Aalborg University Copenhagen
A. C Meyers Vænge 15
2450 København SV
Secretary: Maiken Keller
Telephone (+45) 9940 2471
mks@staff.aau.dk

Title:
Privacy-aware Identity and Access Control System for the Enterprise

Theme:
Master Thesis

Project period:
September - December 2018

Project group:
Group ICTE4SER

Member:
Stefan Reinholdt Jørgensen

Supervisor:
Henning Olesen

No. of Pages: 88

No. of Appendix Pages: 9

Total no. of pages: 119

Finished: December 6, 2018

Abstract:

In the project, the concept of privacy-enhancing identity and access management technologies is explored. In the report, several of such technologies are identified, described, analysed for use in an enterprise context where digital interactions with individuals are present. The goal is to establish if such technologies can enhance the individual's privacy in this relation. The main goal of the problems is to establish if such technologies are feasible to use in this context and how they can be utilised. The method to conclude if this is possible so with the use of a series of scenarios. In the report two scenarios with several use cases are created, where interactions between individuals and enterprises exist. These scenarios are used in the implementation to showcase a feasible solution for the specified problem field. The analysis is consequently done only according to the specified use cases and may be considered inadequate for a practical, real-life implementation. Furthermore, the proposed architecture is made to demonstrate the solution. Overall, the goal is not an actual design of a system that can be used in practice, but a study showcasing the usefulness of such technologies for this context.

When uploading this document to Digital Exam each group member confirms that all have participated equally in the project work and that they collectively are responsible for the content of the project report. Furthermore each group member is liable for that there is no plagiarism in the report.

Aalborg University

Master Thesis

Privacy-aware Identity and Access Control System for the Enterprise

A thesis submitted in fulfilment of the requirements for the degree of M.Sc. In Engineering of Innovative Communication Technologies and Entrepreneurship

in the

Center for Communication, Media and Information Technologies
Dept. of Electronic Systems

December 6, 2018

AALBORG UNIVERSITY

Abstract

Technical Faculty of IT and Design
Dept. of Electronic Systems

M.Sc. In Engineering of Innovative Communication Technologies and Entrepreneurship

Privacy-aware Identity and Access Control System for the Enterprise

by Stefan Reinholdt Jørgensen

In the project, the concept of privacy-enhancing identity and access management technologies is explored. In the report, several of such technologies are identified, described, analysed for use in an enterprise context where digital interactions with individuals are present. The goal is to establish if such technologies can enhance the individual's privacy in this relation. The main goal of the problems is to establish if such technologies are feasible to use in this context and how they can be utilised. The method to conclude if this is possible so with the use of a series of scenarios. In the report two scenarios with several use cases are created, where interactions between individuals and enterprises exist. These scenarios are used in the implementation to showcase a feasible solution for the specified problem field. The analysis is consequently done only according to the specified use cases and may be considered inadequate for a practical, real-life implementation. Furthermore, the proposed architecture is made to demonstrate the solution. Overall, the goal is not an actual design of a system that can be used in practice, but a study showcasing the usefulness of such technologies for this context.

