# Methodology for Data Loss Prevention Technology Evaluation for Protecting Sensitive Information

**López G.*; Richardson N.**; Carvajal J.***

**Escuela Politécnica Nacional, Facultad de Ingeniería Eléctrica y Electrónica, Quito, Ecuador*
*e-mail: {gabriel.lopez, jorge.carvajal}@epn.edu.ec*
*** Sheffield Hallam University, Faculty of Arts, Computing, Engineering and Sciences , Sheffield, UK*
*e-mail: n.richardson@shu.uc.uk*

**Resumen:** El presente trabajo de investigación propone una metodología para la evaluación de un sistema que previene la fuga de información sensible para una organización. La metodología posee un ámbito para abarcar tanto software libre como comercial en instituciones públicas o privadas. La propuesta de metodología está basada en buenas prácticas de la ISO 27001, investigaciones relacionadas, y características de tecnología líder en prevención de fuga de información. El propósito principal de artículo académico es cubrir el faltante de una investigación que proponga una evaluación integral de la tecnología de DLP utilizando el método científico, y además relacionarlo con las leyes ecuatorianas referentes a la privacidad y la nueva matriz productiva del buen vivir de la República del Ecuador.

**Palabras clave**: Prevención de fuga de información, seguridad de TI, ISO 27001, DLP, información sensible.

**Abstract:** This investigation proposes a methodology for the evaluation of Data Loss Prevention Systems in order to secure sensitive information for the organizations. The methodology will be able to cover open source and commercial software in public or private institutions. The methodology proposal is based on the recommendations of the ISO 27001, related investigations, and characteristics of leading technology in data loss prevention (DLP). The main contribution of the academic paper is to cover the flaw of the state of the art in DLP technology evaluation, since no other investigation related specifically to this topic has used the scientific method. Also, the criteria used to develop the proposed methodology in this paper is based on the Ecuadorian laws related to information privacy and the new productive matrix of the Republic of Ecuador.

**Keywords**: Data loss prevention, IT security, ISO 27001, DLP, sensitive information.

## 1. INTRODUCTION

To begin, data leaks is a risk that has been increasing during the last years [1] and the industry should take into account that this represents the loss of one of the most important assets for a company [2]. The information is an asset that the employees may have generated, have access or could be in charge of its storage, so the question is if the companies should trust that the employees will follow the data security policy or should the companies use an enforcement tool to help to make sure that the users comply with the security policies. The truth in the real world is that the employees make mistakes and there is difficult to assure that a company will have 100% honest employees. Some examples of data loss are WikiLeaks founded by Julian Assange which exposed top secret information from governments and military organizations causing a big impact for political regimes and private companies [3]. Also, the case of Edward Snowden, who has been responsible for the most important leak in the NSA's history [4]. Consequently, the authors considered that Data Leaks is a visible problem and investigating Data Loss Prevention (DLP) Technology will help the industry to increase the compliance with information security policies.

In addition, the authors looked for previous studies about open source and commercial DLP technological solutions, and it was found that there is a few academic work in DLP evaluation [5, 6, 7, 8], so it will be a really useful investigation since it is going to present the state of the art of DLP technology evaluation to the academia. It is important to mention that the current academic studies about DLP focus in the latest techniques to perform data loss prevention and it is not related to DLP technology evaluation, so the main contribution of this paper is to present an academic piece of work to propose a methodology of how to evaluate DLP technology based on the criteria of sensitive information related to the new productive matrix of the Republic of Ecuador and the Ecuadorian Law, common and evasion data file extensions, DLP characteristics like performance, capabilities and user experience, policy violation related to sensitive data according to the Ecuadorian Law. It was found work related to DLP evaluation technology, but it did not follow a scientific methodology nor presented details of the evaluation, so it had a high probability of a bias investigation. The authors recognized this flaw in the state of

the art of DLP technology evaluation and realized the necessity of an academic work in this area.

The academic investigation is divided in the following sections: first presents the literature review of DLP technology, then it is developed the methodology proposed for DLP technology evaluation, and finally the conclusion and future work.

## 2. LITERATURE REVIEW

### *2.1 DLP Definition*

DLP is the current technological solution to protect an organization from data leaks. DLP technology enforces the criteria of how the information will flow in and out of the company's electronic network including audit trials, notifications, and response actions [8]. DLP which was first presented to the market in 2006, is a solution that is able to inspect the content of the electronic data of the organization specializing in looking for sensitive or valuable information, which is moving without permission through the company. The DLP software after capturing the data uses file cracking technology for content analysis, which is used to read and understand the information that is inside the file [8]. When sensitive information is identified, DLP triggers alerts and actions in order to prevent this issue. The critical information should have been previously analyzed and identified by the company in a risk analysis, where it were applied controls to the risks with unacceptable level [2]. The unacceptable risks related to data loss will be reduced to an acceptable level by the DLP deployment. In addition, DLP technology supports the protection of sensitive data not only from intruders, it also protects from internal personnel, who may be braking a process. For instance, if an employee stores sensitive information in his computer and it is not being encrypted, a DLP will be able to recognize this non conformity.

