Project Acronym: STORM CLOUDS

Grant Agreement number: 621089

Project Title: STORM CLOUDS – Surfing Towards the Opportunity of Real Migration to CLOUD-based public Services

Deliverable 4.1.2
Ethical issues and data protection report

Work Package: WP4
Version: 1.0
Date: 07/08/2015
Status: WP leader accepted
Nature: REPORT
Dissemination Level: PUBLIC

Editor:	Alkiviadis Giannakoulias (European Dynamics SA)
Authors:	Alkiviadis Giannakoulias (European Dynamics SA)
	Pedro Geirinhas (Alfamicro-Sistemas De Computadores Lda)
Reviewed by:	Eva García Muntión (Research, Technology Development and Innovation, S.L.)

Legal Notice and Disclaimer

This work was partially funded by the European Commission within the 7th Framework Program in the context of the CIP project STORM CLOUDS (Grant Agreement No. 621089). The views and conclusions contained here are those of the authors and should not be interpreted as necessarily representing the official policies or endorsements, either expressed or implied, of the STORM CLOUDS project or the European Commission. The European Commission is not liable for any use that may be made of the information contained therein.

The Members of the STORMS CLOUDS Consortium make no warranty of any kind with regard to this document, including, but not limited to, the implied warranties of merchantability and fitness for a particular purpose. The Members of the STORMS CLOUDS Consortium shall not be held liable for errors contained herein or direct, indirect, special, incidental or consequential damages in connection with the furnishing, performance, or use of this material.

Version Control

Modified by	Date	Version	Comments
Alkiviadis Giannakoulias	26/9/2014	0.1	Initial version
Alkiviadis Giannakoulias	17/10/2014	0.2	Updated version
Alkiviadis Giannakoulias	04/11/2014	0.3	Updated version
Alkiviadis Giannakoulias	22/12/2014	0.4	Updated to new document template
Alkiviadis Giannakoulias	05/02/2015	0.5	Added material on Identity Management, Future of Data Protection, updated VM backup, Removed VM Replication, Live Migration, PRIME,
Alkiviadis Giannakoulias	15/04/2015	0.6	Added Applying Encryption,
Alkiviadis Giannakoulias	08/05/2015	0.7	Added Certifications
Alkiviadis Giannakoulias	08/06/2015	0.8	Added DRM, Executive Summary, Summary and Conclusions
Pedro Geirinhas	17/07/2015	0.10	Added Ethics part
Alkiviadis Giannakoulias	29/07/2015	0.11	Removed Data Integrity Using Encryption Updated conclusions
Alkiviadis Giannakoulias	30/07/2015	0.12	Ready for Review
Eva García Muntión	31/07/2015	0.13	Reviewed
Alkiviadis Giannakoulias	03/08/2015	1.0	Ready for submission Removed: Encryption Key Management

Executive Summary

Work Package 4 (WP4) of the STORM CLOUDS project aims to address the issues of privacy and security, interoperability and multilingualism that affect the operation of all the services in the pilots.

The main focus of this document is to consider the ethical and data protection issues surrounding the requirements and the specification of cloud computing in general, with particular focus on the STORM CLOUDS Platform.

The content includes the ethical context of cloud computing, data protection including legislation along with an extensive list of with recommendations and best practices focusing on the range of elements involved in cloud computing.

Section 2 focuses on the ethical issues that arise when public sector organisations consider transitioning to cloud computing, including recommendations for ethical issues in cloud computing.

Section 3 focuses on data protection. Data protection is the process of safeguarding important information from corruption and/or loss. As more systems, applications and data are moved into cloud service provider (CSP) environments the possibility of a data breach increases, increasing the need to address the issues of data protection and other legal/ethical issues. Hence this document provides a level of detail and discussion on these areas.

It reviews the issues surrounding data protection today, considers why current technologies don't deliver what is required, and proposes a set of key recommendations and best practices including:

- Data Breach Management
- Software updates
- SQL injection prevention
- Decommissioning of unnecessary services
- Password storage
- Configuration of SSL and TLS
- Inappropriate locations for processing data
- Identity Management
- Digital Rights Management

together with strategies for achieving strong and flexible future data protection solutions.

Table of Contents

List of Figures

Abbreviations

Acronym	Description
API	Application Programming Interface
CC	Cloud Computing
CDP	Continuous Data Protection
CMDB	Configuration Management Database
COW	Copy-On-Write
CSP	Cloud Service Provider
DAC	Discretionary Access Control
DBMS	Data Base Management System
DLP	Data Lost Prevention
DPA	Data Protection Authority
DRM	Digital Rights Management
EC	Erasure Coding
HSM	Hardware Security Module
IaaS	Infrastructure as a Service
IBE	Identity-Based Encryption
IoT	Internet of Things
IT	Information Technology
ICT	Information Communication Technology
KMIP	Key Management Interoperability Protocol
MAC	Mandatory Access Control
NIST	National Institute of Standards and Technology
PaaS	Platform as a Service
PC	Personal Computer
PDP	Policy Decision Point
PEPs	Policy Enforcement Points
PAP	Policy Administration Point
ROW	Redirect-On-Write
RRI	Responsible Research and Innovation
RRI ICT	Responsible Research and Innovation in Information Communication Technologies
RTO	Recovery Time Objective
R&I	Research & Innovation
SaaS	Software as a Service
SOA	Service oriented Architecture
SSL	Secure Socket Layer
TCP-IP	Transmission Control Protocol – Internet Protocol

	(suite)
TLS	Transport Layer Security
VHD	Virtual Hard Disk
VMDK	VMware Virtual Machine Disk File
WLAN	Wireless Local Area Network

1 Cloud Computing

"Cloud computing is an evolving paradigm." [2]

The National Institute of Standards and Technology (NIST) in September 2011 released a Special Publication SP 800-145, in which it defined cloud computing as:

Cloud computing is a model for enabling ubiquitous, convenient, on-demand network access to a shared pool of configurable computing resources (e.g., networks, servers, storage, applications, and services) that can be rapidly provisioned and released with minimal management effort or service provider interaction. This cloud model is composed of five essential characteristics, three service models, and four deployment models. [2]

Cloud computing intends to provide a baseline for discussion from what is cloud computing to how to best use cloud computing. [2]

Cloud computing is far more dynamic than traditional data processing. The location where data processing takes place can change dramatically. The current location of data and where it is processed can depend on a variety of factors to which end users and data controllers traditionally have given little thought, and into which they do not necessarily have the insight or ability to control. [6]

Cloud computing differs from conventional computing in a number of ways. As Figure 1-1 illustrates there are five characteristics of cloud computing and are defined as follows:

- **On-demand self-service.** Ability for users to request and use resources as needed;
- **Broad network access.** Network accessibility from various types of hardware and software;
- **Resource pooling.** Sharing of cloud resources by several services, and allocation as needed to meet changing demands;
- **Rapid elasticity.** Expansion and reduction to meet varying resources demand;
- **Measured service.** Charge according to metered resource usage.

On-Demand Self-Service
Broad Network Access
Resource Pooling
Rapid Elasticity
Measured Service

Figure 1-1: The five characteristics of cloud computing [17]

Furthermore, NIST identifies a simple and unambiguous taxonomy of three "service models" available to cloud Consumers (Infrastructure-as-a-Service (IaaS), Platform-as-a Service (PaaS), Software-as-a-Service (SaaS)) and four "cloud deployment modes" (Public, Private, Community, and Hybrid) that together categorize ways to deliver cloud services.

2 Ethical Issues

2.1 Definition

Ethics is the branch of philosophy that involves systematizing, defending, and recommending concepts of right and wrong conduct. [7] Ethics seeks to resolve questions of human morality, by defining concepts such as good and evil, right and wrong, virtue and vice, justice and crime.

In practice, ethics refers to those standards that impose the reasonable obligations to refrain from rape, stealing, murder, assault, slander, and fraud. Ethical standards also include those that enjoin virtues of honesty, compassion, and loyalty. And, ethical standards include standards relating to rights, such as the right to life, the right to freedom from injury, and the right to privacy. Such standards are adequate standards of ethics because they are supported by consistent and well-founded reasons.

Moreover, ethics refers to the study and development of one's ethical standards. As mentioned above, feelings, laws, and social norms can deviate from what is ethical. So it is necessary to constantly examine one's standards to ensure that they are reasonable and well-founded. Ethics also means, then, the continuous effort of studying our own moral beliefs and our moral conduct, and striving to ensure that we, and the institutions we help to shape, live up to standards that are reasonable and solidly-based. [7]

That being said, when moving from traditional servers to a cloud paradigm, the technological foundations change as well as the implication regarding Ethical issues. The ones identified throughout this chapter are based upon Timmermans' research [9][7] and seek to identify those issues that arise from the nature of technology itself rather than specific circumstances. Indeed, they are derived from three underlying technological developments generated by cloud computing:

- The shift of control from technology users (traditional servers) to third party servicing (cloud providers);
- Storage of data from multiple physical locations to worldwide servers from several distinct organizations;
- Interconnection of multiple services across the cloud.

Through identification of ethical issues arising from these new functionalities, it is possible to inform and raise awareness to vendors, users or system designers of ethical questions, in order for them to be proactive in assessing their role in specific implementations and uses. The main challenges from ethical issues are:

2.2 Challenges

2.2.1 Control

According Haeberlen [7] and Kandukuri & Rakshit [7], Cloud computing entails the outsourcing of Information Communication Technologies (ICT) tasks to third party service providers. As such, information that once used to be stored in local premises are now stored in the cloud. Users therefore place data on machines that are not directly controllable and therefore renounce control these resources and data.

As mentioned by Paquette [12] risks associated with this change of control in cloud computing mainly rely in data corruption, infrastructure or system architecture failure or unavailability/outing and unauthorized access by third-parties. Ethical problems arise in times of disaster or simply if something goes wrong. In fact, it is hard to distinguish the entity that has originated the problem, to the point that, as mentioned by Haeberlen [10], it is almost impossible to hold someone accountable and responsible for a problem in a dispute, when lacking strong supporting evidence. In addition, the de-parameterisation[1] shadows the border of organizations IT infrastructure and consequently disguises their accountability.

As an example, in an unfortunate circumstance where a cloud provider has poorly designed network architecture with a single point of failure that brings the system down. Even though the company has the ethical duty to be honest with affected customers, in a networked organizational and technological structure it becomes increasingly difficult to accredit consequences of actions to a single person or organization [9].

2.2.2 Responsibility

As cited by Haeberlen [10], since responsibilities are divided between customer and provider of the service, neither is in position to address these problems.

In cloud computing a service delivered to a user depends on another system that also depends on other systems. A cloud service to the end-users may use service-oriented architecture (SOA) where functionalities aggregate services into larger applications. Once again, ethical problems arise in times of disaster or simply if something goes wrong. By having a highly multifaceted structure of cloud services, it is most certainly difficult to determine who is responsible in case of an undesirable event. This lead to a severe ethical problem called the “Problems of many hands”, that dictates that in a complex chain of systems where people a share in an action that leads to undesirable consequences, many people have also had the opportunity to prevent these consequences, and therefore no-one can be held responsible. [13]

2.2.3 Accountability

According to Pieters [7] in a de-parameterised world the border of an organization accountability blurs and becomes less evident. Personal data stored in the cloud should be managed accordingly, as not doing so would not be ethical by all persons involved in that process. Users of cloud should be empowered by being able to check whether the cloud is performing the agreed the provision of accountability transparency and clear allocation of responsibility, as when recorded, these elements can be used to decide who is responsible whenever a problem occurs or dispute arises.

As already identified in D4.1.1, in 2010, the Article 29 Data Protection Working Party issued an Opinion on the principle of accountability in which it elaborated upon the possibility of including a general provision on accountability in the revised Data Protection Directive. Accountability is a concept with many different dimensions, but in its core meaning, accountability refers to the existence of a relationship whereby one entity has the ability to call upon another entity and demand an explanation and/or justification for its conduct. [79]

[1] “the disappearing of boundaries between systems and organisations, which are becoming connected and fragmented at the same time” [7]

Within the STORM CLOUDS project, municipality of Agueda reported that both for the 'reactive' approach[2] and the 'proactive' approach[3] all complaints are received and managed by the Quality Service, according to the ISO 27001 certification they acquired. Similarly any question regarding data protection or the accountability of actions related with information policy and security are forwarded to the internal Quality Service.

2.2.4 Ownership

The storing of data in different location premises also raises the question of who owns the data a user stores in the cloud. By doing so, the IT admins, engineers, and troubleshooting agents of a provider of cloud services all have access to this information. [14] Moreover, the cloud also generates data itself for different purposes, such as providing accountability, improving services provided, or security performance or security. Digital interactions and tracks are thus being gathered together through unique identifiers and algorithms, which leaves a trail of personal information. There is an ethical duty to not access this information with harmful intent or reckless behaviour, either by providers or third-parties such as hackers (fraudulent use), or it may be accessed and used in ways that individuals did not envisioned.

Also, information stored with a third party can be of easy access to Government agencies and private litigants more easily than from the original owner or creator of the content. This causes a severe ethical issue has to whether it righteous or not to do so, even by Public Authorities figures.

Ownership problems also incur in situations related with infringements on copyrights, since access to massive computing storage, cloud services might facilitate sharing copyrighted material. [15]

2.2.5 Lock-in

According to Nelson [15], if only a limited number of companies are able to achieve a dominant position in the market for cloud services due to economies of scale, this might lead to abuse user needs. Users would become dependent on certain cloud service providers, be it infrastructural or intermediaries. Several ethical risks might exist from these unwanted dependencies on cloud service providers and vendor lock-ins. With little emphasis on interfaces that guarantee data and service portability users may face difficulties migrating from one provider to another or to migrate their data and services back to an in-house IT environment. Similarly, if a service provider ends its operation in the market, not along the data privacy that will be mentioned at a later stage, the possibility to migrate data must be possible. Ethically, such concerns are of vital importance and must be tackled in order to introduce independence from a particular cloud providers and vice versa.

2.2.6 Legal

Providers also need to take into account the laws a specific country follows in terms of data privacy. It is ethically correct to respect customers' laws and companies should might store data in

[2] Which actors are held accountable once their activities have resulted in a complaint

[3] Which actors may called upon to demonstrate their compliance without being alleged of a violation of data protection principle

jurisdictions that may not respect the rights of their users and customers. Favourable privacy laws represent important challenges that need to be faced ethically.

2.2.7 Privacy

As stated above, many companies providing cloud services collect data, much of it consists of sensitive personal information, which is then stored in data centres in countries around the world. Whenever ethical issues arise concerning information about persons they are typically cast in terms of privacy. [16] Privacy aims to constrain access to certain types of personal data and prevent persons to acquire and use information about other persons. Consumers need to trust their cloud provider that certain personal information will not be exposed, as according to their terms that have been previously accepted by the users.

2.3 Recommendations

Several effects and undesirable consequences of cloud computing are still hard to identify. Its complexity and current developments do not alleviate the task. However, this does not mean that the conception of ethical issues in cloud computing and subsequent implementation should be disregarded. Au contraire, they should be engaged since the beginning in order to avoid them at a later stage. In this chapter 2.3, three recommendations for ethical issues in cloud computing are listed.

2.3.1 Pro-activity

It is urgent that all parties involved in cloud computing are proactive, in order to anticipate unforeseeable consequences. As seen in the previous chapter, due to all ethical and uncertainty risks, not doing will only injure and burden possible evidence that should have been collected by them. In particular, these players may never use uncertainty to refrain from designing and providing services that invite moral sound use and inhibit undesirable or controversial actions, which is also argued by Pieters. [13] It is thus recommended as ethical for Cloud providers to have a Terms and Conditions available and for users to know Terms and Conditions of providers.

2.3.2 Regulations & Policies

In addition all technology should be subject to regulation arrangements at least just enough to have innovation leading towards the benefit of society and not enough to have it limit innovation. In any case, regulations can have ethics integrated into technological development and use. It is vital that governance arrangements are more conducive to the inclusion of ethics, including regulations for private companies, which are usually much less subject to ethics-related oversight and more towards profit generation. [10] Such regulations will adapt as cloud computing evolves, similar to what happened with labour law year ago. In the latter case, it is important to remember the core definition of corporate responsibility and follow policies defined by the European Union, such as the ISO26000.

2.3.3 Responsible Research and Innovation

The question remains of how ethics can be incorporated into cloud computing that is comprised of such different structures and processes. The answer might rely on the recognition of the relevance of ethics in technology and having all actors and stakeholders involved in the different stages of technology development agreeing to use ethics. This leads us to Responsible Research and Innovation in ICT (RRI-ICT).

In May 2010, the European Union (EU) has launched its Digital Agenda for Europe flagship initiative, aiming at "rebooting Europe's economy and helping Europe's citizens and businesses to get the most out of digital technologies". It is then necessary to thoroughly explore the two-way interactions between technology and society, and, putting the human at the centre of the analysis, to explore how the digital age can be a true success factor not only for EU's competitiveness but also for EU's values in the DAE area.