Contents

List of Figures

List of Tables

List of Abbreviations

ABAC	**A**ttribute **B**ased **A**ccess **C**control
AC	**A**ccess **C**ontrol
ACM	**A**ccess **C**ontrol **M**odel
API	**A**pplication **P**rogramming **I**nterface
APT	**A**dvanced **P**ersistent **T**hreat
AS	**A**uthentication **S**erver
CA	**C**ertificate **A**uthorities
CA	**C**ertificate **A**uthority
CRL	**C**ertificate **R**evocation **L**ist
CRM	**C**ustomer **R**elationship **M**anagement
CSP	**C**redential **S**ervice **P**rovider
DAC	**D**iscretionary **A**ccess **C**ontrol
DID	**D**ecentralised **ID**entifiers
DLT	**D**istributed **L**edger **T**echnology
DMS	**D**ocument **M**anagement **S**ystem
DNS	**D**omain **N**anme **S**ervice
DS	**D**irectory **S**ervice
EIdMS	**E**nterprise **I**dentity **M**anagement **S**ystem
EPR	**E**lectronic **P**rivacy **R**egulation
ERP	**E**nterprise **R**esource **P**lanning
GDPR	**G**eneral **D**ata **P**rotection **R**egelation
IAM	**I**dentity and **A**ccess **M**anagement
IDM	**I**dentity **M**anagement
IDP	**Id**entity **P**rovider
IEEE	The **I**nstitute of **E**lectrical and **E**lectronics **TE**ngineers
IETF	The **I**nternet **E**ngineering and **T**ask **F**orce
MAC	**M**andatory **A**ccess **C**ontrol
NIST	**N**ational **I**nstitute of **S**tandards and **T**echnology
PA	**P**ermission **A**ssignment
PA	**P**ermission **A**ssignment
PDP	**P**olicy **D**ecision **P**oint
PGP	**P**retty **G**ood **P**rivacy
PKI	**P**ublic **K**ey **I**nfrastructure
POS	**P**oint **O**f **S**ale
RA	**R**evocation **A**uthority
RA	**R**egistration **A**uthority
RBAC	**R**ole **B**ased **A**ccess **C**control
RDF	**R**esource **D**escription **F**ramework
RFC	**R**equest **F**or **C**hange
RP	**R**elying **P**arty

RQP	**R**e**q**uesting **P**arty
RS	**R**esource **S**erver
SAML	**S**ecurity **A**ssertion **M**arkup **L**anguage
SOD	**S**eparation **O**f **D**uty
SP	**S**ervice **P**rovider
SSL	**S**ecure **S**ockets **L**ayer
SSO	**S**ingle **S**ign-**O**n
TCB	**T**rusted **C**omputing **B**ase
TCP	**T**ransmission **C**ontrol **P**rotocol
TLS	**T**ransport **L**ayer **S**ecurity
UA	**U**ser **A**ssignment
UA	**U**ser **A**ssignment
UML	**U**nified **M**odeling **L**anguage
VA	**V**alidation **A**uthority
VoIP	**V**oice **O**ver **IP**
XML	E**X**xtensible **M**arkup **L**anguage

Chapter 1

Introduction

In today's constant-changing business landscape there exists a massive focus on Information Technology and digitalisation in nearly all areas and industries. Companies are forced to rethink how their IT-strategy and business model correlate frequently. In order to be competitive and navigate safely in a world where traditional businesses are disrupted by newcomers, due to advancements in technology that can deliver better service and products. Customers, partners and employees expect and increasingly rely upon well working IT solutions as an absolute necessity. This tendency has moved IT and digitalisation from a support function for the organisations into being a vital part of the enterprise's strategy with the introduction of positions such as CIO, CPO and CISO that are shaping decisions and strategy with an understanding of relevant technology[11]. IT, Privacy and Cybersecurity concerns are complex and at the same time mandatory to understand if a company want to succeed in the Digital realm. This trend is lead by a new generation of companies that have focused on disrupting traditional markets with new technologies. Examples of this is Internet of Things (IoT) for remote monitoring and control, Social media(SoMe), Robots in both physical and virtual forms are going to substitute or assist manual labour and first and foremost the introduction of cloud computing service models. This is a subset of technologies that paves the way in order meet customers demand, deliver better services with lower costs. These technologies have helped new and technology-focused companies disputing the traditional business models and in many cases substantial part of the market. Even though usage of such technologies has been able to bring something new to the table, it has often been with a compromise with for the endusers privacy.

Many of such companies have built their business model with a combination of increased automatisation and collection of data about user identities, attributes and interactions at their core. With this data in their custody, it is possible to create automatised customised services but has left the user with either no or little control and knowledge about which information is collected, spread or sold to third parties. Examples of such companies include international giants like Google, Facebook, Uber, Microsoft, Expedia and Airbnb. However, also national examples such as MobilePay, COOP, NETS and Nordea has considerable amounts of data about their users created as a by-product of their service[12, 13, 14].

