Data Loss Prevention in Healthcare

Release Date: 10/05/2023

Health Information Management Working Group
Abstract

1. The growth of digital services in healthcare has increased the challenges of safeguarding medical data.

2. Healthcare IT professionals need effective security measures and insights into user activities to prevent data loss.

3. A strategy focused on understanding user intent and monitoring data movement is essential, and the implementation of Data Loss Prevention (DLP) in healthcare is crucial.

4. While healthcare process automation enhances hospital efficiency, it also brings data loss risks, such as information theft, data leaks, manipulation, and unauthorized data sharing.
Introduction

1. This paper discusses the application of DLP in the healthcare sector, particularly for transferring patient data.

2. With increasing data migration to the Cloud, detecting data exfiltration becomes challenging.

3. The healthcare industry faces significant risks, such as ensuring patient information compliance across regions and increased vulnerability and costs associated with offsite data storage, especially given the sector’s susceptibility to ransomware attacks.

4. While healthcare process automation enhances hospital efficiency, it also brings data loss risks, such as information theft, data leaks, manipulation, and unauthorized data sharing.

5. It’s vital not only to protect the data but also to monitor its usage and movement.

6. In-depth studies are essential to comprehend these risks and implement effective protection measures.
Data Loss Prevention (DLP) in Healthcare

1. DLP in healthcare focuses on preventing unauthorized or accidental exposure of sensitive patient data.

2. The compromise of such data can lead to serious issues like identity theft or medical mishaps.

3. As organizations increasingly use cloud computing, especially after the COVID-19 pandemic’s push, new threats to healthcare data emerge.

4. This includes potential unauthorized access and technical problems like accidental deletion of encryption keys.

5. It’s crucial for healthcare entities to implement a robust DLP strategy to safeguard their data.
**Figure 1: Cloud is the predominant means for transferring and sharing data**

Organizations today are transferring and sharing data primarily through the cloud—a trend catalyzed by the COVID-19 global pandemic. Organizations use a variety and different methods. The most common way is cloud storage applications (46%) such as OneDrive, Box or Dropbox. Other common methods include cloud-to-cloud (39%), which is used slightly more than email (38%) or cloud collaboration and messaging applications (31%) such as Slack or Teams. Regardless of the method, organizations are clearly trusting the cloud with even their most sensitive data.

**Figure 2: Types of Cloud Applications Organizations Are Most Concerned about Being Attacked**

- **Email**: 36%
- **Authentication platform**: 36%
- **Storage and file sharing**: 35%
- **Customer relationship management**: 33%
- **Enterprise business intelligence**: 30%

*Figure 2: Types of Cloud Applications Organizations are Most Concerned about Being Attacked*
How DLP Works

DLP uses a suite of security tools, such as firewalls, AI, and machine learning, to detect anomalies and prevent data breaches. Its techniques include:

1. **Rule-based matching:** Flags data matching specific known patterns, like the format of Social Security numbers.

2. **Database fingerprinting:** Searches for exact matches of structured data within files.

3. **Exact file matching:** Uses cryptographic hashes of file contents for matching.

4. **Contextual Analysis:** Considers factors like user role and location to enhance policy enforcement accuracy and reduce false positives.

5. **Partial document matching:** Identifies files that align with predefined patterns partially.

6. **Statistical analysis:** Utilizes methods like Bayesian analysis to detect violations.

7. **Pre-built categories:** Uses preset rules tailored for Healthcare Delivery Organizations to categorize data based on sensitivity.

8. **User behavior analytics:** Highlights unusual end user actions.
DLP Types

DLP solutions, designed for various data security needs, come in three main types:

1. **Network DLP:** Monitors and blocks data transfers, like malicious insiders emailing sensitive information.

2. **Endpoint DLP:** Stops employees from copying data to removable devices and scans data at rest on local systems, following specific endpoint data loss policies.

3. **Cloud DLP:** Guards against unauthorized data transfers or unintended sharing in the cloud. Due to threats like reputational damage, loss of intellectual property, legal repercussions, and potential fines, organizations view data loss as the major concern in cloud-based attacks. As a solution, many are shifting from perimeter-based networks to user- and application-focused networks, transitioning from on-premise to cloud-based DLP to inspect outgoing data without hampering the user experience.
In the healthcare sector, there's a need to regularly update medical classifications as per the World Health Organization (WHO).

Protecting transfers of documents with FDA-approved drugs, CPT/ICD-10 codes, and medical diagnostic terms requires constant updates to DLP with medical terms.

Two features assisting this are Exact Data Matching (EDM) and Indexed Document Matching (IDM).

EDM monitors structured data, focusing on row and column (or CSV) formats, while IDM deals with unstructured data, matching indexed documents and images.
Exact Data Matching and Indexed Document Matching (EDM/IDM)

**PHI Data – Exact Data Matching**

<table>
<thead>
<tr>
<th></th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>Names</td>
<td>Health Plan Beneficiary Numbers</td>
</tr>
<tr>
<td>Addresses</td>
<td>Device Identifiers and Serial Numbers</td>
</tr>
<tr>
<td>Dates</td>
<td>Certificate / license numbers</td>
</tr>
<tr>
<td>Phone and Fax</td>
<td>Account Numbers</td>
</tr>
<tr>
<td>Email Addresses</td>
<td>Vehicle Identifiers</td>
</tr>
<tr>
<td>IP Addresses</td>
<td>Website URLs</td>
</tr>
<tr>
<td>SSN</td>
<td>Full face photos</td>
</tr>
<tr>
<td>Medical Record Numbers</td>
<td>Biometric Identifiers</td>
</tr>
</tbody>
</table>

**Figure 3: PHI - Exact Data Matching**

<table>
<thead>
<tr>
<th>Code</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>1001</td>
<td>Colchicine Serum/Plasma 6095X</td>
</tr>
<tr>
<td>1002</td>
<td>HANDWRITTEN ORDER</td>
</tr>
<tr>
<td>1003</td>
<td>GFAP (Alexander Disease) Sequencing Test</td>
</tr>
</tbody>
</table>

**Figure 4: EDM Sample, e.g., Medical code & Desc.**

**IDM – Indexed Document Matching**

- Protects “unstructured” documents – PDF, Word, Powerpoint, etc. by looking for a 10-90% match in text content
- Can protect images or other files with 100% binary file match

Most common healthcare policies:

- Billing forms, patient registration forms, other common patient correspondence templates – usually broken down by department or subsidiary
- Patient data (scans, medical images – usually very tactical and not global)
- Intellectual Property – designs, procedures, etc.

**Figure 5: Indexed Document Matching**

**Figure 6: IDM Sample, e.g., Order Form**
False Positives

1. False positives are common in DLP deployments due to strict policies that may need adjustments.

2. It’s beneficial to have individuals train the DLP system by marking these incidents accordingly.

3. Often, these false positives arise from blank forms with keywords like "Patient ID" or numerical values resembling credit card or Social Security Numbers, which the DLP system flags.
Data classification is vital for data security, playing a pivotal role in recognizing and safeguarding sensitive information, particularly when used with DLP solutions.

Many DLP vendors enhance their security offerings through data classification, either natively or via third-party integrations. These classification standards determine the confidentiality level of data, forming the foundation for measuring data risk.

When data is correctly labeled, the DLP process becomes streamlined, allowing for easy differentiation between sensitive and non-sensitive data. This boosts DLP efficiency, making data classification an essential aspect of many DLP programs.

By classifying data internally, there's clearer insight into its usage and transfer. With DLP, once a file is classified, an identifier is attached to it (e.g., "confidential"), which can be integrated into DLP policies to manage the file's external transmission, ensuring prevention of potential data leaks.
DLP & Threat Intel

1. Collaborating with the Threat Intel team can enhance DLP because they specialize in analyzing user behavior, helping identify data movements that may lead to potential exfiltration.

2. DLP policies can monitor users deemed as risks or those leaving the organization, alerting on possible insider threats.

3. Measures can be put in place to audit or block their data transfers, set file size limits that, if exceeded, trigger alerts, or detect patterns of data exfiltration by individual users.
Data Usage by Locale

1. Data usage varies by location due to different local and international regulations, such as CPA, GDPR, and FDBR.

2. While advanced DLP solutions can automatically manage this data, it’s crucial to understand the importance of protecting all personal data and complying with regional policies to avoid significant consequences for breaches.