Even though, the companies have awareness programs to socialize the security policy, guidelines, and procedures, the user's actions cannot be controlled. It is a fact that user's actions are the major cause of malware infection in companies [9]. This is why it is proposed the investigation of DLP technologies in order to understand its support to enforce the security rules of the company. DLP does not only detects or identifies problems, it prevents outbreaks in order to decrease the data loss of the company.

### *2.2 Underlying Mechanisms of DLP*

After the DLP has access to the content of the information, it generally uses regular expressions, file fingerprinting and dictionaries, or a mix employing a context understanding to detect policy violations related to sensitive information [9]; [8].

Regular expressions define patterns of possible inputs strings [10], which will be related to sensitive information. For example if it is needed to look for the word 'secret' or 'Secret', the regular expression will be '[S/s]ecret'. This will identify files or traffic with the word secret in its content/metadata or payload respectively. Regular expressions are powerful, but they could get really complex as well. IDS technology, such as SNORT [11] uses regular expressions for detection too [12]. Even though regular expressions are a helpful mechanism to detect critical information, they are only useful when the sensitive information target is structured data, i.e. it has a defined format, like a national insurance number, passport number, or credit card numbers, but when the sensitive information does not have a regular pattern or it is unstructured, the only option is to fingerprint the data since it is an undefined format [8].

The fingerprint is a secure hash that is saved in the database of the DLP system. The hash will be compared with the files hash in order to verify the possible existence of classified data. The DLP combines its mechanism of detection with the context of data location in order to check if the data is in a place that should not be. For instance, if personal data of students is in the secure data base system of the university, it will be not an incident, but if the same data is found in the secretary's computer of the faculty, it should raise a flag and alert about the incident. The fingerprinting technology has the option to perform an exact document matching or a partial document matching [8]. The DLP policies can specified the area where it is going to be applied, i.e. staff network or shared folders.

Also, the DLP solutions used dictionaries to detect sensitive data. Dictionaries are lists of key terms predefined or user defined, which are related to a sensitive category, such as personal data, strategic documentation, confidentiality terms, etc. [13]. In some cases the DLP technology is able to read these terms in real time from a database giving a more accurate source of information [14].

### *2.3 Types of DLP*

DLP technology focuses in three main areas: data in motion, data at rest, and data at end points [9, 15].

To begin, data in motion monitors data loss, which may be going through the network channel. This information may be leaked using protocols, such as FTP, P2P, HTTP, HTTPS, SMTP, SSH, etc. This type of DLP can have two types of architectures. One will be on a SPAN connection, which is a sniffer with the interface in promiscuous mode, so it can receive all the traffic from the network. The other architecture is the DLP in inline mode, which means that all the traffic will be physically passing through the DLP system. The main difference is that the SPAN mode can only detect, but not block data loss. The inline mode since it is in de middle of the traffic, it will be able to stop data leaks (Websense 2014). See Fig. 1 for a diagram related to SPAN mode, and Fig. 2 for inline mode.

Common data at rest is information at databases or data at content management, such as DSPACE [16] or ALFRESCO [17], which is available for staff and clients. DLP data at rest will look for sensitive information that is stored in this repositories or network folders, in order to prevent the access

to sensitive information by unauthorized users [9]. For instance, DLP data at rest is very useful when auditing or

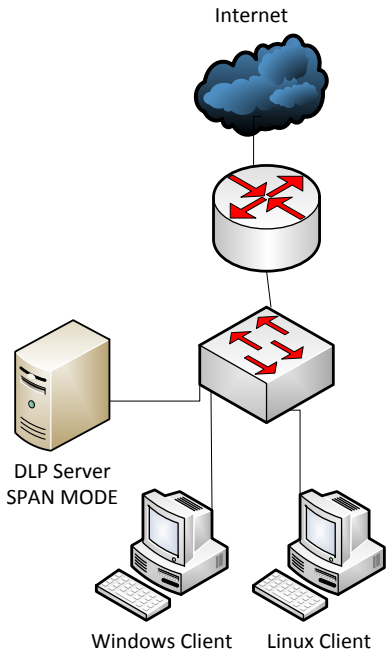


**Figure 1**. SPAN mode

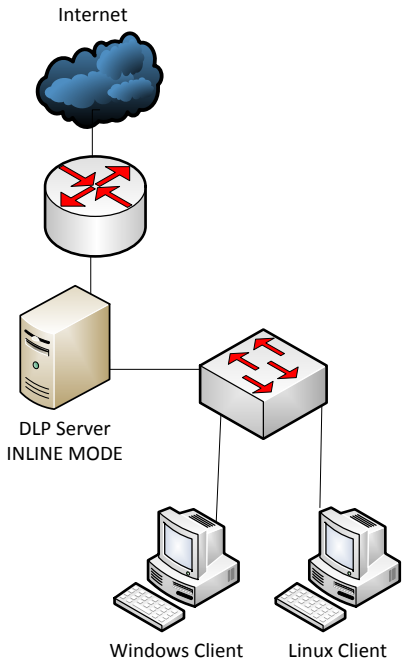


**Figure 2**: INLINE mode

checking for classified information in the DMZ of the network, where servers like public DNS, Email or Web are hosted. Sensitive information should not be stored in this type of services since they do not guaranteed the security of the information, as in the internal servers like a database server, which normally has the protection of firewalls, IDS, and antivirus.

