipped with a separate processor and storage unit that contains the key and sensitive data as well as provides a series of services such as the integrity measurement, data security protection and identity authentication [2]. The Fig. (**1**) is about the structure of TPM.

A typical application of TPM is to measure the boot sequence of the recording system. As the system starts up, the right of control over the machine will be transferred between BIOS, Boot loader, the core of operating system, the external program and application program. If the malicious code is able to intercept the right of control at any link of the boot sequence, it will also manage to fiddle with the following part of the boot sequence. For example, a malicious program of Boot loader like GRUG will seek to load a tampered Linux kernel image without even being noticed by the user. After the startup of this Linux kernel image, the attacker will have access to the right of control over the whole platform and thus compromise its integrity and confidentiality and even steal the sensitive data. In other words, the damage on any link of the boot sequence will put all or part of applications of the platform under threat. Therefore, we have to work out an effective trust mechanism so as to examine whether the system is fiddled by attacker during the course of startup. What is more, the integrity measurement mechanism provided by TPM hardware can also conduct trust evaluation during the startup so as to enable the user to make sound judgment after the system is fiddled.

3.2. Trusted Chain

The core part of the evaluation mechanism of TMP upon boot sequence is its trusted chain [3]. According to the theory of trusted chain, a trusted link will be used to evaluate the security of the next chain before handing over the right of control to the next chain and so forth. In the trusted chain, the right of control will be passed on between different trusted objects.

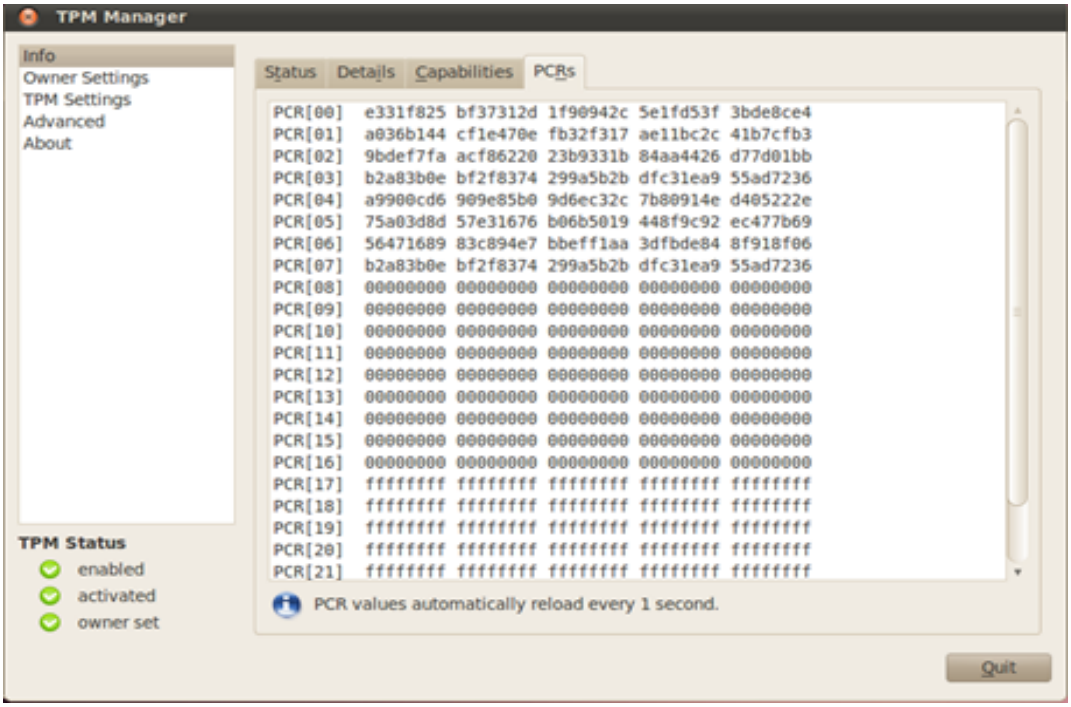

Fig. (2). Measurement value of TPM.

