Data Leakage Prevention Adoption Model & DLP Maturity Level Assessment

Mohammed A. Alsuwaie
DFIRe Lab, School of Computer Science
University College Dublin, Ireland
+966 (50) 582 9833
mohammed.alsuwaie@ucdconnect.ie
suwaiema@yahoo.com

Babak Habibnia
Centre of Excellence in Terrorism, Resilience, Intelligence and Organised Crime research (CENTRIC)
Sheffield Hallam University, UK
+353 87 229 5052
b.habibnia@shu.ac.uk

Pavel Gladyshev
DFIRe Lab, School of Computer Science,
University College Dublin, Ireland
+353 (87) 924 4157
Pavel.gladyshev@ucd.ie

2021 International Symposium on Computer Science and Intelligent Controls (ISCSIC) | 978-1-6654-1627-6/21/$31.00 ©2021 IEEE | DOI: 10.1109/ISCSIC54682.2021.00077

***Abstract*—Data is the most valuable resource organizations possess. Nowadays, intentional and unintentional data leakages to unauthorized entities have dramatically increased in the private and government sectors for different reasons and purposes. Data leak poses a severe threat to the data security of these organizations. In the private sector, such leakage can damage the reputation of a business and weaken its competitiveness, while a leakage in the government sector can severely impact its public and diplomatic relationships and even threaten national security. Therefore, private and government sectors have recently strived to adopt the Data Leakage Prevention (DLP) program to enhance data security and minimize data leakage. However, some organizations limit their DLP programs to DLP technology implementation. DLP technology represents only a part of the whole process and does not include other essential elements such as planning, policy, process, data ownership and classification, data access control, training, and awareness. A comprehensive adoption of these elements might prevent data leakage. This research paper aims to provide a model for DLP adoption and address the most critical elements of the DLP Adoption Model, which are correlated with the DLP Maturity Level Assessment, the "DLP Maturity Level Grid," to measure the success of the model.**

Key words-Organizational Security; Incident Response; Information Security; Risk Management; Adoption; Data Leakage Prevention; DLP; Detection; Prevention

I. INTRODUCTION

Data is an essential asset to organizations where large amounts of data are handled daily in many ways and through many people. IT technologies, such as computers, the internet, e-mail, portable devices, external data storage, and social media, increase the likelihood of unauthorized disclosure and loss of data that leave an organization's boundary open to unauthorized entities. Data leakage creates a significant risk to the private and government sectors. It may negatively impact the credibility of the private sector in the market, competitive advantages, partnerships, and customer credentials, while in the government sector, data leakage may affect public and diplomatic relationships and national security.

The importance of data drives the need to have a robust security system and strategy to protect data sensitivity. Network security systems such as firewalls and Intrusion Detection and Prevention Systems (IDPS) were introduced long ago to prevent outbound threats; however, the same risk may come from inbound sources through data leakage. Thus, the Data Leakage Prevention adoption has been introduced to overcome the deficiency of internal threats. However, despite the potential of data leakage prevention (DLP) programs, very few organizations succeed in following the correct strategy for adoption.

The importance of having a successful data leakage prevention adoption strategy motivates the proposal for a model of Data Leakage Prevention (DLP) adoption as work developed from the current research studies and surveys conducted by academic and professional researchers from the perspective of data protection. This research paper provides a model that should serve as a starting point for organizations to successfully adopt data leakage prevention and addresses the eleven elements of the DLP Adoption Model. However, since the model might encounter several challenges during implementation, these must be considered and addressed before proceeding with the model adoption.

This research paper is structured as follows: Section 2 discusses the review of the past research studies and surveys regarding the adoption of data leakage prevention, along with an analysis of their shortcomings. In the same section, the problem statement and the results of the interviewed IT professionals are furnished. Section 3 describes the proposed solution, which is a comprehensive model for data leakage prevention adoption. Section 4 assesses the DLP Adoption Model evaluation and evaluates the DLP Adoption Model success factor using the DLP Maturity Level Grid explained in Section 5. Section 6 presents the evaluation findings with the screenshots of the assessment results before and after using the model, including the pros and cons, and challenges

978-1-6654-1627-6/21/$31.00 ©2021 IEEE
DOI 10.1109/ISCSIC54682.2021.00077

encountered during the model evaluation. Finally, section 7 presents the conclusion and suggestions for future work.

II. LITERATURE REVIEW

In the early days of data leakage detection and prevention, academic researchers focused on protecting the distributed copy by introducing a watermarking technique. Watermarking is embedded as an algorithm code in the distributed document given to the data distributor agents.

Karthik, Ramkumar, and Sundaram (2014) enhanced the embedded code algorithm by proposing a guilt model that relies on data allocation strategies to advance the probability of identifying the guilty agent who leaks the distributed data and detects the leakage of sensitive information. The guilt model is based on distributing the data intelligently among the agents using mock objects to identify the guilty agents through the invisible watermarking technique. The guilt model increases the probability of identifying guilty agents when the sum objective is minimized [1]. While the watermarking technique may continue in data protection, the growth of the internet has increased the risk of data leakage, so that a new strategy for data leakage prevention and associated technology becomes necessary.

Raman, Kayacık, and Somayaji (2011) pointed out that data leakage technologies available from different vendors address only the threat of data leakage, which is part of the comprehensive strategy of data leakage detection and prevention implementation. Their study also indicated that as data leakage prevention implementation is complicated, determining the type of data to protect, identifying the use of data, and how the data may be leaked requires understanding the business functions before starting the implementation. Furthermore, other crucial comprehensiveness elements of data leakage prevention adoption are discarded, such as the policy, process, data classification, data control, training, awareness, and notification. The authors suggest more research in the data leakage prevention field to resolve the associated problems [2].

