Mag. iur. Dr. techn. Michael Sonntag

Data Loss Prevention

Institute for Information Processing and
Microprocessor Technology (FIM)
Johannes Kepler University Linz, Austria

E-Mail: sonntag@fim.uni-linz.ac.at
http://www.fim.uni-linz.ac.at/staff/sonntag.htm

Agenda

- What is data loss prevention?
  - → Why is it needed?
- Simple DLP vs. working DLP
  - → How useful is it?
- Potential attacks:
  - → Simple
  - → Social engineering
  - → Where it is almost impossible to prevent
- Practical examples:
  - → Scanning E-Mails
  - → USB disk restrictions

What is data loss prevention (DLP)?

- Prevent data from getting “lost”: Actually from “disclosure”
  - → **Not**: Backup, RAID, …
  - → **But**: Encrypted storage, scanning all outgoing data, …
- Prevent sensitive data from being “exported” out of the “acceptable location” to somewhere/someone else
  - → Typically: Prevent storing it on some kind of media and transporting it out of the company, sending it by E-Mail, web, IM, P2P to the outside
- Requires an exact definition what is “inside”, and “who” is authorized to do “what” with all kinds of data
  - → Typically additionally requires: Why, when, from where, how often, logging
- Other names:
  - → Data|Information Loss|Leak|Leakage Prevention|Protection

What is data loss prevention (DLP)?

- One definition (Securosis):

  “Products that, based on central policies, identify, monitor, and protect data at rest, in motion, and in use through deep content analysis”

- Main aspects:
  - → Central policy: Organizational help
    - » “Everyone decides for him/herself, hopefully correctly” is too error-prone (and might be the attacker him-/herself)!
  - → Data at rest, in motion, and in use
    - » Examples: Incorrect server share/folder, sending as mail attachment, copying via the clipboard
  - → Content analysis
    - » The system will identify data itself and not (solely) depend on the user marking it correctly

Why DLP?

- Tipping off the press
  - → Scandals (environment, less pay for women, spying on employees, …) which might be illegal/immoral/bad publicity
- Espionage: Less by secret service but competitors
  - → Financial information, customer lists, blueprints, …
  - → Especially: Employees which are (soon) leaving the company
- Legal requirements: Liability/sanctions on data loss
  - → Especially regarding military/state secrets
  - → Compliance rules (legal or contractual)
    - » Health, financial (e.g. SOX, PCI-DSS)
  - → Data breach notification
    - » USA, but Germany/Austria too; EU soon

Main application areas

- Carelessness or ignorance
  - → Internal employees try to do something dangerous or forbidden, but don't know the rules, forgot, …
    - » DLP can be very helpful there, as no circumvention is tried!
- Internal attackers
  - → DLP may be useful, depending on the expertise
  - → Note: You must get it right fast, or trying will set off alarms
- External attackers
  - → DLP is probably not very useful, as they must have managed to get into your system already!
    - » They might use the same way to get out, reconfigure or DLP, …
  - → Even more dangerous: If the DLP is subverted, you get information where what confidential data exists, and the DLP might even help you collecting it!

Simple DLP …

- Assign security permissions on files and assign users to groups who possess rights on them (or not)
  - → I.e., normal file server security
- Install software to scan all E-Mails for certain words
  - → Or compare attachments to internal documents
- Prohibit the use of portable data storage devices
  - → Note: E.g. iPod will often not be recognized as such a device!
- Ensure employees are not local administrators but “normal users” and that they cannot install software
- Use whole disk encryption on file servers
- Ensure temporary files are not created (or overwritten)
- All cameras/mobile phones with cameras must be handed in at the entrance