During the startup, BIOS module as the source of the trusted chain will input the self-checking configuration information into Storage Measurement Log (SML) before measuring the self-checking BIOS, whose value will be extended into the PCR corresponding to the TPM [4]. After the measurement over the BIOS is done, the right of control implementation will be handed over to the self-checking BISO. In the next stage, the self-checking BISO will first measure the Boot loader such as GRUB. This alternate implementation will ensure that each part of the startup should be measured with its measurement value recorded in each PCR of TPM and configuration information about each startup module stored in the measurement log. The Fig. (**2**) is measurement value recorded in each PCR of TPM.

3.3. Platform Configuration Register

The platform configuration register (PCR) is located inside the TPM and used for recording the operating condition of the system. More specifically, the operating condition of the platform has involved a large amount of information including kernel image, process information list and binary executable program. It should be noted, however, the storage capacity of the TPM is so limited that it is only able to store the abstract of the configuration, which means that the information stored in the PCR is the Hash value calculated by the SHA-1 algorithm. SHA-1 is also the cryptographic hash functions used by TCG and can produce a fixed-length (16bits) output value (Hash value) upon receipt of any input value with arbitrary length. What is more, even a difference as small as 1 bit in the input value will lead to a totally different Hash value. What makes it even better is that Hash function is one-way. Therefore, it is quite easy to determine the Hash vale according to the input value. But it is impossible to make the reverse operation, which also ensures the reliability of the operating condition recorded in PCR. In addition, the PCR has also recorded a range of values about the configuration of the software and hardware, which will also make use of several PCR during the process of startup. In this paper, the PCR[i] denotes the ith PCR. As to the usage of each PCR defined by TCG, it will be illustrated in Table **1**.

In order to protect the configuration value recorded in PCR from tampering or forging by the malicious code, TPM has put a reasonable limitation on the behavior of PCR. As the PCR is located in the TPM, its data is also protected by TPM accordingly. What is more, the read operation toward

Table 1. Platform configuration register standard usage.

PCR	Use
0	Core BIOS,POST BIOS,Embedded Option ROMS
1	Motherboard configuration
2	Option ROM code
3	Option ROM configuration data
4	IPL(Initial Program Loader) code
5	IPL configuration data
6	State transition(sleep,hibernate,and so on)
7	Reserved for OEM
8~15	Not assigned