In this context, Responsible Research and Innovation (RRI) has a particular importance since it can be defined as an inclusive approach to Research & Innovation (R&I), aiming at better aligning both the process and outcomes of R&I with the values, needs, and expectations of the society, notably through reinforcing public engagement, open access, gender dimension, ethical issues, and (formal and informal science) education.

The contribution of Social Sciences and Humanities (SSH) to a RRI approach in ICT is logically critical since SSH can precisely monitor economic, legal, and social issues related to technological developments and update the concepts, meanings, and expectations arising from the deployment of ICT. The need for an RRI approach had been identified by European Commission's DG Research & Innovation in the early 00's, through its "Science and Society Action Plan" encouraging a better connection between science and European citizens. Today, RRI has become a cross-cutting issue in the EU Horizon 2020 framework programme for research and innovation covering the period 2014-2020.

For further details on the approach of European Commission's DG Connect (in charge of the Digital Agenda for Europe) to go for RRI and SSH in ICT-related parts of the first work programme of Horizon 2020 (for 2014-2015) go to http://ec.europa.eu/digital-agenda/en/news/how-go-about-responsible-research-and-innovation-and-social-sciences-and-humanities-ict-related.

It is recommended to register in the RRI-ICT Forum in order to integrate and participate in RRI. The STORM CLOUDS project is represented in the RRI-ICT Forum by Alfamicro.

3 Data Protection

The seventh data protection principle states: “Appropriate technical and organisational measures shall be taken against unauthorised or unlawful processing of personal data and against accidental loss or destruction of, or damage to, personal data.” [66]

3.1 Definition

Data protection is the level of availability or confidence in being able to access important data. There are many options to ensure data protection. These options often come in the form of RAID, data replication, and data archiving. Depending on the data itself, companies may replay on any number of these methods. Data that is changed frequently may be stored at certain intervals using SAN replays, a form of incremental backups. [51]

Old data is often stored as archive data so that it may be referred to if and when it is ever needed. It's heavily compressed using a data de-duplication algorithm. It may be stored on any available storage medium. [51]

Data protection is the process of safeguarding important information from corruption and/or loss. [38]

The term data protection is used to describe both operational backup of data and disaster recovery/business continuity (BC/DR). A data protection strategy should include data lifecycle management (DLM), a process that automates the movement of critical data to online and offline storage and information lifecycle management (ILM), a comprehensive strategy for valuing, cataloging and protecting information assets from application/user errors, malware/virus attacks, machine failure or facility outages/disruptions. [38]

The interpretation of the seventh data protection principle (Part II of Schedule 1 of the DPA) states: “Having regard to the state of technological development and the cost of implementing any measures, the measures must ensure a level of security appropriate to: (a) the harm that might result from such unauthorised or unlawful processing or accidental loss, destruction or damage as are mentioned in the seventh principle, and. (b) the nature of the data to be protected.” [4]

This means that appropriate security measures will vary from case to case, depending on the circumstances.

According to [52] three events caused the development of early data protection mechanisms:

1. The development of the Personal Computer, that resulted in data dispersion on small devices that were susceptible to theft;
2. The commercialization of the Internet, resulting in financial transactions usage and transmission of intellectual property;
3. The development of public key cryptography and the development of technologies to protect data or rather provide enough confidence in its protection.

[4] http://www.legislation.gov.uk/ukpga/1998/29/schedule/1/part/II

3.2 Cloud-Specific Security Threats

In addition to the generic information security threats to **confidentiality**, **integrity** and **availability**, cloud-specific threats are influenced by many factors, including the chosen service and deployment models. Seven significant threats are proposed in [17]:

1. **Data security**. Data placed on the cloud, presents limited control because data storage, transmission and copying are managed by the cloud provider, especially in public cloud service models where the cloud is outside the organisation's direct control.
2. **Cloud management**. Cloud management interfaces can be vulnerable to attacks, resulting in data loss, denial of service and financial costs. This is true especially in public cloud service models because they can be accessed by a wide user base.
3. **Cloud provider insiders**. Cloud provider employees with malicious intent can access and manipulate customer data.
4. **Unknown risk profile**. Unclear (if any) security measures implemented by the cloud provider to protect customer data can lead to data loss.
5. **On-demand self-service**. The unlimited computing power, especially relevant in the IaaS model, given to users can potentially allow them to perform large scale attacks.
6. **Forensics**. Clouds present inherent difficulties in collecting sufficient forensic and auditing information. This is the case with public clouds because they are external to the consumer organisation.
7. **Multi-tenancy**. Shared infrastructure resources amongst various tenants, can lead to vulnerabilities, where a tenant might gain access to another tenant's data. This threat is prominent in IaaS and PaaS models, because of user involvement in low-level management.

3.3 Challenges

As more systems, applications and data are moved into cloud service provider (CSP) environments the possibility of a data breach increases.

Multi-tenancy, a condition where numerous virtual machines are housed on a single physical system, can lead to vulnerabilities, since a tenant might gain access to another tenant's data. This threat is prominent in IaaS and PaaS models, because of user involvement in low-level management.

This multi-tenant environment poses the following challenges:

- **Policy**. Ensuring that proper security policy is applied to sensitive data, systems, and applications is not feasible into a multi-tenant environment such as the cloud. Additionally when migrating systems and applications to a CSP, it might be difficult ensuring internal policies related to data handling and access control.
- **Encryption**. Implementing encryption in the cloud can be challenging due to key management and maintenance, performance issues, and access controls. [39] See 3.6.14
- **Data Loss Prevention (DLP)**. According to [39] to be effective a number of distinct technologies and processes are required:
  - Sensitive data needs to be identified so DLP monitoring tools can recognize them.

  - A centralized policy creation and implementation infrastructure needs to be in place to push policy to DLP monitoring tools, and these monitoring tools need to be in place to inspect traffic on network segments and critical host systems alike.
  - Quarantine and response measures should be implemented to take a variety of actions when a potential policy violation is detected.
- **Monitoring**. Techniques such as intrusion detection, network flow analysis tools, and host-based agents might not be allowed or supported by a CSP, although they might offer it as-a-service. [39]

Another critical element of data protection in the cloud involves the **data lifecycle**. [39] A reasonable lifecycle approach should include the following:

- **Retention**. CSPs should state how long they retain data that relates to customer instances and applications, such as information that potentially contains sensitive details about customer activities. In the first version of the deliverable under section 4.4.1.4 we suggested what the appropriate personal data retention period should be.
- **Disposal**. CSPs should be contractually forced to state that disposal of data is done in a secure manner. The deletion of personal data is an important activity in data protection, given Article 6(e) requirements. In the first version of the deliverable under section 4.5.12 we presented appropriate personal data destruction methods.
- **Classification**. Data segmentation can be used for highly sensitive data. CSPs can offer virtual private clouds or standalone cloud servers.

Since the arrival of cryptography in the late 1990s, there has been little progress in data protection. In the following sub-sections we will take a look at some of the reasons for this lack of progress.

3.3.1 Vendor Recalcitrance

IT and IT security vendors find it easier and cheaper to continue selling existing capability as long as possible, and when changes are required, it is also easier and cheaper to slowly evolve them. [52] This is one of the reasons for the slow adoption of security technologies such as PKI, DNSSEC, secure routing protocols, and the revised network protocol, IPv6.

3.3.2 Application Data Binding

In many large/complex applications, the data is so tied to a specific application that its uses are completely controlled by the application. This means that without support from the application it is difficult to apply standards-based, interoperable data protection. Even some modern operating systems have no concept of user-accessible files independent of their supporting applications (Apple's IOs). This is understandable because vendors will need to continue to include internal security capability as long as they have customers who do not have centralized or enterprise-level authentication and authorization services.

3.3.3 Identity and Access Control Standards

The lack of good quality, pervasive identity standards and common access control mechanisms provides another challenge for information protection. [52]

3.3.4 Metadata

Metadata is required in order to support automated access decisions. A policy decision system needs a deterministic way of finding information, needed by an access decision system, and the solution is to attach this metadata either directly to the data itself or by a link to the file.

The most limiting factor is the creation and management of this metadata. [52] Ideally it would be done without human intervention and with the usage of standard profiles of metadata.

3.3.5 The Human Policy Decision Point

Policy creation needs to be distinct from policy decision. [52] Unfortunately this is still perceived as losing control. Once this is addressed it would be easier to accept a human role as a policy.

3.3.6 Over-Connected

In his latest book OVERconnected, author William H. Davidow describes that being connected has made us more efficient, but there is now the risk of reacting so quickly that we don't give the thought we might have given to data protection, even twenty years ago. Looking at the 2008 financial crisis, Davidow writes:

'It is impossible to really understand what went on in the worldwide economic crisis of 2008 without examining the role that the Internet played in supercharging it. Without the Internet, the credit mess would have undoubtedly caused a recession of some magnitude. While we can never measure the Internet's full effects, we know that it made the current crisis larger, more widespread, and more virulent. It not only carried the information, it helped spread what is known as a "thought contagion." That is, the rate at which greed and fear mongering took place – via instant access to news and online rumours – was accelerated to unprecedented levels.'[5]

3.3.7 Internet of Things (IoT)

According to ABI Research more than 30 billion devices will be wirelessly connected to the Internet of Things (Internet of Everything) by 2020[6]

3.3.8 Unauthorized Data Mining

Programs such as PRISM, are used by organizations, both commercial and government, collecting vast amounts of electronic data.

3.3.9 Data Contextualization

Despite of data being redacted and aggregated, there is still enough information in data warehouses to enable a person to reasonably guess the source of the data. We need appropriate data protection enforcement controls close to the data to protect against re-identification i.e. the ability for someone to reverse-engineer the context, to determine the source.

[5] http://www.inc.com/articles/2010/12/book-review-overconnected.html

[6] https://www.abiresearch.com/press/more-than-30-billion-devices-will-wirelessly-conne

3.4 Protection Issues with Current Technologies

Current information security technologies and practices neither adequately secure information nor allow it to be shared when needed, since they were designed for an earlier era. The result is that we use some combination of the current data protection techniques that come with their associated issues, such as lack of granularity and absence of a comprehensive data protection plan, which is usually too late, costly, and does not scale.

Network-level protection, such as firewalls, routers encrypting traffic and network-based access control systems, can protect large segments of our intranet, including the data that resides therein. Although this low cost solution offers scalability, and manageability it falls out of favour for several reasons, including:

- A network breach can expose all of the data and resources due to absence of granularity;
- Limited granularity in the amount of protection – it's either off or on.

Operating system protection, such as file permission settings, Microsoft® Active Directory account settings, or global disk encryption, although they provide more local control, are not granular enough to protect data. Similarly to network-based controls when breaches occur they tend to expose all data under control of the operating system or application.

The most common tools for directly protecting data are currently limited to file encryption. However encryption is either on or off and once data is decrypted there are no controls to further limit its use, duplication, or redistribution. Other tools include:

- Discretionary Access Control (DAC) systems, whereby data protection is controlled by written policies (guidelines) issued to users, with the burden of selecting and applying the correct level of protection falling directly on the user;
- Mandatory Access Control (MAC) systems, whereby data protection is controlled by systems that use machine-readable policies, information about the user, the data, the environment, and the desired action. This information is then utilized by the systems to render decisions.

Enduring data protection requires two key changes to current approaches: [52]

- Protection must move closer to the data itself;
- Data protection must be automatically applied using policy-driven protection services.

3.4.1 Protection is too Global and Remote

Most data protection is enforced by network devices, operating systems, applications or DBMS that their primary goal is not security. Their inclusion as part of the enforcement mechanism for protecting data potentially broadens the attack surface significantly.

Another issue with using global protection mechanisms is that they encompass too much data, resulting in large quantities of data being exposed, in case of security breach.

Even specialized security systems, as they add functionality, require many more components as they add capability. This is the case with a system involving the use of secret key algorithm for encryption as shown below:

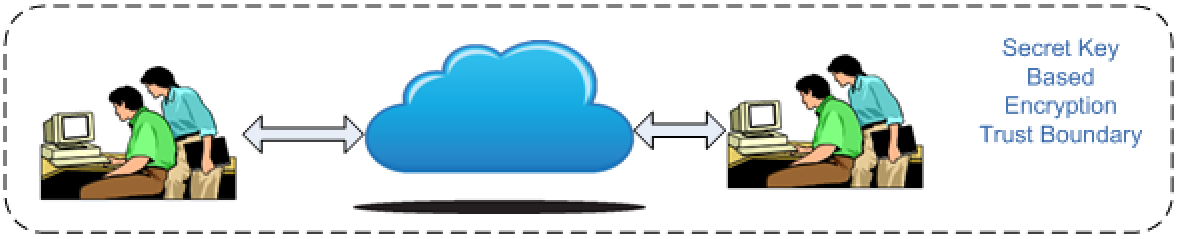


Figure 3-1: Secret Key Encryption Trust Boundary

However, secret key encryption has key distribution issues and is generally point-to-point, the result being lack of scalability and good protection for the secret keys involved. On the other hand public key encryption offers more data protection services, scales better, and allows ad hoc communication. But it does this by adding more devices and services inside the trust boundary, making it easier to circumvent the encryption without actually breaking the algorithms themselves. A simple expansion to our model looks like this: [52]

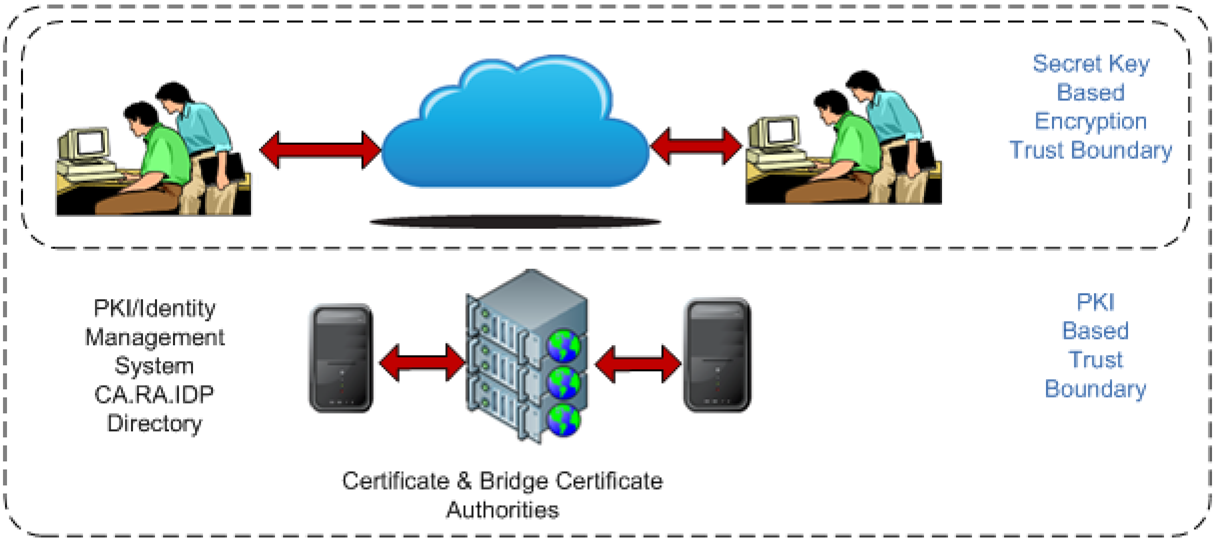


Figure 3-2: PKI Trust Boundary

As the following figure illustrates PKI requires directories, certificate servers, registration authorities, and other components. Specialized data protection tools and services, such as DRM and DLP, add in even more devices and services. [52] The result is a system where the strong protection provided by the cryptography can be nullified by attacks against the other components.