Finally, data at end points is the area, which corresponds to the user's computer or portable devices like laptops or tablets (Bunker and Fraser-King, 2009). The end point is a critical location since here may be the place where information is being generated, visualized, transferred or edited. In general, DLP for end points needs to install an agent in the clients in order to make the user's computer comply with the policies. For instance, some policies will prevent data loss going through channels, such as removable devices, word processing applications, email, printing, or CD burning. Sometimes users need to keep sensitive data in their devices. For instance, a manager needs to work and review a marketing strategy. Since this employee needs to keep sensitive data in his computer, it is recommended to protect this information against robbery or lost. For this situation is helpful the DLP at end points since some technological solutions offer to encrypt the hard disc and the data transferred to removable devices.

### *2.4 Previous investigation of DLP Technologies*

#### *2.4.1 Evaluation of blocking data leaks at the endpoint*

The investigation done by [6] tested tree DLP vendors Identity finder, Websense, and TrendMicro focusing in the evaluation of blocking data leaks at the endpoint. They justified the investigation by presenting the problem, that it is not possible to trust 100% to an employee who has access to sensitive data. This fact is also supported by [9]. The investigators did not justify why those vendors were invited to participate in the evaluation, but the author checked that the majority of them were present in the Magic Quadrant of Content Monitoring and Filtering and Data Loss Prevention according to [18]. The investigators agreed that the purpose of a DLP should be to allow only the activities that the user is supposed to perform. The investigators performed 588 tests to check if endpoint DLP was able to stop sensitive data from leaving its safe environment. In addition, the investigator's methodology was to test the data discovery differences, fingerprinting functionality, actions against a violation, installation and configuration experience, performance, and agent systems resources used.

The Data discovery evaluation focused in checking for network shared folders, and endpoints data. This approach covers the scope for common users, and check a high risk location, which is a shared network folder, but it was not considered the discovery of databases, which may be helpful in order to check if they are holding the information that they are supposed to store. For instance, if information like credit card numbers that according to the internal security policy of a company should be encrypted or protected, but in practice it may not. For this test the investigators discovered that Data Endpoint and LeakProof were able to discover network shared folders and endpoints data with the help of an agent.

Fingerprinting was tested and it worked for Data Endpoint, and LeakProof, but it was evaluated only performing minor changes in the document. It could have been useful to test the effectiveness of the functionality of the DLP by changing a large portion of the document, but maintaining the sensitive data.

The actions against a violation was recognized by the investigators as the hardest decision since it may represent an obtrusiveness in the regular job perform by employees. It was highlighted that Data Endpoint and LeakProof could block, request the user's justification, notify the administrator, and log the incident. Data Endpoint had the singular capability of running custom scripts to respond to a violation. It is also make it clear that the agent can be completely silent to the user, when blocking an action, but it was not discussed if it is positive or negative to educate the user and prevent further leak of sensitive data.

It was identified a problem with the Data Endpoint's application centric policy configuration since it can only detect incidents for the defined applications in the policy, if there is a new application in the client, this may bypass the filters. Thus, it is an issue if the user has the privileges to install new software.

The investigators also describe the installation experience. For Websense they presented that for the server installation, it was required a manual installation of Oracle and MS SQL, and that it had a built in utility for agent software generation. In the configuration stage, it was tested the centralized management and the easy to use interface.

The investigators concluded that Websense does not have a total centralized management console, it was required support for its configuration, which was provided, and they highlighted the multiples templates given by industry and locality.

The author considered important to mention, that in the configuration test it is only specified for Data Endpoint its predefined templates and their default configuration, but for TrendMicro it is not specified if the configuration was left as default, and the investigators mentioned that for the Identity finder's configuration they performed a custom configuration, thus the DLP solutions may not have been in the same conditions if some were left as default and others were configured by the investigators, this could left the vendors in a not fair competition.

The performance tests were done by operating system using Windows XP, Windows Vista, Windows Server 2003, and Windows Server 2008. It has not been described by the investigators why it was not tested a Linux system in this test. The by Protected File Type test included HIPAA, PCI, Source Code, Classified data, Legal information, Media, File Name, and standards.

The authors considers that an example of data could have been useful to understand the nature of the information that was tested. The by Exfiltration Method test was done using: USB drive, CD, network drive, network printer, webmail, open source mail, Microsoft mail, P2P sharing, instant messenger, pasting.

The authors identified that it was not specified the software and versions that were used for the testing. The author considers this as crucial information in order to understand the real capabilities and limitations of the software.

Finally, the agent systems resources tested used of memory, processor, and hard disk usage. In this section the author considers that it should have been useful if the investigators show for which task they did the measurement of memory and processor utilization, and also how and which tool was used to get the results. Also it was not specified if the disk usage reading was just after the installation of the agent or after a certain number of incidents were saved in the log's database of the endpoint.

*2.4.2 Perimeter DLP tools evaluation*

The investigators Blakely, and Evans [5] performed a previous evaluation very similar to the one done by Blakely, Rabe, and Duffy [6], but related to perimeter DLP tools. Here they set up a small network were DLP was installed in-line between the internal and the external network, and it was configured a set of 10 rules.

