International Conference on Information and Communication Technology and Its Applications (ICTA 2016)

Federal University of Technology, Minna, Nigeria

November 28 – 30, 2016

Data Loss Prevention and Challenges Faced in their Deployments

Victor O. Waziri, Ismaila Idris, John K. Alhassan, and Bolaji O. Adedayo
Department of Cyber Security Science, Federal University of Technology, Minna, Nigeria
victor.waziri@futminna.edu.ng

***Abstract*—The technology world has greatly evolved over the past three decades and it is at a pace where an average user's laptop can accommodate up to a terabyte of data, where a tiny SD card can store an entire database of an organization, where file transferring has become less complex, and where users can easily connect to any wireless network (Private or Public) within the range of their wireless devices to exchange sensitive information. This evolvement has led to one of the greatest challenges organizations are faced with, which is in the area of adequately protecting their sensitive information from being lost or leaked. Data Loss Prevention (DLP) techniques was created in preventing these breaches on data loss, when these breaches occur in an organization. DLP systems has gained popularity over the last decade and is now referred as a matured technology, and with the alarming rate at which digitally stored assets is growing, the need for DLP systems has also increased. This paper discusses some of DLP concepts and trends, as well as the some of the challenges these various DLPs face and proffer a solution for a successful implementation.**

Keywords-Data loss prevention; Data loss; Data protection; Data security

I. INTRODUCTION

Data loss can be defined as the unauthorized transfer of sensitive or confidential information about an organization from a workstation or from the organization data center to the outside world or to an untrusted environment. This can be achieved through various channels of communications or by using storage devices or simply by memorizing the information displayed on the screen [1].These information could be either a regular data (Debit card data, Bank verification number and Health care data) or organization secrets (Financial information, intellectual property and trade secrets) [2].

Over the last decades there has been major data loss with serious impacts on organizations and this losses is on an increase in recent years. According to DataLossDB [3], their report shows that2015 surpassed the year 2012 all-time record, for the number of reported data loss incidents worldwide. Over 736 million records were exposed in the 3,930 reported incidents in 2015. An estimate of 50% out of those records experienced data loss in the business sector, 20% in the government sector and the remaining 30% occurring in the education and health sectors. It is also important to mention that private users are also victims of data loss and it is hard to know the extent or amount of data loss that has occurred.

There have been some notable data loss incidents in recent years that has cost organizations millions of dollars in the process. An estimated forecast indicates that an average cost of a data loss will be over $150 million by 2020 and a global annual cost forecast to be $2.1 trillion [4].In March 2016, LulsZec Philippines uploaded COMELEC's entire database on Facebook, after their website had been hacked[5], while in October 2015, TalkTalk a British telecommunications provider suffered a data loss of over 4 million of their customer's details, thereby causing their stock to fall drastically[6]. In February 2015, over 80 million records were lost due to data loss in Anthem, these records included social security numbers and very sensitive information[7]. Adobe Systems revealed in October 2013 that there was a data loss of over 130 million user records to a hack group due to insider assistance.These kinds of incidents has caused organizations major financial losses, damages to their reputation, loss of their customer confidence, legal prosecution, productivity and morale of employee and loss of business opportunities[8].

One of the biggest challenges in mitigating data loss, is that there are so many reasons attributed to data loss in an organization and there is no tool or a simple solution that adequately address these various data losses. However to be able to address the risks faced, a solution must be developed to incorporate the causes of data loss, which are can be classified as people, processes and technology[9].

- *People*: Data loss can be caused by people through their lack of awareness of the security issues relating to sensitive information that are to be securedand most times are not been accountable for protecting these information.
- *Process*: The process of securing these sensitive information can be caused by inadequate data usage policies, no proper data transmission process and lack of data monitoring usage.
- *Technology*: Lack of flexibility and communication platform in technology deployed for the protection of data, makes it difficult for the user, thereby making the user to look for an alternative.

As data loss is one of the major problems been faced by organization and if not properly managed can cost the organization millions in terms of finance. This problem can be mitigated by using various types of Data Loss Prevention

(DLP) methods and techniques. DLP can be defined as a system, which is designed to detect and prevent any potential data breach both intentionally or unintentionally[10].Most organizations combine two or more DLPs to effectively control the potential data loss they might be faced with. DLPsystems differs from the conventional security as it has the ability to analyze the content of the confidential data and the context surrounding those data and it also has the ability to protect those confidential data in all data states.

