Lexmark™

Lexmark Secure Document Monitor Solution overview

Monitor documents

Extract content

Discover leaks

December 2017

Table of contents

Solution overview

Securing paper documents

When it comes to security, most organizations focus on protecting information from outside hackers and attacks. However, as businesses operate across increasing numbers of locations and sources, insider threat—both malicious and non-malicious—is a growing concern.

And with insider threat protection, one gap is especially present: the collection and monitoring of hardcopy data—that information that is printed, copied, scanned or faxed. Lexmark Secure Document Monitor (LSDM) fills the hardcopy monitoring gap.

LSDM resides on your Lexmark MFPs to automatically and discreetly capture content and user data from every document that passes through. Put simply, it provides real-time capture, without interrupting or delaying processes and performance.

From your Lexmark device, captured data is routed seamlessly to Optical Character Recognition (OCR) where the actual content from each page and on to your Data Loss Prevention (DLP) or monitoring system for review. Lexmark can provide the OCR capability, you may use your existing OCR engine or some DLP systems contain their own OCR. LSDM enables you to apply the same monitoring capability to your hard copy documents as you do to email, web activity, and files residing on shared and hard drives.

Use case: Tracking document transactions

1. A user obtains a confidential document with or without authorization.
2. The user logs into the Lexmark MFP with an employee badge or other form of authentication.
3. Using the e-Task touch screen, the user scans the document and makes a copy about "Project Alpha".
4. Lexmark Secure Document Monitor automatically:
   - Captures a digital image of the document
   - Collects the user's ID, date and time of the transaction and the device ID
5. The job is converted into a TIFF image or multiple TIFF images (if the document is more than one page) and sent to an OCR engine, which is either provided by Lexmark, the customer or the customer's existing DLP.
6. Each document's image, metadata and full-text results (extracted during OCR) are then pushed into a DLP server.
7. The content is submitted against the monitoring parameters set up in the DLP by the organization.

Within the customer's DLP

8. The organization was worried about information leaking about Classified "Project Alpha" and had set up an alert for any document containing that phrase.
9. The security officer receives a notification that a document was copied by an employee that contains phrase, "Project Alpha".
10. The security officer proceeds with an investigation, using the digital document images and specific transaction details as evidence.

Solution benefits

- Audit trail - Users must authenticate at the MFP before performing tasks and document metadata is captured with each transaction.
- Cost and time saving - Investigations into unauthorized document distribution done in minutes instead of days or weeks.
- Search full document content - the automatic OCR of each document lets you search actual document content for important confidential or secret words.
- Easy DLP integration - As long as your DLP or other backend system can open an API call via web services, LSDM can send the document for consumption and processing against your established monitoring rules.

* MFP or printer with Lexmark eTask user interface. A list of supported devices is provided in Appendix A.

Solution building blocks

Lexmark Secure Document Monitor takes advantage of Lexmark's unique ability to tightly integrate advanced MFP functionality with a wide variety of electronic backend document monitoring and management systems such as DLP and Enterprise Content Management (ECM).

Software components

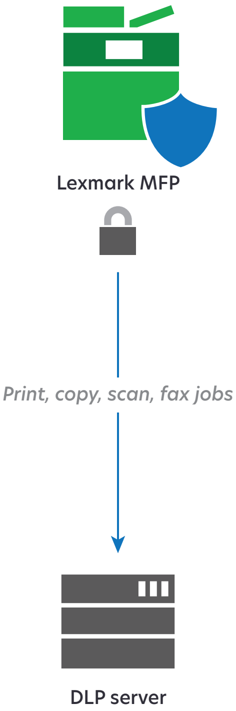
- Lexmark Document Distributor (If customer requires Lexmark to perform OCR of documents)
- Lexmark Secure Document Monitor eSF Application licensed for each enrolled print device