Shabtai, Elovici, and Rokach (2012) described data leakage prevention as a solution that detects and prevents data leaks. Their studies characterized data leakage prevention solutions according to the source of leakage, the location of the data when leakage occurs (e.g., data in use, in transit, and at rest), the channels of leakage, the approaches of detection and prevention, and the actions taken when such leakage occurs. Shabtai et al. (2012) discussed the detection and prevention approaches, such as data access control, filtering, fingerprinting, tagging, encryption, policy, and awareness. The survey in their study describes and analyses data leakage detection and prevention from the viewpoint of commercial solutions and academic research. The authors call attention to the Gartner report [Outllet, 2009], according to which large organizations understand the importance of implementing the data leakage prevention (DLP) technology as an element of data leakage prevention solution, to minimize the possibility of an intentional or unintentional leak of data [3].

Kumaresan (2014) listed the key considerations that can help organizations successfully adopt data leakage prevention as planning, efficient process, risk assessment, data classification, and compliance with data privacy regulations. The most crucial consideration in Kumaresan's paper is implementing data leakage prevention in a phased manner and avoiding poor policy and procedure. In general, the key considerations mentioned by Kumaresan are constructive and may be considered in the planning of adopting data leakage detection and prevention [4].

Shey and Kindervag (2015), in the article "Rethinking DLP," (published by Forrester Research Inc.), emphasize that most of the companies fail in data leakage prevention adoption because they do not have the policies, process, and data classification in place before starting DLP deployment. Their study suggests five process stages that organizations should consider before reaching DLP maturity, viz., data discovery, data classification, data consolidation, DLP policy designing, and DLP policy enforcement. The study also points to the necessity of using the Forrester DLP maturity grid as a process guide to ensure the success of a DLP strategy. However, the grid does not provide precise details about the DLP policy and process [5].

Alneyadi, Sithirasenan, and Muthukkumarasamy (2016) discussed data leakage detection and prevention as a study of the typical DLP system from an academic research perspective. They analyzed data leakage prevention techniques; however, there is no actual perception or suggested model on successfully adapting the data leakage detection and prevention as a comprehensive strategy for protecting data from disclosure and how the strategy helps digital investigators discover the evidence efficiently [6].

Further, the published literature of research firms like Gartner Group Inc., EY (Ernst & Young), Forrester Research, Inc., and Information Systems Audit and Control Association (ISACA) discuss the data leakage detection and prevention from the management and commercial point of view.

In summary, most research papers in the field of data leakage detection and prevention aver that organizations sometimes misunderstand the adoption of DLP due to poor planning and ineffective process, which causes failure in detecting and preventing data leakage; further, there is no specific methodology to follow in DLP adoption.

A. Problem Statement

Preserving the confidentiality, integrity, and availability of data are the primary goals of any organization. Data could reside on numerous media, such as computer hard drives, websites, USB Flash drives, mobile devices, office documents, and e-mails. Data leakage arises from intentional and unintentional disclosures, stolen devices, hacking, and illegal practices of internal employees [7]. Therefore, all organizations strive to protect data from vulnerability, alteration, and leakage by unauthorized entities [8].

Many organizations implement Data Leakage Prevention (DLP) technology obtained from many vendors. Some organizations dive straightaway into implementing a state-of-art DLP technology and discard the other comprehensiveness elements. Thus, adopting a data leakage prevention program is challenging for any organization if it is not systematically adopted [9]. Organizations should answer the following six

fundamental questions before adopting a data leakage prevention program:

- What is critical data for the business?
- What are the types of data classification?
- Who can access classified data?
- Where is data stored?
- How is the business data used?
- What is the current mechanism for detecting and preventing data leaks (e.g., policy rules, technology, and access control)?

B. Interview Result

Eight IT professionals from different organizations were interviewed to understand the deficiency in data leakage prevention adoption. The following are the key findings from the interviews:

- Retailers and educational organizations have a low level of understanding of data leakage threats.
- Retailers and educational organizations lack the adoption strategy for data leakage detection and prevention.
- Retailers and educational organizations have a low level of visibility and awareness of data leakage.
- The financial and gas & oil industries know the importance of data leakage threats and have a matured implementation of the DLP strategy.
- The financial and gas & oil industries have greater visibility and awareness of data leakages.
- Figures 1 and 2 illustrate the results of interviews with the private and government sectors:

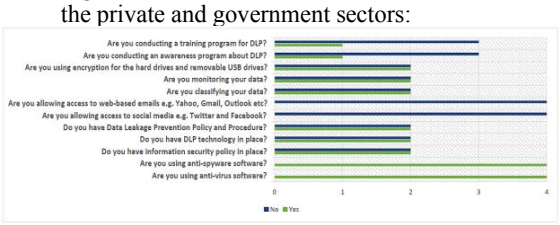


Figure 1. Interview result for the private sector.

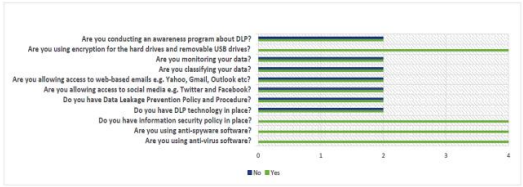


Figure 2. Interview result for government sector.

III. DATA LEAKAGE PREVENTION ADOPTION MODEL

Most research papers on data leakage detection and prevention assumed that organizations might misunderstand the adoption of data leakage prevention due to ineffective planning, policy, process, data classification, data access control, training, and awareness, which lead to failure in the detection and prevention of data leakage [10].

In addition to the efforts of industries and professional researchers to develop new strategies to detect and prevent unauthorized entities from accessing sensitive data, this paper proposes a comprehensive solution called a Model of Data Leakage Prevention Adoption (DLP Adoption Model), which describes the process for the effective implementation of data leakage prevention. In the following section, the DLP Adoption Model describes the data leakage detection and prevention elements, viz., planning, policy, process, data leakage channels, data states, data ownership and classification, data access control, technology, data log files analysis, reporting, notification and escalation, training, and awareness. The DLP Maturity Level Assessment also measures the efficiency and success of the DLP Adoption Model implementation.