Problems with storing large amounts of personal data in these service silo's are numerous. First, it might violate the privacy of users, that employees at such companies have access to sensitive information.
Subsequently if one of these companies fail to protect the information that they have stored, by getting hacked, or carelessly handle the data, it ends up getting leaked. Both cases might be able to reveal sensitive information about the affected users that are irreversible.
Recent scandals have shown that such collection of personal information combined with lousy cybersecurity has led to a loss of privacy, not only for the few but the masses. One of the reactions due to the increased compromised user privacy is strengthened privacy laws. In Europe legislations such as General data protection regulation (GDPR) and the upcoming E-privacy directive has come into action that regulates the level of personal data that enterprises can use, save, distribute and handle. These laws set high and new requirements for businesses which want to deal with personal identities and information.

Many of the earlier mentioned companies have built their solution around the internet. However after 50 years since the internet was invented [15], and with more then 45% of the world connected to it [16], there is still not solution that solves the problem with distribution and validation of identities online. In the offline world, a person can easily present his verifiable identity with the use of passport, driver license or birth certificate. However, in the online world, these entities do not exist which is the problem, that this thesis focus on and it has been around for as long as online commerce.

This chapter will contain an introduction to the basic principles regarding how enterprises deal with identity, access control and privacy. The problem field highlighted in this section will be further analysed and a problem statement will be presented.

1.1 Identity and Trust

Identity is a basic human characteristic, which embraces all elements that make each human distinct from each other[17]. Furthermore, all characteristics that indicate membership to a particular group or culture, establishing the status within that group, should also be seen as a part of a persons identity [17]. Defining identity is not a tedious task since different concepts exists within various scientific disciplines. Besides the definition used by security researchers, which this project is based on. Several philosophical researchers have tried to model how identities are perceived or comprehended[17].

One of these ways of defining identity in philosophy is described by Ricoeur [18]. He refers to these two aspects as **ipse** and **idem**.

The **Ipse** "I" is identity recognised as self hood, an identity that is close to our individual, a kind of inexpressible inner core that labels us as what we are[19]. **IIdem**, on the other hand, is the "Me" identity understood as unity, it is a more outside possibility of identifying the self as self despite the loss or mutability of the attributions of that self in time. Ricoeur further proposes that **Ipse** identifies the "who", the self is and **idem** the "what", the self is made of[19]. His way of modelling identity can be separated into the following three fundamental perspectives[1]:

- ***The "I"*** *The indeterminate first-person perspective*
- ***The Implicit "Me"*** *How the person perceives herself*
- ***The Explicit "Me"*** *How this person is perceived and represented*

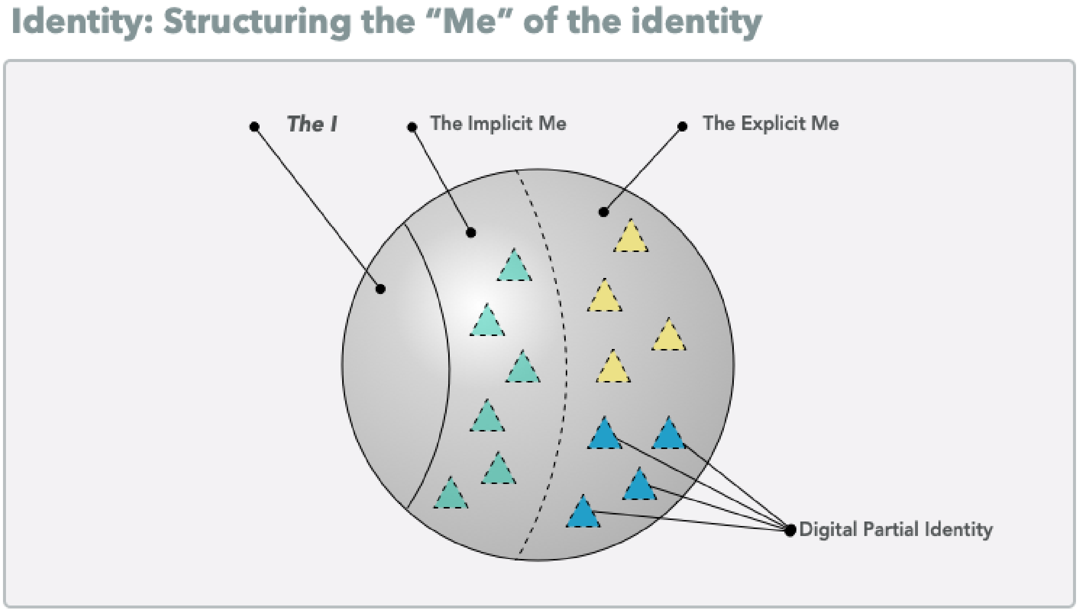


Figure 1.1: Identity: Structuring the "Me" of the identity proposed by [1]