They tested the speed of the network after applying the filters, and the test was basically using an external computer to get files from an internal machine. It was used 1000 files with sensitive data and some inoffensive data. The vendors used were Fidelis Security Systems, Palisade Systems, Code Green Networks and GTB Technologies. The results were presented in the same way as in the Blakely, Rabe, and Duffy [6] investigation.

The author considers important to comment that the investigators did not justified why they chose the tested vendors, and it was not stated what kind of sensitive data was being transmitted. It may be useful in order to understand the context of the sensitive data, which is trying to be protected. For instance, credit cards numbers (PCI) or personal data (Ley de Comercio Electrónico, Firmas Electrónicas y mensajes de Datos) [19]. It is unknown in which context the vendors are capable of stopping data leaks.

*2.4.3 DLP Evaluation that can protect both endpoint, and perimeter*

The most recent investigation performed by Blakely, Rabe, and Duffy [7], which completed their series of DLP reviews is about DLP that can protect both endpoint, and perimeter. They performed the test for the vendors McAfee, and Sophos.
The tested network was bigger than the used before in the previous evaluations. The test evaluated the installation, configuration/functionality, device control, remediation capabilities, monitoring, notification, and workflow. The author considers important to mention that this test did not provided the type of sensitive data used nor ether the policies tested. It was only mentioned that it was found templates for HIPAA, PCI, and personally identifiable information. It was positive that for this test the investigators performed an out of the box policy compliance evaluation.

All of these tests were performed at the Iowa State University Internet-Scale Event and Attack Generation Environment (ISEAGE) Laboratory. It is important to mention that the investigation done by Blakely, Rabe, and Duffy [6] gives a good contribution to the present study since it has clearer metrics than the other researches, it is related to end point DLP, and it evaluates commercial vendors leaders in the market like Websense.

## 3. METHODOLOGY FOR DLP EVALUATION

### *3.1 General Description*

The authors considered suitable the used of the inductive method [20] in order to test the capabilities of DLP technology. Since technological solutions are based on the requirements of the industry, the author considered viable to develop the methodology using a specific necessity of all private and public companies [21]. If the investigator does not have a clear idea of what to protect or which law has to be complied, testing will be complex and it may face resource limitation problems since it is aiming to test everything. The private information related to employees or customers has been chosen due to support the new productive matrix of the Republic of Ecuador, where it is imply that services should be improved, so information security will be the starting point to assure that the service will be delivered correctly [22]. Consequently, the authors will proposed a methodology to test the capabilities of DLP technology by checking how it can prevent the loss of sensitive person's data, which according to the "Ley de Comercio Electronico, Firmas y Mensajes de Datos" from Ecuador[19], must be protected because it is private information. The methodology will provide the capabilities of the DLP technology for this specific scenario, but the capabilities of the DLP technology can be generalized using the inductive method [23].

In addition, the specific to general approach will be used in combination with a quantitative method [24], since the methodology will propose to gather exact data related to the capabilities of the software, which help to understand the level of accuracy to detect sensitive data. The templates proposed for the quantitative measurement (See section 3.2) shows the independent variables of this study because they are data, which is controlled by the investigator [25]. For example, the sensitive data formats will be the independent variables (See appendix 1) and the results of DLP capability will be the dependent variables since they will be reacting depending on the scenario that is given to the DLP software.

### *3.2 Sensitive Information*

The sensitive information is related to personal information that employees or customers may supply to organizations. In general this information includes contact details, such as identification number, name, gender, date of birth, address, phone, and email.

Furthermore, the common data formats with sensitive content, which is proposed in the DLP testing methodology are the most common recognized based on the extensions of Microsoft Office, [26], and Open Office (OpenWith 2009). The common data formats are: txt, doc, docx, rtf, xls, xlsx, csv, ppt, pptx, pdf, odt, ods, odp, odg, odf, pub, html, and xml. Also, it is going to be used a filter evasion data type in order to test the limitations of the DLP technology. They were chosen by the author as a type that the users may use in order to bypass the DLP filter, based on extension of images (File Extensions 2014), cryptography, hash, encoding [27], and compression [28]. The filter evasion data formats are: png, jpg, bmp, bin, iso, sha-512, AES-256, tar, gz, tar.gz, 7zip, rar, zip, zip_password, and base64. All the encrypted, hash, compressed, and encoded data is proposed to use as an input a file with sensitive information in a txt format.

### *3.3 DLP testing and measurement tools*

Firstly, the authors designed to begin the testing methodology with a characterization of the DLP technology using parameters based on the investigation of Blakely, Rabe and Duffy [6], Mogull [8], and the datasheets of the Open Source tools selected OpenDLP [29], and MyDLP [30]; and one of the commercial leaders in DLP technologies according to Gartner, Websense. In this stage, it is proposed to get the capabilities of the DLP software in: resources required, variety of filters type, deployment flexibility, functionality, granularity of policy templates, and granularity of reactions against an incident. The authors considered important to include in this section metrics related to the user experience when using the software as well [6]. See the complete tables for DLP software characterization in Appendix 2. In this stage the methodology obtains the general characteristics of the DLP software in order to give a context of the technical features of the DLP.