A. DLP Adoption Model

An organization must have efficient DLP planning to mitigate the possibility and impact of data leakage before acquiring data leakage prevention (DLP) technology. DLP Technology can enforce DLP policy and process efficiently; however, relying solely on technology does not mitigate the risk of data leakage. Organizations must adopt a comprehensive solution to detect and prevent any potential data leak while data is in use, transit, and rest. Figure 3 shows the diagram of the proposed Data Leakage Prevention Adoption Model (DLP Adoption Model):

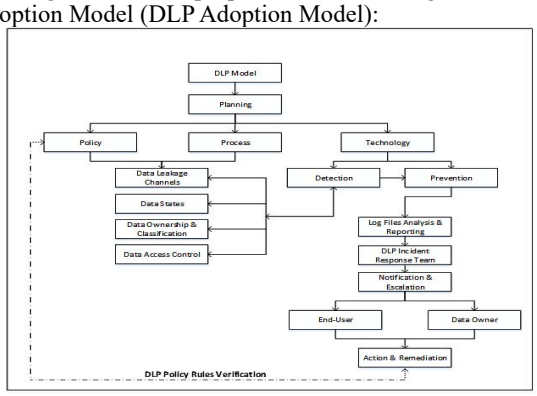


Figure 3. Data leakage prevention model.

1) Planning: The planning phase in the DLP Adoption Model is the roadmap and microscope of data leakage detection and prevention. This phase constitutes three significant elements: DLP Policy, DLP Process, and DLP technology. In this phase, the organization should carry out the following before adopting the model:

- Identify the objectives of data leakage prevention (DLP) implementation.
- Involve stakeholders in the planning process.
- Determine the DLP's scope of work and ensure data leakage detection and prevention requirements are specific, realistic, and measurable.
- Identify the data owner and custodian.
- Define and constitute the DLP Incident Response Team (IRT).

- Classify data according to its sensitivity and security impact.
- Determine proper controls for data access and authorization.
- Identify the type of data leakage channels.
- Identify the tools used to collect and analyze the log files.
- Determine the mechanism, the text body of notification and escalation, and the recipient in case of a data leakage incident.

2) DLP Policy: A typical deficiency in adopting DLP is establishing DLP Policy because it is an initial point in strategy before data classification and data access controls. DLP Policy contains multiple rules with conditions and actions for handling by technology based on specific DLP incidents. The data owner is responsible for identifying the data to be protected and defining effective DLP policy rules to monitor and prevent data from leaving the organization's boundaries. DLP policy rules should be examined and approved by the data owner before being deployed in the production environment, to eliminate false positives. Further, DLP Policy rules should be examined and reviewed regularly to cope with any changes in technology and new security threats. Figure 4 shows the DLP Policy rules lifecycle.

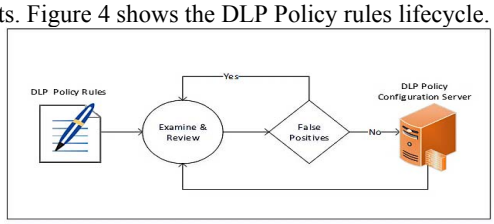


Figure 4. DLP policy rules lifecycle.

3) DLP Process: The organization should consider DLP Processes such as people, data governance, risk assessment, detection, and prevention. Figure 5 shows the typical DLP process [9]:

Figure 5. DLP process.

a) People: This includes data owners, data custodians, and end-users in an organization. The data owner holds data ownership of being accountable for data classification according to its sensitivity and security impact. The data custodian is responsible for implementing DLP policy rules in DLP technology to prevent data from reaching unauthorized entities. Finally, the end-user is responsible for using the data according to the organization's policies and procedures.

b) Data Governance: It is a process within the organization to govern the DLP plan, DLP policy, and DLP strategy to ensure confidentiality, integrity, and data availability. Data governing requires identifying the regulatory and security controls requirements the organization should follow to comply with information security and ensure the policies and procedures are in place.

c) Risk Assessment: It is a process to identify the threats associated with using the organization's data. First, the organization must determine and evaluate the sensitive data. Thereafter, the sensitive data can turn into rules, and would be fed to the content of DLP discovery technology as words, patterns, and finance symbols.

d) Detect: It is a process that deploys in network devices or end-user devices to flag sensitive data that should not leave an organization. The flagging mechanism is performed by implementing a set of DLP policy rules to fulfill the purpose of DLP detection technology.

e) Prevent: It is a transition from proactive to reactive, for taking appropriate action on the flagged sensitive data by detection technology. The prevention process enforces DLP policy rules to take the action of blocking, quarantine, deleting, encrypting, notifying, and acknowledging the data transmitted, as appropriate.

4) Data Leakage Channels: The scope of data leakage channels is broad and not limited to just e-mail, social media, and webmail. Data Leakage can originate from a malicious attack, laptop theft, or a lost or stolen removable storage device. Therefore, the monitoring of transport data channels that may leak data is critical to mitigate the risk of data leakage. The following are a few examples to show the concept that data can travel through [7]:

Endpoint	**Network**	**Storage**
• Hard Drives.	• E-mail.	• Web Servers.
• USB Flash Drives.	• Webmail.	• SharePoint.
• Removable Storage Devices.	• File Transfer Protocol (FTP).	• File Servers.
• Printers and Fax machines.	• Social Media, such as Twitter, Facebook, WhatsApp, and Snapchat.	

5) Data States: Knowing where data is stored, where it is in transit, and accessing data are critical steps in data leakage detection and prevention adoption. Data states provide the insight needed to build an effective DLP policy and process that protect data and help in digital forensics investigation. The data states come in three stages [3]:

a) Data in Use (Endpoint): Data is located on the end-user workstation, laptop, USB flash drive, removable hard

drive, printer, and fax machine. In addition, the endpoint agent of DLP technology is commonly installed on the end-user device to monitor the use and transfer of data [3] [4].

b) Data in transit (Network): Data is flown on the network, such as the internet, social media, e-mail, webmail, and FTP. The discovery and monitoring feature of DLP technology is used to detect and inspect the data traffic flowing on the network [3] [4].

c) Data at rest (Storage): Data is stored in the database storage, file servers, web servers, file sharing, and external data storage. The discovery and monitoring feature of DLP technology is used to encrypt and control data access [3] [4].