3. Not every piece of data should be shared, but when it’s necessary, the organization must guarantee its security.

4. Using Identity and Access Management (IAM) is vital for data security.

5. At its core, a DLP solution should monitor data usage to prevent the loss or misuse of sensitive information. DLP software categorizes regulated data and detects policy breaches, often driven by regulatory standards like HIPAA, PCI-DSS, or GDPR.

6. Organizations should think about employing Information Rights Management (IRM), a tool that shields documents with sensitive details from unauthorized access.
DLP Protect

1. DLP solutions offered by vendors range from simple content filtering to advanced network- and host-based monitoring.

2. Companies need to oversee and safeguard their sensitive information, watching for various exfiltration methods like email, browsers, cloud sharing, social media, and more.

3. DLP’s mechanism, which inspects plaintext packets, conflicts with encryption as the data must be unencrypted for inspection, potentially creating vulnerabilities if misconfigured.
DLP Protect

Encryption ensures data remains unreadable without the correct decryption key. Techniques include:

1. **Symmetric Encryption**: One key for both encryption and decryption.

2. **Data Shredding in Cloud**: Removes the decryption keys, making data unreadable.

3. **Tokenization and Partial Homomorphic (PHE)**: Replaces sensitive data with surrogate values.

4. **Polymorphic Encryption**: Uses various encrypted forms, each with different keys.

5. **Homomorphic Encryption**: Enables data manipulation without decryption.
## DLP Protect

Additional protective measures include:

<table>
<thead>
<tr>
<th></th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td><strong>Access Controls</strong>: Ensures only authorized users can access data.</td>
</tr>
<tr>
<td>2</td>
<td><strong>User Behavior Analytics</strong>: Monitors user activity to detect suspicious patterns.</td>
</tr>
<tr>
<td>3</td>
<td><strong>Regular Audits</strong>: Identifies vulnerabilities in the system.</td>
</tr>
<tr>
<td>4</td>
<td><strong>Data Monitoring</strong>: Real-time content analysis to prevent breaches.</td>
</tr>
<tr>
<td>5</td>
<td><strong>Training</strong>: Educates employees on security, with specialized sessions for those who repeatedly make errors.</td>
</tr>
<tr>
<td>6</td>
<td><strong>Secure Communication</strong>: Uses encrypted channels to prevent data interception.</td>
</tr>
<tr>
<td>7</td>
<td><strong>Choosing a Reputable Cloud Provider</strong>: Ensures strong security measures are in place.</td>
</tr>
</tbody>
</table>
DLP Best Practices for Enhanced Data Security

1. Centralized DLP Program: Adopt a unified DLP approach for consistent data monitoring and improved visibility, fortifying data security throughout the company.

2. Evaluate Resources: Ensure the presence of knowledgeable staff versed in DLP risk analysis, breach response, legal regulations, and training.

3. Inventory and Assessment: Before implementing DLP, assess the nature and value of data. Identify data types, their compliance needs (e.g., HIPAA, PCI, PII), business sensitivity, processing, transmission, and storage protocols.

4. Phased Implementation: Recognize DLP as an ongoing journey and roll it out in stages for best results.

5. Data Classification System: Before enforcing DLP policies, establish a robust classification framework for data.

6. Data Handling and Remediation Policies: Based on the classification system, develop or refine policies to manage different data types.

7. Employee Education: Equip employees with the knowledge and importance of security policies, ensuring their commitment to safeguarding data.
Clarify the Objective: While many HDOs adopt DLP for compliance standards like HIPAA, a versatile DLP solution offers more, such as data protection and incident response. Customize the solution based on organizational needs.

Alignment with Security Architecture: During DLP design and implementation, evaluate pre-existing security tools like firewalls to ensure the DLP solution integrates seamlessly with the existing security infrastructure.

Continuous Security Review: Regularly assess the DLP solution’s performance. As new features are added, they should be reviewed for their applicability and efficiency.

Change Management Protocols: Have a standardized configuration agreed upon by all stakeholders, with any changes needing consensus before being rolled out.