A basic DLP system consist of three stages which include discover, monitor and protect[11]. This stages are vital in setting up an effective DPL system. The discovery stage locates where your confidential data are been stored, by takin a detailed inventory of this classified data and then regrouping these sensitive data in terms of priorities. In the monitoring stageit monitors how the confidential data are used, by understanding the content and context of this sensitive data and by analyzing when a breach occurs. The last stage which is the protect stage, basically describes the ways for protecting data loss and this is done by been proactive in protecting these confidential data or by enforcing the data loss policies created.

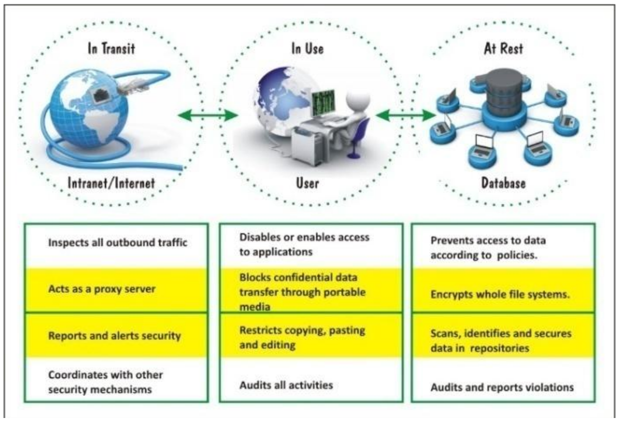
In Transit
In Use
At Rest
Intranet/Internet
User
Database
Inspects all outbound traffic
Acts as a proxy server
Reports and alerts security
Coordinates with other security mechanisms
Disables or enables access to applications
Blocks confidential data transfer through portable media
Restricts copying, pasting and editing
Audits all activities
Prevents access to data according to policies.
Encrypts whole file systems.
Scans, identifies and secures data in repositories
Audits and reports violations

Figure 1. The different Data States and DLPs functionalities

For a DLP system to be effectively deployed in an organization, the data life cycle of the organization is considered. A data life cycle is a detailed outline of the phases involved in effectively preserving and managing of data to be used and reused. This stages include data at rest (data in storage), data in use (data flowing through internal network) and data in transit (data that are been accessed). Figure 1 shows a summary of these phases and how the DLPs prevents data loss at those phases. In the protection of targeted data, DLPs can take many forms during its deployment, which are mostly based on the data state [12].

II. COMPARISON OF DLP TOOLS

In evaluating some of these tools designed by various vendors, we were able to make a comparison for some of the top DLP tools from various security vendors such as CA Technologies (A), Code Green Networks (B), Digital Guardian (C), Forcepoint (D), McAfee (E), Palisade Systems (F), RSA (G), Trend Micro (H), Trustwave (I) and Symantec (J), it was observed that these security vendors offer protection for various data states. Though some of these tools are specialized in their design (don't perform other security task), they can be integrated to support other technologies like identity access management or encryption. The table 1 summarizes the features for each of this DLP vendors and figure 2 also shows the performance based on these features.