6) Data Ownership and Classification: It is crucial to identify which data is valuable to an organization, to ensure that data receive a proper level of protection commensurate to its value. The data value should drive the structure of data ownership and classification [11]. A process must be developed for data ownership and classification structure to align data sensitivity and value with the business functions. The proper identification of data ensures DLP technology functioning as designed and knowing what to detect and prevent. The technology verifies the data classification through policy rules to detect and prevent data from being transferred outside the organization without appropriate approval.

Notably, there is no global standard for data classification because it is based on data sensitivity and value to the organization. However, it is challenging to classify data according to monetary value. Thus, data is usually classified using a scale like Low, Medium, and High [12].

Data classification needs to be regularly reviewed, and the data classification structure should be able to handle the data classification scale changes easily. Table I illustrates the typical scheme for data classification used by many organizations [12].

TABLE I. DATA CLASSIFICATION DEFINITION AND VALUE SCALE

Data Classification	Sensitivity & Value	Definition
Non-Business Data	Low	Information is published publicly and does not harm the organization's business.
Business General Data	Medium	Information is used in day-to-day business and has already been approved to be shared with others.
Confidential Data	High	Information includes financial documents, intellectual property, patents, design documents, and data related to an organization.

7) Data Access Controls: Data access controls are the safeguards or countermeasures that mitigate risk to data confidentiality and integrity. The conventional methods in this regard are authentication and authorization. Authentication is verifying the request data access, while authorization is determining the eligibility of data access. Data leaks are difficult to detect and prevent when data access control is not correctly implemented and maintained. Consequently, the digital forensics investigation process would be complicated and useless. Additionally, some organizations misuse the data access control, so the least privileged access to end-users is essential to control data access, but this should not hinder productivity [6].

An organization must decide which data access control is the most appropriate based on data classification and business requirements. It is always preferable to implement multiple data access controls called Defence-in-Depth to avoid single data access control failure. The conventional forms of data access control are [6]:

- Ensure that antivirus software is installed and updated daily.
- Ensure that appropriate security controls are installed on mobile devices such as smartphones and tablets.
- Implement two-factor authentication.
- Implement an audit trail to log the data access activities.
- Employ Segregation of Duty (SoD) methodology to segregate data access privilege into what data can be accessed, who can access it, and when it can be accessed.
- Monitor data that flows through corporate e-mail, FTP, and the internet, and apply the policy rules of auditing, such as block, encrypt, and quarantine.
- Limit access to webmail services, such as Yahoo, Gmail, and Outlook to avoid downloading malicious attachments or receiving phishing e-mails.
- Prevent end-user from uploading data on Internet-based storage and file sharing, such as Dropbox, Google Drive, RapidShare, and SkyDrive, because the organization cannot control the access to these services.
- Limit the copy and access to the USB removable drive without getting proper permission.
- Establish DLP policies and standards.

8) DLP Technology: Different DLP technologies are available in the market from many vendors, with each vendor's technology having different functionalities and capabilities. However, they have the same concepts that

apply to DLP technology. As illustrated in Figure 3, DLP technology has two main modules to handle data leakage incidents: detection module and prevention module. In general, these two modules comprise the following hardware and software components to perform the detection and prevention operations [6] [13]:

a) Endpoint: It is a software agent installed on desktop workstations, laptops, and servers to detect and prevent unauthorized transmission, printing, modification, copying, and pasting data. Further, it provides the reports required for analysis [6] [13].

b) Network: It is a device installed on the network to monitor, analyze, and block any suspicious network traffic through real-time monitoring [6] [13].

c) Discovery: It is a device installed on a network to discover the activities of the end-user based on the policy rules defined and take proper action against violation [6] [13].

The success of DLP technology depends on well-defined DLP policy rules and data classification used by DLP technology to detect and prevent leakage. Figure 6 illustrates DLP Technology Solutions Architecture:

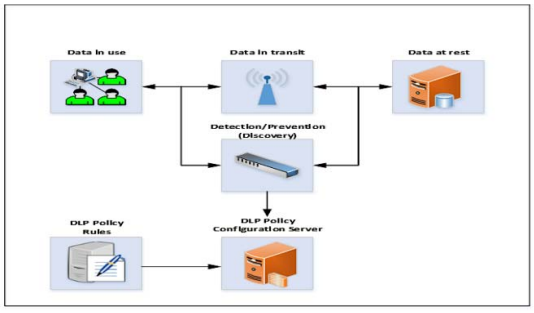


Figure 6. DLP technology solution architecture.

9) DLP Notification & Escalation: DLP technology usually generates incidents potentially impacting the organization's data leaving the organization's boundaries. These incidents should be reviewed and analyzed by the DLP Incident Response Team or Digital Forensic Investigators for appropriate action. The remedy for incidents varies based on the negative impact on the organization's business. Hence, an effective and responsive process must be in place to help the team review and analyze the generated incidents. Further, end-user and data owners should be notified if there is a violation of the DLP policy rule. The notification should explain how the breach occurred and the suggested remediation, such as allowing, encrypting, quarantining, or blocking. Further, the data owner requires being informed of the violation. Figure 7 illustrates the lifecycle of the DLP incidents process [4]:

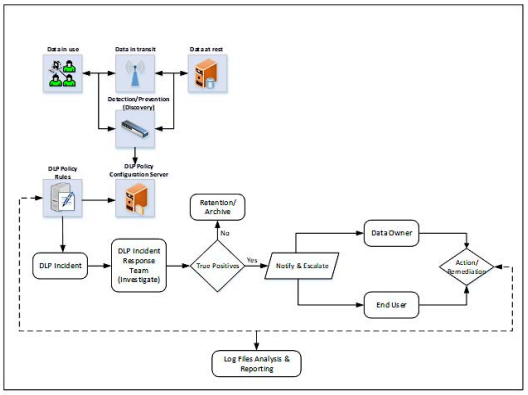


Figure 7. DLP incident process lifecycle.