This separation concept is illustrated in figure 1.1. The figure depicts a partial digital identity, that point to a subset of the whole identities attributes, this is the identity that is usually referred to in digital research. Several things can be concluded based on this theory of identity. This separation address the imperfection of identities. Identity will always be a manageable subset of all of a person's attributes. As the physical identity evolves throughout the life of any entity. There exists a technical challenge in this imbalance between recorded attributes and the physical attributes as well as one between the implicit and explicit attributes. Tools that can align and correct this imbalance need to be found in order for a system to solve the identity problem[8].

1.1.1 Control of Identities

In the previous section, the psychical identity and how implicit and explicit identity map to each other and their relation to digital identities was described. However, this separation also introduces the question of whom that control a person's identity. Examples of external attributes that are provided by third-parties such as governments; are identifiable person numbers such as the Danish CPR or Tax identification numbers (TIN) used in many countries to report earnings. Other entities such as companies or persons can also give external attributes. This can be customer numbers, bank accounts or nicknames[17]. For these external controlled attributes, the person only has limited power to change them. Several of these attributes are governed by laws, which restrict or control the rights to access or change. For example, legal names, where the law directs the naming of persons[1] or the controversial attribute gender that points to a social construct that might not necessarily be the same as the person's sex attribute, but still has two distinct set of rules that dictate change[2]. This is two examples where individual does not control all of its attributes. In the privates sphere external attributes without control also exists. Another example is where the individual have limited control over their own identity is companies which save records of their business transaction. These can either be due to convenience of marketing, analytics or due to being regulated by law. In these examples the users might not be informed about which information the business store about them.

In Duran article [22], he suggest a separation of these internal and external attributes, can be used to categorise various attributes based on the power of control. He suggests that attributes should be divided into three tiers of identity[22] [17]:

- ***Tier 1*** *- "My" Own Identity (Personal)*
  The real personal digital identity that is controlled entirely by the person to which they belong, for his sole purpose or benefit. These attributes are timeless and unconditional, and might exists for people as well as devices, programs or objects.

- ***Tier 2*** *- "Our" Assigned Identity (Corporate)*
  Corporate or enterprise established the identity as a relation to a personal identity. These typically include names, titles, roles, customer and employee-numbers or phone numbers. This kind of identities are conditional and temporarily assigned to a particular context like a customer relation or an employee relation. Information can be revoked by the controller and/or the person itself. This includes both identifiable attributes and the ones which are not.

- ***Tier 3*** *- "Their" Abstracted Identity (Marketing)*
  An abstracted identity that identifies persons through his demographic

[1] Navne loven[20]

[2] Lov om Det Centrale Personregister[21]

and other reputation like attributes, but does not need to do so in a one-to-one manner. These identities are diffuse and a result of profiling. Companies aggregate attributes into categories or personas for the purpose of advertising, communication or commerce typically based on previous interactions. An example of this is CRM systems and most online advertising.

Usually, enterprises only deal with Tier 2 attributes, which is used for identification or authorisation purposes. However, this categorisation of attributes, add yet another aspect of identity.
The Tier 3 attributes which the person might not even have any knowledge about. Attributes that are a result of profiling by third-party. These attributes are in general weak because they are typically collected without precise knowledge about the subject or a person can be identified to be part of a group, in marketing terms a so-called Buyer Persona which is defined as "A persona is a semi-fictional representation of your ideal customer based on market research and real data about your existing customers"[23].
Another way of categorising identities is presented by Satchell [24] that suggests that attributes can be summarised into the following five categories:

- **Intrinsic** