Data Leakage Prevention

A newsletter for IT Professionals

Issue 5

I. Background of Data Leakage Prevention

Data leakage refers to unauthorised transmission of data from within an organisation to an external destination or recipient. The transmission can be done both electronically and physically and the types of data leaked usually include:

- Confidential / Sensitive Information
- Intellectual property
- Customer / Student Data
- Health Records

Given today's strict regulatory and legal compliance requirement on intellectual and personal data protection, organisations, including universities, have invested a great deal of time and resources in safeguarding their information from potential unauthorised access and disclosure. Security vendors and researchers also developed various counter measures to fight against data leakage issues, which are collectively known as Data Leakage Prevention (DLP) solutions. A key distinguishing feature of DLP that contrasts with non-DLP security tools, such as data encryption, is deep content analysis based on pre-defined security policies.

In general, DLP refers to any systems or tools that identify, monitor, and protect the following type of data:

- **Data in Motion** – Any data that is moving through the network to the outside via the Internet. This feature applies to all data transmitted on wire or wirelessly. E.g. Examination results sent to students over the Internet.

- **Data in Use** – Data at the endpoints of the network (e.g. data on USB devices, external drivers, MP3 players, laptops, and other highly-mobile devices). E.g. Patent information stored on portable hard disks.

- **Data at Rest** – Data that resides in files system, databases and other storage methods. E.g. A university's financial data stored on the financial application server.

In response to the above types of data having exposure to potential leakage problem, specific DLP systems / tools have been engineered to mitigate the risks or detect any security violations:

Reference:
http://www.getadvanced.net/pdfs/SANS%20Institute%20Data_Loss_Prevention.pdf
KPMG Publication – Data Leakage Prevention

I. Background of Data Leakage Prevention (cont'd)

1 Network DLP

Network DLP is designed to detect any leakage incidents related to data in motion, by detecting if particular important data files are being transferred through universities' networks. This kind of DLP devices usually supports multiple protocols such as HTTP, FTP, P2P and SMTP, and is commonly attached to network equipments (e.g. routers, switches), where all traffic leaving universities' internal network can be captured for inspection.

Nowadays, most universities have already implemented certain network traffic filtering systems, such as e-mail and web activity monitoring programs, which can achieve part of the functionalities of Network DLP. Some more specialised Network DLP tools include McAfee Network DLP Manager, RSA DLP Network, and Symantec Data Loss Prevention Network series.

2 Endpoint DLP

Endpoint DLP products are agents or software that usually reside on end user terminals such as mobile devices and laptops. The common use of Endpoint DLP is to prevent users from storing sensitive information on removable media devices such as USB flash drives and CD/ROM discs and to protect against unauthorised transmission of sensitive information when a user is not connected the universities' own networks (e.g. public free Wi-Fi spot). An Endpoint DLP software can also utilise disk encryption, which prevents unauthorsied access to information on a lost or stolen laptop.

Popular Endpoint DLP products currently on the market include NextLabs Enterprise DLP, Symantec Protection Suite Enterprise and McAfee Host Data Loss Prevention.

3 Embedded DLP

Universities are also given a less expensive choice to implement "partial" DLP solutions instead of setting up a comprehensive data leakage management infrastructure. Such solutions are commonly known as Embedded DLP.

Embedded DLP are planted within specific applications to effectively monitor the data outflows, identify keywords or related patterns belong to sensitive information and block any suspicious data leakage attempts. For instances, scanning and rejecting outgoing e-mails for sensitive keywords or attachments, restricting printing of copyrighted softcopy documents.

The design and implementation of Embedded DLP can be performed within Universities or acquired from existing security vendors. Cisco's IronPort e-mail security technology provides functionalities to detect sensitive content, patterns or images in a message body or within attachments. Websense Web Security Gateway Solutions incorporated Websense TruWeb DLP capability offers embedded DLP over outbound communications to destinations like web mail and social networks.