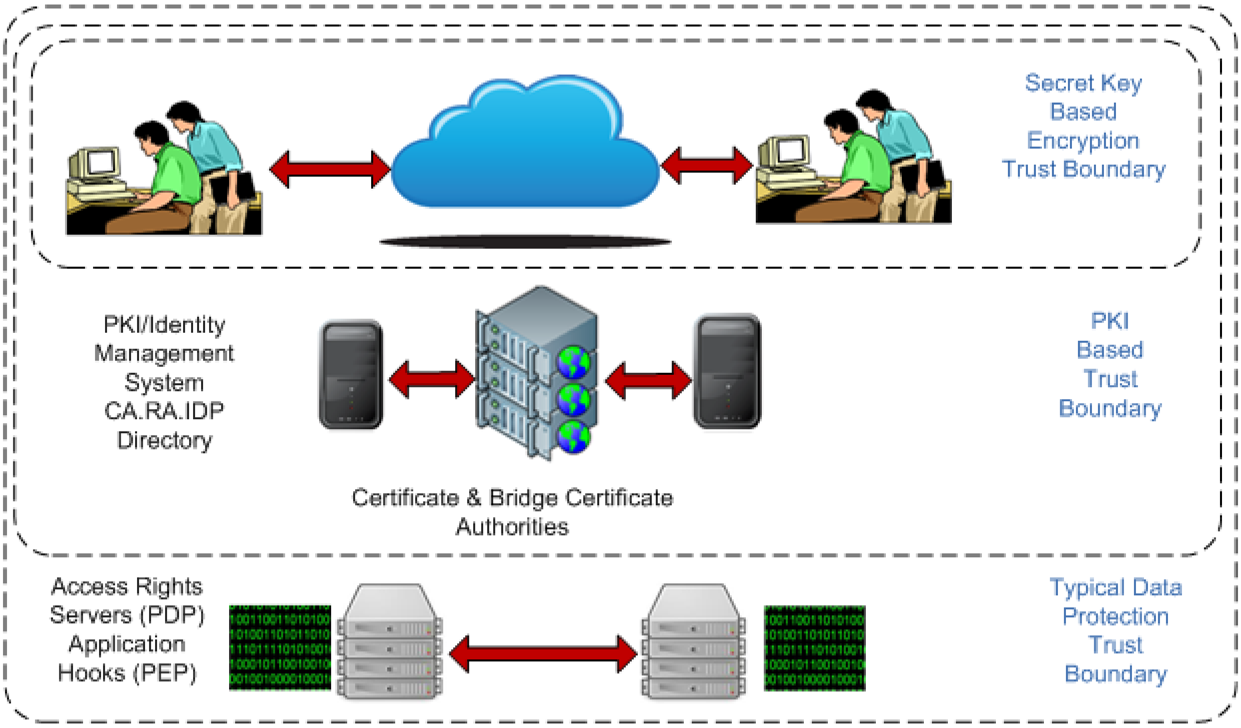

Secret Key Based Encryption Trust Boundary
PKI/Identity Management System CA.RA.IDP Directory
PKI Based Trust Boundary
Certificate & Bridge Certificate Authorities
Access Rights Servers (PDP) Application Hooks (PEP)
Typical Data Protection Trust Boundary

Figure 3-3: Typical DRM Trust Boundary

3.4.2 Protection is neither Granular nor Efficient Enough

Encryption is the only portable, interoperable technology that can be applied to data independent of its location or environment. However, encryption lacks persistence. It is either applied or not and the data being protected has no influence on its application. [52]

However, current Internet-based uses for data do not match the protection model of encryption.

Control is another issue, especially when data is being shared with or created and owned by multiple individuals/organizations. Many current information protection technologies are built on the assumption that only one entity will control access polices and enforcement mechanisms. Viable protection mechanisms need some form of negotiation between the entities involved that enables them each to adjust their risk or level of comfort appropriately. [52]

3.4.3 Protection is not Integrated with Centralized Authorization Services

Centralized authorization services require that the functions making the decision, namely Policy Decision Point (PDP), need to be separated from the enforcing functions, namely Policy Enforcement Points (PEPs). Here, data should be treated as any other resource in our cloud environment, and protection of that data should be driven by an enhanced set of authorization decisions based on not just user identity, but the relevant identity and attributes of all entities[7] involved in the transaction chain. According to [52] Current encryption and DRM-based technologies lack the connections to be driven from centralized PDPs or decision services and operate either as standalone services or as a feature of an application, such as email, document management, or a document office suite.

[7] According to [50] the are five entity types: People, Devices, Organizations, Code, and Agents

3.4.4 Protection is enforced by Weak Security Services

As a general principle we can state that “weak things should not be used to protect stronger things”. When using encryption the biggest vulnerability key protection using applications or operating system components.

For example current encryption and rights management products leverage centralized identity management systems which are based on centralized services and directories. These, in turn, are running on servers that run operating systems. The attack surface has now been expanded from the cryptographic algorithms to a set of servers running applications. We can now bypass the cryptography and instead obtain the identity of someone with access to the data by leveraging vulnerabilities in the identity management system or the platforms on which it resides. [52]

3.5 Access Control

Before designing a data protection system, it is important to provide some background on access control, the differences between access control and related security technologies, and the architectural options for deploying access controls.

As Figure 3-4 illustrates, at its simplest, access control is just the insertion of a guard between the user and the desired resource that mediates specific actions. In the IT world this is presented in the form of a simple access control structure embedded in either the operating system or an application. For example in UNIX there is a differentiation between individuals and groups, allowing discrete actions (read, write, execute, and navigate) on data or program execution. Data access control decisions are made by users setting a group of permission bits associated with a file and the operating system enforcing those permissions when access is requested.

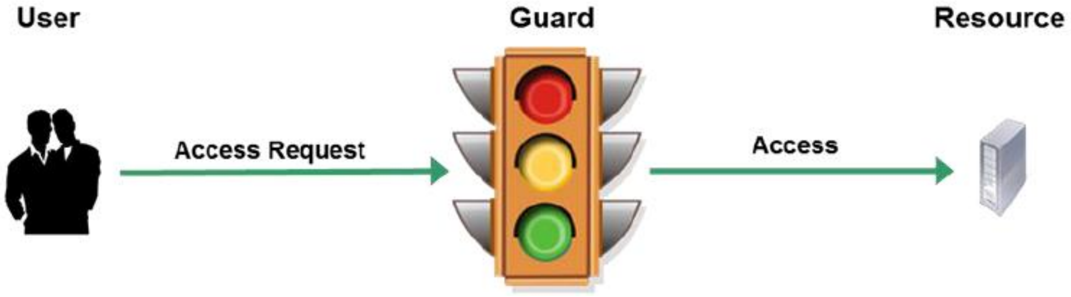


Figure 3-4: A Simple Access Control Model [52]

However, it’s better to separate the enforcement functions from the decision functions that the guard performs. Decision should be managed centrally and enforcement closer to the resource, allowing for better scalability and flexibility when designing systems. Enforcement functions are high performance and designed for volume, while decision functions are more intelligent requiring much less bandwidth since they just respond to requests from the enforcement functions.

As Figure 3-5 illustrates, in a Discretionary Access Control (DAC) system, the decision function is performed by a person and the operating system or application enforces that decision. Nevertheless, humans are not very good access control decision-makers, prone to mistakes leading to security breaches.

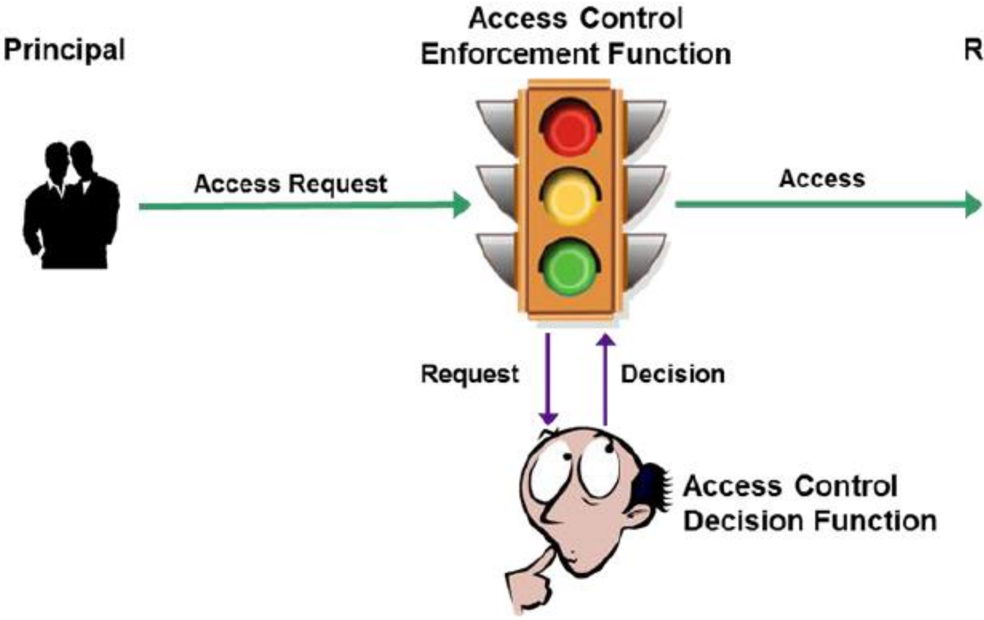

Principal
Access Control Enforcement Function
Resource
Access Request
Access
Request
Decision
Access Control Decision Function

Figure 3-5: A Discretionary Access Control Model [52]

As shown in Figure 3-6, in a Mandatory Access Control (MAC) system these decisions are made by the security system itself. In a MAC system people create a set of machine-readable access control policies, while the system makes individual access control decisions based on those policies. According to [52] MAC systems have a lot of potential advantages:

- Humans are excluded from the day-to-day decision-making process, thus improving accuracy and reliability;
- Can scale both in complexity and volume;
- Harder to compromise, since the policy administrator and the user are two different roles;
- React faster to policy changes or new attacks.

Unfortunately the technology to build MAC systems has been limited and expensive, resulting in being used only to the most critical information, such as national security controlled information. Only recently technology has become available making feasible affordable deployment.

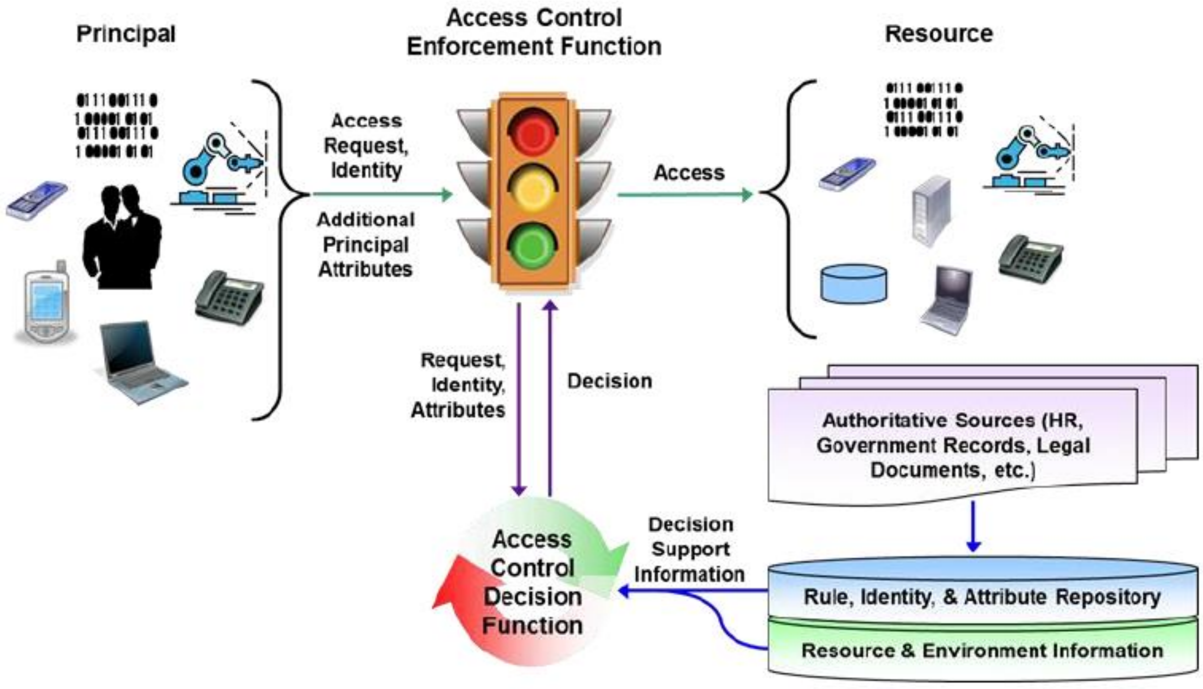

Principal
Access Control Enforcement Function
Resource
Access Request, Identity
Additional Principal Attributes
Access
Request, Identity, Attributes
Decision
Authoritative Sources (HR, Government Records, Legal Documents, etc.)
Access Control Decision Function
Decision Support Information
Rule, Identity, & Attribute Repository
Resource & Environment Information

Figure 3-6: A Mandatory Access Control Model [52]

As shown in Figure 3-6, the major components of a MAC access control system are:

1. **Principal**: A generally accepted industry term to describe any entity whose identity can be authenticated – and therefore can legitimately access resources (also known as “actors” and “initiators”). As shown the principal feeds with information (referred to as a set of attributes) the access control system, used in making access control decisions.
2. **Resource**: It includes computers, industrial control systems, telephones, mobile devices, etc. as well as data. A number of metadata, about the data (or other resources), are then feed to the access control system in order to make the correct decision.
3. **Enforcement Function**: Also referred to as Policy Enforcement Point (PEP). They differ depending on the layer they operate at. The enforcement function receives access requests from users and then forwards these requests to the decision function. Depending on the response it either allows the user to access the resource or not.
4. **Decision Function**: Also referred to as Policy Decision Point (PDP), is the brains of the access control system. It receives requests from the enforcement function and evaluates the request against a set of rules (or policy), metadata about the request object, user and environment, and information about the nature of the request (read, write, delete, etc.). The response is then sent back to the enforcement function.
5. **Administration**: Also referred to as Policy Administration Point (PAP), is the administration function (human-accessible GUI communicating with the decision function) that allows translation of access policies, entered by a person, into machine language that the decision function understands.
6. **Metadata**: Information consumed by the decision function, roughly grouped into four categories [52]:
   - Information attributes about the resource – in this case metadata about the data;
   - Information attributes about the requestor (the user or principal);

  - Information attributes about the desired action(s) on the resource;
  - Relevant environmental information attributes (time of day, physical location, etc.)

7. **Protocols and Standards**: Required to facilitate communication between the access control components. Protocols include:
   a. OASIS eXtensible Access Control Markup Language (XACML) protocol as the standard protocol for initiating and evaluating access control requests.
   b. Lightweight Directory Access Protocol (LDAP) protocol for obtaining metadata information from stored data.
   c. OASIS Security Assertion Markup Language (SAML) protocol for federated authentication
   d. Trusted Computing Group Interface for Metadata Access Points (IF-MAP) protocol for network-layer access controls.

In summary, the right strategy is to implement an architecture that: [52]

- Replaces DAC with MAC;
- Separates the access decisions from the enforcement functions;
- Connects the components with standard protocols which allows the transition to data-centric information security;

This architecture allows data protection to be applied to the data itself, independent of the environment. It is the confluence of two distinct, but necessary changes:

(a) the shift to dynamic, policy-driven access control, and

(b) the movement of data protection enforcement closer to the data itself

3.6 Recommendations and Best Practices

3.6.1 Three-Pronged Approach

As already identified in the first version of this deliverable – chapter 4.4.1.5 – we should make sure that we have the right physical and technical security, backed up by robust policies and procedures and reliable, well-trained staff. Regarding the latter Thessaloniki has reported that the personnel supporting the application users has been initially trained and is trained again after significant application changes. Moreover, since Agueda supports encryption keys rotation (see 3.6.14.2) training is limited to the usage of keys pairs for authentication, in order to access the server.

People and human behaviour need to be taken into account for effective IT governance. People's skills, abilities, morale, capabilities and even ethics are factors to consider when addressing data protection. [55]

Meanwhile technology offers us the tools to ensure that data leaves the enterprise encrypted, with the right security measures. Next-generation firewalls can assist us in understanding what kind of data is leaving the perimeter and, if it does not have the right security attributes or is going to an unfriendly destination, we can enforce encryption and make sure that data is protected. [55]

Finally, robust policies, procedures and processes are very important since they will allow people to understand what is acceptable and what is not. The right policies will educate users on which technology is safe to use, which is not safe to use, and the appropriate behaviour for maximum data security. [55]

We should use the triad of people – process – technology to make sure cloud services are used in the right way. [55]

At the time of writing Thessaloniki application has a "Terms of use" page in place, while Agueda and Valladolid haven’t documented any policies, procedures or processes.

3.6.2 Data Protection Principles

According to [50] the following principles should be taken into consideration when designing data protection mechanisms:

- **Render Data Useless Out of Context**. Data should be unintelligible[8] except when an authorized individual performs authorized operations on that data.
- **Enforce Protection at the Data Itself**. Access enforcement for data should be done as close to the data as possible, without excluding additional protection offered by network, applications, and other environmental controls. Protection should be specific to the information being protected, ideally should be carried with the data, making it consistent across any environment in which the data resides.
- **Use Consistent Security Throughout the Information Life Cycle**. Data should be protected throughout its life cycle and ideally also while in use. Moreover, if the sensitivity of the data changes, mechanisms should ensure that protection level is also updated, without gaps in the protection enforcement during these policy changes.
- **Standardize Data Access Decisions**. In an technological evolving environment, standards should be used to provide continued data access, including:
  - A standard container for encapsulating or protecting the information;
  - A standard programming interface for manipulating the protection around the data;
  - A standard protocol for communicating relevant data rights between data consumers and data owners;
  - A standard classification system for capturing the meta data necessary to process data access decisions.