Routine Testing and Audits: Ensure the DLP system functions as desired by regularly testing its effectiveness and identifying potential areas of improvement.
DLP in Real-Time

1. Real-Time Data Monitoring: DLP solutions actively monitor data as it travels within an HDO’s network and systems, ensuring that sensitive data, like file transfers, are halted in real-time. This includes monitoring points like network exits, user devices, and cloud apps.

2. Content Inspection and Analysis: The data passing through is scrutinized by the DLP. If sensitive data is found, its transmission is stopped. The identification process uses methods like keyword search, regular expressions, data fingerprinting, and machine learning based on set policies.

3. Contextual Analysis: By understanding the context of data movement or access (e.g., user role, location, device), the DLP ensures accurate actions and minimizes false positives. Unauthorized transmission of sensitive information, for instance, is halted.

4. Policy Enforcement: DLP solutions proactively enforce policies when breaches are detected. Depending on the settings, actions might range from stopping data transmission to encrypting data, notifying stakeholders, or logging the event.

5. Incident Response: Upon detecting a policy violation, the DLP initiates an incident response process, which may involve notifying security teams, escalating concerns, or implementing measures to counteract data exposure.

6. Reporting: Through DLP, organizations can generate comprehensive reports, such as compliance assessments, breach analyses, and trend evaluations, shedding light on their data security stance.

7. Continuous Updates and Improvement: To remain effective against ever-evolving threats and regulations, DLP solutions regularly update their databases and detection methods.
Measuring DLP Maturity

Critical Elements of Data Loss Prevention Program Maturity

1. **Program Governance**: It involves the organizational structure, processes, and policies guiding decisions related to DLP uses, implementations, and policies.

2. **Enterprise Coverage**: This measures the extent of DLP protections across the network, users, and data within the organization.

3. **Policy Coverage**: Evaluates the comprehensiveness of DLP protections for all data types relevant to the organization.

4. **Incident Remediation**: Refers to how the organization responds to violations of DLP policies, encompassing processes, training, and personnel.

5. **Security Awareness**: Focuses on the knowledge level of end-users about security policies, the value of DLP tools, and the security implications of data.

6. **Metrics and Reporting**: Emphasizes the importance of having measurable outcomes for DLP programs to demonstrate security improvement and return on investment.
Homomorphic Encryption and CMMI Maturity Levels

As a result no information is lost by getting encrypted results and decryption at its own end.

Figure 7: A systematic review of homomorphic encryption in Healthcare

<table>
<thead>
<tr>
<th></th>
<th>Initial</th>
<th>Managed</th>
<th>Defined</th>
<th>Quantitatively Managed</th>
<th>Optimizing</th>
</tr>
</thead>
<tbody>
<tr>
<td>Program Governance</td>
<td></td>
<td>X</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Enterprise Coverage</td>
<td></td>
<td></td>
<td>X</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Policy</td>
<td>X</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Coverage Incident Remediation</td>
<td>X</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Security Awareness</td>
<td></td>
<td></td>
<td>X</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Metrics and Reporting</td>
<td></td>
<td></td>
<td></td>
<td>X</td>
<td></td>
</tr>
</tbody>
</table>

Figure 8: CMMI Maturity Levels
Measuring DLP Maturity

Maturity Rating Using CMMI

1. **Initial**: At this level, processes are chaotic, poorly managed, and reactive, leading to increased risks.

2. **Managed**: Here, processes are project-based but still reactive.

3. **Defined**: The organization has well-defined processes and acts more proactively. There are set standards guiding the processes.

4. **Quantitatively Managed**: Processes are not just managed but also measured. The organization relies on data to ensure processes align with its goals.

5. **Optimizing**: The highest maturity level, where processes are both stable and adaptable, with the focus on continuous improvement and adaptability.

6. By identifying these critical elements and then applying the CMMI maturity levels, HDOs can get a clearer picture of their DLP program’s effectiveness and areas of improvement.
Conclusion

1. Data protection is vital for all, particularly in the digital age with numerous communication channels.

2. While the healthcare industry stands to gain significantly from digitalization, it also faces heightened risks if adequate security measures and awareness are lacking.

3. Continuously monitoring data boundaries and collaborating with data owners can enhance preventative strategies and reduce vulnerabilities, ensuring data safety.