Reference:
http://www.ironport.com/technology/
http://investor.websense.com/releasedetail.cfm?ReleaseID=466193

I. Background of Data Leakage Prevention (cont'd)

Key Benefits Achieved through Data Leakage Prevention

- **Prevent Data Leakage** – Preventing accidental or malicious loss of data by insiders (e.g. employees, students and contractors) or outsiders (e.g. hackers) is the main purpose of all DLP solutions. With appropriate implemented DLP mechanism, universities' can control the access to their sensitive data by external parties.

- **Reduce Cost of Investigation and Damage to Reputation** – Leakage of sensitive data of universities usually means economic loss or damage to the reputation. Implementing DLP solutions within universities' network or information systems can put control over the outflow of sensitive data and thus effectively reduce the risk of unauthorised disclosure. In case when data leakage incidents do occur, DLP tools or software can also assist the investigation by providing useful information on system activity history.

- **Facilitate Early Risk Detection and Mitigation** – DLP solutions require universities to perform a series of preparation work, including data classification, risk assessment, research on regulatory and privacy requirement, development of policies, standards and procedures for data protection, through which a number of around vulnerabilities of data leakage can be noticed within management radar and makes early mitigation possible.

- **Increase Comfort Level of Senior Management** – Data leakage is one of the most critical issues facing universities' senior management because various sensitive data is stored and processed by universities' information systems, such as student / employee records, confidential research data and patent. Universities must properly secure such information to comply with regulatory and legal requirement, maintain competitive advantage and protect their reputation. Where DLP controls are implemented and operating effectively, senior management is able to concentrate on other critical issues.

Statistical Report

Data leakage prevention growing 10 percent annually

According to a recent report from Network World, data leakage prevention is currently growing at 10 percent a year. While this figure is lower than what many experts anticipated, it still represents one of the better percentages among security technologies. Companies looking to remain compliant with regulatory authorities requirements have been fuelling the increase in the technology's adoption.

See the article:
(http://www.messagingarchitects.com/resources/security-compliance-news/email-security/report-data-leak-prevention-growing-10-percent-annually-19920447.html)

Reference:
http://www.foundstone.com/us/resources/whitepapers/wp_dlp_program.pdf

II. Risk of Data Leakage Prevention in Universities

Implementation of DLP solutions encompasses a variety of complex IT areas such as data classification, risk assessment, compilation of policies, standards and procedures. If not designed and managed adequately, DLP solutions can result in a number of risks to universities. Many of these risks can directly impact universities' normal operations or expose them to even greater threats. Examples of these risks are listed below.

1 Excessive Reporting and False Positives

Similar to an improperly configured Intrusion Detection System (IDS), DLP solutions may generate significant number of false positives that overwhelm universities' IT security resources and obscure valid hits. Trying to monitor too much data volume or too many keywords / data patterns can easily exhaust limited resources.

2 Conflicts with Software or System Performance

DLP solutions, especially those Endpoint DLP products, can cause compatibility issues when conflicting with other systems and software. For example, some application software cannot run properly on encrypted hard drive. Applications errors or performance degradation are two common results of such conflicts. In worst case, the compatibility issues may cause the abnormal termination of other security controls and expose universities' information system to even great risks.

3 Improperly Configured Network DLP Module

When a Network DLP is not able to handle the amount of network traffic, due to insufficient consideration of traffic volume during the design stage or increased network traffic over time, some network packets may be missed or dropped, allowing certain data to pass uninspected. It may render Network DLP ineffective when unauthorised transmission of sensitive data to external parties is ignored.

4 Improperly Tuned Network DLP Module

Universities must pay particular attention to strike a balance between permitted and prohibited disclosure of sensitive data. Otherwise, inadequate tuned Network DLP solutions may cause disruption of universities' operation, waste of staff or students' time, damage to relationship with external parties such as contractors and general public. E.g. Blocking employees sending sensitive data to authorised external parties; disrupting normal e-mail services used by universities.

5 Changes in Processes or IT Infrastructure

DLP solutions are complex in nature and must be carefully configured or customised to cope with universities information system and network environment. If DLP is not maintained regularly and timely, any changes to the set of application software used, network architectures or the operational procedures may weaken the DLP effectiveness or introduce other problems like compatibility issues, and disruption of operation.