TABLE I. DATA LOSS PREVENTION PRODUCT MATRIX

Vendor	A	B	C	D	E	F	G	H	I	J
Mobile/tablet	X	X	✓	✓	X	X	✓	✓	X	✓
Laptop/Desktop/Workstation	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓
Local network	X	✓	X	X	✓	X	X	X	X	X
Server	✓	✓	✓	X	✓	✓	X	✓	X	X
Cloud/SaaS	✓	✓	✓	✓	✓	X	✓	✓	X	✓
Detection technologies										
biometric signatures	X	X	X	X	✓	X	X	X	X	X
classification	✓	✓	✓	X	✓	X	X	X	X	X
context analysis	✓	X	✓	✓	✓	X	✓	X	X	X
data matching	X	✓	X	X	X	X	X	X	✓	✓
flagging	X	✓	X	X	X	X	X	X	X	X
dictionaries/lexicons	X	✓	X	X	✓	X	X	X	X	X
data discovery	✓	✓	✓	X	X	✓	✓	✓	✓	✓
file type detection/classification	X	X	✓	✓	✓	X	X	✓	X	✓
machine learning/pattern recognition	X	✓	X	X	X	X	✓	✓	X	✓
Optical Character Recognition	X	X	X	✓	✓	X	X	X	X	X
regular expressions/pattern matching	X	✓	X	X	✓	✓	X	✓	X	✓
Enforcement technologies										
block	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓
encrypt	✓	✓	✓	X	✓	X	X	✓	✓	✓
fingerprinting	✓	✓	✓	X	✓	X	X	X	X	✓
move/remove	✓	✓	✓	X	X	X	X	X	X	X
notify/alert	✓	✓	✓	X	X	✓	X	✓	✓	X
quarantine	✓	✓	X	X	✓	X	✓	✓	✓	✓
Software integration										
Databases (e.g. SQL Server)	X	✓	X	X	✓	X	✓	X	X	✓
Email client	X	✓	✓	✓	✓	✓	✓	✓	✓	✓
file sharing	X	✓	X	X	✓	✓	✓	X	✓	✓
instant messaging	X	✓	X	X	✓	✓	X	✓	✓	✓
web 2.0	X	✓	X	✓	✓	✓	X	X	✓	✓
webmail	X	✓	X	X	X	✓	✓	✓	X	X
Hardware integration										
CD/DVD	X	X	X	X	✓	✓	✓	✓	X	✓
external/removable HD	X	✓	✓	X	✓	X	X	✓	X	X
Printer	X	X	X	X	✓	X	✓	✓	X	X
USB drives	X	✓	X	✓	✓	✓	✓	✓	X	✓

wireless devices	X	✓	X	X	X	X	X	X	X	X
Monitoring										
centralized	X	✓	✓	✓	✓	✓	✓	✓	X	✓
offline	X	✓	X	✓	✓	X	X	X	X	X
real-time	X	X	✓	✓	✓	✓	✓	✓	✓	✓

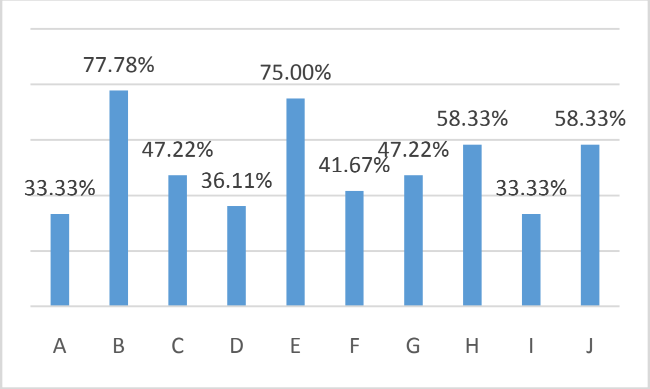

Figure 2. Overall performance of the DLP vendors

Though this doesn't necessary imply that Code Green Networks has the best DLP system, it simply means it compensates in the areas where it lacks. When choosing the overall DLP vendor for your organization, it is ideal to check the features that best suites the implantation, such as the ease of installation, scalability, control features as well as its maintenance.

III. CHALLENGES ON DLP SYSTEMS DEPLOYMENT

In protecting sensitive data from loss, DLP systems faces many challenges and like other security mechanism these challenges can render the system ineffective. In a review conducted by researchers in the area of both industrial and academic DPL systems, it was discovered that there were seven common challenges been identified[13]. These are Leaking Channels, The Human Factor, Access Rights, Encryption and Steganography, Data Modification, Scalability and Integration, and Data Classification. For an effective DLP system to be implemented, these various challenges must be addressed. In the following sections we will discuss those challenges faced and try to suggest possible solution for each of them.

A. *Leaking Channels*

Everyday there is need to share and access data between different medium and users, and this done with the assistance of intermediate channels. In an ideal scenario these channels are used to legitimately exchange data from one end to another, however these channels can also create a major treat in the leakage of sensitive data. These channels cannot be totally blocked, as it is an important aspect in the sharing of data, which requires some or even all these channels to be open. As technology keeps growing at a fast pace and more channels becoming available, it has become hard to keep pace with securing these channels[14]. The figure 3 shows some of the commonly used channels used for data exchange, as not all these channels are very easy to secure. Some of them will require a great number of techniques and monitoring to adequately secure them.