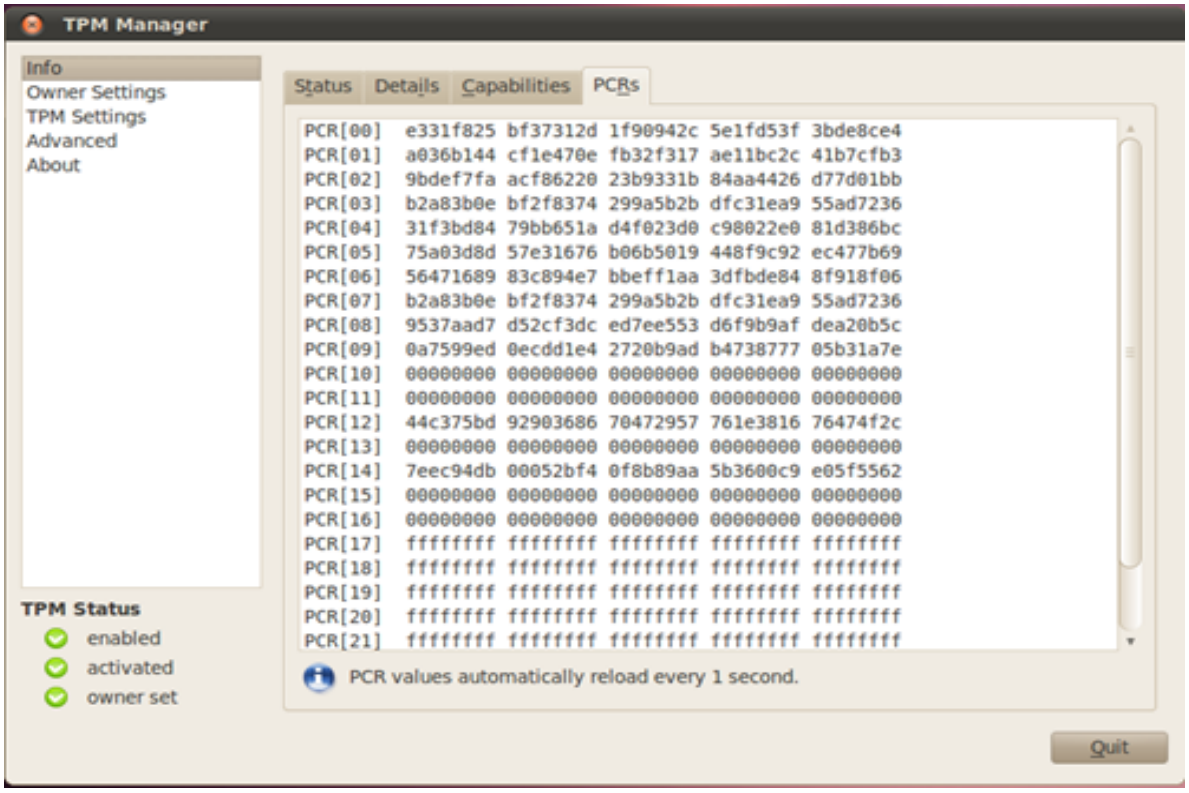

Fig. (3). Measurement value after use extention operation.

PCR content is not subject to this limitation like the modification operation. As to the modification of PCR value, there are only two operations allowed by TPM, which are reset and extension. The reset operation will take place after the machine shuts down or restart, with the PCR subjected to zero clearing in an automatic manner. After the physical machine is powered again, the extension operation will enable the modification of PCR value, only if the physical attack over TPM is not observed. In other words, after the current value of PCR is interfaced with new value, the newly calculated Hash value will be stored in PCR as the new integrity measurement value.The Fig. (**3**) is measurement value after use extention operation.

New PCRi=Hash (Old PCRi ||New Value)

Among them, Old PCRi denotes the PCR value before extension while New PCRi means the one after extension. As to the New Value, it is referred to as the value of integrity measurement [5]. The Hash( ) in here stands for Cryptographic hash algorithm.

The Hash function has the following properties.

If A ≠B, then Hash (A) ≠Hash (B)

If A ≠B, then Hash (A||B) ≠Hash (B||A)

Therefore, by making use of extension, PCR is able to record a infinite-length series of measurement value, which also reflects the change in system state and the integrity of current measurement software [6, 7]. In case that one single measurement value in this extended series is modified, the following part of the series will be affected as well.

3.4. Data Safety Protection

According to TCG, there are a total of seven kinds of key in TPM, some of which are storage key, bind key and sealing key.

- Storage key is a asymmetrical key and used for the encryption of other keys and TPM external data.
- Bind key is used to encrypt a small amount of data in certain platform, such as the asymmetrical key and decrypt it in another TPM platform, since it has to use a specific key to conduct encryption, it should be bound with the platform.
- Sealing key is not only bound with the system, but also connected with certain hardware or software, which means that the encryption or decryption can be initiated only after this condition matches.

On the whole, the proper use of the keys above will enhance the data safety greatly. Moreover, since TPM and RTS is mainly targeted at the storage of keys and TPM sensitive data, such as the PCR value, RTS is not suitable for the safe storage of large-amount user's information.

4. TECTION TECHNOLOGY

According to the data, up to 90 percent of data leakage is caused by the intentional behavior of the internal staff. If those sensitive data can only be read in the original computer and not accessed by other computers even with the key, the data safety will be preserved accordingly. So far, the

computer cryptography has grown into an important measure to protect the sensitive data. Given that the encryption and decryption does not rely on the concrete computer end and application environment, this undesired flexibility has also brought about the safety risk during the process of processing sensitive data. Fortunately, the data seal proposed by TCG has sought to bind the sensitive data with the specific computing environment so as to remove this risk [8]. Similarly, we can also apply it into the data leakage prevention for the sake of data protection.

4.1. TPM Data Protection Measures

Although TPM has enabled the creation, storage and management of the asymmetrical key in a convenient way, it is also time-consuming to encrypt it. Therefore, if a large amount of data is required to be encrypted, TPM will not come off as an ideal solution. Given that the private data stored in the device is in many cases huge in number, it has also made the encryption over the asymmetrical key far less effective [9]. Fortunately, some tools such as 3DES, AES and Blowfish have been able to decrypt and encrypt the symmetrical key in a speed hundreds of times faster than asymmetrical key and thus are considered as the suitable option to deal with a large amount of data.