3.6.3 Data Residency

The data residency requirements also apply to cloud providers with data centres around the world, which in the normal course of operation may transfer and store data in countries that do not meet European privacy rules. [47]

[8] Just appear as a random bit pattern

3.6.4 Data Minimization

Data processing systems should be designed in accordance with the aim of collecting, processing or using no personal data at all or as few personal data as possible.

Regarding the selected STORM CLOUDS applications, municipalities have confirmed that they are consistent with the data protection principle of "data minimization".

3.6.5 Data Breach Management

Article 17 of the European Data Protection Directive imposes the obligation upon data controllers and processors to apply technical and organisational measures to protect data against accidental or unlawful destruction loss disclosure, and other forms of unlawful processing. Many organisations take the view that one of those measures might be the adoption of a policy on dealing with a data security breach. [48]

According to [48] there are four important elements to any breach management plan:

1. Containment and recovery
2. Assessment of ongoing risk
3. Notification of breach
4. Evaluation and response

At the time of writing only Agueda reported that their policy on dealing with a data security breach is under development. However, their draft policy covers:

i. What is a breach

ii. How to report the breach

iii. Investigation and Risk Assessment

iv. Containment and Recovery actions

3.6.5.1 Containment and Recovery

In case of a security breach apart from the initial response required to investigate and contain the situation, a recovery plan including, where necessary, procedures for damage limitation is also required. According to [48] we should consider:

a) Decide on who should lead the breach investigation, ensuring that they have the appropriate resources.

b) Establish who needs to be informed and what it is expected from them in order to contain the breach, such as isolating or closing a compromised section of the network.

c) Establish what can be done to recover any losses and limit the damage the breach can cause. This could involve the use of back up tapes to restore lost or damaged data or mechanisms to detect when stolen data are being used to access accounts.

d) Where appropriate, inform the police.

3.6.5.2 Risk Assessment

Since the risks associated with each data breach are different, before deciding on what steps are necessary for breach containment, a breach risk assessment should be performed. The assessment should address the potential adverse consequences for individuals, how serious or substantial these are and how likely they are to happen. The following points (as suggested in [48]) may be helpful in making this assessment:

- Type of data involved.
- Categorization of data in terms of sensitivity. For example is it a health record or a bank account?
- Data protection scheme in place, such as encryption.
- Type of data breach, since it results in different type and level of risk. Is data stolen or damaged?
- Categorization of data in terms of "real" usage. Can the data tell a third party about the individual? Can this information be used to build up a detailed picture of the individual (identity theft)?
- Amount of individuals' personal data affected by the breach. This does not imply a linear connection between the amount of data affected and the risk magnitude, but it is an important determining factor in the overall risk assessment.
- Identification of individual type, whose data has been breached. For example staff, customer, client, supplier. This to some extent will determine the level of risk posed by the breach and, therefore, the actions needed to mitigate those risks.
- Categorization of harm type/level. For example physical safety or reputation, financial loss, combination of these.
- Other wider consequences to consider, such as risk to public health, loss of public confidence in the provided service.

3.6.5.3 Breach Notification

Data security breach notification can be an important element in the breach management strategy. It should be clear who needs to be notified and why.

A draft version of the new regulation would require organisations to notify EU regulatory authorities within 24 hours of a data breach, even if the breach occurs in a third-party cloud service. The problem arises from the fact that many cloud providers expressly put the responsibility on the customer to detect breaches and this can be an impossible task. [47]

Some existing regulations, including the UK General Data Protection Regulation and France Data Protection Act (DPA), allow organisations to circumvent breach notification requirements if data is made inaccessible to third parties using encryption. Unfortunately, only 1.2 percent of cloud providers today provide the tenant-managed encryption keys required to do so. [47]

According to [48] the following questions will assist in deciding whether to notify:

? Are there any legal or contractual requirements?

? Can notification help meet security obligations with regard to the seventh data protection principle?

? Can notification help the individual? For example an individual could change their password or cancel their credit card.

? Should relevant Data Protection Authority (DPA) be informed? In case a large number of people are affected, or there are very serious consequences we should inform the DPA.

? Is there a need to notify all individuals, if only a portion of them is affected by the breach? “Over notification” can result in increased workload.

Deciding who and how to notify is also an important activity. Although this, to a large extend, depends on the nature of the breach, the following points (as suggested in [48]) may assist this process:

- Ensure that the appropriate regulatory body is notified. This could involve notifying only the sector specific regulatory body, or the DPA in the case of personal data breach. In the latter case details of the security measures in place (encryption) and, where appropriate, details of the security procedures in place at the time of the breach should be included.
- Ensure that the most appropriate type of notification is used, taking into account the security of the medium as well as the urgency of the situation.
- Include a description in the notification body of how and when the breach occurred, what data was involved, what actions were taken to address the risks posed by the breach.
- Assist individuals on the steps they can take to protect themselves, by giving them specific and clear advice.
- Provide a way, for affected individuals, to get further information or ask what has occurred. This for example can be a hotline or a web page.
- Consider notifying third parties such as the police, insurers, professional bodies, bank or credit card companies who can assist in reducing the risk of financial loss.

3.6.5.4 Evaluation and Response

It is important to investigate the causes of the breach and also evaluate the effectiveness of the response to it. While doing so we need to identify where improvements can be made and update our policies and procedures accordingly. The following points (as suggested in [48]) can assist:

- Identify what personal data is held and where and how it is stored.
- Establish where the biggest risks lie. For example, how much sensitive personal data is held? Is data stored across the business or is it concentrated in one location?
- Ensure not only the security of the transmission medium but also that only the minimum amount of data necessary is shared or disclosed. By doing this, even if a breach occurs, the risks are reduced.
- Identify weak points in the existing security measures (portable storage devices, access to public networks, …)
- Continuous staff training.
- At a very least identify a group of people responsible for reacting to reported breaches of security.

3.6.6 Software Security Updates

We should apply all security updates to the system's software, whenever they are made available. It is therefore important to ensure that **all** software components used to process personal data are subject to an appropriate security updates policy, including operating systems, applications, libraries and development frameworks.

Although it may not be practical to apply security updates as soon as they are made available, we should ensure that these updates do get applied within a reasonable time.

Additionally, we must also ensure that no relevant components are ignored. This is a common risk where responsibility for updates is split between multiple people, or where third-party libraries or frameworks are used. [44]

By default, only supported software should be used, i.e. software for which updates are still being provided. [44]. Within STORM CLOUD all applications are based/using supported software.

In addition we should define and adhere to an appropriate software updates policy for systems that process personal data otherwise we could breach the seventh data protection principle. The following considerations can help us define a software updates policy and keep to it [44]:

- Group multiple systems together that have similar requirements;
- Consider whether it is practical to enable automatic updating, as long as we trust the distributor of the software in question;
- Where possible use the operating system's native update process or package management system. In addition to the ease with which updates will be performed, this approach offers improved security during the install process, and better assurance of compatibility of different software components;
- If necessary, we can prioritise updates according to severity of the security flaws that they fix. However, here we should be cautious about deferring updates for supposedly lower risk vulnerabilities since:
    a) The longer a “low risk vulnerability” is left unfixed, the greater the total risk exposure;
    b) Classifications change as more information is gained about the vulnerability;

Finally we should make sure that the relevant updates are being applied in accordance with our policy.

At the time of writing, municipalities have reported that the actor appointed for this task is the IT department and/or the person responsible for the deployment. Additionally Agueda has reported that they have an appropriate software update policy that includes:

i. System software updates provided by Ubuntu

ii. Applications updates triggered either:

   a) Upon request to add new features or fix bugs.

   b) After a breach report. According to their data security breach policy “containment and recovery actions” may require updates to the application.

3.6.7 SQL Injection Prevention

SQL injection affects applications that pass user input to databases in an insecure manner. SQL injection flaws allow an attacker to inject database **instructions** using an input method which was only intended to receive **information**. [44]

SQL injection is particularly relevant since:

- The storage of personal data is concentrated within an SQL database (PostgreSQL for Municipio de Águeda and Ayuntamiento de Valladolid applications and MySQL for Thessaloniki application);
- Dynamic content is provided through a connection to a back-end database.

STORM CLOUDS applications can be both externally-maintained and bespoke, internally-maintained applications. For the former case in order to **prevent and detect** SQL injection we should consider the following [44]:

- Regularly check for software security updates, of the external organisation software product;
- Automated vulnerability 'scanning';
- Regular penetration testing;

For the later case in order to **prevent and detect** SQL injection we should consider the following [44]:

- Use of a set of coding standards which include specific guidance and training on avoiding SQL injection;
- Review the source code before it is put into production. An audit trail should be maintained in order to keep track of when the review took place, whether or not changes were made and who made them;
- Use of automated code review tools and vulnerability assessment;
- Penetration testing;
- Ensure that municipalities technical departments have the capability to accept bug reports and fix security flaws in code they are responsible for.

To **remediate** existing SQL injection flaws we should: [44]

- Address them promptly and comprehensively;
- Ensure that responsible developer uses the secure parameterized methods provided by the API in use, since they are designed to make sure that information is never treated as a set of instructions;
- Ensure that developers do not rely on:
    - Solutions using a blacklisting approach. Blacklisting can be useful as a short-term fix or for intrusion detection purposes. However, there are many ways in which an attacker could evade blacklisting.
    - Client side solution implementation, such as JavaScript, since client-side scripting cannot be relied upon for security.
- Review the rest of the application's code to find similar flaws, since an insecure input method used once is likely to have been re-used elsewhere, especially in code by the same author.

Regarding the selected applications, Thessaloniki application escapes every value that is sent to the database, while Agueda application SQL injection is not possible by design.

3.6.8 Decommissioning of Unnecessary Services

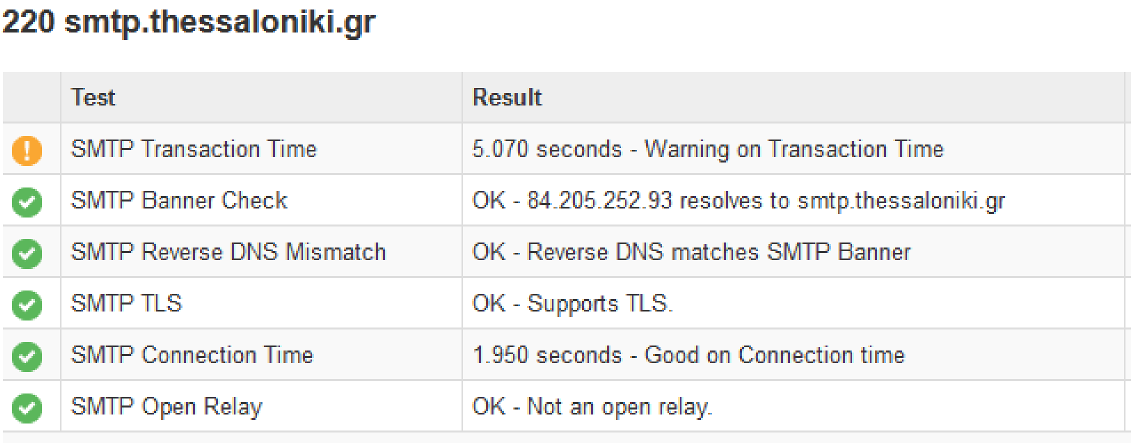
An important principle in network security is to only run the services that are absolutely necessary, thus reducing the number of ways an attacker might compromise our systems. It's important that we periodically review the necessity of the services provided and completely decommission any service that is not necessary.

Moreover we should avoid insecure services, wherever possible, since they can be can be exploited by an attacker. Such services include:

- telnet; instead we should use SSH;
- Plain FTP;
- Open mail (SMTP) relays

Both the Participação Pública (PPGIS – Public Participation) and Thessaloniki Virtual City Market applications from Municipio de Águeda and Municipality of Thessaloniki, use open mail SMTP mail server. However, they are not acting as an open relay, hence we should not experience any issues, as Figure 3-7 and Figure 3-8 indicate.

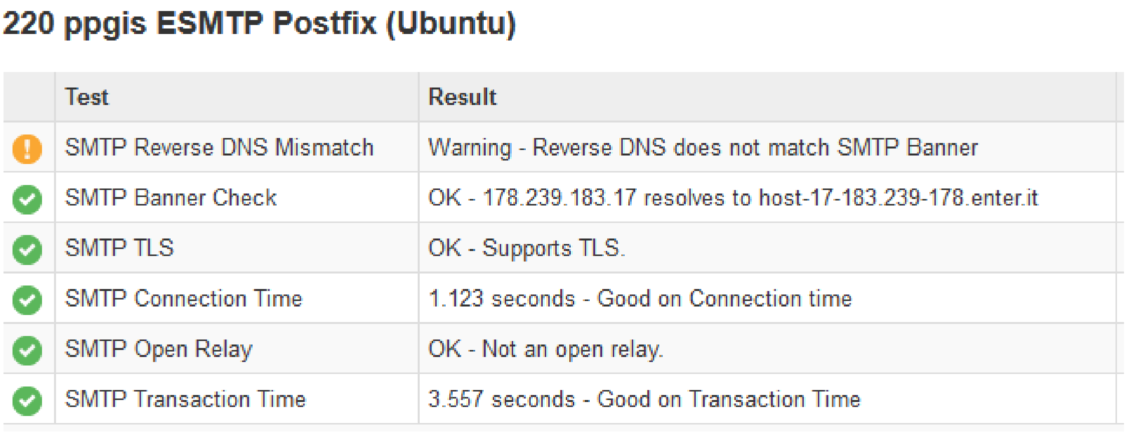
Additionally as Figure 3-9, Figure 3-10 and Figure 3-11 show insecure services are blocked.



220 smtp.thessaloniki.gr

	Test	Result
!	SMTP Transaction Time	5.070 seconds - Warning on Transaction Time
✓	SMTP Banner Check	OK - 84.205.252.93 resolves to smtp.thessaloniki.gr
✓	SMTP Reverse DNS Mismatch	OK - Reverse DNS matches SMTP Banner
✓	SMTP TLS	OK - Supports TLS.
✓	SMTP Connection Time	1.950 seconds - Good on Connection time
✓	SMTP Open Relay	OK - Not an open relay.

Figure 3-7: Thessaloniki Email Server Test Result (mxtoolbox.com)



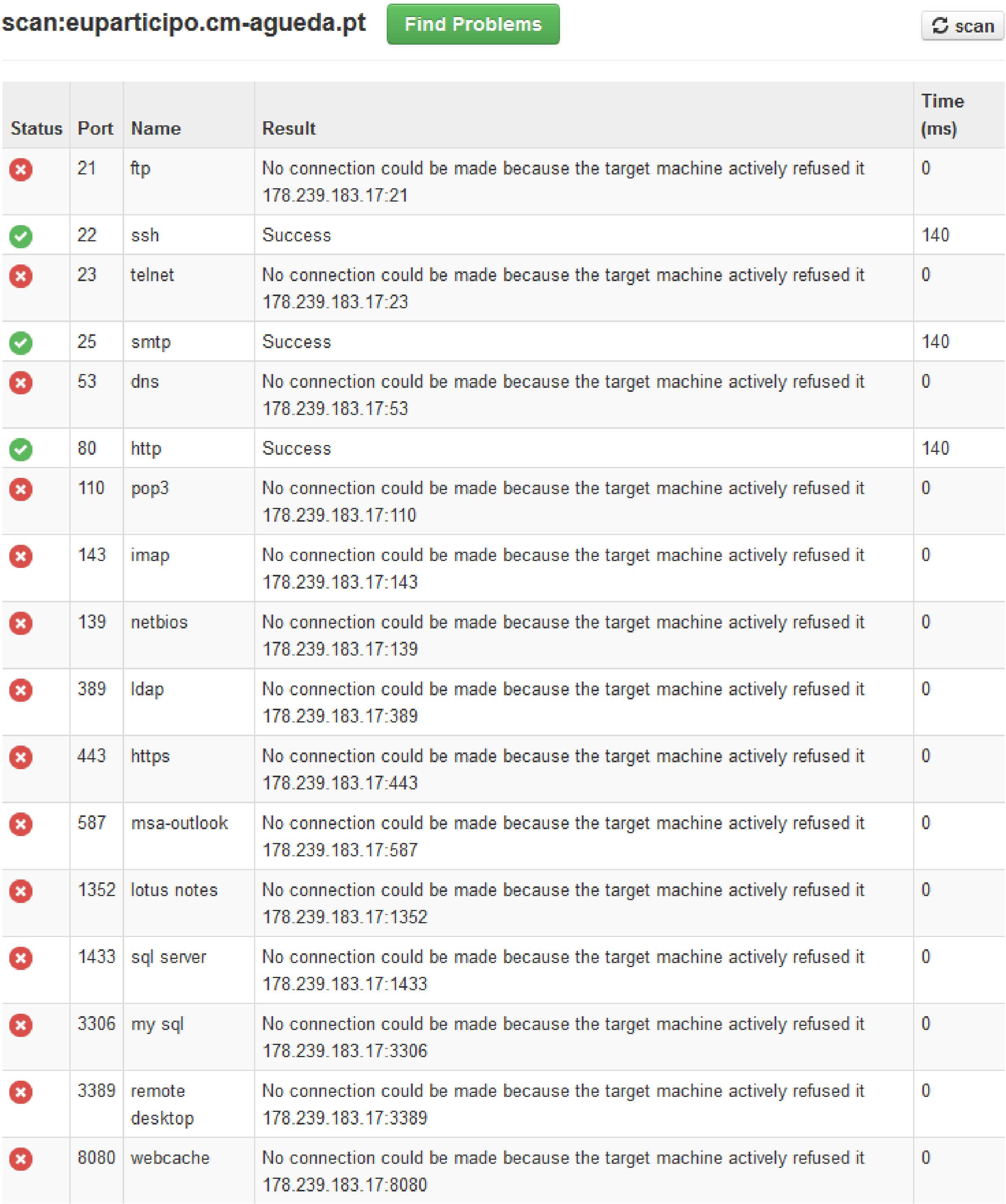
220 ppgis ESMTP Postfix (Ubuntu)