10) DLP Training and Awareness: Humans are the weakest link in the chain of data leakage threats. Therefore, an organization must educate end-users on recognizing the threat of data leakage and to operate securely. Training and awareness of the end-user are an essential part of the success of data leakage prevention and can add a wall defense to data protection. DLP Policy should provide clear guidance to the end-user on the risk of data leakage. The training and awareness sessions are proactive approaches to data leakage prevention. Proper training and awareness o end-users are crucial factors to resolve the concerns and ensure the success of data leakage prevention. Therefore, it is imperative to conduct periodic training and awareness sessions for the organization's end-users and supplemental training such as computer-based training, posters, and SMS text messages [4].

11) DLP Log Analysis & Reporting: Gaining insight into outbound and inbound data leakage activities and keeping abreast of DLP Technology log files is challenging, as the DLP Technology generates a massive quantity of audit and log files daily. Therefore, it is necessary to use an audit and log files analyzer tool to gain insight into the threats and channels of data leakage to improve the data leakage prevention posture strategy and create visualized reports [4]. Thereby, log analysis and reporting are the evaluation of audit and log files of data leakage activities [4] [14].

Access control should be applied to the collected log files to protect them from alteration and unauthorized access. DLP policy should address the preservation and retention of log files. Log files are needed in generating reports and as evidence in the intentional and unintentional data leakages. The log file analysis and reporting also help the organization mitigate the risk of data leakage and comply with the internal policy, regulations, and rapid response to the incident. The collected log files can be used to generate the required reports by the organization's management. The following are examples of reports that are usually needed by the management:

- A total number of open DLP Incidents.
- A total number of DLP Incident triggers by intentional and unintentional data leakage for a

particular channel of Data Leakage (e.g., e-mail, USB).

- A total number of DLP Incident triggers by intentional data leakage for a particular channel of Data Leakage (e.g., e-mail, USB).
- A total number of DLP Incident triggers by unintentional data leakage for a particular channel of Data Leakage (e.g., e-mail, USB).
- A total number of resolved DLP incidents.
- An average of the total number of DLP Incident triggers by unintentional data leakage. The KPI formula is as follows:

$$\frac{\text{Total Number of Unintentional Data Leakage Incidents}}{\text{Total Number of Intentional and Unintentional Incidents}} \times 100 \quad (1)$$

- An average of the total number of DLP Incident triggers by intentional data leakage. The KPI formula is as follows:

$$\frac{\text{Total Number of Intentional Data Leakage Incidents}}{\text{Total Number of Intentional and Unintentional Incidents}} \times 100 \quad (2)$$

- DLP Incident Average Resolution Time. The KPI formula is as follows:

$$\frac{\text{Incident End Time} - \text{Incident Start Time}}{\text{Total Time of resolved Incident}} \quad (3)$$

IV. EVALUATION OF DLP ADOPTION MODEL

The purpose of this evaluation is to test the proposed solution to the Data Leakage Prevention Adoption Model (DLP Adoption Model). The proposed solution was tested on the IT Infrastructure of Market Trading Company, a Saudi retail business established in 1980, operating in food industries. The company has a big warehouse supporting fourteen hypermarkets and a centralized IT department to provide IT services for all its branches in Saudi Arabia.

The company has an intranet used by employees and an extranet used by customers and product sales. In addition, the company's website allows customers and partners to perform financial transactions, which require manipulating sensitive data. Therefore, sensitive data drives the need to have robust security systems and strategy. The company has not yet implemented the data leakage prevention program but are plans to implement a robust security system and strategy for data leakage prevention to overcome the outbound and inbound threats.

The company agreed to use the proposed solution to provide a model for Data Leakage Prevention Adoption and address the essential elements of the DLP Adoption Model. Further, DLP Maturity Level Grid is used to measure the success of the DLP Adoption Model.

A. Evaluation Scope

The scope of the evaluation is to address the weakness and strengths of the DLP Adoption Model. In addition, DLP Maturity Level Grid is used to assess the overall progress of implementing each element of the DLP Adoption Model.

B. Evaluation Objectives

The objective of the evaluation is to identify the success factors for implementing each element of the DLP Adoption Model.

V. DLP MATURITY LEVEL ASSESSMENT

The DLP Adoption Model can be assessed and measured using the DLP Maturity Level Grid to evaluate the overall progress of implementing each element of the DLP Adoption Model. DLP Maturity Level Grid identifies and addresses any gaps before and after the implementation of each element. The assessment ensures that the implementation process accurately reflects the success of the DLP Adoption Model. The DLP Maturity Level Grid comprises three maturity levels: Weak implementation, Moderate implementation, and Mature Implementation.

Weak Implementation

This means that the DLP Adoption Model is unachievable (Weak). Therefore, it is strongly recommended to review and re-examine elements that score "Low", to achieve maturity.

Moderate Implementation

This means that the DLP Adoption Model is moderated. Therefore, it is recommended to enhance elements that score "Low or Medium" to achieve maturity.

Mature Implementation

This indicates that the DLP Adoption Model is achievable. All elements are implemented efficiently. The following section describes the specification details of the DLP Maturity Level Grid.

A. DLP Maturity Level Grid

DLP Maturity Level Grid is developed by using a Microsoft Excel worksheet divided into eleven rows (Elements of DLP) and three columns (Levels 1, 2, and 3).