It should be noted that TPM does not support the encryption over symmetrical key but enables the management over the encryption and safety issue [8, 10]. What is more, the bind and seal symmetrical key can not only guard against the unauthorized access, but also store the key safely when it is not used.

4.2. Data Seal

Data seal is another method provided by the trusted technology to encrypt the data. It can not only bind the data with the specific trusted platform, but also associate the data with the specific state of the platform [11, 12]. The seal operation will integrate the sensitive data, PCR value and encrypted key into a whole so as to encrypt the data, which will rely on the state of the trusted platform and can only be decrypted in the trusted condition.

Seal: Supposed that M is a piece of information and K is a symmetrical key, then VPCR-x will be a group of PCR GPCR-x in the certain TPM the value . The public key and private key of the asymmetrical key of the TPM will be KPUB-S and KPRI-S respectively.

The definition of the seal will be illustrated as follows:

- Use symmetrical key K to encrypt the information M and lead to *K {M}*.
- Use KPUB-S to encrypt VPCR-x and K, resulting in *KPUB-S{VPCR-x,K}*.

If the M is to be obtained, the K should be acquired first, which means that TPM must use KPUB-S to decrypt the KPUB-S{VPCR-x,K}. After that, TPM will also examine whether the value of GPCR-x is equal to that of VPCR-x. If so, the K will be provided by TPM; or otherwise, it will be denied.

VPCR-x, the value of GPCR-x in the group of PCR, is also referred to as a state or configuration of PCR. And the sealing operation will also associate the M with a state of TPM. Only if the state of TPM is consistent with the sealing state, the sealed K {M} will be decrypted into the original M. In other words, only when the state of sealing TPM is consistent with that of TMP in decryption, the decryption will be initiated.

5. DLP DATA PROTECTION MODEL

5.1. Existing Data Protection Methods

So far, the cryptography has been recognized as the major measure to protect the sensitive data. During the application, the encryption key and decryption key are both stored in the specified file, which is only accessible to the system administrator for key operation.

5.2. Underlying Problem

The data safety also stands for the safety of data storage, which requires that the important data should be stored in the disk after encryption so as to prevent the data leakage and object reuse. What is more, the storage, management and access of the encryption key and decryption key should be also protected as well. However, the current system still leaves something to be desired in the key management and protection. That is, when the key is stored in the form of file, it will be also under attack from malicious virus or Trojan and then leads to data leakage. What is more, it also fails to prevent the administrator from disclosing important data through normal channels.

5.3. Solution

When the plain data is under encryption, it might just take a lot of time for the asymmetrical algorithm to encrypt the original data due to its large amount. Therefore, the traditional symmetrical algorithm is preferred. In order to ensure that the symmetrical key will not be let out during the process of transmission and storage, the RSA public key provided by TPM should be used to encrypt the symmetrical key. After that, the RSA private key should be used in the receiving end to decrypt the cipher text of the symmetrical key. It is only after getting the symmetrical that the receiver will be able to decrypt the cipher text and get the plain text. Since the asymmetrical key should be provided by TPM, it will be very hard for those viruses or Trojan to cause harm.

In some cases, the precaution should be made to prevent the authorized user from disclosing important data. For example, if the administrator wants to copy the encrypted data into another device through the mobile storage device or network, it can also lead to data leakage. A feasible solution is to seal the symmetrical key through data sealing technology and bind the sealed data with the platform configuration.

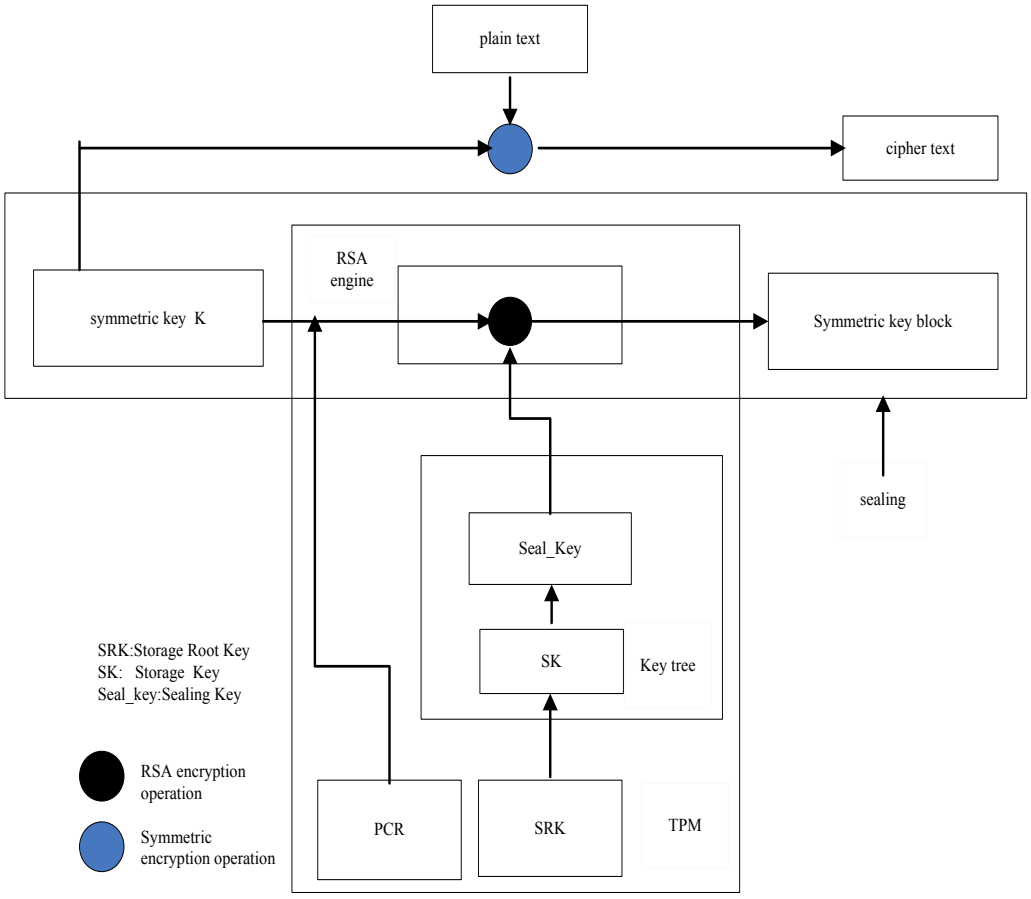

Fig. (4). Data protection model.

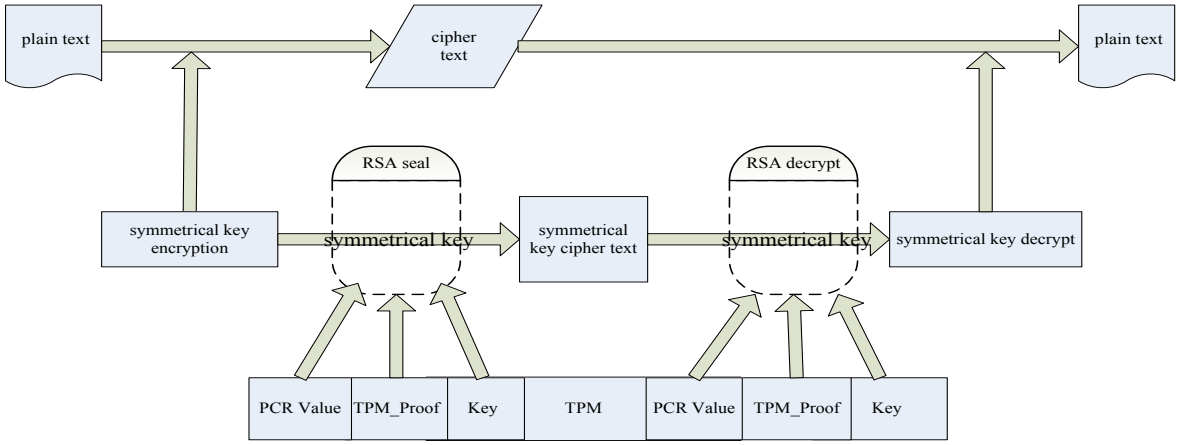

Fig. (5). Sealing process.

