Current status and future prospects of data leakage prevention technology: A brief review

Lv Xuming, Cao Lina, Ji Peng, Gao Xiao, Chen Shuo

1State Grid Liaoning Electric Power Co., Ltd., Shenyang, Liaoning 110006, China
xuming_lv@126.com

Abstract. With the rapid development of mobile Internet, the application environment of data information is becoming more and more complex. How to prevent the leakage of sensitive data has become an urgent practical problem to be solved in the field of information security. In this situation, data leakage protection (DLP) technology emerges, its aim is to protect sensitive data in all directions and to establish a security protection mechanism covering the whole data life cycle. Starting from the research background and basic concepts of data leakage protection, this paper reviews and summarizes the current research status and development trend of data leakage protection technology.

1. Introduction
As a comprehensive platform for carrying mobile services, mobile intelligent terminals bring revolutionary changes to people’s work and life. Users can install mobile applications on mobile intelligent terminals by themselves, the services provided by application software cover all aspects of people’s work and life. However, behind the rapid development of mobile Internet, is the increasingly complex mobile security situation. In the application process of mobile interconnection, information security problems constantly arise, such as operating system security vulnerabilities, endless emergence of malicious software, pseudo base station or dangerous Wi-Fi access, payment transaction information, personal privacy information, confidential information and other information security issues. While improving work efficiency, information systems also put forward security requirements for information storage and access control of computer terminals and servers in the systems[1]. With the emergence of new technologies and services (such as mobile Internet, cloud computing, mobile smart terminals and applications), the problems of communication network security, mobile phone security and information security brought about by these technologies have attracted the attention of the whole information industry and even the whole society. Mobile Internet has become the focus of attention in the field of information security in various countries.

Whether it is military, government, enterprises or individuals, the leakage of sensitive data will pose a serious threat to information security. In order to prevent data leakage, government agencies, enterprises and scientific research institutions often spend lots of money to install firewalls, intrusion detection, anti-virus, information encryption and vulnerability scanning and other security products. These products have largely helped them fend off attacks from the outside, but they are powerless against internal threats. Because internal authorized users are familiar with system privilege settings and business processes, they can bypass the detection and access control mechanism, use known system vulnerabilities and their own privileges to launch internal attacks through fully legitimate operations, leading to the leakage of sensitive enterprise information. All of these make internal threats become the main threat of data leakage. With the wide application of distributed and cloud computing technology,
the problem of data leakage protection caused by internal threats will become more prominent. According to the data of the National Computer Information Security Evaluation Center of China, in the incidents that caused heavy losses due to internal important secrets leaked through the network, only 1% were stolen by hackers, while 99% were caused by intentional or unintentional leaks from internal employees.

In recent years, the concept of Data Leakage Prevention (DLP) has been put forward in order to solve the problem of data leakage in enterprises. By integrating various existing security technologies and strategies, its goal is to prevent internal personnel from leaking data from enterprises in the form of violating security policies [2]. Nowadays, research on data leakage protection technology that can deal with the internal threats has become a hot issue in academia and industry. This paper first introduces the basic concepts of data leakage protection, then reviews and summarizes the current research status and development trend of data leakage protection technology, with a view to benefiting both relevant academics and professionals.

2. Basic concepts of data leakage protection

Data leakage, also known as information leakage, has always been an important concept in computer system security research. With the widespread application of the Internet, the scope of data leakage has become more and more extensive, including not only personal or enterprise information leakage caused by e-mail or web services, but also data leakage caused by laptop theft, hacker intrusion, backup disk loss and so on. In the new application environment, the focus of data leakage also extends from the inside of closed system to the inter-system and network environment. American Association for Systems and Network Security defines data leakage as the transfer of data or information within an enterprise or organization to external storage or recipients through unauthorized channels, either electronically or physically. The term “unauthorized” here does not refer specifically to intentional or malicious operation, data leakage caused by unintentional or illegal operations can also be considered unauthorized.