The following are the specifications of the DLP Maturity Level Grid:

- Each level has a single point (point of 0, or 0-1, and 0-2) and Element Score, which is the result of multiplication:

$$\text{Element Score} = \text{Point} \times \text{Level (\#)} \quad (4)$$

- Each Element has a Total Score, which is the sum of the Element Scores under each level:

$$\text{Total Score} = \sum \text{Element Score} \quad (5)$$

- The Total Score has a Score Scale of three (Low, Medium, and High).
- The Overall Score is the sum of the Total Scores of all the elements that indicate the maturity level (Weak Implementation, Moderate Implementation, or Mature Implementation).

$$\text{Overall Score} = \sum \text{Total Score} \quad (6)$$

Table II illustrates the formulas of Key Performance Indicators (KPIs) used to measure the Maturity Level of the DLP Adoption Model:

TABLE II. FORMULAS OF KEY PERFORMANCE INDICATORS (KPIS)

Key Performance Indicators (KPIs) of Maturity Level Metrics of DLP Adoption Model			
Element Score	*Element Score = Point × Level (#)*		
Total Score	*Total Score = ∑ Element Score*		
Score Scale	**Low**	**Medium**	**High**
	0	>= 1 < 6	>= 6
Overall Score	*Overall Score = ∑ Total Score*		
Maturity Level	**Weak**	**Moderate**	**Mature**
	< 20	>= 20 < 66	>= 66

VI. EVALUATION FINDINGS

The following sections describe the evaluation findings:

A. Preliminary Analysis

Preliminary analysis was conducted to answer six fundamental questions before adopting the DLP Adoption Model. The answers to the questions reflect the position of the client in terms of the data leakage prevention implementation; Table III shows the questions and answers of the client:

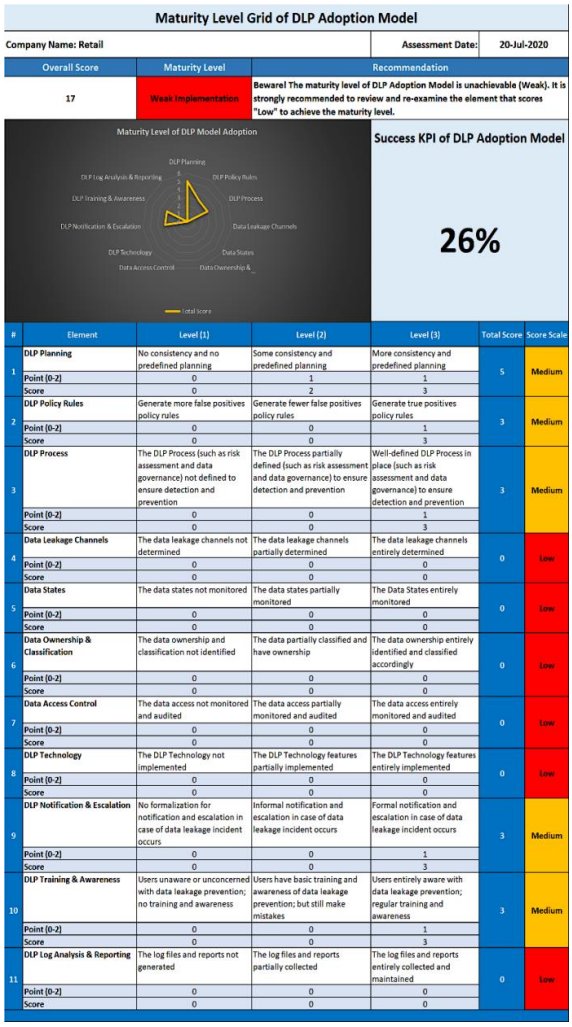
TABLE III. DLP ADOPTION MODEL: QUESTIONS AND ANSWERS

#	Question	Answer
1	What is critical data for business?	Trade secrets, finance, sales, marketing, and personal information of employees and customers.
2	What are the types of data classification?	None.
3	Who can access classified data?	All users, based on the functional role of the employees.
4	Where is data stored?	In the database servers and users' workstations.
5	How is business data used?	In-transit, In-use, and at Rest.
6	What is the mechanism for detecting and preventing data leaks currently used?	Only Security Policy and Instructions.

B. Assessment Result Before using DLP Adoption Model

The DLP Adoption Model can be assessed and measured using the DLP Maturity Level Grid to evaluate the overall progress of implementing each element of the DLP Adoption Model.

Figure 8 illustrates DLP Maturity Level Grid that identifies and addresses any gaps before the implementation of each element.



Maturity Level Grid of DLP Adoption Model

Company Name: Retail		Assessment Date:	20-Jul-2020
Overall Score	**Maturity Level**	**Recommendation**	
17	Weak Implementation	Beware! The maturity level of DLP Adoption Model is unachievable (Weak). It is strongly recommended to review and re-examine the element that scores "Low" to achieve the maturity level.	