5.4. Data Protection Model

According to the schema above, we have proposed a DLP data protection model, as seen in Fig. (**4**).

The Fig. (**5**) is sealing process. The sealing data in Fig.5 is symmetrical key. PCR Value means the platform configuration during the data encapsulation; TPM_Proof means the unique identification of the TPM while key mean the encapsulated key.

After releasing the data, the TPM will make the following two comparisons. If they are consistent with each other, the encapsulated data will be output, which is also known as symmetrical key.

- Whether the PCR value of the current platform configuration is equal to the PCR value after decryption or not.
- Whether TPM_Proof in the TPM is equal to the TPM_Proof after decryption or not.

5.5. Encrypting Files

Now let's use the above methods for encrypted file.What we'd like to do is have a function that takes a path to a file and a TPM key handle and creates a new symmetric key, encrypts that file using the symmetric key, and seals the symmetric key with the TPM key. We'd like to invoke the function like this:

```
MyFunc_SealFile(FILE_PATH, hKey, NULL, 0);
```

or to seal the symmetric key to PCRs 8, 9, and 12:

```
UINT32 pcrs[ ] = { 8, 9, 12 };
UINT32 num_pcrs = 3;
MyFunc_SealFile(FILE_PATH,hKey,pcrs, um_pcrs);
```

We'll just write the encrypted file out to FILE_PATH.enc, and we'll write the encrypted symmetric key blob to FILE_PATH. key. Also, we'll use a dummy function to do the symmetric encryption, since that will have to be provided by an external library.

```
Int MyFunc_SealFile(char *file, TSS_HKEY hKey,
UINT32 *pcrs,UINT32 num_pcrs)
{
FILE *fin, *fout;
UINT32 inSize, outSize, encSymKeySize =
SYM_KEY_SZ;
BYTE in[DES_KEY_SZ], out[DES_KEY_SZ],
encDesKey[DES_KEY_SZ];
TSS_HPCRS hPcrs;
BYTE *outFileName;
DES_cblock key;
DES_key_schedule schedule;

DES_random_key(&key);
DES_set_key_checked(&key,&schedule);

inSize = outSize = DES_KEY_SZ;
outFileName = malloc(strlen(file) + 5);
sprintf(outFileName, "%s.enc", file);
fin = fopen( file,"r" );
fout = fopen( outFileName,"w" );
while ((inSize = read(fileno(fin), in, inSize) ==
DES_KEY_SZ))
{
/* Call our external library to do the bulk encryption
* using the symmetric key */
DES_ecb_encrypt(&in,&out,&schedule,DES_ENCRY
PT);
/* Write the encrypted file back out */
write(fileno(fout), out, outSize);
}
fclose(fin);
fclose(fout);
/* Create the PCR composite object to seal the sym-
metric key with */
MyFunc_CreatePcrs(num_pcrs, pcrs, &hPcrs);
/* Now seal the symmetric key using our TPM key */
MyFunc_SealData(hKey,hPcrs,SYM_KEY_SIZE,
symKey,&encSymKeySize, &encSymKey);
/* Write out the encrypted symmetric key blob */
sprintf(outFileName,"%s.key", file);
fout = fopen( outFileName,"w" );
write(fileno(fout), encSymKey, encSymKeySize);
fclose(fout);
return 0;
}
```

This file unable to oopen that not in the same platform, also ensure the security of data.

5.6. Experiment Analysis

The system hardwires required in the experiment will be presented as followed.

- CPU: Intel Core i3
- Memory 2G
- Software experimental environment: Fedora Linux 8,Operation system kernel version: 2.6.24.
- Trusted platform module version 1.2.
- The Tspi_Data_Seal and Tspi_Data_Unseal will be achieved in the core as well.

The paper will also make a comparison between the proposed method and the existing methods in the field of data leakage prevention from several aspects, as seen in Table **2**.

According to the test results, the TPM-based method of data leakage prevention can prevent the authorized users from actively disclosing important leakage as well as solve the problems within the data encryption in terms of software and hardware.