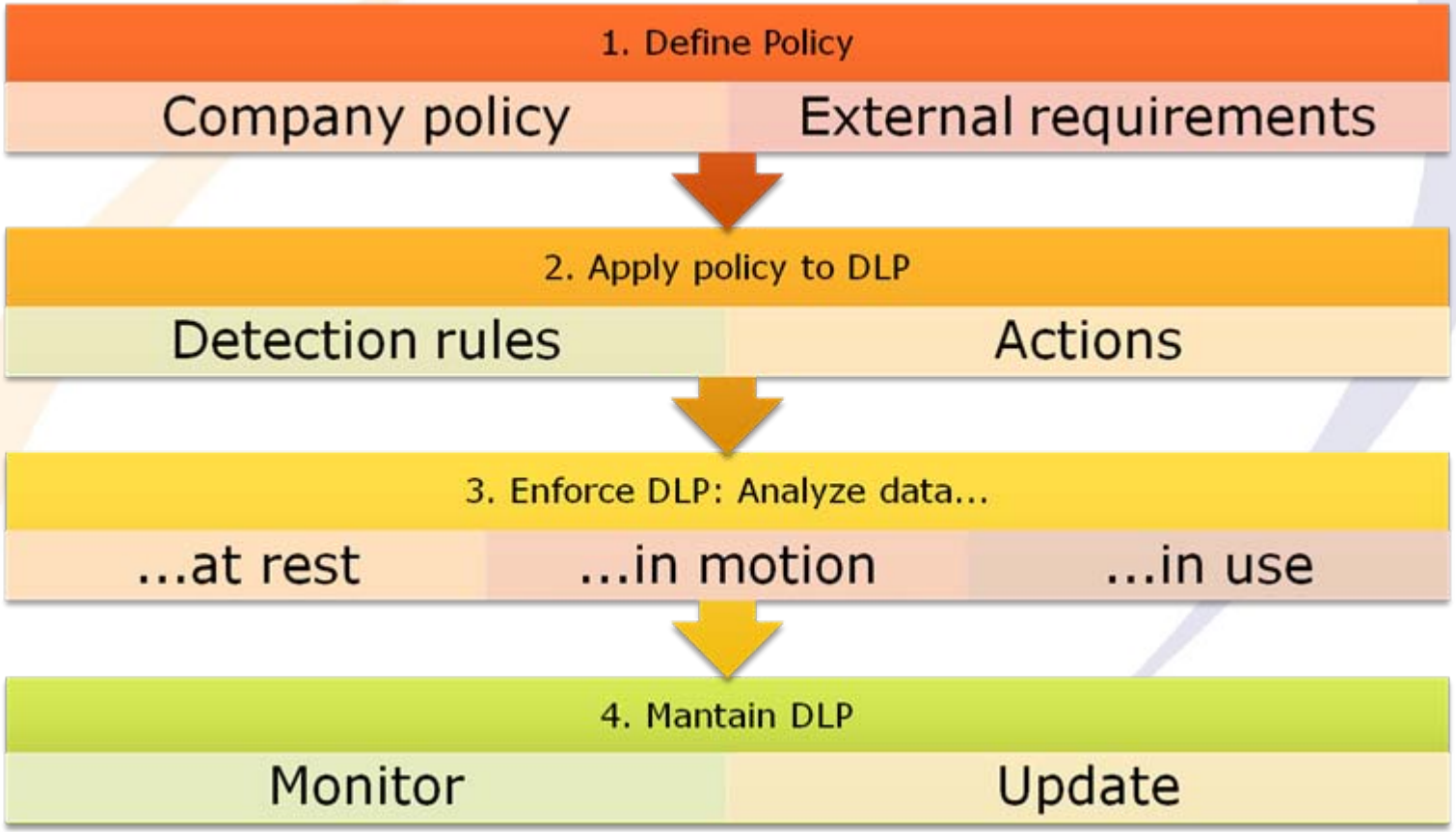
… and working DLP

- Computer and network security is a requirement
  - → Against the administrator there is little protection
    - » Possible, but VERY complex and expensive
      - – Two persons required, hardware security modules, or similar!
- It must be absolutely impossible to install any (!) software
  - → Otherwise it could be some kind of encrypting SW!
- Any used media must be disposed of securely
- Any transmission must be encrypted
- Storage must be physically secure or (better: and) encrypted
- All persons must be identified, authorized, …
- Problems:
  - → Huge effort required
  - → Difficult to cope with changes (new employees, changing area of work, teleworkers, mobile devices, …)
  - → False alarms or easy to subvert

How useful is it?

- If DLP is very strict and therefore working, it is very intrusive
  - → Privacy concerns must be addressed (→ work council)
    - » Everything is monitored!
  - → If it hinders the work, it will be circumvented
- Where does it work well?
  - → Military-/Public administration-type organizations
    - » Clearly defined hierarchy
    - » Clear division of duty, responsibilities, and permissions
    - » Effective measures to enforce guidelines
    - » Little importance for very quick and agile responses, but big focus on correct procedure
    - » Static organization/work → Time for **THE** solution
  - → Centralized data storage and processing
    - » One big file server/server + dumb terminals
  - → Not required to work perfectly, just somehow

Basic workflow



Enforcing happens where?