Maturity Level of DLP Model Adoption

Success KPI of DLP Adoption Model: 26%

#	Element	Level (1)	Level (2)	Level (3)	Total Score	Score Scale
1	DLP Planning	No consistency and no predefined planning	Some consistency and predefined planning	More consistency and predefined planning	5	Medium
	Point (0-2)	0	1	1		
	Score	0	2	3		
2	DLP Policy Rules	Generate more false positives policy rules	Generate fewer false positives policy rules	Generate true positives policy rules	3	Medium
	Point (0-2)	0	0	1		
	Score	0	0	3		
3	DLP Process	The DLP Process (such as risk assessment and data governance) not defined to ensure detection and prevention	The DLP Process partially defined (such as risk assessment and data governance) to ensure detection and prevention	Well-defined DLP Process in place (such as risk assessment and data governance) to ensure detection and prevention	3	Medium
	Point (0-2)	0	0	1		
	Score	0	0	3		
4	Data Leakage Channels	The data leakage channels not determined	The data leakage channels partially determined	The data leakage channels entirely determined	0	Low
	Point (0-2)	0	0	0		
	Score	0	0	0		
5	Data States	The data states not monitored	The data states partially monitored	The Data States entirely monitored	0	Low
	Point (0-2)	0	0	0		
	Score	0	0	0		
6	Data Ownership & Classification	The data ownership and classification not identified	The data partially classified and have ownership	The data ownership entirely identified and classified accordingly	0	Low
	Point (0-2)	0	0	0		
	Score	0	0	0		
7	Data Access Control	The data access not monitored and audited	The data access partially monitored and audited	The data access entirely monitored and audited	0	Low
	Point (0-2)	0	0	0		
	Score	0	0	0		
8	DLP Technology	The DLP Technology not implemented	The DLP Technology features partially implemented	The DLP Technology features entirely implemented	0	Low
	Point (0-2)	0	0	0		
	Score	0	0	0		
9	DLP Notification & Escalation	No formalization for notification and escalation in case of data leakage incident occurs	Informal notification and escalation in case of data leakage incident occurs	Formal notification and escalation in case of data leakage incident occurs	3	Medium
	Point (0-2)	0	0	1		
	Score	0	0	3		
10	DLP Training & Awareness	Users unaware or unconcerned with data leakage prevention; no training and awareness	Users have basic training and awareness of data leakage prevention; but still make mistakes	Users entirely aware with data leakage prevention; regular training and awareness	3	Medium
	Point (0-2)	0	0	1		
	Score	0	0	3		
11	DLP Log Analysis & Reporting	The log files and reports not generated	The log files and reports partially collected	The log files and reports entirely collected and maintained	0	Low
	Point (0-2)	0	0	0		
	Score	0	0	0		

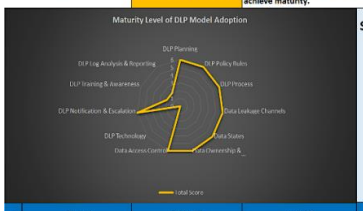
Figure 8. Assessment result before using DLP adoption model.

C. Assessment Result After using DLP Adoption Model

Figure 9 illustrates DLP Maturity Level Grid after the corrections of any gaps in each element.

Maturity Level Grid of DLP Adoption Model

Company Name: Retail		Assessment Date:	20-Jul-2020
Overall Score	**Maturity Level**	**Recommendation**	
52	Moderate Implementation	The maturity level of the DLP Adoption Model is moderated. Therefore, it is recommended to enhance the element that scores "Low or Medium" to achieve maturity.	



Success KPI of DLP Adoption Model

79%

#	Element	Level (1)	Level (2)	Level (3)	Total Score	Score Scale
1	DLP Planning	No consistency and no predefined planning	Some consistency and predefined planning	More consistency and predefined planning	6	High
	Point (0-2)	0	0	2		
	Score	0	0	6		
2	DLP Policy Rules	Generate more false positives policy rules	Generate fewer false positives policy rules	Generate true positives policy rules	6	High
	Point (0-2)	0	0	2		
	Score	0	0	6		
3	DLP Process	The DLP Process (such as risk assessment and data governance) not defined to ensure detection and prevention	The DLP Process partially defined (such as risk assessment and data governance) to ensure detection and prevention	Well-defined DLP Process in place (such as risk assessment and data governance) to ensure detection and prevention	6	High
	Point (0-2)	0	0	2		
	Score	0	0	6		
4	Data Leakage Channels	The data leakage channels not determined	The data leakage channels partially determined	The data leakage channels entirely determined	6	High
	Point (0-2)	0	0	2		
	Score	0	0	6		
5	Data States	The data states not monitored	The data states partially monitored	The Data States entirely monitored	6	High
	Point (0-2)	0	0	2		
	Score	0	0	6		
6	Data Ownership & Classification	The data ownership and classification not identified	The data partially classified and have ownership	The data ownership entirely identified and classified accordingly	6	High
	Point (0-2)	0	0	2		
	Score	0	0	6		
7	Data Access Control	The data access not monitored and audited	The data access partially monitored and audited	The data access entirely monitored and audited	6	High
	Point (0-2)	0	0	2		
	Score	0	0	6		
8	DLP Technology	The DLP Technology not implemented	The DLP Technology features partially implemented	The DLP Technology features entirely implemented	0	Low
	Point (0-2)	0	0	0		
	Score	0	0	0		
9	DLP Notification & Escalation	No formalization for notification and escalation in case of data leakage incident occurs	Informal notification and escalation in case of data leakage incident occurs	Formal notification and escalation in case of data leakage incident occurs	6	High
	Point (0-2)	0	0	2		
	Score	0	0	6		
10	DLP Training & Awareness	Users unaware or unconcerned with data leakage prevention; no training and awareness	Users have basic training and awareness of data leakage prevention; but still make mistakes	Users entirely aware with data leakage prevention; regular training and awareness	2	Medium
	Point (0-2)	0	1	0		
	Score	0	2	0		
11	DLP Log Analysis & Reporting	The log files and reports not generated	The log files and reports partially collected	The log files and reports entirely collected and maintained	2	Medium
	Point (0-2)	0	1	0		
	Score	0	2	0		

Figure 9. Assessment Result After using DLP Adoption Model.

D. DLP Adoption Model Advantages and Disadvantages

The following are the essential pros and cons of the DLP Adoption Model:

1) Advantages of using the DLP Adoption Model:

- Enhance compliance with data regulation and standards.
- Offer complete data visibility and control.
- Maintain evidence for digital investigation.
- Proactive for data leakage.
- Understand where data is stored, where data is transiting, and who can access data.
- Review and monitor DLP implementation progress and learn from mistakes.

2) Disadvantages of not using the DLP Adoption Model:

- Potential for unauthorized transfer of data through any data leakage channel.
- Improper data access control.
- Insufficient response to data leakage and confusing the digital investigation.
- Improper data classification.
- Intentional and unintentional transmission of sensitive data.