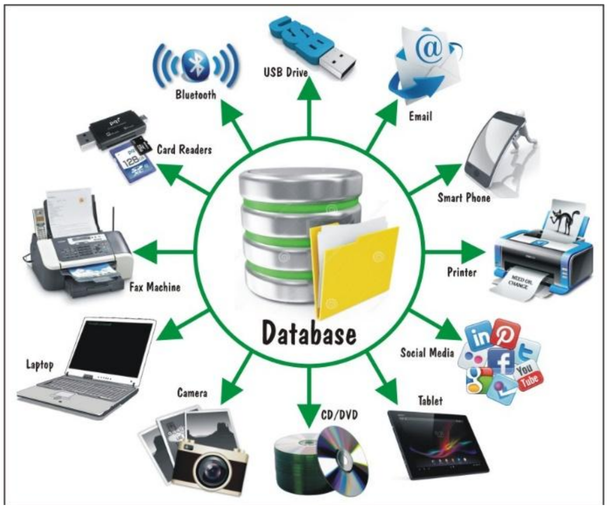

Figure 3. Some common data leaking channels

Sensitive data that are either '*at rest*' or '*in use*' can be compromised using these channels, which could include USB ports, CD/DVD drives, printed documents and even through web services. Though data leakages can be mitigated using host DLPs for CD/DVD drives and USB port channels, it isn't adequate enough to prevent data leakages from other channels such as Instant Messaging (IM) and emails that are always made available [15]. Even when the access right are restrict to confidential data, some of these data can still be still be accessed in a printable format. While channels like file sharing and web services associated with data '*in transit*' has been one of the biggest challenges in migrating, as these channels cannot be blocked and the serve as the backbones of the organization in terms of data exchange. To effectively maintain a maximum security in these channels, an intensive filtering traffic is to be done. The DLP system to be deployed should always try to create a balance in security without affecting the interconnectivity in these channels.

B. *The Human Factor*

Humans are generally a complex being, as their behaviors and motives are usually hard to predict or to determine, as they are been influenced by many factors, which could be psychological or sociological. Decision makings such as granting of access to a set of users, defining the confidentiality level of a data and setting a threshold level for DLP systems, is basically affected by human actions. It must also be noted here that, even when organization's security policies are in place to mitigate such data loss, it doesn't mean it is guarantee to tackle the problem. Almost all human interactions with data occurs at the data '*in use*' state, which simply means the user needs an endpoint terminal to access this confidential datafor there to be data leakages [16]. Though a typical DLP system will tend to put some restrictions on data leakages by the user, in the form of disabling some aspect of the system, such as CD/DVD drives, USB ports and removable drives. But this user restrictions can easily be bypassed, by the sharing of access rights by users either intentionally (by trust) or unintentionally (by social engineering), there by compromising the security of the confidentiality of that data that would be accessed. Users can make use of mobile gadgets to snap pictures of sensitive information or even use hidden cameras to record the entire classified documents and transmit it remotely. The human factor will always be a major challenge when deploying a DLP systems, as long as in is human interactions with the system[17].

C. *Access Rights*

Access right has always been a key feature in the deployment of any security mechanism including DLP systems. Therefore it is of great importance to be able to categorize these access rights properly and to be able to separate each category of users from each other based on their level of permission. DLP systems won’t be able to prevent illegitimate users from accessing confidential information, if there is not a proper categorization of access right in place. Access rights is of great importance in preventing data loss in an organization and should always be updated regularly[18]. As an obsolete access right can have a huge impact on the entire system negatively, thereby making the system vulnerable to data loss. For instance, when a user is downgraded or dismissed from the organization and the access rights are not updated according, it leaves the system vulnerable to data leakages, as the DLP system won’t be able to detect any data breach when that user tries to access information he/she is not permitted to access.

A data leak can also occur by a legitimate user with the right access rights either intentionally (due to so many factors such as financial gain or whistle blower) or accidentally. For efficiency, a DLP system should be able maintain and control access rights of the organization while also performing the function of protection of data from intentional and unintentional leakages.

D. *Encryption and Steganography*

Encryption is another major challenge been faced by network based DLP systems, as these systems uses different forms of analytical techniques in identifying copies of the sensitive data and comparing it with the original data that are been classified as confidential. But with complex encryption of the confidential data by the user it makes it hard for the DLP system to be able to analyze such data content, thereby creating a major vulnerability in the system. The implication of this, is that a confidential document can bypass the DLP system detection mechanism when the user encrypts the documents, thereby allowing the user to be able to send the confidential document through his/her email as an attachment[19]. Stenography is another type of challenge similar to encryption, but it is more challenging to mitigate and even impossible to detect when used by the user. The user uses stenography tools to hide classified documents within other media, these media could be digital photos, audio files and video files. This becomes a challenge to DLP system as it won’t be able to detect the confidential data inside those media[20]. In some instances a document can be compressed, or converted to a different format, thereby making the system unable to detect such documents, as it won’t be able to analyze such documents.