Data leakage protection, understood from a certain point of view, is actually a kind of confidentiality protection with subjective expectations. Usually, there is a range of expected data to be protected, which is collectively referred to as the security protection domain. Outside the security protection domain is the non-protection domain. The goal of data leakage protection is to prevent data from flowing from the secure domain to the unprotected domain. The security protection domain of data leakage protection does not specifically refer to the security domain of intranet, but can be any virtual or real area with data leakage protection requirements, such as a disk partition or a file storage area. As shown in Figure 1, data leakage protection is a three-dimensional hierarchical system, including the protection requirements of different layers from network layer, storage layer, system layer to application layer, and each level is closely related. Therefore, research on data leakage prevention technology requires comprehensive consideration of protection requirements at all levels.

The leakage protection at the network layer usually divides the security protection domain by physical LAN of an enterprise or virtual network consisting of network nodes with application associations. Sensitive information can flow freely between network nodes within the protection domain, but cannot leak to network nodes outside the protection domain. Therefore, data leakage protection at the network layer needs to focus on restricting the exports of physical or logical networks. For example, deploying a content filtering firewall on the gateway for content analysis and filtering, deploying a “waterproof wall” on the LAN to disable the external network port and so on.

The leakage protection at the storage layer mainly divides the security protection domain by the storage space, and the goal is to ensure that the data in the storage space can only be used in a specific environment such as disk encryption and decryption, intranet mobile storage device management, etc. With the development and application of distributed technology and cloud computing technology, the protection domain at the storage layer is no longer confined to a storage device, but may also span different network nodes. Although the boundary between storage and application has changed, the need for leak protection remains the same.
The leakage protection at the system layer and application layer are aimed at the protection requirement of data in use, and the protection domain is delineated according to the usage range of data. The protection of system layer usually restricts the usage range of sensitive data through file system, process management, network management and port control. And the protection of application layer is usually provided by the application itself. The application is equivalent to building a container for the data, preventing the spread of data content by constraining the interface and operations of the program itself.

3. Research status of data leakage protection technology

The traditional data leakage protection technology is mainly based on the idea of sub-domain control, and is realized through the terminal security control mechanism. According to the principle of realization, there are three main ways at present:

(1) Content Filtering Monitoring: Filter the content of the transmitted data at the egress of the host or intranet, prevent data leakage with sensitive content[3]. Although this method is convenient to implement, it cannot process encrypted data and data transmitted by hidden technology, so the application range is limited.

(2) Domain-based access monitoring: Different security domains are classified according to data protection requirements, the input and output of the security domain are centrally managed and controlled according to the access control policy, and data leakage is prevented by restricting information flow between security domains. Typical examples include Data Tethers structure which divides security domains based on environmental factors[4], Bhilare’s intellectual property and privacy protection system in a campus network environment through active directory[5], and the terminal data leakage prevention scheme which realized through device and port control[6]. Although this method can effectively block data leakage, it also reduces the availability of the system, because inter-domain communication is essential for many applications.

(3) Encryption and decryption control: Ensuring data storage security through data encryption. At present, transparent encryption mode is widely used, such as Zhang’s Intranet Data Leakage Protection System at file system filter layer[7], and TrueCrypt, BitLocker, File Vault and other file protection systems that use Full-Disk Encryption (FDE) mode. These encryption schemes can effectively prevent data leakage caused by disk theft or loss, but since legitimate users can freely use data, there is still a risk that users will actively leak data, thus these schemes are usually regarded as the basis of other security schemes. In addition, many information security product providers have launched data leakage protection products and solutions for enterprise application needs, such as McAfee, Verdasys, etc. Some
companies have also proposed “waterproof wall”, “fortress system”, “intranet security management” and other data leakage protection solutions.