	Test	Result
!	SMTP Reverse DNS Mismatch	Warning - Reverse DNS does not match SMTP Banner
✓	SMTP Banner Check	OK - 178.239.183.17 resolves to host-17-183.239-178.enter.it
✓	SMTP TLS	OK - Supports TLS.
✓	SMTP Connection Time	1.123 seconds - Good on Connection time
✓	SMTP Open Relay	OK - Not an open relay.
✓	SMTP Transaction Time	3.557 seconds - Good on Transaction Time

Figure 3-8: Agueda Email Server Test Result (mxtoolbox.com)

Moreover, we should use periodic port-scanning to check for unnecessary services which have been inadvertently enabled and ensure that services intended for local use only are not made publicly-available, such as :

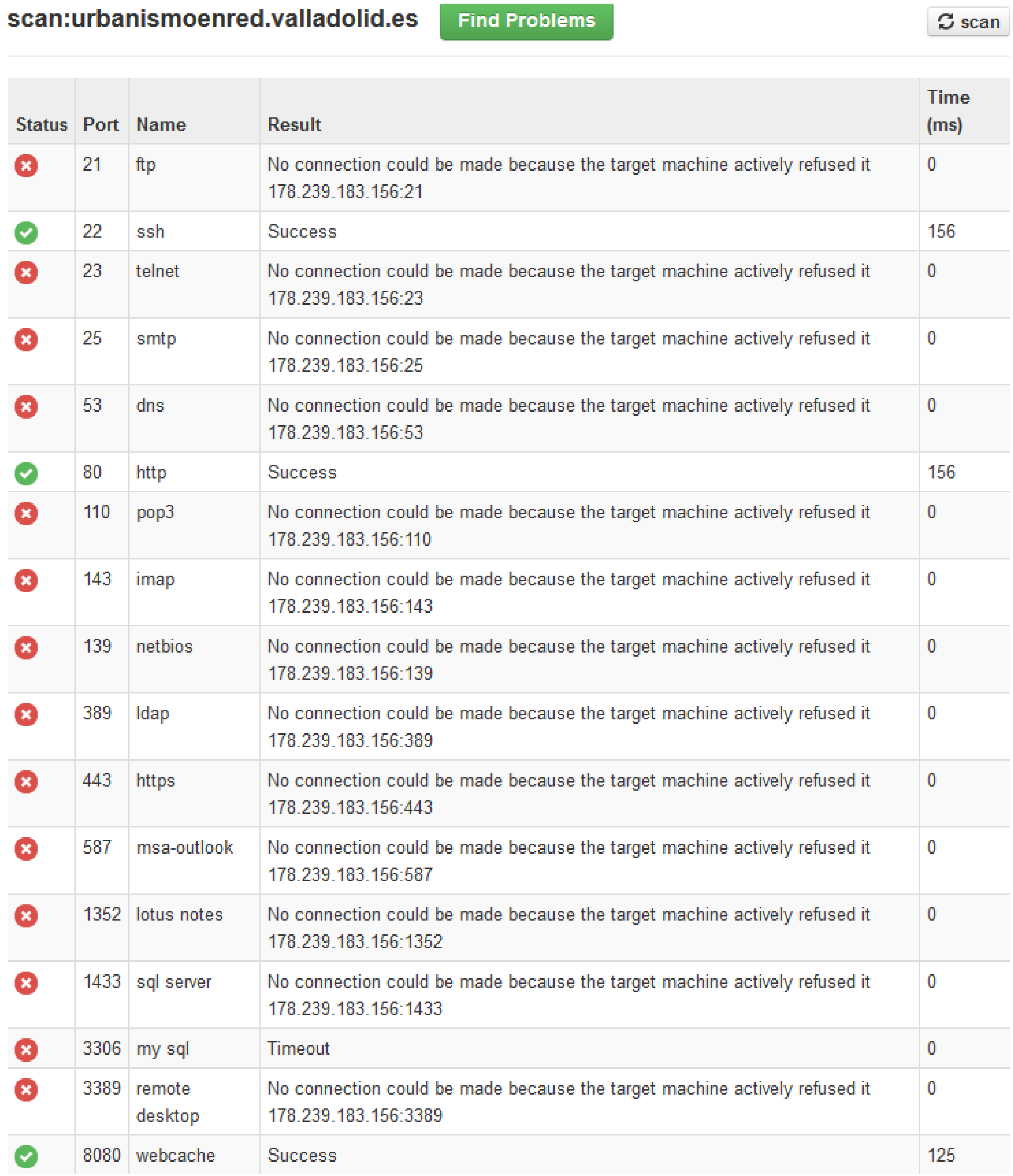
- ✓ file- and print-sharing services (SAMBA, NFS, CUPS);
- ✓ memcached;
- ✓ direct database access;
- ✓ Universal Plug 'n' Play (UPnP);
- ✓ Remote Procedure Calls (RPC);
- ✓ Simple Network Management Protocol (SNMP);



scan:euparticipo.cm-agueda.pt

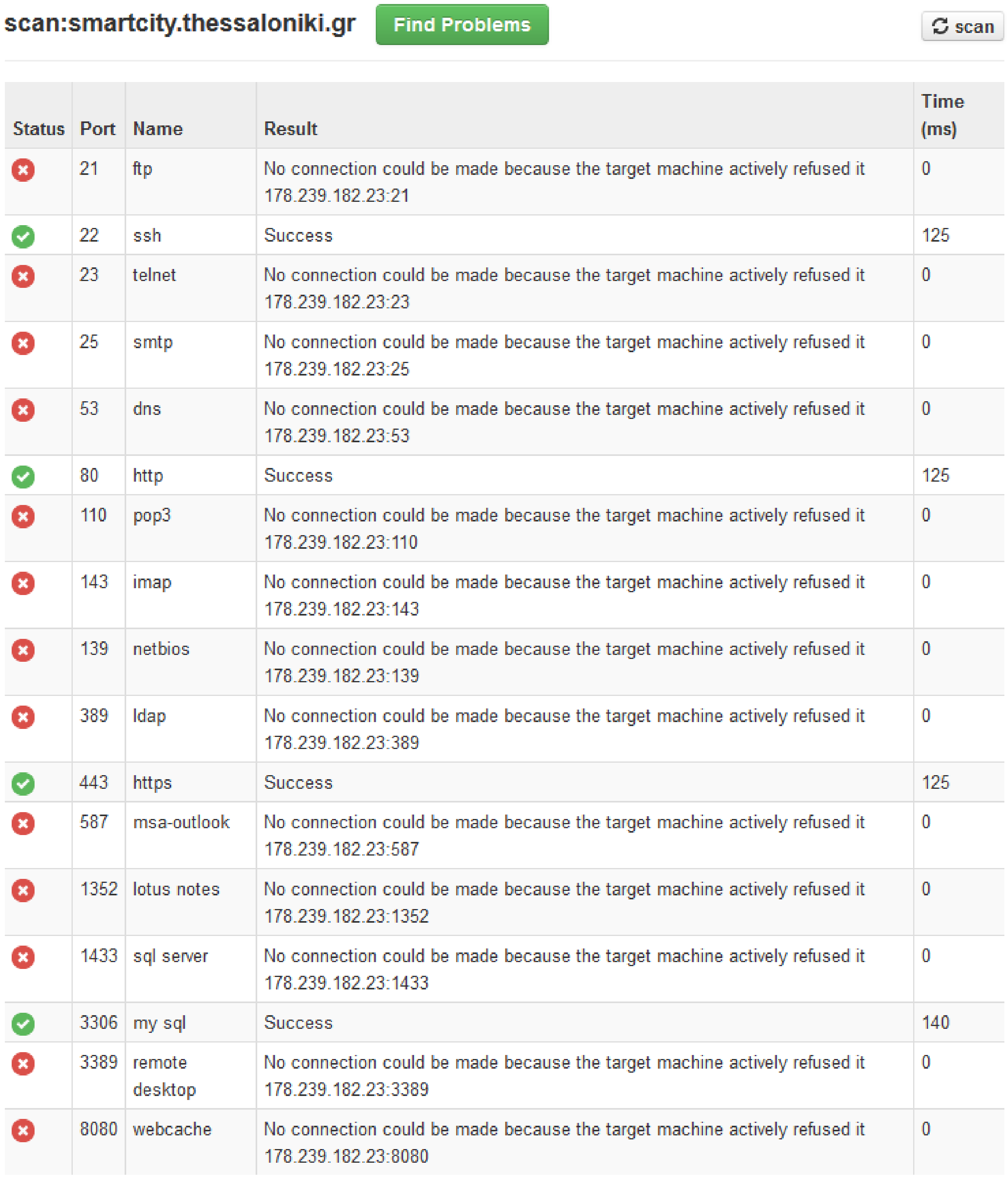
Status	Port	Name	Result	Time (ms)
✗	21	ftp	No connection could be made because the target machine actively refused it 178.239.183.17:21	0
✓	22	ssh	Success	140
✗	23	telnet	No connection could be made because the target machine actively refused it 178.239.183.17:23	0
✓	25	smtp	Success	140
✗	53	dns	No connection could be made because the target machine actively refused it 178.239.183.17:53	0
✓	80	http	Success	140
✗	110	pop3	No connection could be made because the target machine actively refused it 178.239.183.17:110	0
✗	143	imap	No connection could be made because the target machine actively refused it 178.239.183.17:143	0
✗	139	netbios	No connection could be made because the target machine actively refused it 178.239.183.17:139	0
✗	389	ldap	No connection could be made because the target machine actively refused it 178.239.183.17:389	0
✗	443	https	No connection could be made because the target machine actively refused it 178.239.183.17:443	0
✗	587	msa-outlook	No connection could be made because the target machine actively refused it 178.239.183.17:587	0
✗	1352	lotus notes	No connection could be made because the target machine actively refused it 178.239.183.17:1352	0
✗	1433	sql server	No connection could be made because the target machine actively refused it 178.239.183.17:1433	0
✗	3306	my sql	No connection could be made because the target machine actively refused it 178.239.183.17:3306	0
✗	3389	remote desktop	No connection could be made because the target machine actively refused it 178.239.183.17:3389	0
✗	8080	webcache	No connection could be made because the target machine actively refused it 178.239.183.17:8080	0

Figure 3-9: Agueda Port Scan (mxtoolbox.com)



Status	Port	Name	Result	Time (ms)
✗	21	ftp	No connection could be made because the target machine actively refused it 178.239.183.156:21	0
✓	22	ssh	Success	156
✗	23	telnet	No connection could be made because the target machine actively refused it 178.239.183.156:23	0
✗	25	smtp	No connection could be made because the target machine actively refused it 178.239.183.156:25	0
✗	53	dns	No connection could be made because the target machine actively refused it 178.239.183.156:53	0
✓	80	http	Success	156
✗	110	pop3	No connection could be made because the target machine actively refused it 178.239.183.156:110	0
✗	143	imap	No connection could be made because the target machine actively refused it 178.239.183.156:143	0
✗	139	netbios	No connection could be made because the target machine actively refused it 178.239.183.156:139	0
✗	389	ldap	No connection could be made because the target machine actively refused it 178.239.183.156:389	0
✗	443	https	No connection could be made because the target machine actively refused it 178.239.183.156:443	0
✗	587	msa-outlook	No connection could be made because the target machine actively refused it 178.239.183.156:587	0
✗	1352	lotus notes	No connection could be made because the target machine actively refused it 178.239.183.156:1352	0
✗	1433	sql server	No connection could be made because the target machine actively refused it 178.239.183.156:1433	0
✗	3306	my sql	Timeout	0
✗	3389	remote desktop	No connection could be made because the target machine actively refused it 178.239.183.156:3389	0
✓	8080	webcache	Success	125

Figure 3-10: Valladolid Port Scan (mxtoolbox.com)



Status	Port	Name	Result	Time (ms)
✗	21	ftp	No connection could be made because the target machine actively refused it 178.239.182.23:21	0
✓	22	ssh	Success	125
✗	23	telnet	No connection could be made because the target machine actively refused it 178.239.182.23:23	0
✗	25	smtp	No connection could be made because the target machine actively refused it 178.239.182.23:25	0
✗	53	dns	No connection could be made because the target machine actively refused it 178.239.182.23:53	0
✓	80	http	Success	125
✗	110	pop3	No connection could be made because the target machine actively refused it 178.239.182.23:110	0
✗	143	imap	No connection could be made because the target machine actively refused it 178.239.182.23:143	0
✗	139	netbios	No connection could be made because the target machine actively refused it 178.239.182.23:139	0
✗	389	ldap	No connection could be made because the target machine actively refused it 178.239.182.23:389	0
✓	443	https	Success	125
✗	587	msa-outlook	No connection could be made because the target machine actively refused it 178.239.182.23:587	0
✗	1352	lotus notes	No connection could be made because the target machine actively refused it 178.239.182.23:1352	0
✗	1433	sql server	No connection could be made because the target machine actively refused it 178.239.182.23:1433	0
✓	3306	my sql	Success	140
✗	3389	remote desktop	No connection could be made because the target machine actively refused it 178.239.182.23:3389	0
✗	8080	webcache	No connection could be made because the target machine actively refused it 178.239.182.23:8080	0

Figure 3-11: Thessaloniki Port Scan (mxtoolbox.com)

Finally we should maintain a list of which services should be made available while periodically review the necessity of the services provided and restrict or decommission unnecessary ones. Additionally we should record any temporarily installed services which will eventually need to be disabled/ decommissioned. Attention should be made in ensuring that the decommissioning procedure has actually succeeded, using for example port scanners.

At the time of writing municipalities reported that they didn't periodically review the necessity of the provided services, or perform port-scanning for unnecessary services. However, after understanding the necessity to run the services that are absolutely necessary, they committed to perform this activity as often as possible, given the limited IT resources they have.

3.6.9 Password Storage

Users' access credentials (e.g. a username and password or passphrase) are particularly valuable to attackers, for a number of different reasons [44] making credentials management an important activity.

An important principle is that passwords should not be recoverable directly, meaning that passwords **should not be stored in plain text** because these are immediately readable. It also means that encryption is not generally appropriate, since an encrypted value needs to be decrypted to retrieve the original password. This action of decryption requires access to a key, which must be securely managed. [44]

Moreover, applications should never remind a user's current password directly in plain text (e.g. by including it in an email), as this is a clear indication of an insecure approach. Thessaloniki and Agueda applications, that require user login, provide a mechanism whereby users receive an email with instructions on how to set a new password.

Instead we should **use a hash function and only store the hashed values** (as described in 4.5.8 of D4.1.1).

A hash function is a one-way method which converts a password into a hashed value, often simply called the 'hash'. When a user first registers with a service and provides a password this is hashed and only this hash value is stored. When a user returns and enters their password, the hash is freshly calculated then compared with the stored hash. If the two hashes match, then the user can be authenticated. The one-way nature of hashes is key: if an attacker somehow obtains a list of hashes, they cannot directly work out what the passwords are, even if they know the particular hash function that was used. [44]

MD5 and SHA-1 should not be used for hashing passwords. The hash function should have appropriate strength to make offline brute-force attacks extremely impractical (delay an offline brute-force attack for at least a matter of months).

We should also periodically review the strength of the hash function and keep up to date with advances in computing power. The best way of achieving this is to use a password hashing scheme with a configurable work factor. [44] Examples include: PBKDF2, bcrypt and scrypt.

The European Union Agency for Network and Information Security Agency (ENISA) has published a detailed assessment of cryptographic measures. The report is entitled Algorithms, Key Sizes and Parameters Report - 2013 Recommendations. In section 3.3 it supports that:

- MD5 is not appropriate; and
- SHA-1 is acceptable for legacy systems but should not be designed into new systems.

Moreover we should also **use salting** to make offline brute-force attacks less effective.