E. *Data Modification*

The design of some DLP systems are created to compare the original sensitive data and inspected traffic flowing through the system by using data signatures and patterns to achieve prevention of data leakages. In this system, detection occurs whenever there is a signature and patterns match to that of the confidential data or when there is a high percentage of similarity to the confidential data. The major challenge of this system design is that confidential data are mostly not sent in a form that will enable the system detect such modification. Data can be modified using various type of techniques, which are readily available online. These confidential documents can be easily modified by removing some vital lines in the documents or adding to it, thereby creating an entire different document before sending such documents over the allowed channels. The user can also entirely change the structure or format of the document, thereby rendering the documents undetectable from the DLP systems [21].

In some other design DLP systems uses data hashing in analyzing the outgoing traffic by comparing the values (including SHA1 and MD5) with the original confidential data. The moment these two values matches each other, then detection of data leak occurs. The problem with hashing design, is it becomes ineffective the moment the confidential documents is extensively modified, which in turns gives a different hash value [22].

F. *Scalability and Integration*

The volume data processed can affect the performance of any security mechanism deployed in securing an organization’s assets. DLP systems can also be a victim of such challenges, which means when deploying them either in a host, network or storage section, it should be effective in performing its function and smoothly incorporated into the system without affecting or causing delay in the entire work flow of the organization’s system. Therefore factors affecting the scalability of a DLP systems such as its computational ability and analyzation techniques should be considered when deploying the system [23].

There is usually some challenges faced when integrating the DLP systems during its deployment, as there are similar function already been handled by other security mechanism like firewalls and intrusion detection. Therefore before deployment, the entire system must be carefully analyzed and implemented to give an effective performance. As there shouldn’t be repetition of functions, as having two similar function on the system can cause a delay in the entire process of the system, thereby reducing the performance of the system.

G. *Data Classification*

Data classification is the process of organizing data into categories or levels for an effective and efficient use [24]. This definition implies that, DLP systems rely entirely on well-defined data classification to enable the system differentiate confidential data from normal data. The main purpose of classification of data is in determining the baseline of security controls to be used in safe guarding data. There are different ways by which data can be classified based on the organization classification, with terms like *confidential*, *secret* and *top secret* been used by the military to classify their data, while for an Institution data can be classified as *restricted data*, *private data* and *public data*[25]. With this classification it becomes easy in identifying those confidential data, thereby making the system adequately equipped to protect those confidential data. However the problem with data classification is determining the level of secrecy of those sensitive data. For there to be a proper classification of these secrecy levels, the owner of the data for protection should be the one responsible for this classification process. However, most times the classification process is left to those people who don't have enough knowledge about all the data. This creates a vulnerability in the DLP system, as those who are not permitted to see certain information are now equipped to having access to such confidential information. It is therefore important to properly have a good classification, as without it the DLP system becomes ineffective.

TABLE II. COMPARISON OF DLP SYSTEM WITH IPS/FIREWALL SYSTEMS

Security Gaps	**IPS/Firewall Systems**	**DLP Systems**
Network partitioning of network security	No	Yes
Load Balancer Integration	No	Limited
Accelerated program delivery	No	Yes
TCP connection pooling	No	Yes
SSL offloading	No	Yes
Built in authentication engine	No	Yes
Validate encrypted sessions	No	Yes (MTA Sensor)
Multiple applications single sign on	No	Yes
Injection attack protection (XSS, SQL)	Limited	Yes
Normalize encoded traffic	No	Yes
Inspect HTTPS traffic	No	Yes (vary from different policy)
Session tampering/ hijacking/ riding protection	No	Yes
Forceful browsing prevention	No	Yes
Data theft protection, cloaking	No	Yes
Brute-force protection	No	Yes
Trojan/Warms/Virus/malware upload protection	Yes (Back Door Detection)	Yes (Block and report users‟act)
Rate control protection	No	Yes
Request, response rewrite	No	Yes
Application access logging and user audit trails	No	Yes (depends on policy rule)