Reference:
http://www.isaca.org/Knowledge-Center/Research/Documents/DLP-WP-14Sept2010-Research.pdf

II. Risk of Data Leakage Prevention in Universities (cont'd)

6 Improper Definition of DLP Needs

DLP solutions can only be effective based on accurate and comprehensive DLP policies. If universities failed to address all potential vectors for data leakage (e.g. identification of sensitive data and required protection level, determination of acceptable use of information resources, relevant regulatory and legal requirements), the DLP tools are either ineffective or has incomplete coverage of all data leakage risks the universities face.

7 Undetected Failure of DLP Modules

Like other application software or systems, DLP solutions rely on technologies implemented over software and hardware infrastructure. Failures of software or hardware often draw less attention from universities IT personnel. Program bug, power failure, environmental hazards may strike the infrastructures that support DLP functions. If the failures go unattended, universities will be completely exposed to data leakage risks.

8 Legal

When universities adopt DLP solutions that monitor the activities performed by their employees, students and contractors, one of the issues they encounter is whether deploying DLP will conflict with legal or employee agreements that protect privacy. Without establishing appropriate policies, disclaimers and agreements to address the necessity and purpose of data monitoring, legal proceedings may be launched against the universities.

Recent Incident

HSBC fined over US$5 million for data security failings

In July 2009, HSBC has received an almost £3.2 million fine from UK's Financial Services Authority (FSA) after three of its firms lost computer discs and posted unencrypted customer details. The UK's biggest bank was fined for the "careless" handling and loss of confidential details of tens of thousands of its customers. In a series of security failings, the bank sent large amounts of "unencrypted" data via post or courier to third parties.

HSBC has taken remedial action to address the problems that FSA identified, including stronger processes to ensure all confidential data that is electronically transmitted or stored and transported on CDs and laptops is encrypted, better training for staff and restricting the ability to download data to portable devices.

See the article: (http://www.cw.com.hk/content/hsbc-fined-over-us5-million-data-security-failings)

Reference:
http://www.isaca.org/Knowledge-Center/Research/Documents/DLP-WP-14Sept2010-Research.pdf
https://365.rsaconference.com/docs/DOC-2174;jsessionid=9C450C7061C6E3814B21538423A222DB.node0

III. Exploitation on Data Leakage

Data leakage can be caused by internal and external parties, either intentionally or accidentally. According to INFOWATCH's "Global Data Leakage Report 2009", 51% of data leakages were resulted from intentional attacks and 43% leakages were due to accidental events, which indicates a strong increase of intentional leakages when comparing to 2007's figures (i.e. 29% intentional and 71% accidental). Several intentional exploitations on data leakage are illustrated below:

1 Code Injection

Poor programming of information systems and applications can leave universities exposed to various code injection attacks, or allow inappropriate information to be retrieved in legitimate database queries.

Structured Query Language (SQL) injection is one of the most common attack techniques for applications or websites that use SQL servers as back-end database. If the applications or websites failed to correctly parse user input and sanitise user input, the content within the database may be stolen or program errors may occur and interrupt the relevant services.

2 Malware

Malware is designed to secretly access a computer system without owner's informed consent. Sophisticated data-stealing malware may take various forms including Trojan, spyware, key loggers, screen scrappers, adware, and backdoors. Users are usually infected during installation of other application software bundled with malware or from malicious web sites. Examples of data-stealing malware are Bancos (steal sensitive banking information) and LegMir (steal personal information such as account name and passwords).

3 Phishing

Another data leakage channel is through the use of phishing sites as a lure to steal sensitive data from users. Phishing spam can be sent to staff or students' e-mail address. Once they are fooled to click the links in the malicious e-mails, their browsers can be re-directed to fraudulent websites that mimic reputable organisations, where users may unnoticeably leak their account name and passwords to hackers. If the login credential to a university's web mail system is leaked, the hacker can authenticate himself or herself as university member and gain full access to any sensitive information stored within the e-mail system. It is also possible that the phishing spam received directs users to a site that uploads malware to their computers.