- Numerous different devices to support
  - → “Bring your own device” is especially problematic!
- All avenues for extrusion must be covered
  - → E.g.: SMS gateways, printers, automated paper mail systems, backup (tape stealing), …

Endpoints

Storage Servers

Internet Gateway

Additionally

Mobile

Print Server

Mail Server

DLP agents

- One important kind of DLP is agent-based
  - → On every client a „watchdog“ software must be installed
  - → Alternative: On server/firewall/listening in the network
    - » Problem: Things might already have happened; encryption
    - » Useable only as an additional layer or enhancement
- Tasks of these agents:
  - → Document what was done with data
  - → Informing users through pop-ups (sensibilization)
  - → Request user credentials and confirmation
  - → Identify all data and assign it to policy classes
  - → Block all access to and all activities with data which are not explicitly allowed in the policy
    - » Open file (embedding), write to new file, copy, move, drap&drop, cut&paste, screenshots, (simultaneously open text editor), …
  - → Send alarms on detected misuse or attempts
  - → Automatically encrypt data if policy requires this

Simple attacks (1)

- Use a custom encoding (“encryption” with your own alg.)
  - → Unknown → Cannot be scanned
  - → Note: This can be programmed by hand in short time and without compilers (need not be “unbreakable”)!
    - » Office Macro changing “e” and “n” to e.g. “q$2” and “*xx”
- Compress the file and give it a password
  - → Send it by mail/IM/on disk/…
    - » Many protocols (e.g. Skype) allow embedded file transmission!
  - → Scanning the content is impossible without breaking the PW!
    - » And preventing is hard (Skype) or undesirable (company use)
- Use “hidden” storage devices: USB sticks in toys, …
  - → Relatively simple to prevent through blocking all USB devices (or only mass storage devices)

Simple attacks

- Direct access to hardware: Remove disk from laptop
  - → Or replace it: Reinstall operating system
  - → Add new computer through your own switch/WLAN
  - → Note: This is probably suspicious activity if good security is already in place!
- Use password/device/laptop “borrowed” from co-worker
  - → Password: Found under keyboard or just ask for it
  - → Use company laptop for family as well
- Bring your own keylogger
  - → Will copy everything you type

Attacks: Social engineering

- Just ask for the password, e.g.:
  - → “Here is the administrator. Because of a security problem on the mailserver we have to inspect all mail accounts. You must give me your password, or your account will be locked.”
  - → Create a sweepstake site and ask for users to register
    - » The password will probably be used somewhere else too!
- Does work astonishingly well!
  - → Prerequisite: Know something about company/person
    - » E.g. full name, position, tasks: Can often be found on public web!
  - → Other variants: Phishing, dumpster diving
- Note: Technological measures will not help here!
  - → And “institutionalized mistrust” is economically bad
  - → Any export of data will be duly authorized, logged, etc.
    - » But still reach an unintended recipient!

Attacks: Techniques not reasonable to prevent

- Make films/photographs of the screen
  - → Special hardware preventing this might exist
  - → But: Screenshots **can** be prevented!
- Non-IT attacks:
  - → Bribing the person who has access to data or the one assigning permissions
    - » Can be made difficult, but this is very hard to prevent!
    - » This is something the very best DLP solutions can prevent!
  - → Memorising data: Typically rarely useful, except for methods
    - » Note: Every day a little → In sum interesting
  - → Writing it by hand on paper or dictating it into another device
    - » Requires a long time, easy to observe
  - → Looking at “foreign” screens (“shoulder surfing”)
    - » Special foils e.g. for laptop screens do exist

Content identification

- Keywords: Scanning the documents for keywords
  - → Requires lists (large part will be company dependent!)
  - → Easy to implement and fast
  - → Inefficient/impossible for data with certain structure, but no determined values (e.g. social security numbers)
- Regular expressions: Search for patterns
  - → Requires patterns (might be difficult to define)
    - » Note: Social security numbers, telephones, etc. are written differently depending on country, language, …
- Hashing: Documents known as secret are hashed
  - → Whenever something is sent/moved/… a hash value is calculated and compared to the list of sensitive files
  - → Automatic classification possible → Everything in this folder
    - » Adding a file automatically leads to addition of hash to DB

Content identification

- Partial hashing: Not only the whole file is hashed, but also parts of it (e.g. paragraphs)
  - → Small changes won't influence the decision
  - → Requires understanding the format (or not so useful)
  - → Threshold needed till when a document is "still the same"
- Machine learning: Different approaches similar to SPAM
  - → Difficult, false positives/negatives are more common
- Predefined categories
  - → Relies on explicit (manual) classification
    - » Take care: Who annotates this (deliberately marked as "free"!)?