IV. COMPARISON WITH OTHER SECURITY TECHNIQUES

In the aspect of security, there are a lot of security systems and security vendors in the market. These security systems can be classified or grouped as network security systems, antivirus systems, monitoring systems, scanning systems, data controlling systems as well as transaction systems. These systems are unique and separated by their functionality. Take for an example, the antivirus system cannot perform encryption of data but it works perfectly in the monitoring of the data source code. For this reasons corporate organization will require many types of security systems in their protection of data. As these security system have specific functions, this makes the DLP system having more edge over the rest, as it has the ability to perform various functions. This reduces the cost of purchase a lot of security systems in the monitoring of the various security gaps. The table 2 summarizes the features of a DLP when compared to similar protection system (Intrusion Prevention System (IPS) and Firewall System).

V. THE WAYS DLP SYSTEMS ANALYZES DATA

Though there are different ways by which DLP systems analyze their data, these analysis can be grouped into two major group. They are context analysis and content analysis. The context focuses on the surroundings of the data while the content focuses on the actual data[26].

- *Context analysis*: This method of analysis actual analyzes the metadata properties with the confidential data. It does this by examining the information about data and keeps track of the data using various attributes of the data such as the size of the document, the source, the destination, when the document was created or modified and other properties. With this metadata attributes of the confidential data, a pattern and signature can be used to form a process in defining how the policies can be created for the detection of data loss [27].
- *Content analysis*: In this method, analysis focuses on the content of the confidential data, which could be text or any multimedia material. It does this by comparing the transmitted data with the original confidential data and detects a breach if there is a high percentage in similarity [28]. This process can be done through basically three techniques: data fingerprinting (identifies patterns with exact or partial match), regular expression (identifies its patterns based on words or text) and statistical analysis (using prerecorded information) [29].

DLP systems could be either preventive or detective, depending on the type methods been used by the organization. The preventive methods includes: Policy and Access Rights, Virtualization and Isolation, Cryptographic Approaches, Quantifying and Limiting; while detective methods includes: Data Identification, Social and Behavioral, Data Mining/Text Clustering, Quantifying and Limiting.

A. *Policy and Access Rights*

This type of method is widely suitable for organizations, as long as there is a proper classification of their data and a well-defined access rights system in place. This becomes easy to manage as the procedures are clearly stated and makes it ideal for data 'at rest' and data 'in use'. This method is constrained by basically improper classification of data and not using the effective access controls. As it is a preventive method, it doesn't have the capability to detect when a breach has occurred [18].

B. Virtualization and Isolation

It is based isolating the activities of the user virtually and only allowing the system process trusted function or data to pass through the system. This method usually requires hardware in its implementation, thereby reducing the amount of administrative functions as it makes use of the existing data classification on the system. However it isn't cost effective and doesn't detect when there is a data leakage[30].

C. Cryptographic Approaches

This approach involves encrypting the confidential information with strong encryption tools to enable it produce a maximum level of security. This approach is almost used in all DLP systems as it has various options to encrypt such files and it is effective for data '*at rest'*. The major challenge is that encryption doesn't hide those confidential documents even though they might be encrypted. It isn't a detective method, making it vulnerable when there is a data leakage[31].

D. Quantifying and Limiting

This method has an added advantage, as it also monitors the channels in which those data travels and blocks any sensitive data from passing through those channels. It can effectively be implemented for data '*in transit'*, '*in use'* and '*at rest'*, thereby making it easy to deploy it for a specific attack on the organization system. As with the other preventive methods, it makes it hard to detect data leakages and if not properly deployed can disrupt the workflow of the entire system. It is also limited to specific scenarios of data leakages thereby making it vulnerable to other data forms of leakages[32].

E. Social and Behavior Analysis

This method involves analyzing the level of interaction between people or in this case users of the organization and measuring this level, by creating adequate guidelines for the protection of sensitive data. When adequately implemented prevents leakages by detecting any relationship that is of malicious intent and it is effective in all data states. As it is difficult to predict such human behaviors, thereby leading to a high percentage of false positives and also requiring the administrator to regularly interact with the DLP system. This method also requires a huge amount of time in profiling the various users and indexing each of their behavioral patterns [33].