Hardware components

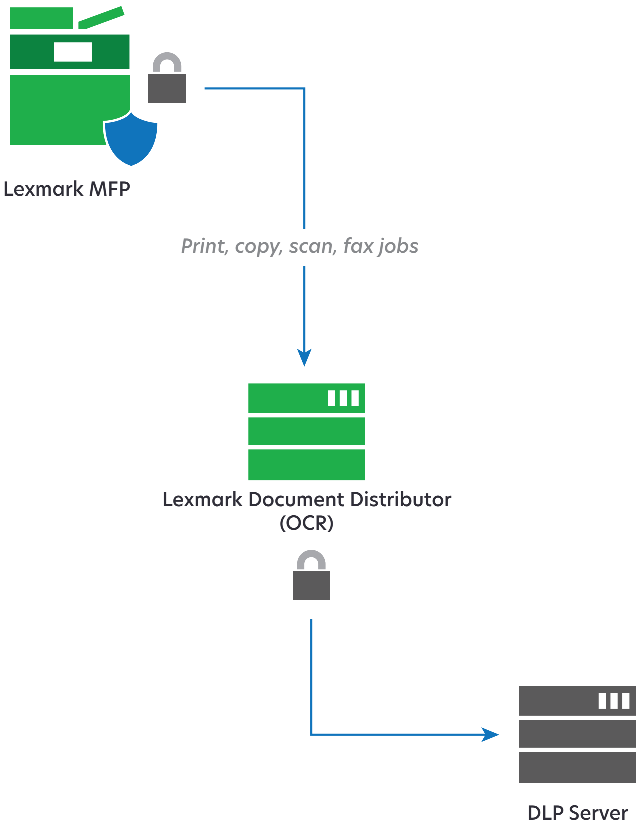
- Lexmark Multifunction Printers (MFPs) or printers with e-Task touch screen interface
    - A list of compatible Lexmark MFPs and printers is found in Appendix A
- Memory: 512 MB minimum Hard disk: required

LSDM architecture

eSF app to 3rd Party DLP



eSF app to LDD to DLP


Lexmark MFP
Print, copy, scan, fax jobs
Lexmark Document Distributor (OCR)
DLP Server

How LSDM eSF application works

An Embedded Solutions Framework (eSF) grabs the TIFF(s) and posts it to a web service of the customer's DLP system or Lexmark Document Distributor (LDD) if the customer needs Lexmark to perform the OCR. A second eSF application on the device will perform identification card authentication. Each app must have a license file applied.

Users must authenticate at the MFP or printer before performing any copy, fax, scan or print retrieval tasks. Authentication is performed by swiping a badge or by manually entering user-name and password credentials.

Once authenticated at the device, a session is created for that user. As each page is processed, firmware writes a copy of the page to disk within the device and sends it (along with metadata such as the user and printer) to an OCR engine where the content of each page is extracted. This backend system can be:

- Lexmark LDD
- A customer's existing OCR system
- A customer's DLP system, which contains OCR capability.

*In order to assure optimal OCR, the solution removes the ability to scan at low resolution (75 dpi).

After the document is OCR'd, that content is submitted against the monitoring policies that have been established in the customer's DLP.

The app can be configured to lock the device if there is an issue with the hard disk or connection to the DLP system. This helps to prevent circumvention. A configurable message is displayed informing the user what to do in the event of an error.

System sizing and scalability

OCR system

There is a combination of factors that are used to determine the demand created upon the OCR system by the LSDM solution.

Peak demand

- Number of people who will print and scan documents
- Number of print jobs expected each day, hour or second
- Typical print/scan job file size
- Length of time job is held

Performance optimization across servers

If load balancing is employed, additional servers may be desired.

Site locations

Organizations which have many locations physically distributed across large areas may experience wait times being exaggerated as print data moves across the WAN. Several distributed server nodes managed by a central node will improve performance when supporting multiple work sites in different geographic locales.

Document storage impacts

Several important factors are considered in determining the number and location of servers required to process jobs efficiently as well as the size and type of database.

Number of pages being processed per second

The system must receive and process pages from the device at a rate that will report potential leaks in a responsive window of time. The total pages per second for all devices will influence the processing speed required for the OCR engine.