Secondly, the testing methodology will check how the DLP technology is able to identify policy violation related to sensitive data. The authors generated a table with the most common methods that DLP solutions detects policy violation related to sensitive data, which are: regular expressions, dictionary lists, and fingerprinting [6, 8], and these techniques are going to be used to detect policy violation related to each sensitive data, that according to the "Ley de Comercio Electrónico, Firmas y Mensajes de Datos", should be protected (See section 3.2). This data corresponds to: ideology, political association, ethnic origin (structured), disability (structured), sexual preference, migratory status, and religion (unstructured) according to the "Ley del Sistema Nacional de Registro de Datos Públicos"[31]. The authors in this stage of the methodology propose to test the capabilities of the DLP solutions to detect policy violation for these specific types of sensitive data, using predefined and user defined policies. The predefined policies of the software will be used with their default configuration in order to verify the effectiveness of the DLP software with the default configuration. Also, it will be generated by the author policies tailored for the specific purpose of the organization, which in most cases will be specified in the information security policy. This section of the methodology is done in order to test if the software is flexible

to support the specific scenario for the industry. This stage will be testing the DLP software using files with sensitive information in txt format copied to a removable device channel since they are one of the simplest format and common channels in organizations for endpoints. If the software only supports data at rest, it will be perform a discovery scan. This part of the methodology does not focus on file formats nor channels because it will be evaluating the methods to detect policy violations related to classified data, and not the file cracking capabilities of the DLP solution. See Appendix 3 for the summary table that is going to be used for testing the method of policy violation detection related to sensitive information.

Thirdly, it is proposed to be performed a functionality test for endpoint and data at rest DLP technology focusing in different data formats files containing sensitive data against the channels that the user may utilize to leak information. To begin, in this stage of the methodology, it is going to be evaluated if a policy is able to prevent a data loss, using all the common and filter evasion data formats defined for sensitive information in section 3.2, such as txt, pdf, doc, bmp, zip, etc., and this will be put in contrast with the channels that the user may use to leak sensitive information out of the company. The channels defined by the author were based on the most common channels used in a company and also it is supported in the investigation done by [6]. The channels proposed to test the capabilities of the DLP technology in preventing data loss are: USB or removable device, printer, Email, HTTP, HTTPS, FTP, SSH, Shared folders or network folders, and CD burning. The outcome of this experiment will present which data format is detected and not detected through the different channels of the data endpoint.

In addition, the testing methodology for data at rest will check if the different data formats can be discovered by the DLP technology in File Systems and Databases verifying if it is required an agent or it can be performed an agentless operation. The discovery of sensitive information in databases will be tested having the classified information stored in plain text, encrypted, and hash. The methodology is able to support any Database management system (DBMS), such as MS SQL Server [13], or MySQL [32], which are one of the most popular DBMS in Windows and Linux Systems respectively.
It is important to mention that the procedure proposed will support clients in Windows and Linux Systems. All the tables related to data formats vs channels testing endpoints and data at rest DLP can be seen in Appendix 1.

### *3.4 Chosen DLP technology and testing environment*

The DLP software, which can be tested could be Open Source or Commercial, since it does not depend on the type of vendor, but it depends on its capabilities and mechanisms.
In addition, it is recommended to use a virtual environment when testing the DLP evaluation methodology due to the flexibility and friendly testing environment.

## 4. CONCLUSION AND FUTURE WORK

It was obtained a methodology to evaluate DLP technologies for either open source or commercial software, for private or public institutions, who are requested by law to secure the personal information of employees or clients. The methodology also took characteristics of leading open source and commercial software related to DLP technologies in order to generate template for DLP evaluation.

The future work is the possibility of testing the methodology against Open Source, and commercial software in order to analyze if the methodology, and the DLP technology is effective.

## REFERENCES


[1] GHOSH, Mahuya (2010). Telecoms fraud. [online]. *Computer fraud & security*, 2010 (7), 14-17. [online]. Last accessed 01 October 2014.

[2] BS ISO/IEC 27001:2013: Information technology. security techniques. information security management systems. requirements. (2013).

[3] WIKILEAKS (2014). Latest Releases/WikiLeaks Archives. [online]. Last accessed 01 October 2014 at: https://www.wikileaks.org/.

[4] THEGUARDIAN (2013). Edward Snowden: the whistleblower behind the NSA surveillance revelations. [online]. Last accessed 01 October 2014 at: http://www.theguardian.com/world/2013/jun/09/edward-snowden-nsa-whistleblower-surveillance.

[5] BLAKELY Benjamin, and EVANS Nate (2009). Perimeter DLP tools require fine tuning to effectively block 'bad' data from escaping the network [online]. Network World, Inc. Network World, Last accessed 01 October 2014 at: http://www.networkworld.com/article/2259775/security/best-data-loss-prevention-tools.html

[6] BLAKELY Benjamin, RABE Mark AND DUFFY Justin (2009). Block data leaks at the endpoint; TrendMicro, websense offer effective protection against insider security breaches. [online]. Network World, Inc. Network World, 44 Last accessed 09 April 2014 at: http://shu.summon.serialssolutions.com/2.0.0/link/0/eLvHCXMwXV1BCgIxDAziCxb0F5U2aUl7Fpd9wH6gSVpv_v9oFjyo54QcZyYMzAAQ3mL4wwRXGcxRLc5pXWrLiF3G4epkMi7yE2PwBfDrAqfxusC-Pvb7Fj79AOHJlAI3S9P51Fnb79XeWlElTWiaxReObCkSw65silYKDWnqg1rTyK40rnD2F3u8AZEOKqU.

