Data Loss Prevention at Research Infrastructures

WISE@STFC 2018-02-28

Urpo Kaila, <urpo.kaila@csc.fi>

EUDAT receives funding from the European Union's Horizon 2020 programme - DG CONNECT e-Infrastructures. Contract No. 654065

What is Data Loss Prevention (DLP)?

- Data Loss Prevention is a set of measures to make sure that users do not send confidential information out from the organisation’s network

- DLP is also called Data Leak Prevention

- DLP is a classic but currently often overlooked security domain

- DLP is one of the fundamental security activities to protect confidentiality of data, systems and services

Why Data Loss Prevention?

- The objective of DLP is to identify and restrict leaks of non-public information, such as
  - Personally Identifiable Information (PII)
  - Confidential information
  - Internal information
- Disclosing non-public information inadvertently or through abuse can put organisations and people at risk
- Compliance requires to protect confidentiality
  - Personal data (GDPR!)
  - Sensitive personal data
  - Information restricted by non-disclosure agreements

General Data Protection Regulation and DLP

The General *Data Protection Regulation* requires to ensure that

- Staff is trained on data protection
- Personal data is used only for purposes it was collected
- The data processor implements appropriate technical and organisational data protection policies
- A code of conducts is defines a compliance mechanism
- Procedures to identify and report data breaches are in place
- Privacy by design is implemented

Available generic solutions for DLP

- There are commercially available technical solutions for DLP to identify and prevent leaks of confidential data
  - Information leak detection and prevention
  - Extrusion Detection Systems
- Solutions typically consist of host-based and/or network based sensors and agent software
- Can cover data-in-motion, data-in-use, and data-at-rest
- DLP solutions can cover both structured and unstructured data
- Rule-based detection and prevention algorithms available

DLP in Digital Research Infrastructures (DRI)

- Compared with commercial IT environments, where the format of data and the data transfer protocols are often well defined, e.g. the financial sector, the DLP within Digital Research infrastructures faces formidable challenges
  - This is due to the complexity of the data, the large variety of data formats and system platforms
  - The dynamic nature of research
  - Unclear or undefined responsibilities on accountability
  - The growing amount of PII in research environments that were originally designed for handling public scientific data but not for PII

DLP Solutions for Digital Research Infrastructures

- DRI's should start by identifying their needs for DLP
  - Should it be based on a centralised or a de-centralised model?
  - How to design a basic security architecture for DLP
    - Where to handle sensitive information
    - Where NOT to handle sensitive information
    - Access controls
    - How to trust internal and external providers
  - Jointly developed open source software and operated services can be better solutions to ensure proper DLP for a reasonable price

How to implement DLP for DRI

- Implementation of DLP requires
  - Identification of data to be protected
  - Guidelines and procedures to dynamically classify non-public data
  - Applying access controls and other security mechanisms, such as layered defence
  - Feasible (software) tools to identify and prevent data leaks
  - Procedures to contain DLP incidents
  - Awareness training and technical training on DLP
- Best practices for DLP should be shared between DRI's

How the Security Landscape differs in DRI’s?

How do we differ from ´normal´ organisations?

- Academic freedom
- Vague definitions in chain of command
- Non-monetary objectives
- Double standards in policies
- Thin/weak management culture
- Focus on personal responsibility
- Staff sometimes create their own IT solutions
- Difficult to apply shared ways of work
- Semi-autonomous institutions, personal NDA´s
- Research often a personal mission, not just work

Joint DLP framework for DRI's through WISE?

- Wise Information Security for e-Infrastructures (WISE) trust community is a global framework where security experts can share information on topics such as risk management or experiences about certification processes
- Current WISE activities:
  - Updating the SCI framework (SCI-WG)
  - Security Training and Awareness (STAA-WG)
  - Risk Assessment WISE (RAW-WG)
- WISE should take up a joint initiative on DLP as a special topic in forthcoming activities