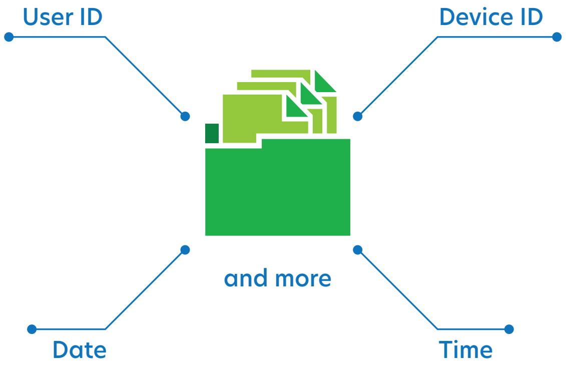
File size and volume

The disk space required to store the content within the DLP depends on the amount of content that must be stored. The storage requirements for scanned images and other documents can grow quite large. As a guide, one million black and white scanned pages (scanned at 300 dpi) will consume 70 GB of storage.

The size of files and the length of time files are held in the content system will impact the storage requirements.

Tracking and reporting

Documents that are sent to a DLP can have a variety of data describing the document (metadata) that is associated with each document.



The following table lists the metadata fields that are examples of the different types of metadata that can be associated with each document that is sent to the DLP.

User name who performed the operation
Time operation occurred
Device name where the operation occurred
Type of document (print, copy, fax, email, FTP)
Number of copies
Printed document file name
Computer name and user name that sent originating print job
Recipient fax, email or FTP address
Sender's fax number or email address
Email subject
Device location, device contact name, and device hostname as specified on embedded web page

Employee stories

Real world examples of insider threats with paper documents

Nghia Hoang Pho (2017)

NSA employee accused of removing classified material in both hard copy and digital form.

Reality Winner (2017)

NSA contractor accused of leaking a classified report on Russian hacking. Told FBI agents that she smuggled classified paper documents out of a high security intelligence facility in her clothes.

Harold Thomas Martin (2016)

NSA contractor, was charged with theft of government property and unauthorized removal and retention of classified materials. Boxes of printed classified documents were found in his car and home.

Hanjuan Jin (2012)

A former Motorola software engineer—was convicted in February 2012 of three counts of stealing trade secrets and attempting to sell them to the Chinese military. Jin went to Motorola at night and electronically downloaded and printed 1,000 pages of trade secrets and schematics. The only reason she was caught was because as she was about to board a flight from Chicago to China, she was subjected to a random search and the hard drives and paper documents were discovered.

Thomas Drake (2010)

A high ranking senior official at the NSA—allegedly, used a printer to attempt to hide his transmission of classified information. In the indictment against Drake it states that he actually copied and pasted classified information onto untitled Word documents, removed the classification markings and brought these hard copies home with him.

Beyond the MFP

If you look at other recent insider threat cases, where the indictment doesn't specifically call out the printing and photocopying of classified information, you have to wonder the incredible effort and expense those government agencies when through during their investigation to discover the actual content of the documents these individuals printed and photocopied:

- Jeffrey Alexander Sterling – CIA (2011)
- Stephen Jin-Woo Kim - Department of State (2009)
- Mary McCarthy – CIA (2006)
- Stephen "Sandy" Berger – National Security Council (2005)

LSDM "user stories"

Project Alpha

A fictional secret mission run by a government agency to get an American operative, codenamed "Unicorn", that has is providing classified intelligence on the enemy.

1. General alerts – "No one print this"

A government agency has just received the green light to engage in Project Alpha. The members of the Project Alpha planning and execution team have been reminded that absolutely nothing is to be printed or copied about this mission. Given the extreme sensitivity of Project Alpha, the Information Assurance Office sets an alert within their DLP for the phrase "Project Alpha" and set up a filter as a top priority. If anyone prints or scans a document that contains the phrase "Project Alpha", they will receive a notification.

2. Exclusionary alerts – fine tuning based on need

The Agency's Project Alpha team is working feverishly to prepare for their mission to be launched and as a result they are creating numerous reports and memos containing the phrase "Project Alpha". While this mission is still top secret, Information Assurance does not want to be distracted by all of the DLP alerts that will be created by team members doing their job. Instead, they want to focus on any documents being printed or copied by people who are not associated with Project Alpha. As a result, Information Assurance creates employee specific DLP alert filter for the phrase "Project Alpha":

"Alert me on any job containing 'Project Alpha' that is not created by employee 234543, 655334, 776543 or 543887."

3. Specific alerts – "Let's watch her"

Agency personnel have reported that one of their colleagues, Amy Richards, is acting strangely. Information Assurance conducts a search on all of the documents that reside in Amy's hard drive, email and internet activity. They also set up an alert within their DLP for all print and copy jobs that Amy creates and commence with watching her activity. Now, everything that Amy prints or copies will automatically trigger an alert.