A 'salt' is a string of random data **unique to each user**. The salt is used by combining it with the user's password, then hashing the result. When a user returns to the service the stored salt and the supplied password are freshly combined and hashed. As in the unsalted method, the new hash and the stored hash are compared to determine if the user should be authenticated. [67]

A typical **salt-length would be 128-bits**, and this length is given as a minimum in NIST's "Recommendation for Password-Based Key Derivation".

Nevertheless we should have a plan of action in case of a password breach. This should include how to reset users' passwords in bulk and how to notify them of what has happened and what they need to do about it. [44]

Even with robust, salted hashing in place, there is still the threat of an attacker using a 'dictionary attack' to guess passwords that are known to be common. Hence we should use a combination of password strength requirements and user-education to ensure that attackers can't simply guess common passwords. [44]

According to [44] users should be encouraged to strengthen their passwords by:

a) creating long password or passphrase;

b) using wide range of characters such as uppercase/lowercase letters, numbers, punctuation, special symbols;

c) avoiding dictionary words and use of patterns derived from the physical keyboard layout such as "qwerty" or "!1@2#3qaz".

To visually assist users, and give them an immediate feedback, a well designed password "strength meter" could be provided. Finally the maximum length or permissible character set of passwords should not be limited as this simply reduces the time needed to mount a successful brute force attack.

At the time of writing Thessaloniki application requires passwords to be at least eight (8) characters long and include at least one (1) number, while Agueda application simply requires long passwords. Unfortunately selected applications don't provide a password "strength meter".

Regarding dictionary attacks, Agueda has reported that on the server side, each authentication request is logged including the IP address of the machine made the request. For each failed attempt, a delay is added to discourage consecutive repeated attempts.

3.6.10 SSL/TLS Configuration

Secure Sockets Layer (SSL) and Transport Layer Security (TLS) are two closely related encryption schemes providing assurance that: [44]

- communication is encrypted; and
- The identity of one or both of the endpoints can be trusted.

Failure to provide encryption results in any sensitive information transmitted being viewable by anyone system on the route between the two systems.

Providing proper transport layer protection can affect the site design. It's easiest to require SSL for the entire site. For performance reasons, some sites use SSL only on private pages. Others use SSL only on 'critical' pages, but this can expose session IDs and cookies, which in turn would let an attacker gain access as if they were a logged-in user. [68]

According to [44] and [68] at a minimum we should do all of the following:

1) Require SSL for all sensitive pages or where transmission of personal data takes place. Non-SSL requests to these pages should be redirected to the SSL page.

2) Use SSL version 3[9], but preferably use TLS with the latest version possible[10].

3) Only enable ciphers with 128-bit strength or greater.

4) Avoid any ciphers with known weaknesses in their design, such as RC4.

5) Set the 'secure' flag on all sensitive cookies.

6) Configure the SSL provider to only support strong (e.g., FIPS 140-2 compliant) algorithms.

7) Ensure that every SSL or TLS service uses a certificate that is valid, not expired, not revoked, and matches all domains used by the site. We should also schedule renewal of all certificates before they expire to ensure the services remain secure.

8) Backend and other connections should also use SSL or other encryption technologies.

9) Use SSL or TLS throughout the entire domain in order to reduce complexity. Remember that any included content such as images, JavaScript or CSS should also be provided over SSL or TLS in order to avoid 'mixed content' warnings in users' browsers.

10) Do not encourage users to ignore SSL or TLS security warnings as this would result in lack of identity assurance that can be exploited using a 'man-in-the-middle' attack[11].

At the time of writing only Thessaloniki supports encryption of their network communications using SSL for the entire site, while ensuring that the certificate is valid, not expired, not revoked, and matches all domains used by the site.

3.6.11 Data Processing Locations

According to [44] a common data breach is caused by personal data being processed in an inappropriate location. The typical causes of this problem are:

- Poor security architecture,
- Accidentally storing personal data in a publicly accessible location.

[9] As of September 2014 due to the Poodle attack SSL version 3.0 is no longer secure. The first and best way to mitigate this problem is to completely disable SSL version 3.0 on all of our servers. The second solution is to support TLS_FALLBACK_SCSV.

[10] At the time of the writing TLS 1.2 is the latest version

[11] A man-in-the middle attacker impersonates the SSL or TLS service that a user expects to see. Regular users of a legitimate service are likely to ignore a security warning about an impostor, potentially setting up an untrusted connection despite SSL or TLS being in use and the communication being encrypted between the user and the attacker. The man-in-the-middle attacker then forwards the communication onto the legitimate service and acts as a go-between, observing the communication unencrypted on its way past (hence the name 'man-in-the-middle'). In this way, both the user and the legitimate service are potentially unaware that the connection is compromised, yet the attacker can easily obtain unencrypted information [44]

3.6.11.1 Security Architecture

Poor design of systems and networks can contribute to compromise of personal data. On the other hand, having well designed systems and networks can provide clarity as to where personal data should be processed internally, and help prevent it being leaked to inappropriate locations. [44]

To that extend we should segregate testing or staging environments from production environments. According to [44] the most significant reasons for this are:

- Production systems won't be affected, while we are developing;
- A separate testing environment gives advanced notice of many operational problems allowing to fix them without affecting the production systems; and
- Example data should be used while developing and testing, avoiding putting personal data at risk.

Redundancy and diversity should be appropriately included within the network, to guard against another aspect of the seventh data protection principle, namely "accidental loss or destruction of, or damage to, personal data".

At the time of writing, two instances of the platform are available:

1. one at Hewlett Packard's premises (SCP@HP), used by project participants for testing the platform, the applications and the 'cloudification process'.
2. the other hosted at EnterTheCloud (SCP@Enter), used by project participants as the production environment where cloudified applications are merely moved from SCP@HP to SCP@Enter, followed by minor application configurations.

3.6.11.2 Publicly Accessible Data Locations

Even with clear security architecture in place, it is possible to inadvertently leave a copy of personal data in an accessible place where unauthorized people can access it. This is made possible either due to an administrative error, or due to failure in realizing that the storage place is widely accessible.

To address this we should make sure that appropriate policies exist, describing how, when and where personal data will be processed. We should avoid:

- Web server with directory browsing enabled, since the entire contents can be indexed within minutes and therefore its contents made available.
- Using a unique URL parameter as part of a dynamic website without any additional access controls (such as log in), since editing the URL manually can make personal data accessible to others.
- Easily-guessable directory names that don't have specific access restrictions. According to [44] easily guessable common directory names include cv, admin, uploads, data, images, docs, icons, service, db, forum, search, cms.

Instead we should:

- Set up an isolated network or use internal firewall policies, if applications need to process particularly sensitive information.
- Ensure web servers are exposing only the intended content.

- Provide awareness of personal data and where and how it should be stored. This can be achieved with good training. To be effective training should be tailored to different roles in the organization.

Although none of the selected applications handle personal data or Personally Identifiable Information (PII) the above recommendations were followed by the project municipalities.

3.6.12 Default Credentials

Default credentials for a wireless router or firewall, an administrator database account or an operating system, can be easily found using a simple web search. Hence, default credentials should be changed as soon as possible, preferably before development or testing, and certainly before production.

When changing default credentials, remember to follow password choice practices described in the last two paragraphs of section 3.6.9

According to [44] credentials:

a) should NOT be transmitted in plain text and

b) should NOT be hard-coded into the software, because they can be easily obtained by simply viewing the source code.

Regarding the selected applications the administrator credentials of the databases and the operating system have been changed. Only Thessaloniki application uses a configuration file, where credentials are hard-coded, but since the apache web browser is configured to deny access to certain file types[12] and paths[13] we can safely assume that credentials cannot be obtained.

3.6.13 Backup Strategy

According to [44] we could use both regular on-site backups (CSP premises) and less frequently off-site backups (municipality premises). On-site backups can guard against data loss, destruction or damage to a subset of the data on-site, while off-site backups can guard against larger-scale common failure modes such as an on-site fire, virus outbreak or insider attack.

However we should note that for STORM CLOUDS IaaS/PaaS deployments once the VM is running, it's up to the VM owner (the application provider) to implement the desired backup strategy/technology.

At the time of writing municipalities, based on the statistics of the data generated by their applications, decided that cloud backups is the most favourable solution. More specifically:

- Within the STORM CLOUDS platform we currently support a dedicated server for backup where only authenticated users and clients have access to it.
- Municipalities take daily incremental on-site backups
- Municipalities take weekly on-site and off-site full backups

[12] http://www.ducea.com/2006/07/21/apache-tips-tricks-deny-access-to-certain-file-types/

[13] http://www.ducea.com/2006/08/11/apache-tips-tricks-deny-access-to-some-folders/

- Data is encrypted both for storage and transmission.
- Data recovery options (based on the backup taken) are tested on a monthly basis

The backup solution is based on **duplicity**[14] a free software package implementing "encrypted bandwidth-efficient backup, using the rsync algorithm".

Moreover VM backups are taken whenever we have a major application/software update, in order to ensure system availability.

3.6.13.1 Virtual Machine Backup

According to [65] VM backup is a critical consideration in ensuring systems are delivered with the highest level of availability. However, it differs from the traditional approach used to back up physical servers and this means thinking differently.

One of the characteristics of virtual environments is consolidation of resources that can cause huge oversubscription to processor, network and storage resources and major bottlenecks when many concurrent backups create high levels of random I/O. This is why VM backup is one of the most pressing tasks facing IT departments today. [34]

Another critical issue is the VM state and the availability of each workload to users. [37]

According to [34] common backup strategies are:

1. Hypervisor-based VM backup. Here, hypervisor features are used to synchronise and extract backup data.
2. Hardware-based VM backup. By utilizing the shared storage infrastructure we can make a copy of virtual machine image files. However, this method is more suited to NFS or SMB-provisioned virtual machines, because the data is easily exportable as a file share snapshot.

Choosing the right virtual machine backup strategy means examining a number of factors that influence the effectiveness of virtual machine backup and restore. [34] More specifically:

- **Application consistency**. This is needed to ensure that data within the application is consistent.

  In agent- or hypervisor-based VM backup, the backup software can request from the operating system and application to flush outstanding data to disk and enter a temporarily inactive or inhibited state (quiesce) while the backup is taken. [34] If the hypervisor cannot quiesce[15] the VM, the machine state on disk may not reflect the current machine state in memory, resulting in poor, essentially un-restorable backup that wastes time and effort. [37]

- **Backup object abstraction**. Backup solutions must be able to back up and restore virtual machines to logical locations rather than physical ones, possibly making hardware-based solutions, much less practical than agent- or hypervisor-based ones. [34]

[14] Available at http://duplicity.nongnu.org/

[15] Temporarily stop from performing any more transactions

- **Granular recovery**. Data protection is all about the efficiency of the restore process, so a key factor is the level of granularity to which restore can be carried out. When only a few files are needed, restoring the entire VM is too cumbersome or time-consuming. [34]
- **Replication**. Replication of the backup data is the process of taking a copy of the data offsite to another location as part of a disaster recovery strategy.
- **Licensing**. Many traditional backup systems base their licensing on physical server (or client) count, making the process of tracking licences cost-prohibitive as VMs come and go on a regular basis. More efficient cost models are based purely on backup capacity, making cost management much easier. [34]
- **Server location identification**. According to [34] the use of physical and virtual servers in the same environment can make it difficult to process a restore request for server X, without knowing which platform it resides on and which backup tool was used to back it up.

  A solution can be given via the naming of servers, or via a more abstract reference, with the use of a Configuration Management Database (CMDB), that indicates the logical location of the server, i.e. which site, which hypervisor, and so on.

  By abstracting the reference to the VM, both backup and restore processes are no longer dependent on the physical host hardware, which provides operational benefits by reducing the work involved in restores for clusters that have been physically or logically reconfigured. [43]

The first step to implementing a backup strategy is determining which data needs to be backed up. [40]

If there is enough disk space available we should consider backing up virtual machines (VMs) to enable easy distribution while still containing the OS and the data; otherwise it's better to back up only the data. To store only data, we should install a backup agent inside the OS to take care of data backup, while using a separate virtual LAN to send the backup data over, creating another security layer.

Moreover we should consider backing up the configuration files as it could minimize the downtime during a system restoration, avoiding reconfiguring some services, which can be time-consuming and labour-intensive to install and configure.

3.6.13.2 Snapshots

There are two discrete methodologies for data snapshots [41]:

1. An exact duplicate copy of the data at a specific point-in-time (PIT);
2. A copy of the data's state or metadata at a PIT.

Complete data duplication takes time depending on the amount of data copied. Clones are commonly synchronous mirrors performed with each write. Many take advantage of triple mirrors where the third mirror is broken to create a point-in-time (PIT) clone, but all variations consume considerable capacity. [41]

This is why state-based or metadata snapshots are more popular than PIT snapshots because they consume a very small amount of capacity. They are known as:

- copy-on-write (COW) snapshot or
- redirect-on-write (ROW) snapshot

COW and ROW snapshots make a copy of the metadata or pointers of what the data looks like at a given PIT. Both provide crash-consistent images of the data, meaning the snapshots look like exact data replicas at a PIT, the same as if the system was just shut down. [41]

However:

1. COW requires capacity reservation for the full amount of the data being snapshot. The data gets copied only when changes are about to be made to that data set. COW theoretically makes a complete copy of the data for each snapshot as the data changes, but only if the data changes. However, if there is a corruption of changed data, any snapshots that follow will also be corrupted. [41]
2. ROW does not require any capacity reservation because it writes all changes to the data separately from the PIT snapshot image and ties them together with pointers. ROW is a bit more complicated than COW in that it requires more intelligence (smart algorithms) in piecing the data together on reads. The added complexity often adds marginal latency to reads as the volume of ROW snapshots increases. [41]
3. ROW does not actually make copies of the data nor does it consume as much capacity as COW. But since there are no actual copies of data, ROW snapshots can be a significant data protection issue. If there is any corruption of the original data, all snapshots that follow are also corrupted. If there is a corruption of changed data, any snapshots that follow will also be corrupted. [41]
4. Another data protection issue with COW and ROW is that the snapshots are not application-aware, which is an issue with structured data, such as database data. For this reason applications running a database, should be quiesced, before getting an application-consistent image.

In conclusion we can say that snapshot technology is best utilized as a part of a data protection strategy, not as the entire data protection solution. ROW consumes the least amount of capacity but has lower reliability and higher potential latency. Cloning has the highest reliability but consumes the most capacity. And COW is somewhere in-between but a bit closer to ROW snapshots than to cloning. [41]

3.6.13.3 Cloud Backup

Cloud backup, also known as online backup, is a service that involves sending a copy of the data to an off-site server. [31]

The way a cloud backup service works is that, according to the type of service agreed in the SLA (daily, weekly, monthly, other), it collects, compresses, encrypts and transfers data to the service provider's servers. To save bandwidth and the time taken to transfer files, we could provide incremental backups after the initial full backup. This process is known as data deduplication, offering a great way to efficiently move data across. However, it presents some significant challenges when data recovery needs to take place [36]:

1. Rehydration or reconstitution of backup data into its application native format. The rehydration process itself is time consuming, but the biggest challenge is the time taken to

retransmit the reconstituted data from the CSP to the CSC, especially when large amounts of data need to be transmitted.

2. Resource contention within the CSP data centre, in the case that there is a massive request for disaster recovery from multiple cloud Customers, bringing recovery operations to an absolute standstill.

However, cloud-based data protection is not always a good thing. According to [42] there are at least five different reasons why:

1. **Data security** due to potential security breaches. Since we don't have direct control over the way that our data is protected, it opens up the possibility that our data could be exposed due to a security breach.
2. **Privacy**, since there is no assurance that CSP will not use our data, for own purposes.
3. **Ongoing cost**. CSP billing/charging system is based on the amount of space consumed, and on the I/O load produced by their cloud Customer. Hence the CSP is going to charge us for any space consumed, month after month.
4. **Long recovery times**, especially when restoration has to be made from the cloud, due to Internet bandwidth limitations.
5. **CSP becoming a single point of failure**, if for example the CSP goes out of business.

3.6.14 Encryption

Cryptography can be used as a means of implemented countermeasures to reduce identified risks. Although, cryptography alone will not address all possible threats, it can be used to help provide various security services, such as user authentication, data origin authentication, data integrity, non-repudiation and confidentiality.

Encryption by itself will not prevent data from being lost or stolen, but makes data manipulation significantly less valuable to anyone other than an authorised user. To a certain extent, encryption can help address a number of threats, such as: [17]

1. **Data security.** When applied at various stages of the data lifecycle encryption can protect against malicious or accidental disclosure of private information.
2. **Cloud management.** Encryption can protect not only the confidentiality of business data, but also that of cloud-management messages and API calls.
3. **Cloud provider insiders** depending on the type of encryption and key management used.
4. **Unknown risk profile**. Encryption can provide some assurance of confidentiality regarding data stored in the cloud.
5. **Forensics.** Encryption can be used to maintain the confidentiality of cloud audit information and control access to it with need-to-know restriction policies.