To sum up, data leakage protection not only needs to rely on traditional anti-virus software, firewall, intrusion detection, identity authentication and file encryption technology to guard against outside, but also needs “waterproof wall” and internal network security management system to “guard against inside” by controlling the network and storage ports. However, relying solely on network and storage port control will not only lead to the contradiction between security and availability, but also be difficult to form a complete solution. At present, some mobile storage devices (such as U-disk, mobile hard disk, writable CD-ROM, MP3 player) and network connection modes (such as Wi-Fi and Bluetooth) that are widely used by internal personnel, may bring serious information leakage threats to enterprises. Especially in the current open network environment, data sharing and distribution are more frequent. To protect data security in a distributed environment, it is necessary to break through the traditional passive methods of patching and plugging vulnerabilities, and propose effective solutions from a new perspective.

4. Development trend of data leakage protection technology

In recent years, with the further development of object-oriented file system, active storage technology, and especially trusted computing platform technology, many scholars have put forward the idea of active defense. Its main objective is to establish a data-centric security protection system, by compressing the boundary of data security protection to the data itself, it makes various security control mechanisms more closely related to data usage, thus changing the concept of data static and passive protection. To this end, researchers integrated some of the data processing and security control functions into the storage device to enhance the initiative of the storage device. The active storage technology represented by Active Disk[8] and the intelligent storage technology represented by IDisks[9] focus on the initiative of storage devices in data computing and management. The core of self-secure disk represented by S4 is to introduce the idea of intrusion detection into storage devices and detect illegal disk access by monitoring storage access interface[10]. Pereira proposes a hardware architecture integrated with security services to meet the needs of data service applications, realizing the integration of security services and hardware[11].

In order to ensure the safe storage and trusted access of data, TCG issued the Trusted Storage (TS) specification, which takes the trusted computing platform module TPM as the trusted root to ensure that data can only be accessed and used by trusted subjects. Therefore, in order to prevent data leakage caused by internal threats, it is necessary to combine other corresponding security assistance measures.

With the development of virtualization technology and the gradual application of cloud computing technology, the application architecture of data center plus virtual terminal is adopted by more and more companies[12]. This architecture establishes a secure data path between the data center and the terminal based on trusted computing, and builds a secure isolation environment for data usage in the terminal through virtualization technology, thereby implementing data leakage protection. Griffin proposes a Trusted Virtual Domains (TVDs) architecture, which can build trusted channels between different trusted domains in distributed service applications[13]. On the basis of TVDs, Berger further grouped virtual machines and underlying data resources according to the security requirements of centralized data services, and proposed the structure of Trusted Virtual Datacenter (TVDc)[12]. Gasmi applied TVDs to Enterprise Rights Management (ERM) and designs a secure and flexible ERM system[14]. Burdonov prevented data leakage by building an isolated environment for sensitive data in an untrusted environment[15]. Catuogno applied TVDs to the security protection of mobile storage devices, and built corresponding trusted virtual domains for different storage devices through centralized management[16]. Yanbo Han analyzed the sensitive and privacy data protection problems faced by Business Process Management (BPM) in cloud environment. A new BPM security architecture based on cloud computing was proposed, which provided distributed user terminals for sensitive data and non-compute-intensive activities to ensure the security of data in the cloud[17].
In addition, Kang also designed a dedicated hardware structure that integrates data and signature management software for the data leakage protection of mobile storage devices[18]. Kuhn applied trusted computing technology to disk encryption and control of safe dormancy to prevent data leakage during work-in-progress[19]. Yin Fan proposes a trustworthiness-based distributed data leak protection model, which can be used to control file usage by calculating user credibility based on user’s historical behavior, thereby preventing leakage of distribution documents[20].

In summary, the combination of trusted computing and virtualization technology has become a major trend in the development of data leakage prevention technology. Because virtualization technology is mostly used to build data isolation environment, and trusted computing is mainly used to ensure the security and credibility of data storage, the combination of these two technologies actually shows that data leakage protection begins to pay more attention to the continuous protection in the data life cycle. The idea of active storage technology enlightens us that it is a good choice to enhance the active protection capability of data itself, in order to ensure the data security in untrusted environments that are out of protected domain.