4 Malicious Insider

Universities' sensitive data are also vulnerable to intentional data leakage performed by their internal users (e.g. employees, students). Motivations are varied, but usually fall into corporate espionage, financial interest, or a grievance with their employers. Sensitive data can be unauthorisedly transferred out through remote access, e-mail, instant messaging or FTP. Even if DLP solutions have been deployed within universities, these malicious insiders, especially IT personnel, can bypass the restrictions through sabotage DLP systems. E.g. altering the DLP configuration to create backdoor; shutdown DLP services; physically cut off the power supply; de-classify sensitive data.

Reference:
http://www.sans.org/reading_room/whitepapers/awareness/data-leakage-threats-mitigation_1931
http://www.infosec.gov.hk/english/anti/recent.html
http://prolog.biz/presse/infowatch/artikel/infowatch_global_data_leakage_report_2009_eng.pdf

IV. Hardening Steps for Data Leakage Prevention – Preparation

A comprehensive DLP solution that protects data in motion, data at rest and data in user require complex and significant amount of preparation activities. Among these activities, data classification, risk assessment and policy development are the most critical ones and involve both the commitment from senior management and IT security personnel in universities.

1 Risk Assessment

The main purpose for a risk assessment is to identify all types of data within the universities and the associated threats and vulnerabilities. Key stakeholders from different parties should be gathered together to discuss and reach agreements on topics including, but not limited to, the following:

- What data should be protected? E.g. Internal, Confidential, Highly Confidential
- What applications or infrastructure should be covered by DLP?
- What regulatory and legal requirement we need to comply with?
- Who are the authorised personnel that can receive data from us?
- What is the reporting and workflow of DLP solutions?
- What are the expected accuracy rates for different kinds of data? E.g. statistical / conceptual analysis or partial database matching?

2 Data Classification

Data classification helps to categorise data based on the value to universities and add additional controls to limit the access and movements of sensitive data. Proper data classification allows universities to determine the order of protection for different types of data and focus DLP capabilities on information with higher priorities. A typical data classification should include the following:

- Develop a standard or policy for data classification
- Identify data type by departments
- Identify administrator/custodian/users for each data type
- Identify systems maintaining, processing, or storing each data type
- Specify the criteria of how the data will be classified and labelled
- Create an user awareness program

3 Develop Policies, Standards and Procedures

Comprehensive policies, standards, and procedures are the basis for an effective DLP solution. By referencing to established policies, standards, and procedures, the following criteria can be defined for DLP tools to meet:

- Target data classification(s) that require protection from DLP
- What actions are permitted to be performed on such data
- What are the security violations that require DLP to prevent and alert
- What are the handling processes for identified violations
- Whom should be informed when there are security violations identified

Developed policies, standards, and procedures should be reviewed and approved by management of relevant parties before finalisation.

Reference:
http://www.isaca.org/Knowledge-Center/Research/Documents/DLP-WP-14Sept2010-Research.pdf
http://www.ironport.com/pdf/ironport_dlp_booklet.pdf

V. Hardening Steps for Data Leakage Prevention — Implementation

A comprehensive DLP solution is usually a combination of Network DLP, Endpoint DLP, Embedded DLP components and employee training program. The following addressed several techniques / processes to mitigate the data leakage threats:

1 Secure Content Management

This technique is often used by Network DLP to analyse the traffic passing through a specific gateway within a university. It examines the content of the messages and looks for specific keywords, patterns or fingerprinting (i.e. hashing of data at rest) that may belong to sensitive data. Examples are:

- Keywords include "Confidential", "Restricted", "Internal Only"
- Regular expressions that match with specified data format, e.g. 1 character following by 7-digit sequence could indicate Hong Kong identification card number
- Outbound files that match with the stored data fingerprints (i.e. hash numbers)

2 Embedded DLP in Applications

Many application software are embedded with DLP functionalities to provide first-tier protection against unauthorised access, copy and printing of sensitive information. Several frequently used applications are listed below:

- For Microsoft Office users, they should enable password protection for confidential documents or spreadsheets through "Save As > Tools > General Options > Password to Open"
- For Adobe Acrobat users, they can go to "Advanced > Security > Encrypt with Password"
- For Microsoft Access users, they can activate password encryption by going to "Database Tool" tab and click "Encrypt with Password"

3 Thin Client

Universities can consider implementing disk-less thin clients as an Endpoint DLP solution to ensure that only necessary data needed by the users to do their jobs is released to them. Disable or removal of USB from the thin clients will also prevent users from copying sensitive data to removable media. Major vendors for thin client solutions include IBM, HP and SUN, Wyse Technology and NComputing.