[7] BLAKELY Benjamin, RABE Mark AND DUFFY Justin (2010). McAfee, Sophos shine in test of data loss prevention tools that can do it all [online]. Network World, Inc. Network World, Last accessed 01 October 2014 at: http://www.networkworld.com/article/2205740/network-security/data-loss-prevention-comes-of-age.html

[8] MOGULL Rich (2010). Understanding and Selecting a Data Loss Prevention Solution. [online]. Websense. Last accessed 01 October 2014 at: https://securosis.com/assets/library/reports/Understanding_and_Selecting_DLP.V2_.Final_.pdf

[9] KANAGASINGHAM Prathaben (2008). Data Loss Prevention. [online]. SANS Institute. Last accessed 09 April 2014 at: https://www.sans.org/reading-room/whitepapers/dlp/data-loss-prevention-32883

[10] LEFEBVRE, William (1999). Regular expressions. *Performance computing*, **17** (11), 49-51. . [online]. Last accessed 06 October 2014 at: http://search.proquest.com.lcproxy.shu.ac.uk/docview/237179854/abstract?accountid=13827#

[11] Loachamín, D. S. G., & Lanchas, V. M. (2014). Arquitectura Distribuida para la Respuesta Automática a Intrusiones en un IRS Basado en Ontologías. Revista Politécnica, 33(1).

[12] BEALE, Jay (2007). Snort IDS and IPS toolkit. Burlington, MA, Syngress Publishing, Inc.

[13] MICROSOFT (2014). Data Loss Prevention. [online]. Last accessed 07 October 2014 at: http://technet.microsoft.com/en-gb/library/jj150527(v=exchg.150).aspx

[14] MyDLP (2013). MyDLP Administration Guide. [online]. Last accessed 07 October 2014 at: http://www.mydlp.com/wp-content/uploads/Myministration-Guide.pdf
[15] BUNKER, Guy and FRASER-KING, Gareth (2009). Data leaks for dummies. Chichester; Hoboken, N.J, Wiley Publishing, Inc.
[16] DSPACE (2014). About DSpace. . [online]. Last accessed 07 October 2014 at: http://www.dspace.org/introducing
[17] ALFRESCO SOFTWARE. (2014). About Alfresco. [online]. Last accessed 02 October 2014 at: http://www.alfresco.com/
[18] Secure Computing (2008). Secure Computing in Leaders Quadrant. [online]. Last accessed 09 October 2014 at: http://www.securecomputing.com/magicquadrantweb2008-gartner.cfm
[19] Ecuador. (2002). Ley de comercio electrónico, firmas electrónicas y mensajes de datos. Corporación de Estudios y Publicaciones.
[20] UPP, Victor (2006). The sage dictionary of social research methods. [online]. Thousand Oaks, Calif; London, SAGE Publications. at: http://shu.summon.serialssolutions.com/2.0.0/link/0/eLvHCXMwY2AwNtIz0EUrE8zMDRLTjExTzRItDVNA2lKTkyySzBJBywKTDZLTUI4xQCrg3YQYmFLzRBlk3FxDnD10izNK46HDGvFJhuagtjuwQy7GwJsIWg6eVwLeNpYiwaCQamZpkWieZJpqYZZiYmKelJgCrBkNUlOBzfUky2Rzc0kGUaxmAQD2fjIP.
[21] GILL, John, JOHNSON, Phil and CLARK, Murray (2010). Research methods for managers. [online]. Los Angeles, [Calif.]; London, SAGE. at: http://shu.summon.serialssolutions.com/2.0.0/link/0/eLvHCXMwY2AwNtIz0EUrEwyNkpONQFWLUZqBmUFiSqqBkWVqinmaZQrosBLwmUmIYwyQCng3IQam1DxRBhk31xBnD93ijNJ46LBGfJKhqTn4LBEjMQbeRNBy8LwS8LaxFAkGBYNksyRw5yUxKc0kKcnMIi3F2Ngy2TLZ1MLMNNHETJJBFKtZADGjMKk.
[22] León, M. (2009). Cambiar la economía para cambiar la vida. El Buen Vivir. Una vía para el desarrollo, 63-74.
[23] SAUNDERS, Mark, LEWIS, Philip and THORNHILL, Adrian (2012). Research methods for business students. [online]. Harlow, Pearson. at: http://shu.summon.serialssolutions.com/2.0.0/link/0/eLvHCXMwY2AwNtIz0EUrE9KSUszMDZNMDZINkoyNUgxMTdNSjIzTLBJT08wNktLMUI4xQCrg3YQYmFLzRBlk3FxDnD10izNK46HDGvFJhuagzgOw_hZj4E0ELQfPKwFvG0uRYFCwTAW2w5PTksyNTRJNDC0MEtPSzFOTkgxNLcwMDVOT0yQZRLGaBQDHiTH4.
[24] GHAURI, Pervez N. and GRØNHAUG, Kjell (2010). Research methods in business studies. [online]. Harlow, Financial Times Prentice Hall. at: http://shu.summon.serialssolutions.com/2.0.0/link/0/eLvHCXMwY2AwNtIz0EUrEyxTjQzTEtOSDE1SUkySTJKSjZJT04BN5VQLSzPDJMNklGMMkAp4NyEGptQ8UQYZN9cQZw_d4ozSeOiwRnySoTnoXFtgHSTGwJsIWg6eVwLeNpYiwaBgnmpiYZlsYJxkCayDzFOTEoH9jrQ0i9REC9C-lERDSQZRrGYBANisMZU.
[25] COLLIS, Jill and HUSSEY, Roger (2014). Business research: A practical guide for undergraduate and postgraduate students. [online]. Basingstoke, Hampshire, Palgrave Macmillan. at: http://shu.summon.serialssolutions.com/2.0.0/link/0/eLvHCXMwY2AwNtIz0EUrE8ySU9KATYE086TkVFMDs1TLlJTkZKO0NMOkxLREY8ilEfBjDJAKeDchBqbUPFEGGTfXEGcP3eKM0njosEZ8kqG5hQmwg2VhJMbAmwhaDp5XAt42liLBoGBpnmxskWRumGhikWYC2u1pbGRgnpYGtNTQzCjNJFWSQRSrWQANaDIS.
[26] File Extensions (2014). Microsoft Office File Extensions. [online]. Last accessed 07 October 2014 at: http://www.file-extensions.org/filetype/extension/name/microsoft-office-files
[27] SCHNEIER, Bruce (1996). Applied cryptography: Protocols, algorithms and source code in C. [online]. Wiley. at: http://shu.summon.serialssolutions.com/2.0.0/link/0/eLvHCXMwY2AwNtIz0EUrEyxSUwzSUk3NEpNTgGnM0tQ0DbTpMTHJwDgx1dIUPLKNOMYAqYB3E2JgSs0TZZBxcw1x9tAtziiNhw5rxCcBayZDc2B1aiTGwJsIWg6eVwLeNpYiwaBgmJScaATadGlkmmaSkgIsg43N0kyAjXfQr
[28] PAVLOV (2014). 7-Zip. [online]. Last accessed 06 October 2014 at: http://www.7-zip.org/
[29] OpenDLP (2012). OpenDLP 0.5.1 README. [online]. Last accessed 06 October 2014 at: https://code.google.com/p/opendlp/downloads/detail?name=README-0.5.1&can=2&q=
[30] MyDLP (2013). MyDLP Administration Guide. [online]. Last accessed 07 October 2014 at: http://www.mydlp.com/wp-content/uploads/Myministration-Guide.pdf
[31] De Transparencia, L. (2004). acceso a la información pública. Registro Oficial, 337.
[32] ORACLE (2014). MySQL. [online]. Last accessed 05 October 2014 at: http://www.mysql.com/