Although encryption can be applied at various points of the data lifecycle, the preferred method is to do it before data is stored on the cloud. According to [17] its good practice not to entrust a CSP, which stores encrypted data, with key management operations. Instead we can be responsible for this task or allocate it to a trusted third party.

According to [69] to analyse both the protection benefits we'd get from encryption and where those benefits are most needed the following steps must be considered:

1. Firstly we must perform **data classification** and **service inventorying**. Since not all our applications process sensitive or confidential information distinguishing which applications are appropriate to apply encryption, to those which is not, is crucial. To do so we must identify and record – in as granular detail as possible – where data that we might want to encrypt resides in the cloud.
2. Secondly we must **evaluate** the specific usage, **determine** whether we will encrypt, and **decide** how we will implement encryption.

Cryptographic mechanisms should meet the following criteria with regards to their three main components – Algorithms (and associated modes of operation), Protocols and Implementation:

1. The cryptographic algorithms and associated modes of operation deployed should have been scrutinized, evaluated, and approved using a review process that is open and includes a wide range of experts in the field. Examples of such approved cryptographic algorithms and associated modes of operation are: FIPS 186-3 for Digital Signatures, FIPS 180-4 for Secure Hash, SP 800-38A for modes of operation and SP 800-56A & SP 800-56B for key establishment. [58]
2. The cryptographic protocols used should have been scrutinized, evaluated, and approved using a review process that is open and includes a wide range of experts in the field. IETF protocol specifications for SSH and TLS are examples that meet these criteria. [58]
3. The implementation of a cryptographic algorithm or protocol should undergo a widely recognized and reputable independent testing for verification of conformance to underlying specifications. NIST's CAVP and CMVP are examples of such independent testing programs. [58]

After analysing the application functionality and data content municipalities have concluded that there is no need for application data encryption. However, Agueda encrypts user passwords (SHA-256), while Thessaloniki hashes them using MD5. Valladolid is not providing any kind of encryption since their application deal only with public data.

Regarding:

a. The backup encryption all municipalities support RSA 2048 public key cryptography using the GNU Privacy Guard (GPG) (https://www.gnupg.org/).

b. Off-site backups (municipality premises) municipalities consider them to be stored in a safe place and hence they are not encrypted.

3.6.14.1 Challenges

Apart from general challenges inherently present by encryption-based security solutions and processes, there is an added complexity brought by cloud computing and additional challenges that need to be addressed when implementing an encryption system.

According to [17] the inherent challenges and considerations that should be made when implementing cloud encryption are:

1. **Encryption policy**. Different encryption policies of various parties and the ever increasing number of involved parties make management and alignment of encryption policies a challenging and difficult task. To ensure that encryption policies cover an entire cloud

encryption workflow a holistic approach should be followed, and assisted by appropriate procedural and technological controls.

2. **Availability of encrypted data**. As already identified its good practice not to allow the cloud service provider to perform encryption on our behalf. If the provider's cryptography services suffer a fault or if the cloud provider loses the cryptographic keys, then availability of encrypted data can become problematic, leading to overall service unavailability or degradation as the data owner is unable to access data in its unencrypted form.
3. **Integrity of encrypted data**. Data integrity is an important phase in cloud computing as it ensures that encrypted data has not been tampered with.
4. **Encryption client security**. Since cloud services are accessed from a variety of devices, including computers and mobiles, with differing levels of security, the potential attack surface is increased to include session communication between a cloud consumer and its provider, possibly providing access to data in unencrypted plaintext form. Therefore the client device should be properly protected.
5. **Compliance with legislation and standards**. Apart from the technological challenges introduced by cloud encryption, legal and compliance issues related to the storage of data and to its encryption must also be addressed when dealing with encryption.
6. **Key management**. Key management is a fundamental process of any encryption system, but also the Achilles' heel of an encryption system and should therefore be carefully considered.

3.6.14.2 Key Rotation

If we are using encryption we should plan to rotate keys. According to [49] the reason why we might consider encryption keys rotation is that a person who had access to the master key left the organization and the company needs to ensure key access is secure once they're gone. Another reason might be a possible master key compromise. Moreover key rotation might be required by internal policies and processes or compliance mandates. Finally another reason is a security incident requiring a change to ensure that the encryption key doesn't fall into the wrong hands.

3.6.14.3 MySQL Encryption

There are several approaches to encrypt MySQL databases in the cloud. We present three of them targeting both the database level and the field level.

1. Full Disk Encryption. Encrypting an entire database contents is a common approach, ensuring protection of data from unauthorized users and hackers while satisfying many compliance requirements. The approach presents the following advantages:

    a. Simplicity and reduced fallibility since the probability of not encrypting important data is small, since we encrypt everything at a single step;

    b. Transparency since it works with our applications without changing the application code.

    However, this approach has a major disadvantage in the form of not being able to configure what we need to encrypt as it is usually "all or nothing".

2. File Encryption (table encryption). This approach uses the fact that MySQL can be configured so that each DB table can be saved into a separate file, thus encrypting only the files that are considered sensitive. This is also called Transparent Data Encryption (TDE), in comparison to the Oracle TDE and the Microsoft SQL Server TDE.
3. Field Specific Encryption (Rows/Columns encryption). This approach uses the SQL language, as implemented in MySQL, ability to encrypt specific rows or specific fields. Obviously this is the most granular approach, however it require modifications at the application-level code. Moreover, it is possible (and encouraged) to use different encryption keys for each field, thus controlling who can access which fields. Additionally field specific encryption may also improve performance, since it allows accessing of non-sensitive information without having to decrypt anything.

Which approach to use depends on our needs. If we require:

1. Configurability then field specific encryption is obviously the most configurable, followed by file-level encryption. If for example we use separate MySQL rows to represent a user's personal data, then for maximum security we could use different encryption keys for each user. However, this comes at an increased cost in terms of complexity, since developers are required to write code or system administrators to configure options.
2. Performance then full disk encryption is the most suitable. However, performance depends on the encryption engine used. For file-level encryption we suggest using the *ecryptfs* engine, while for full disk encryption we suggest the *cryptsetup/dmcrypt* engine.
3. Simplicity of Security then the appropriate approach depends on our needs.
4. Finely granular authorization and access then both full disk and file-level encryption are the most appropriate.

3.6.14.4 Encryption Solution Selection

For IaaS deployments we have the greatest flexibility in terms of the available encryption solutions. Full disk encryption, filesystem (agent-based and agent-less) level encryption, hypervisor, and database encryption solution can be used depending upon specific situations. [77]

For PaaS deployments we can use the native database encryption as well as row/column level encryption solutions. [77]

At the time of writing municipalities felt that this was not applicable to their applications.

3.6.15 Erasure Coding

In order to guarantee the integrity of data stored in the cloud, it is natural to store replicas in multiple disks, such that data losses can be tolerated as long as there is at least one replica available. However, replicas can significantly reduce the storage efficiency.

Erasure coding (EC) is a method of data protection in which data is broken into fragments, expanded and encoded with redundant data pieces, and stored across a set of different locations, such as disks, storage nodes or geographic locations. [25]

The goal of erasure coding is to allow corrupted data to be reconstructed by using information about the data that's stored elsewhere in the array.

The key concept behind erasure codes is the *polynomial interpolation* or *oversampling*. In mathematical terms, the protection offered by erasure coding can be represented in simple form by the following equation: n = k + m. The variable "k" is the original amount of data or symbols. The variable "m" stands for the extra or redundant symbols that are added to provide protection from failures. The variable "n" is the total number of symbols created after the erasure coding process. [25]

At the time of writing, EC presents the following potential benefits:

a. The biggest benefit of EC is that the 3x overhead inherent in OpenStack Swift can be reduced to a lower number[16]. The downside is that object storage, retrieval, and error correction is significantly more computationally intense, making EC ideally suited for objects less frequently accessed ("warm" storage). [27]

   Based on the nature of the STORM CLOUDS applications we don't expect stored data access on a regular basis. Hence, there is minimal downside to utilizing erasure coding for cold[17] data, and a significant potential for massive reduction in total cost of ownership[18]. This greatly enhances the archival value of the Swift platform, as it allows us to achieve tape-like costs, and "warm" storage performance characteristics. [27]

b. Another potential benefit is greater durability for EC data. [27]

   For example with a traditional 3x overhead of Swift's replication model, 3x replicas would result in data loss with as few as 3 drive failures (ignoring Swift's built in data protection). At the same time EC data could be structured for example in 30 "slices," any 10 of which could be used to recreate the object. This means that the EC object would require 21 drive failures before the data would be lost, meaning that we are facing a datacenter-level catastrophe with significantly greater odds than the odds of data loss through hardware failure. [27]

3.6.16 Identity Management

Identity management is to manage the life cycle of the user's identity. An identity management (IDM) mechanism can help authenticate users and services based on credentials and characteristics [59] instead of the traditional security tokens. An identity management identifies a particular person on the basis of claimed values (name, email address, credit card number) and prevents unauthorized access of resources. By doing so it helps consumers to make a proactive choice about how and what personal information they disclose.

Customers can either use the offered cloud service provider identity management services or develop an identity management system or adopt open standards available in the market for managing their identity. [74] . The three known identity management systems are discussed in the following paragraphs. These systems give users control of their own identifying information and minimise the personal information required by service providers. They prevent multiple

[16] Probably 1.2x to 1.5x depending on durability requirements

[17] Rarely accessed data

[18] e.g. in an environment where 90 percent of the data is cold, and a 1.2x parity level is acceptable, this would yield over a 50 percent reduction in TCO

organisations from linking together information about specific individuals, and allow users to provide anonymous "credentials" that prove various attributes (such as permission to drive or buy age-related products) without revealing any identifying information. [75]

Regarding both the private (SCP@HP) and the public (SCP@Enter) platform environments, platform administrators use the OpenStack Keystone identity service, for managing who can access the application and which actions each person can perform as well as providing authentication, policy management, registering of tenants and users, tokens for authorization, and policies creation spanning all users and services. Regarding the selected applications, Thessaloniki reported that they have their own identity management system, while Agueda reported that their application is using the external library "Hello"[19] to connect to OAuth 2.0 services.

3.6.16.1 OpenID

OpenID is used for accessing many web applications using only one username and one password. Every user authenticates to an OpenID server to get their OpenID token, which is then used to authenticate them to web applications.

With OpenID users do not need to provide sensitive information such as user name, credit card information, email address etc. to a cloud service provider.

It is a decentralized approach meaning that there is no need for a central authority to approve or register cloud service providers. An end user can freely choose the OpenID Provider (OP) and can preserve their Identifier if they switch OpenID Providers.

However OpenID is highly susceptible to phishing attacks!

3.6.16.2 OAuth

As with OpenID, OAuth is used to grant access to resources that are controlled by multiple applications in a distributed environment.

The increased popularity of social media apps, mobile apps and cloud services has lead to the OAuth standard. In this model, at least three entities are involved: the user, the client application and the service provider. This is referred to as the three-legged OAuth model. The user is the owner of the resource and it grants client application access to its resources that are controlled by the service provider. OAuth standard enables the user to grant client application/service access to its resources without ever sharing its username/password with the client application. [64]

The OAuth standard enables websites to access user profile data outside their domain of control without the need for users pass login credentials (username, password) to the site. [64]

Figure 3-12 shows an OAuth architecture deployment, where we are leveraging a cloud based identity management system (GMail) to control access to STORM CLOUD applications. As the figure illustrates:

1. User attempts to access STORM CLOUDS applications
2. User is redirected to GMail to provide credentials for authentication (email & password)

[19] https://github.com/MrSwitch/hello.js/

3. Upon successful authentication to user is redirected back to the STORM CLOUDS applications with OAuth credentials
4. STORM CLOUDS applications that receive the user's OAuth credentials validate with GMail before granting access to application services

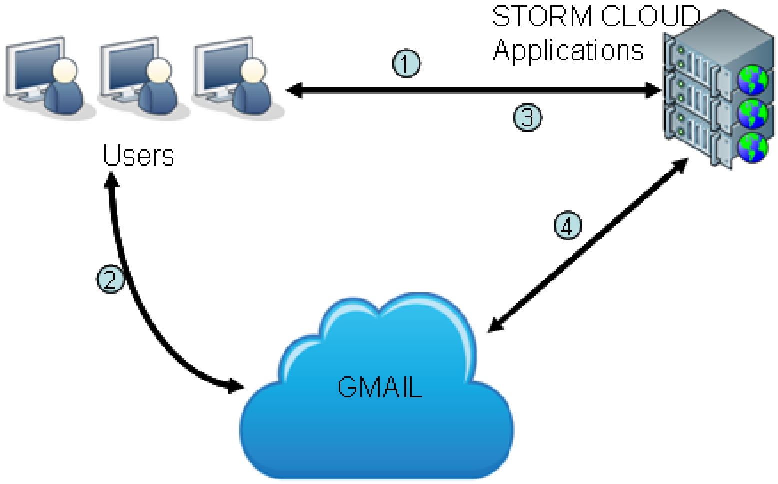


Figure 3-12: OAuth Architecture Deployment

However, according to [64], cloud-based access control architecture poses several challenges:

- STORM CLOUDS applications require modification to be OAuth enabled;
- Over time, scalability becomes an issue. As new applications are deployed, they must be integrated and tested with OAuth, which requires time and resources;
- No centralized monitoring and enforcement is offered;
- Usage of SSL for exchanging credentials with Google cloud, results in performance degradation.

3.6.17 Certification

More on certification can be found in chapter 3.2.2 of D4.3.

3.6.18 Digital Rights Management

According to [56] digital rights management (DRM) is one of the best security technologies to provide the pervasive protection required against insider threats[20]. Traditional IT security technologies have weaknesses and/or limitations. For example:

- Firewalls and intrusion detection/prevention systems (IDS/IPS) are protecting from outsiders but provide little help when it comes to mitigating the insider threat;
- Identity and access management systems (IAM) can distinguish between insiders and outsiders while making clear who the users are and what are their access rights and privileges, but they don't protect data from misuse.

[20] The problem of legitimate users accidentally or intentionally leaking confidential data.

- Strong authentication used for mitigating the issues related to IAM credentials, does not protect from the insider threat, since malicious insiders often use someone else's credentials to cover their tracks.
- Network traffic inspection systems (NIS) are used to detect unwanted attention of outsiders and any data access problems when data is in motion, leaving data unprotected when in use or at rest.
- Data loss prevention systems (DLP) can protect only where deployed, meaning that once files are outside our direct control there is no way to monitor them.
- Encryption can protect information at rest and in transit over networks, but does nothing to protect once a file is decrypted, leaving unencrypted data vulnerable to malicious insiders that can now copy, share, take screenshots, cut and paste, print and so on.

To effectively address the insider threat protection needs to:

✓ be data-centric rather than system-centric,

✓ protect data at rest, in transit and in use,

✓ enforce security policies,

✓ maintain audit trails,

✓ enable legitimate users (insiders) to access data

With a DRM system all documents are classified from the moment of creation and monitored throughout their life cycle. Policy is controlled via an online server, which is referred to each time a sensitive document is accessed and an audit trail of who has done what to a document, and when, is collected and managed. [56] According to a report from Quocirca[21], audit trails are considered as one of the most important security technologies.

A DRM system enables insiders, through granular access controls, to use data safely, ensuring they cannot over-exploit their access rights.[56]

However, there are a number of reasons why DRM is not as widely adopted as other security technologies:

1. Mitigating insider threats is an area that many organizations are satisfied.
2. Misconceptions about DRM capabilities, including scalability, intrusiveness and user acceptance.

At the time of writing the selected STORM CLOUDS applications don't deal with files containing personal data. Hence this approach can be applied once municipalities deploy applications dealing with personal data in order to provide the pervasive protection required against insider threats.

[21] "The adoption of cloud-based services", July 2013, http://quocirca.com/content/adoption-cloud-based-services

4 Future of Data Protection

4.1.1 Business Requirements

Data protection mechanisms must be affordable, usable, and manageable, and must be aligned with and facilitate explicit business objectives. They should make clear which threat they are mitigating and improve user workflow, rather than constraining the user's task or adding significant task overhead.

Enterprise culture must change so that anyone who creates and handles information should quickly realize the inherent value of information and internalize safe handling practices.

Moreover there needs to be a structured entitlement process at a business level. These entitlements, i.e. the decisions governing data access, are often based on corporate cultural habits. A formal process for creating, documenting, and disseminating these entitlements is a necessary input to architecting and designing a set of capable information protection services. [52]

4.1.2 Security Requirements

Protection policies should be consistent throughout the lifecycle[22] of the data and should only change when explicitly and authoritatively requested. If the sensitivity of the data changes during the lifecycle, mechanisms should allow for the protection level to adapt accordingly, without any gaps in the protection enforcement during these policy changes. Ideally the data should also be used in its encrypted form, using techniques such as 'homomorphic encryption'. However, such solutions are not yet matured enough

Protection policies should be the same regardless of its location or environment. However, protection policies should be adjusted as risk, threats, or legal jurisdiction change.