Acknowledgement
Sponsored by State Grid Corporation of China Technology Program (Contract No. 2019YF-20)

References
[1] M. E. Whitman, and H. J. Mattord, Principles of information security: Cengage Learning, 2011.
[2] S. Alneyadi, E. Sithirasenan, and V. Muthukumarasamy, “A survey on data leakage prevention systems,” Journal of Network and Computer Applications, vol. 62, pp. 137-152, 2016.
[3] D. W. Oard, and G. Marchionini, “A conceptual framework for text filtering,” Submitted to User Modeling and User-Adapted Interaction, 1996.
[4] M. Beaumont-Gay, K. Eustice, and P. Reiher, "Information protection via environmental data tethers." pp. 67-73.
[5] D. S. Bhilare, A. K. Ramani, and S. K. Tanwani, "Protecting intellectual property and sensitive information in academic campuses from trusted insiders: leveraging active directory." pp. 99-104.
[6] M. Fabian, "Endpoint security: managing USB-based removable devices with the advent of portable applications." p. 24.
[7] X. Zhang, F. Liu, T. Chen, and H. Li, "Research and application of the transparent data encryption in intranet data leakage prevention." pp. 376-379.
[8] H. Lim, V. Kapoor, C. Wighe, and D. H.-C. Du, "Active disk file system: A distributed, scalable file system." pp. 101-101.
[9] K. Keeton, D. A. Patterson, and J. M. Hellerstein, “A case for intelligent disks (IDISKs),” ACM SIGMOD Record, vol. 27, no. 3, pp. 42-52, 1998.
[10] J. D. Strunk, G. R. Goodson, M. L. Scheinholtz, C. A. Soules, and G. R. Ganger, "Self-securing storage: protecting data in compromised system." p. 12.
[11] F. D. Pereira, and E. D. M. Ordonez, "A Hardware Architecture for Integrated-Security Services," Transactions on Computational Science IV, pp. 215-229: Springer, 2009.
[12] S. Berger, R. Cáceres, D. Pendarakis, R. Sailer, E. Valdez, R. Perez, W. Schildhauer, and D. Srinivasan, “TVDC: managing security in the trusted virtual datacenter,” ACM SIGOPS Operating Systems Review, vol. 42, no. 1, pp. 40-47, 2008.
[13] J. L. Griffin, T. Jaeger, R. Perez, R. Sailer, L. Van Doorn, and R. Cáceres, “Trusted virtual domains: Toward secure distributed services.” pp. 12-17.
[14] Y. Gasmi, A.-R. Sadeghi, P. Stewin, M. Unger, M. Winandy, R. Hussein, and C. Stübke, "Flexible and secure enterprise rights management based on trusted virtual domains," pp. 71-80.
[15] I. Burdonov, A. Kosachev, and P. Iakovenko, "Virtualization-based separation of privilege: working with sensitive data in untrusted environment.” pp. 1-6.
[16] L. Catuogno, H. Löhr, M. Manulis, A.-R. Sadeghi, and M. Winandy, "Transparent mobile storage
protection in trusted virtual domains." pp. 159-172.

[17] Y.-B. Han, J.-Y. Sun, G.-L. Wang, and H.-F. Li, “A cloud-based BPM architecture with user-end distribution of non-compute-intensive activities and sensitive data,” Journal of Computer Science and Technology, vol. 25, no. 6, pp. 1157-1167, 2010.

[18] D. Kang, B. Jung, and K. Kim, "Hardware Based Data Inspection for USB Data Leakage Prevention." pp. 57-63.

[19] U. Kühn, and C. Stüble, "User-friendly and secure tpm-based hard disk key management," Future of Trust in Computing, pp. 171-177: Springer, 2009.

[20] Y. Fan, Y. Wang, L. Wang, and R. Yu, “A trustworthiness-based distribution model for data leakage prevention,” Wuhan University Journal of Natural Sciences, vol. 15, no. 3, pp. 205-209, 2010.