4. Parallel alerts – "To me, this is okay. To you, this is bad"

Information Assurance team member Johnson has been tasked with watching for the term "Unicorn"—the code name of a terrorist informant. Another team member, Nelson, without the knowledge of the other team members, has been tasked with watching for any documents containing "Project Alpha" and "Unicorn" (in this case), the code name of the Agency's operative.

Agent, Amy Richards, prints a document...

Information Assurance Agent Johnson receives a hit on his DLP alert for "Unicorn". He reviews this document and sees "Unicorn" highlighted in the document. It appears to be a copy of the Richards family's annual Christmas card and it reads,

"Little Annie continues to love ***Unicorns*** *and hopes to own one sometime before she gets married. She's so cute!"*

To Agent Johnson, this appears to be benign and takes no further action.

However, Information Assurance Agent Nelson also receives a hit on this same document in his DLP alert. He reviews the document and sees "Unicorn" and "Project Alpha" highlighted in the document. To Agent Nelson, this also appears to be a Christmas letter and to him it reads,

"Justin has been working hard on building his tree house ***project. Alpha****, our faithful dog watches him all day and is keeping a careful eye on him. Little Annie continues to love* ***Unicorns*** *and hopes to own one sometime before she gets married. She's so cute!"*

To Nelson, this is concerning and he routs it forward for further investigation.

The Team Lead of the Information Assurance investigation team, Agent Vincent, receives a notification that there's a new item for him to look at. Vincent opens up the document and reads the Christmas letter and begins his investigation. In addition to **"project Alpha"** and **"Unicorns"** He quickly notices two things:

1. The term "Little Annie", which has been used to describe a terror cell leader—who he thought had been killed last year.
2. This document was printed by Amy Richards, who is currently being monitored.

Vincent decides that he better notify the Director.

3. Auditability - photo bomb

In order to ensure that the information assurance team is on their toes and not missing any hits, the Information Assurance Audit team creates a false employee and commences printing and copying documents under her name, Amy Richards. One of these documents would seemingly appear to be a simple Christmas letter. However, the audit team has included the words "Project Alpha", "Unicorn" and "Little Annie".

Later that day, Information Assurance Agent Nelson, receives a hit in his DLP alert and forwards it to his investigations team for further investigation.

Agent Nelson has passed this test.

Appendix A: device support (December 2017)

Multifunction printers (MFPs)			
Monochrome		Color	
MX912	Yes	XS955	Yes
MX911	Yes	X954	Yes
MX910	Yes	X952	Yes
MX812	Yes	X950	Yes
MX811	Yes	X945	No
MX810	Yes	X940	No
MX711	Yes	X925	Yes
MX710	Yes	XS796	Yes
MX611	Yes	X792	Yes
MX610	Yes	X782	No
MX511	Yes	X748	Yes
MX510	Yes	X746	Yes
MX410	Yes	X738	Yes
MX310	No	X736	Yes
X864	Yes	X734	Yes
X862	Yes	X548	Yes
X860	Yes	X546	No
X854	No	X544	No
X852	No	CX924	Yes
X850	No	CX923	Yes
X658	Yes	CX922	Yes
X656	Yes	CX921	Yes
X654	Yes	CX860	Yes
X652	Yes	CX825	Yes
X651	Yes	CX820	Yes
X646	No	CX725	Yes
X644	No	CX510	Yes
X642	No	CX410	No
X466	Yes	CX310	No
X464	Yes		
X463	Yes		

Single function printers (SFPs)			
Monochrome		Color	
MS911	Yes	CS923	No
MS812	Yes	CS921	No
MS811	Yes	CS825	No
MS810	Yes	CS820	No
MS610	Yes	CS725	No
MS510	No	CS796	Yes
MS410	No	CS510	Yes
MS310	No	CS410	No
W854	No	CS310	No
W852	No	C950	Yes
W850	No	C925	Yes
W840	No	C792	Yes
T656	No	C782	No
T654	No	C780	No
T652	No	C748	Yes
T650	No	C746	No
T644	No	C736	No
T642	No	C734	No
T640	No	C546	No
		C544	No
		C543	No
		C540	No