In order to effectively leverage data, these protections must be applied at the data layer itself, with the scope of the protection being specific to the information being protected. Ideally the protection will be part of the data, making it consistent across any environment in which the data resides. [52]

This does not mean that other protection mechanisms should be discarded. On the contrary we should use them for the primary role that where designed. For example perimeter security services should be used for protecting the availability of IT services. Platform security services should be used for providing a safe environment for the use of applications and their related data.

Since data protection systems are being hosted and interacted by other applications and operating system components, we should not trust them equally. Strong authentication trust roots, utilizing trusted hardware and processes should be used, while at the same time being independent of the applications and operating systems. Leveraging a Trusted Platform Module (TPM), smart card, or other secure cryptographic device for controlling the identity of entities accessing the data, we could severely limit the compromises that can come from infiltrating the applications and operating systems hosting the data protection systems.

[22] Includes creation, modification, aggregation, distribution, archive, and destruction

In addition, we should use secure standard protocols, allowing enterprises with dissimilar IT infrastructures to share information securely. We should avoid use of proprietary protection mechanisms as this increases vendor lock-in, decreases interoperability, impedes e-commerce, and weakens security.

To ensure data protection systems integrity they should be administered securely. According to [52] a two-person system could be used, where one administrator creates roles and polices and a different administrator enrols people in those roles. Finally, audit records need to be kept of all access as well as failed access attempts. The logs themselves need to be protected from unauthorized access. Unusual events should trigger real-time alerts.

4.1.3 Architectural Requirements

Data protection systems should support different data types such as:

- Structured data, such as RDBMS data, data managed by complex applications (email, ERP, SharePoint, WebEx, etc.), and data held in sophisticated storage management systems and
- Unstructured data, such as office files, system logs, configuration files, generic text files, and other discrete digital data objects.

Data protection systems should be able to distinguish discrete activities or actions applied to the data and enforce permissions based on those actions. Actions include modification, deletion, read, infer from, distribution, format conversion, archiving, manipulation, etc.

Data protection systems apart from confidentiality, they should support other attributes such as integrity, location, access instances, and evidence of secure deletion.

Finally data protection systems should support standardisation in the areas of:

- Container for encapsulating or protecting the information;
- Programming interface for manipulating the protection around the data;
- Communication of relevant data rights between data consumers and data owners;
- System classification for metadata capturing.

4.1.4 Move from Extrinsic to Intrinsic Data Protection Mechanisms

According to [52] the move involves 3 steps:

1. *Ad hoc* data protection, displaying a varying protection level, mainly affected by the environment or location of the data, meaning that it's up to those entities that create and manipulate the data to apply protection. Unfortunately this is true for the majority of data today.

   When a protection boundary is added to the data, it travels with the data. This is typically some form of encryption system (encryption or encryption enhanced by digital rights technologies) that is often integrated with other non-security systems that are used to manage data.

2. Mandatory Access Control (MAC) data protection that, as already identified in 0, enforces data protection to be applied to the data itself, based on properties of the data, and access request types and entitlements granted to the user of the data.

   This can be made possible through the addition of:

- Metadata or labels, containing data attributes. This gives data access decisions applications some of the information necessary to grant specific accesses to the data. Rights management systems represent the most common products of this type.
- Access rules as part of the metadata, for interrogation by a local PDP. These rules might carry additional requirements for access, such as nationality, project relevance, physical location, enhancing the decision process.

3. Smart data, data protection. Adding intelligence to the data by giving it some measure of ability to participate in the access control decisions, results in what is often called smart data. According to [53] smart data is “Data that remains appropriately protected when outside an entity’s direct locus of control (Jericho Forum Commandment #9) and notes that “it is not sufficient to only establish data as ‘smart’ – we also need mechanisms to verify whether that data remains “smart”, and also to enable smart data to promote awareness that it is “smart”. Put simply, it is data that is enabled to look after itself.

   This can be made possible through the addition of:

   - A signature, used to validate the safety and integrity of the environment before opening up or allowing access to the protected data.
   - A small PDP attached to the data, ideally small enough to not add too much overhead to the data and also to undergo formal testing for proof of correctness. It could be an add-on or plug-in where needed. OASIS XACML, is the only real contender for constructing a portable rules engine.

While all mechanisms expose data to their environment when the data is accessed, the more sophisticated they are the more they need confirmation of the state of the environment before allowing access. *Ad hoc* data protection already relies on the environment and MAC-based policy access control can consider environment security as one of the attributes that the PDP takes into account when determining access, but smart data is more isolated and will require communication with a trusted entity to determine how much to trust the environment. [52]

These evolutionary steps are progressive meaning that in order to implement any given step the previous steps have to be implemented first.

5 Summary and Conclusions

The rapid growth in data volume, complexity of usage, and business information criticality requires that new approaches, risk-based and business-focused, need to be developed for protecting information. They also need to be flexible while providing consistent protection. Today's regulatory compliance and daily reliance upon critical data makes data protection more important than ever. But data protection has become far more complex and sophisticated to deploy and manage because of VM sprawl and the increased transformation of environments from physical, on-premises infrastructure to virtualized and cloud-based architectures.

Several effects and undesirable consequences of cloud computing are still hard to identify. Its complexity and current developments do not alleviate the task. However, this does not mean that the conception of ethical issues in cloud computing and subsequent implementation should be disregarded. Au contraire, they should be engaged since the beginning in order to avoid them at a later stage.

As the report relates to the seventh data protection principle, it is intended to inform municipalities about appropriate measures to safeguard personal data being processed by their applications. The report takes advantage of existing knowledge already available in the field of information security, while specifically focusing on the most significant threats to data protection.

The report provided good practices for how to guard against frequent data protection issues common in online platforms, under the condition that they should be simple to implement. Data protection issues include software updates, SQL injection, unnecessary services, password storage, configuration of SSL and TLS, inappropriate locations for processing data and default credentials

The report also defined some of the things an organisation needs to consider in the event of a security breach, by suggesting the four important elements of a breach management plan. This should assist municipalities in deciding on an appropriate course of action if a breach occurs.

Current information security technologies and practices neither adequately secure information nor allow it to be shared when needed, since they were designed for an earlier era. The result is that we use some combination of the current data protection techniques that come with their associated issues, such as lack of granularity and absence of a comprehensive data protection plan, which is usually too late, costly, and does not scale. Hence new approaches need to be developed for protecting information. These approaches need to be risk-based and business-focused. They also need to be flexible and yet provide consistent protection commensurate with the data value and threat potential. This necessitates both the movement of protection enforcement closer to the data and a shift from user-applied to policy-driven protection models. This must also lay the foundation for the development of smart data.

Key choices in virtual machine backup were presented, since backups of virtual machines using one agent per virtual server is not the most efficient solution, as it can result in contention at the network and storage layers.

Moreover snapshot technology, as a data protection strategy, was analysed. Like all data protection technologies, snapshots have flaws, meaning that snapshot technology is best utilized as a part of a data protection strategy, not as the entire data protection solution.

When municipalities consider adding cloud based backups to their backup strategy, the motivating factors are the lower cost and the convenience of outsourcing at least a piece of the backup problem to someone else. Even so, cloud-based data protection is not always a good thing, and as Phil Goodwin mentioned in [35] we shouldn't expect on-premises service levels at off-premises prices.

Even though implementation and management of encryption is a challenging task, its virtues should not be overlooked because it can help us achieve information security in a cloud environment. To analyse both the protection benefits we'd get from encryption and where those benefits are most needed we should consider performing data classification and service inventorying, in order evaluate the specific usage, and determine whether we will encrypt, and decide how we will implement it. When encryption is used properly, cloud computing should be even more attractive for users concerned about data security. Potential key management solutions were also proposed.

Although it is unlikely that the threats, challenges and best practices discussed here will apply to every cloud-based scenario, we have highlighted some of the considerations required when developing a cloud computing use case.

References

[1] STORM CLOUDS Consortium, "Surfing Towards the Opportunity of Real Migration to CLOUD-based public Services", November 2013

[2] National Institute of Standards and Technology (NIST), Special Publication 800-145, The NIST Definition of Cloud Computing, September 2011.

[3] Data Protection and Data Security Issues Related to Cloud Computing in the EU, Paolo Balboni, Tilburg Law School Research Paper No. 022/2010, August 21, 2010.

[4] Balboni, Paolo, Mccorry, Kieran & Snead, David: Cloud Computing – Key Legal Issues. In: Cloud Computing Risk Assessment. European Networks and Information Security Agency (ENISA), 2009, p. 97 – 111.

[5] Opinion 3/2010 on the principle of accountability, ARTICLE 29 DATA PROTECTION WORKING PARTY, 00062/10/EN, WP 173, Adopted on 13 July 2010

[6] Working paper on cloud computing – privacy and data protection issues – Sopot Memorandum / International Working Group on Data Protection in Telecommunications, 24/4/2012.

[7] Wikipedia. (2015, July 2). Ethics. Retrieved from Internet Encyclopedia of Philosophy: https://en.wikipedia.org/wiki/Ethics

[8] Manuel Velasquez, C. A. (2014). What is Ethics? Retrieved from Santa Clara University: http://www.scu.edu/ethics/practicing/decision/whatisethics.html#sthash.2ofmpBoS.dpuf

[9] Timmermans, J., Stahl, B. C., Ikonen, V., & Bozdag, E. (2010). The Ethics of Cloud Computing: A Conceptual Review

[10] Haeberlen, A. (2010). case for the accountable cloud. *SIGOPS Oper. Syst.*

[11] Kandukuri, B. R., & Rakshit, A. (2009). Cloud Security Issues. *IEEE international Conference on Services Computing* (pp. 517-520). Washington: IEEE Computer Society

[12] S. Paquette, P. T. (2010). Identifying the security risks associated with governmental use of cloud computing. *Government Information Quarterly*

[13] Pieters, van Cleeff, A. a. (2009). Security Implications of Virtualization: A Literature Study. *IEEE International Conference on Computational Science and Engineering* (pp. 353-358). Vancouver: IEEE Computer Society.

[14] Murley, D. (2009). Law Libraries in the Cloud. *Law Library Journal*

[15] Nelson, M. R. (2009). The Cloud, the Crowd, and Public Policy. *Issues in Science and Technology*

[16] B. C. Stahl, R. H. (n.d.). *Identifying the Ethics of Emerging Information and Communication Technologies: An Essay on Issues*, Concepts and Method International Journal of Technoethics

[17] Encryption in the cloud – The challenges of providing effective encryption to protect data stored in the cloud, Daniel Cuschieri, MSc in information security (Royal Holloway) and Po Wah Yau, ISG, Royal Holloway

[18] Data Integrity Verification in Cloud Storage without using Trusted Third Party Auditor, Rana M Pir, 2014 IJEDR, Volume 2, Issue 1

[19] Protocols for Secure Cloud Computing, School on Applied Cryptographic Protocols, Christian Cachin, IBM Research – Zurich

[20] Cloud Computing Security using Federated Key Management, International Journal Of Engineering And Computer Science ISSN:2319-7242, Volume 3 Issue 2 February, 2014 Page No. 3978-3981

[21] Doing it right: Cloud encryption key management best practices, Dave Shackleford, SearchCloudSecurity

[22] Do you need a hardware security module to protect your information?, Chris Moyer.

[23] Cryptographic Key Management – Issues & Challenges in Cloud Services, Ramaswamy Chandramouli, Michaela Iorga, Santosh Chokhani, NISTIR 7956, September 2013.

[24] Cloud tokenization: Why it might replace cloud encryption, Dave Shackleford

[25] Erasure Coding, SearchStorage

[26] Erasure Coding for Cloud Storage Systems: A Survey, Jun Li and Baochun Li, TSINGHUA SCIENCE AND TECHNOLOGY, June 2013

[27] Object Storage Tomorrow: Erasure Coding, Jonathan Kelly, April 2014, Rackspace – Cloud Industry Insights

[28] Replication & Erasure Coding Is the Future for Cloud Storage & Big Data, Paul Carpentier, June 2013, Cloud Computing Journal

[29] Can cloud providers' disaster recovery capabilities stand up to critical apps?, Stephen J. Bigelow

[30] Cloud data backup is not cloud disaster recovery, Biff Myre

[31] cloud backup (online backup), Margaret Rouse, February 2013

[32] Why You Need Backup and Disaster Recovery, N-able Technologies, white paper, July 2014

[33] How virtual machine replication to the cloud is changing data protection, Brien Posey, July 2014

[34] Key choices in virtual machine backup, Chris Evans, computerweekly.com June 2014

[35] The pros and cons of cloud-based backup services, by Phil Goodwin

[36] The pros and cons of cloud-based backup, by Colm Keegan.

[37] Update your approach to virtual server backups, by Stephen J. Bigelow, October 2013

[38] data protection, by Margaret Rouse

[39] Challenges with data protection in the cloud, by Dave Shackleford, Voodoo Security

[40] Crafting a secure data backup strategy on a private cloud, by Dejan Lukan, October 2014

[41] The truth about using snapshot technology as data protection, by Marc Staimer, August 2014

[42] Cloud-based data protection not always a good thing, by Brien Posey

[43] Challenges in virtual server backup and how to tackle them, by Chris Evans

[44] Protecting Personal Data in Online Services: Learning from the mistakes of others, UK Information Commissioner’s Office, May 2014

[45] Algorithms, Key Sizes and Parameters Report 2013 recommendations, ENISA, version 1.0 – October 2013

[46] Recommendation for Password-Based Key Derivation Part 1: Storage Applications, NIST, December 2010

[47] Only 1 in 100 cloud providers meet proposed EU Data Protection requirements, August 2014, net-security.org

[48] Guidance on data security breach management, UK Information Commissioner's Office, December 2012

[49] The benefits of encryption key rotation for cloud users, by Dave Shackleford, November 2014

[50] Data Protection - Problem Statement and Requirements for Future Solutions, Jericho Forum®, 2012

[51] Data Security vs. Data Protection, http://www.skullbox.net/data-security_vs_data-protection.php

[52] Protecting Information – Steps for a Secure Data Future, A White Paper by: Members of the Security Forum, a forum of The Open Group, January 2014

[53] Smart Data for Secure Business Collaboration, A White Paper by: Members of the Security Forum, a forum of The Open Group, January 2014

[54] Data security in cloud computing: Data encryption controls, Ed Moyle, SearchCloudSecurity

[55] Security Think Tank: Three-pronged approach to cloud security

[56] The insider threat: solved with DRM, Quocirca Ltd, June 2014

[57] 5 Strategies for Modern Data Protection, CommVault, 2014

[58] "Cryptographic Key Management Issues & Challenges in Cloud Services", Ramaswamy Chandramouli, Michaela Iorga, Santosh Chokhani, NISTIR 7956, September 2013-NIST

[59] "Security and privacy challenges in cloud computing environments", Takabi, H., Joshi, J. B., & Ahn, G. J.

[60] Bhargav.B et.all. (2011) "Privacy in Cloud Computing through identity Management" Electrical and computer engineering, Purude University.

[61] PRIME Framework V3, https://www.primeproject.eu

[62] PRIME White Paper V3, https://www.primeproject.eu/

[63] OpenID Explained, http://openidexplained.com/

[64] "Cloud-based Enterprise Identity Management using OAuth", Forum Systems, White Paper, 2014

[65] "Top five issues in VM backup", Antony Adshead, January 2015

[66] "The guide to data protection", Feb 2015, UK Information Commissioner's Office

[67] "ICO outlines technical detail of IT security best practices", May 2014, out-law.com

[68] "Top 10 2010-A9-Insufficient Transport Layer Protection", OWASP Top 10 Application Security Risks - 2010

[69] "Data security in cloud computing: Data encryption controls", Ed Moyle, 2014, SearchCloudSecurity

[70] "Privacy-Preserving Public Auditing for Secure Cloud Storage", Cong Wang, Student Member, IEEE, Sherman S.-M. Chow, Qian Wang, Student Member, IEEE, Kui Ren, Member, IEEE, and Wenjing Lou, Member, IEEE

[71] "Cloud Encryption: The Pros of Encryption and Tokenization", ciphercloud, 2013

[72] "Tokenization vs. Encryption – Which One is Best for Your Business?", Alex Pezold, 2013, TokenEx News

[73] "Identity Management issues in Cloud Computing", Smita Saini, Deep Mann, 2014

[74] "Cloud Computing – Managing Security", Harpreet Singh Sidana, 2012

[75] "The challenges to European data protection laws and principles" Ian Brown, Oxford Internet Institute, University of Oxford, Working paper, 2010

[76] "Data Privacy & Compliance in the Cloud", Perspecsys Whitepaper, 2014

[77] "Encryption Solutions in Cloud", Rafeeq Rehman, April 2013, Verizon Cloud

[78] "MySQL in the cloud: database encryption options and cloud encryption", Porticor, 2012

[79] "The accountability principle in data protection regulation: Origin, development and future directions", Joseph Alhadeff, Brendan Van Alsenoy and J. Dumortier