Protection: Scanning E-Mails

- Happens on the mail gateway
- Requires unpacking all attachments
  - → This can be difficult and requires time and resources
    - » „ZIP-bombs“: Tiny files with huge compression ratio
  - → Requires understanding all relevant file formats and prohibiting anything else
    - » Just because it looks like a video it doesn’t need to be one: The complete file must be decoded and checked whether it actually shows something (and which should not be text!)
- Typical actions: Block, quarantine, encrypt
- Works quite well for unintentional disclosure, …
  - → Accidentally entered the wrong E-Mail address as recipient
  - → Mention a bit of information in a unrelated/private mail

… but works very bad regarding intentional subversion!

Protection: Scanning E-Mails

- Problems:
  - → Encryption: Must take place on the gateway or requires knowledge of all keys (!!!)
  - → Works moderately well for text, but is problematic for images
    - » Scans of paper → OCR, image recognition
  - → Mail server must be able to identify content as restricted
    - » Often: Dictionaries and RegExp (e.g. social security numbers)
  - → Interface for allowed E-Mails → Un-quarantining
  - → Language and character sets
    - » Professional software does cope with this!
- Example (Cisco IronPort Email DLP):
  - → Germany: The following numbers are detected:
    - » IBAN/BIC, Drivers License, National Identification, Passport
    - » Seems to be more a regulatory issue than real protection …

Protection: Restricting access to USB disks

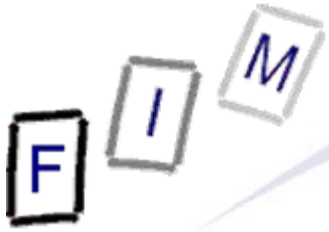
- Generally: Disable USB in BIOS and protect it by password
  - → Potential problem: USB keyboards/mice!
- Windows:
  - → Prevent USB storage devices to be installed
    - » Assign the users “Deny” permissions on
      - – %SystemRoot%\Inf\Usbstor.pnf
      - – %SystemRoot%\Inf\Usbstor.inf
  - → Prevent already installed USB storage devices from working
    - » HKLM\SYSTEM\CurrentControlSet\Services\UsbStor: Set “Start” to “4”
  - → Attention: This only blocks “storage devices” → Everything else (or other functionality of a device) will still work!
  - → Special software exists to only prevent writing
    - » Still a security issue, as data can be imported to the company!
      - – E.g. software, false data, …
  - → Special software also allows unblocking specific devices
    - » Based on their serial number (which is public)

Protection: Restricting access to USB disks

- Linux:
  - → Unload the USB storage module
    - » modprobe -r usb_storage
    - » Or: Delete these files
  - → And ensure it is not loaded again (→blacklist)
    - » E.g. through udev rules
      - – Allows also setting permissions, i.e. who is allowed to do this
    - » E.g. /etc/modprobe.conf → “alias usb_storage off”
  - → Debian/Ubuntu: Disabling in Grub possible
    - » Add “nousb” to boot parameters
- Important: Check which is working for your system first!

Conclusions

- DLP was a hype some years ago and still is to some degree
  - → What it can do is very limited, unless you invest a huge sum and effort and cope with a lot of restrictions!
  - → You need to have a very tight security before DLP can bring any real added value
    - » So better improve general security first!
- Works reasonably well regarding unintentional disclosure!
  - → Preventing data loss accidents of all kinds
- In some sectors (financial, health) legal regulations require the implementation of some kind of DLP
  - → There's no option then – But try to use a sensible amount
- Suggestion: Improve general security and try to improve employee security awareness (social engineering!)



Questions?

Thank you for your attention!

References

- How to disable USB sticks and limit access to USB storage devices on Windows systems http://diaryproducts.net/about/operating_systems/windows/disable_usb_sticks
- How can I prevent users from connecting to a USB storage device? http://support.microsoft.com/default.aspx?scid=kb;en-us;823732