**Appendix 1**. Data Formats vs Channels Testing for Endpoints and Data at Rest
DLP
-Data Formats vs channels functionality test for endpoint (common data type)

| | Sensitive Data Format | Channel - Data at Endpoint | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|
| | | Removable Device | Printer | Email | HTTP/ HTTPS | FTP | SSH | Shared Folder | CD Burning |
| Functionality Test for Endpoint (common data format) | .txt | | | | | | | | |
| | .doc | | | | | | | | |
| | .docx | | | | | | | | |
| | .rtf | | | | | | | | |
| | .xls | | | | | | | | |
| | .xlsx | | | | | | | | |
| | .csv | | | | | | | | |
| | .ppt | | | | | | | | |
| | .pptx | | | | | | | | |
| | .pdf | | | | | | | | |
| | .odt | | | | | | | | |
| | .ods | | | | | | | | |
| | .odp | | | | | | | | |
| | .odg | | | | | | | | |
| | .odf | | | | | | | | |
| | .pub | | | | | | | | |
| | .html | | | | | | | | |
| | .xml | | | | | | | | |

- Data formats vs channels functionality test for endpoint (filter evasion data type)

| | Sensitive Data Format | Channel - Data at Endpoint | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|
| | | Removable Device | Printer | Email | HTTP/ HTTPS | FTP | SSH | Share Folder | CD Burning |
| Functionality Test for Endpoint (filter evasion data format) | .png | | | | | | | | |
| | .jpg | | | | | | | | |
| | .bmp | | | | | | | | |
| | .bin | | | | | | | | |
| | .iso | | | | | | | | |
| | sha-512 (.txt) | | | | | | | | |
| | AES-256 (txt) | | | | | | | | |
| | .tar (txt) | | | | | | | | |
| | .gz (txt) | | | | | | | | |
| | .tar.gz (txt) | | | | | | | | |
| | .7z (txt) | | | | | | | | |
| | .rar (txt) | | | | | | | | |
| | .zip (txt) | | | | | | | | |
| | .zip_password (txt) | | | | | | | | |
| | .zip (x 2) | | | | | | | | |
| | MS SQL Backup | | | | | | | | |
| | MySQL Backup | | | | | | | | |
| | Base64 Encoded | | | | | | | | |