F. Data Identification

This methods uses a mechanism that compares data traffic flowing through the system with that of the original confidential documents and tries to prevent such data from been leaked when there is match. This method produces a very low false positive, when using fingerprinting in its analysis. However this method can easily be bypassed by extremely modifying those data, making it impossible to detect it [34].

G. Data Mining and Text Clustering

This method involves the ability to be able to predict when a data leakage will occur by learning about the data process and data leakages patterns over time. It is effective in detecting unstructured documents, making it less dependable on administrative interfacing, which makes the method easy to integrate. The method is faced with a very high false positive as it requires a learning phase to work, thereby requiring a huge amount of processing power [35].

VI. SOLUTION FOR A SUCCESSFUL DLP IMPLEMENTATION

For there to be a proper implementation of any DLP systems, there are ten key steps we have considered and if this steps are followed would help an organization to adequately implement the DLP systems for protection of their confidential data. These steps are as follows:

Step 1: Implementation of a universal technique and value proposal for DLP centered on a risk assessment

Step 2: Involve the right people with the right organization model

Step 3: Identify sensitive data and understand how they are handled

Step 4: Provide a phased implementation based on progress

Step 5: Minimize the impact to system performance and business operations

Step 6: Create meaningful DLP policies and policy management processes

Step 7: Implement effective event review and investigation mechanisms

Step 8: Provide analysis and meaningful reporting

Step 9: Implement security and compliance measures

Step 10: Implement an organizational data flow and oversight mechanism

VII. CONCLUSION

Many of organizations have given a great deal of attention in protecting their sensitive data from been lost accidentally or intentionally. DLP systems cannot function effectively in isolation, this implies that for a DLP system to effectively function it requires linking other security information process. However, before implementing any DLP system, there is need to adequately understand what confidential data the organization wants to hold, where does confidential data are to be stored in terms of locations as where those data are been stored are vital in its protection and the destination and the channels this information will pass through.

There are several challenges associated with DLP systems, before they are deployed it is necessary and as well as important to adequately have a deep understanding and be able to analyze these various challenges associated with the system. It is also important to make the system easy to be used and managed, so as to avoid any form of complexity, as the more complex a DLP system, the more likelihood the system will be compromised by the user.

As new technology are been developed and the ways this technologies communicates changes as well, it is of great importance an organizations must keep pace with these increasing technology advancements by identifying new and better ways in protecting data from been lost by unauthorized users.

REFERENCES

[1] N. Kumaresan, "Key consideration in protecting sesitive data leakage using Data Loss Prevention Tools," *ISACA Journal,* vol. 1, pp. 1-5, 2014.

[2] E. Bergstrom and R. M. Ahlfedt, "Information Classification Issues," *Sprin International Publishing,* pp. 27-41, 2014.

[3] DataLossDB. (2016). *2015 Reported data breaches surpasses all previous years*. Available: http://blog.datalossdb.org

[4] IBM and Ponemon Institute LLC, "2015 Cost of Data Breach Study: Global Analysis," Ponemon Institute LLC Research Department 2308 US 31 North Traverse City, Michigan 49686 USA 2015.

[5] Trend Micro. (2016). *Data Protection Mishap leavees 55M Philippine Voters at Risk*. Available: http://blog.trendmicro.com/treandlabs-security-intelligence/55m-registered-voters-risk-philippine-commission-elections-hacked

[6] BBC NEWS. (2015). *TalkTalk hack 'affected 157,000 customers*. Available: http://www.bbc.com/news/business-34743185

[7] C. Osborne. (2015). *Health insurer Anthem hit by hackers, up to 80 million records exposed*. Available: http://www.zdnet.com/article/health-insurer-anthem-hit-by-hackers-up-to-80-million-records-exposed

[8] T. Seals. (2016). *Data Breach Trends to Evolve in 2016*. Available: http://infosecurity-magazine.com/news/data]breache-trends-to-evolve-in

[9] EYGM Limited. (2011). *Data Loss Prevention: Keeping your sensitive data out of the public domain*. Available: http://www.ey.com

[10] R. R. Tahboub and Y. Saleh, "Data Leakage/Loss Prevention Systems (DLP)," *ResearchGate,* 2014.

[11] Jonathan Jesse and ITS Partners. (2015). *Symantec DLP Overview*. Available: http://www.symantec.com/en/uk/business/theme.jsp?th

[12] Price Waterhouse Coopers, "Data Loss Prevention: Keeping sensitive data out of the wrong hands*," pp. 1-16, 2008.