4 Restriction on Removable Media

To prevent data from being copied to removable media like CD, DVD, portable hard drive and USB stick, universities should establish corresponding policy or standards, stating that only authorised personnel are allowed to do so. By default, all computers and laptops should have their CD/DVD writers and USB ports removed or disabled. For laptops from which the CD/DVD writers cannot be removed, universities should uninstall relevant drivers and software for CD/DVD burning, and monitor whether unauthorised installation of burning tools by users.

Reference:
http://www.sans.org/reading_room/whitepapers/awareness/data-leakage-threats-mitigation_1931
http://www.isaca.org/Knowledge-Center/Research/Documents/DLP-WP-14Sept2010-Research.pdf

V. Hardening Steps for Data Leakage Prevention — Implementation (cont'd)

5 Application Proxy Firewalls

Unlike stateful firewalls that only examine transport and network layers, application proxy firewalls work on all 7 layers of the OSI model. They strip down the network traffic and re-assemble it again, analyse specific commands or payloads carried by the packets. For examples, a university may configure its application proxy firewall to filter FTP commands "APPEND", "MKDIR" and "PUT" in order to prevent uploading of sensitive data through FTP programs. The university can also utilise the keyword searching function to examine outgoing e-mails and reject any e-mails containing keywords, regular expressions or patterns of data possibly classified as internal, restricted or confidential.

6 Secured Data Transmission via Internet

Secured method should also be implemented by universities when sensitive data is required to be transmitted over the Internet or to be accessed remotely by authorised external parties. A popular means is to deploy Virtual Private Network (VPN) with Secure Sockets Layer (SSL) capability, which creates a virtual "tunnel" connecting two endpoints and the network traffic traverse through the "tunnel" is encrypted. One popular VPN product is Cisco Easy VPN, which provides various VPN solutions for small/medium organisations to large enterprise.

7 Training and Awareness

As almost half of the data leakages are accidental because of human negligence, it is critical for universities' members to have a strong awareness of the acceptable use of information resources and necessary preventive measures towards data leakage threats and vulnerabilities. The awareness training should typically include the following topics:

- **Classification and Handling of Universities Information Asset** – before implementing DLP, users must know how to distinguish sensitive data and the respective protection required.

- **Risks and Consequences of Data Leakage** – Users should be aware of the risks and serious consequences of leaking sensitive data to unauthorised parties. Examples on loss of patent secrets or loss of personal privacy data are recommended to be used during the training.

- **Policies and Procedures for DLP** – In this section, users are informed of the policies and procedures established by the universities to enforce DLP. The major components include DLP techniques, useful tools approved by the universities, DOs & DON'Ts, reporting and escalation of data leakage incidents.

Reference:
http://www.cisco.com/en/US/products/sw/secursw/ps5299/products_white_paper09186a00800a4b36.shtml
http://www.sans.org/reading_room/whitepapers/awareness/data-leakage-threats-mitigation_1931

VII. Summary

The development of networking and mobile computing technologies has posed serious threats to the data security of organisations including universities. As the capabilities of data transmission and storage are being continuously improved nowadays, data leakage incidents may result in more significant damages, diminishing organisations' value and reputations.

In developing DLP solutions, management should consider all types of data (i.e. data in motion, data at rest and data in use) and work closely with IT professionals and general users to determine the user requirements and suitable DLP products.

A comprehensive and effective DLP solution requires the commitment from both the management and general users to carefully determine the system specifications, functional requirements and data coverage, so that the solution can best fit in the university's existing IT infrastructure and operational process and would not introduce inefficiencies and incompatibilities.