- Data formats vs channels functionality test for file systems - data at rest (common data type)

| | Sensitive Data Format (common) | Windows | | Linux | |
|---|---|---|---|---|---|
| | | Agent | Agentless | Agent | Agentless |
| Functionality Test File Systems for data at rest (common data format) | .txt | | | | |
| | .doc | | | | |
| | .docx | | | | |
| | .rtf | | | | |
| | .xls | | | | |
| | .xlsx | | | | |
| | .csv | | | | |
| | .ppt | | | | |
| | .pptx | | | | |
| | .pdf | | | | |
| | .odt | | | | |
| | .ods | | | | |
| | .odp | | | | |
| | .odg | | | | |
| | .odf | | | | |
| | .pub | | | | |
| | .html | | | | |
| | .xml | | | | |

- Data formats vs channels functionality test for file systems - data at rest (filter evasion data type)

| | Sensitive Data Format (evasion) | Windows | | Linux | |
|---|---|---|---|---|---|
| | | Agent | Agentless | Agent | Agentless |
| Functionality Test File Systems for data at rest (filter evasion data format) | .png | | | | |
| | .jpg | | | | |
| | .bmp | | | | |
| | .bin | | | | |
| | .iso | | | | |
| | sha-512 (.txt) | | | | |
| | AES-256 (txt) | | | | |
| | .tar (txt) | | | | |
| | .gz (txt) | | | | |
| | .tar.gz (txt) | | | | |
| | .7z (txt) | | | | |
| | .rar (txt) | | | | |
| | .zip (txt) | | | | |
| | .zip_password (txt) | | | | |
| | .zip (x 2) | | | | |
| | Base64 Encoded | | | | |

- Data formats vs channels functionality test for data bases - data at rest

| | Sensitive Data | MySQL (Agentless) | SQL Server (Agentless) |
|---|---|---|---|
| Functionality Test Data Bases for data at rest | Plain text | | |
| | Encrypted | | |
| | Hash | | |
| | Encoded Base 64 | | |

| Type | Description | Level |
|---|---|---|
| User Experience | Availability of documentation | |
| | Ease of installation | |
| | Ease of configuration | |
| | Customizable Dashboard | |
| | Ease of Reporting | |
| | Ease of policy editing | |
| | Support | |
| | Total | |
| | Result (%) | |

Key Score:

| Description | Level |
|---|---|
| Very Low | 1 |
| Low | 2 |
| Medium | 3 |
| High | 4 |
| Very High | 5 |

- Resources required

| Type | Description | | Level |
|---|---|---|---|
| Resources required | Agent/ Agentless | CPU Usage for (one full discovery at client /agent) | |
| | | Memory Usage for (one full discovery at client/agent) | |
| | | Approximate time of scanning (data at rest only) | |
| | Disk Usage for installation in server | | |
| | Disk Usage for installation in client for agent | | |

- Capabilities

| Type | Description | Support |
|---|---|---|
| Variety of filters type | Endpoint | |
| | Data at Rest | |
| Deployment Flexibility | Supports Windows Endpoint | |
| | Microsoft Active Directory Integration | |
| | Supports Linux Endpoint | |
| | Supports Windows Agent | |
| | Supports Linux Agent | |
| | Workstation Discovery | |
| | ICAP Integration | |
| | SMTP Gateway Integration | |
| | Native Syslog Integration | |
| | Supports Virtualization | |
| Functionality | Offline Endpoint Protection | |
| | Protection after license expiration | |

**Appendix 2.** DLP Software Characterization Tables

- User experience

| Type | Description | Support |
|---|---|---|
| Granularity of policy templates | Supports Regular Expressions | |
| | Predefined Legal Compliance | |
| | Monitors Web Traffic | |
| | Monitors Mail | |
| | Monitors Removable storage devices | |
| | Removable Storage Inbound Data Monitor | |
| | Printer Protection | |
| | Screenshot protection | |
| | Supports keywords | |
| | Predefined Dictionaries | |
| | Predefined Policies | |
| | Predefined Regular Expressions | |
| | Partial (Approximate) Document Matching | |
| | Document hashes | |
| | Predefined Data Formats | |
| | Predefined Source Code Identification | |
| | Mail BCC Protection | |
| | Mail Recipient Details | |
| | Level to check the fingerprint mutation | |
| Granularity of reactions against an incident | Quarantine Actions | |
| | Encrypt data | |
| | Data scrubbing and sanitization | |
| | Email Notification | |
| | Exporting incidents | |
| | Save sensitive data file | |
| | Removable Storage Inbound Archive | |
| | Escalate the incident | |
| | No trust user oriented | |

**Appendix 3.** Testing the method of policy violation detection related to sensitive information that the Ecuadorian Law "Ley del Sistema Nacional de Registro de Datos Públicos" implies to protect.

| Software | DLP Type | Sensitive Data | Method Applied in Policy | | | | | |
|---|---|---|---|---|---|---|---|---|
| | | | Regular Expression | | Dictionary | | Fingerprinting | |
| | | | Predefined | User Defined | Predefined | User Defined | Files | Database Table |
| | Data at Rest/ Data at Endpoint | Ideology | | | | | | |
| | | Political Association | | | | | | |
| | | Ethnic Origin | | | | | | |
| | | Disability | | | | | | |
| | | Sexual Preference | | | | | | |
| | | Migratory Status | | | | | | |
| | | Religion | | | | | | |