E. DLP Adoption Model: Challenges

During the implementation, the model may encounter several challenges leading to the unsuccessful execution of some model elements. These challenges must be considered and addressed before using the model to ensure the success of Data Leakage Prevention (DLP) adoption. These challenges fall into the following broad categories:

- Lack of Commitment from the organization's management.
- Lack of Organizational ownership of data.
- Lack of DLP Training & Awareness.
- Inefficient Technology Implementation.
- Ineffective Incident Response Team.
- Users resist the new changes regarding the data access controls.
- Ineffective Data Governance.
- No Asset Inventory Management Solution.
- There is no IT Service Management Solution implementation to log and track Incident, Problem, Change Management, and Configuration Management Database (CMDB).

VII. CONCLUSION AND FUTURE WORK

In conclusion, the successful adoption of data leakage prevention continues to be a challenge to organizations. Data Leakage Prevention (DLP) technology implementation alone does not guarantee the success of data leakage prevention. However, planning, policy, process, data leakage channels, data states, data ownership and classification, data access control, technology, data log files analysis and reporting, notification and escalation, training, and awareness ensure the likelihood of data leakage prevention. The result of interviews with IT professionals showed that most organizations have some pieces to the jigsaw puzzle of data leakage prevention, but not all of them. Some organizations struggle to find the best way to mitigate data leakage risks. Therefore, this research paper provides a model that should serve as a starting point for the private sector and government organizations toward successfully adopting data leakage prevention. An organization should not rely solely on DLP technology implementation to prevent data leaks and discard the other essential elements, such as DLP planning, DLP policy, DLP process, data classification, data access control, training, and awareness. The organization should have precise planning and identify the data assets to protect before starting data leakage prevention. Further, the effectiveness of the DLP Adoption Model can be measured by using the DLP Maturity Level Grid to ensure the success of the DLP Adoption Model implementation.

Future work will focus on detecting and preventing data leaks that reside in smartphone devices and cloud-based systems as organizations move day-to-day operations toward smartphone applications and cloud-based systems.

REFERENCES

[1] Karthik. R, Ramkumar.S, and Sundaram. K, 'Data Leakage Identification and Blocking Fake Agents Using Pattern Discovery Algorithm,' International Journal of Innovative Research in Computer and Communication Engineering, vol. 2, no. 9, pp. 5660–5667, 2014.

[2] Raman P, Kayacık HG, & Somayaji A. Understanding data leak prevention. Paper presented at the 6th annual symposium on information assurance (ASIA'11). Albany, New York, USA; 2011. p. 27–31. Available at: https://homeostasis.scs.carleton.ca/~soma/pubs/raman-asia2011.pdf. Last accessed 12/02/2020.

[3] Shabtai A., Elovici Y., Rokach L. (2012) A Taxonomy of Data Leakage Prevention Solutions. In: A Survey of Data Leakage Detection and Prevention Solutions. SpringerBriefs in Computer Science. Springer, Boston, MA.

[4] Nageswaran Kumaresan. 'Key Considerations in Protecting Sensitive Data Leakage Using Data Loss Prevention Tools' (2014) 1. Available at: https://www.isaca.org/Journal/archives/2014/Volume-1/Pages/Key-Considerations-in-Protecting-Sensitive-Data-Leakage-Using-Data-Loss-Prevention-Tools.aspx. Last accessed 16/02/2020.

[5] Heidi Shey and John Kindervag. (2015) 'Rethinking DLP: Introducing The Forrester DLP Maturity Grid | Boldon James.'. Available at: https://www.boldonjames.com/resources/complimentary-forrester-research-inc-report-2/. Last accessed 24/02/2020.

[6] Alneyadi S., Sithirasenan E., & Muthukkumarasamy V., (2016). 'A Survey on Data Leakage Prevention Systems.' Journal of Network and Computer Applications, ISSN: 1084-8045, Vol: 62, Page: 137-152. Available at: https://www-sciencedirect-com.ucd.idm.oclc.org/science/article/pii/S1084804516000102?via%3Dihub. Last accessed 07/02/2020.

[7] B, R. K., Gvs, R. K., & Srinivas, Y. (2017). A Survey on Data Leakage Techniques. International Journal of Advanced Research in Computer Science, 8(7). Available at: https://ucd.idm.oclc.org/login?url=https://search-proquest-com.ucd.idm.oclc.org/docview/1931130506?accountid=14507. Last accessed 25/02/2020.

[8] Mrs. Grinal Tuscano et al. 'A Survey on Data Leakage Detection.' Int. Journal of Engineering Research and Applications ISSN: 2248-9622, Vol. 5, Issue 4, (Part -6) April 2015, pp.153-158. Available at: http://www.ijera.com/papers/Vol5_issue4/Part%20-%206/Y50406153158.pdf. Last accessed 07/02/2020.

[9] 'Maximizing the Value of a Data Protection Program' (EY June 2014). Available at: http://www.ey.com/Publication/vwLUAssets/EY_-_Maximizing_the_value_of_a_data_protection_program/$FILE/EY-insights-on-grc-data-protection.pdf. Last accessed 07/02/2020.

[10] Segall L, 'Mark Zuckerberg in His Own Words: The CNN Interview' (CNNMoney, 21 March 2018). Available at: http://money.cnn.com/2018/03/21/technology/mark-zuckerberg-cnn-interview-transcript/index.html. Last accessed 22/04/2020.

[11] National Institute of Standards and Technology, (2004). 'FIPS 199, Standards for Security Categorization of Federal Information and Information Systems. Available at: https://csrc.nist.gov/publications/detail/fips/199/final. Last accessed 11/02/2020.

[12] 'Identify Data Sensitivity | Data ONE.' Available at: https://www.dataone.org/best-practices/identify-data-sensitivity. Last accessed 11/02/2020.

[13] Intuate Group (2011). 'Components of a Data Leak Prevention System | ITWeb.' Available at: https://www.itweb.co.za/content/o1Jr5MxjpbRvKdWL

[14] Zhang E, 'What Is Log Analysis? Use Cases, Best Practices, and More' (Digital Guardian, 16 October 2017). Available at: https://digitalguardian.com/blog/what-log-analysis-use-cases-best-practices-and-more. Last accessed 11/04/2020.