CONCLUSION

By integrating the basic safety functions such as encryption, decryption and authentication into the hardware chip, the trusted computing platform can prevent the data con-

Table 2. Data leakage prevention technology features contrast.

Data Protection Method	Prevention Initiative Leak	Passive Leak	Application Compatibility	Easy to Bypass	Data Corruption	Change the Hard-ware Architecture	Affect System Efficiency
Access control method	no	no	general	general	no	no	not affect
File encryption method	no	yes	general	hard	yes	no	small affect
TPM data protection method	yes	yes	excellent	hardest	no	yes	small affect

tained in the chip from being accessed by the external program. Moreover, the combination of trusted computing and data leakage prevention has also ensured the data safety. What is more, the paper has sought to integrate the technology of sealing data with the traditional encryption method and also bound the encrypted data with the platform configuration so as to prevent the internal users from disclosing important data. In addition, the paper has also proposed a new model of data protection to approach the data leakage in the system. It should be noted that the technology of data sealing can enhance the safety greatly, though it is not quit flexible during application. Although the use of non-migratable key can upgrade the protection level, the key and associated data will not be accessible or restored when the system or TPM breaks down, which just remains to be further studied.

CONFLICT OF INTEREST

The author confirms that this article content has no conflict of interest.

ACKNOWLEDGEMENTS

The authors would like to thankWang Zheng, Wang Ying, Fu Donglai, and Bian jing for their insightful comments and suggestions. The anonymous reviewers also provided valuable feedback. This paper is supported by the Natural Science Foundation of Shanxi Province under Grant No. 2009011022-2 and Shanxi Scholarship Council of China Grant No. 2009-28.

REFERENCES

[1] W.B. Mao, F. Yan, and C.R. Chen, "Daonity: Grid security with behaviour conformity from trusted computing", In: *Proceedings of the first ACM Workshop on Scalable Trusted Computing,* New York: ACM Press, 2006, pp. 43-46

[2] D. Challener and K. Yoder, *A Practical Guide to Trusted Computing*, IBM press, Indiana, 2007.

[3] H. Zhang, J. Luo, G. Jin, Z. Zhu, F. Yu, and F. Yan, "Development of trusted computing research", *Journal of Wuhan University (Natural Science Edition)*, vol. 52, no. 5, pp. 513-518, 2006

[4] M. Xu, J. He, B. Zhang, and H. Zhang, "A new data protecting scheme based on TPM", In: *Proceedings of the 8th ACIS International Conference on Software Engineering, Artificial Intelligence, Networking and Parallel/Distributed Computing*, IEEE Computer Society, 2007.

[5] L. Liu and J. Peng, "Research on Distributed and Dynamic Trust Transfer and Measurement", In: *Proceedings of the 2009 International Conference on Networks Security, Wireless Communications and Trusted Computing*. Washington DC USA: IEEE Computer Society, 2009.

[6] J. Lu, S. Yang, and S. Lu, "A property-based sealed storage solution in trusted computing", *Information Technology*, vol. 1, no. 1, pp. 124, 2008.

[7] Trusted Computing Group. TCG Software Stack (TSS) Specification. Reference: Available from http://www.trustedcomputinggroup.org/resources/tcg_software_stack_tss_specification

[8] X. Li, D. Feng, Z. Chen, "Model for attribute based access control", *Journal on Communications*, vol. 29, no. 4, pp. 90-98, 2008.

[9] A. Liu, and Y. Han, "Test and analysis of trusted platform module data protection", *Journal of Computer Applications*, vol. 30, no. 5, pp. 1243-1245, 2010

[10] Trusted Computing Group. TPM Main Specification. Reference: Available from http://www.trustedcomputinggroup.org/resources/tpm_main_specification

[11] M. Strasser, and H. Stamer, "A Software-Based Trusted Platform Module Emulator", In: *The First International Conference on Trusted Computing and Trust in Information Technologies*, Berlin Germany: Springer, 2008.

[12] Y. Cui, and X. Zhang, "Improving credibility of systems integrity measurement of property remote attestation", In: *Asia-Pacific Conference on Computational Intelligence and Industrial Applications*, 2009.

Received: September 22, 2014 Revised: November 03, 2014 Accepted: November 06, 2014