[13] N. Lord, "Experts on the Data Loss Prevention (DLP) Market in 2016 & Beyond," ed, 2016.

[14] V. Shaj and K. P. Kaliyamurthie, "A review of Data Leakage Detection," *IJCSMC Journal,* vol. 2, pp. 577-581, 2013.

[15] T. T. T. Huong and J. Corner, "The impact of communication channels on mobile banking adoption," *International Journal of Banking Marketing,* vol. 34, pp. 78-109, 2014.

[16] I. Ponemon, "The Human Factor in Data Protection " *Trend Micro,* pp. 1-27, 2012.

[17] T. Pepper. (2016). *The people problem: How to manage the human factor to shore up security*. Available: http://www.scmagazineuk.com/the-people-problem-how-to-manage-the-human-factor-to-shore-up-security/article/494638

[18] D. Gibson, "What's missing from Data Loss Prevention," *Data Center Journal,* 2012.

[19] S. R. Raj, A. Cherian, and A. Abraham, "A Survey on Data Loss Prevention Techniques," *International Journal of Science and Research,* vol. 2, pp. 240-241, 2013.

[20] N. B. Pamula, M. S. Naga, and P. K. Deepthi, "Preventing Data Leakage in Distributive Strategies by Steganography Technique," *International Journal of Computer Science and Information Technologies,* vol. 4, pp. 220-223, 2013.

[21] S. W. Ahmad and G. R. Bamnote, "Data Leakage Detection and Data Prevention using Algorithm," *International Journal of Computer Science and Application,* vol. 6, pp. 394-399, 2013.

[22] M. Hart, P. Manadhata, and R. Johnson, "Text Classification for Data Loss Prevention," in *Privacy Enhancing Technologies*, ed Waterloo, ON, Canada: Springer Berlin Heidelberg, 2011, pp. 18-37.

[23] J. Thorkelson. (2010). *Data Loss Prevention: Simplified*. Available: http://www.codegreennetworks.com

[24] M. Rouse. (2015). *Data Classification*. Available: http://searchdatamanagement.techtarget.com/data-classification

[25] R. Bragg, "Data Classification," in *CISSP Training Guide,* 1st ed 800 East 96th Street, Indianapolis, Idiana: Pearson IT Certification, 2002, pp. 48-51.

[26] A. Bryman, *Social Research Methods*, 2nd ed. Great Clarendon Street, Oxford, United Kingdom: Oxford University Press, 2004.

[27] S. A. Kale and S. V. Kulkari, "Data Leakage Detection," *International Journal of Advanced Research in Computer and Communication Engineering,* vol. 1, pp. 668-678, 2012.

[28] K. A. Neuendorf, *The Content Analysis Guidebook*. Thousand Oaks, Ca.: Sage Publication Inc., 2002.

[29] K. Krippendorf, *Content Analysis: An introduction to its methodology*. Thousand Oaks, Ca.: Sage Publication Inc., 2004.

[30] J. N. Mathews, W. Hu, M. Hapuarachchi, and T. Deshane, "Quantifying the performanceof IsolationProperties of Virualization Systems," *ACM,* pp. 1-9, 2007.

[31] K. Scarfone. (2013). *How to help DLP and Encryption Coexist*. Available: http://www.statetechmagazine.com/article/2013/11/how-help-dlp-and-encryption-coexist-state

[32] S. Vavilis, M. Petkovic, and N. Zannone, "Data Leakage Quantification," presented at the Data Applications Security and Privacy XXVIII: 28th Annual IFIP WG 11.3 Vienna, Austria, 2014.

[33] J. M. Kizza, *Computer Network Security and Cyber Ethic,* 4th ed. Jefferson, North Carolina: McFarland & Company, Inc., 2014.

[34] M. Tu, K. Spoa-Harty, and L. Xiao, "Data Loss Prevention Management and Control: Inside Activity Incident Monitoring,Identification, and Tracking in Healthcare Enterprise Environments," *The Journal of Digital Forensics, Security and Law,* vol. 10, pp. 27-44, 2015.

[35] I. H. Witten and E. Frank, "Classification rule," in *Data Mining Practical Machine Learning Tools and Techniques*, 2nd ed 500 Sansome Street, Suite 400, San Francisco, CA 94111: Morgan Kaufmann Publisher, 2005, pp. 200-213.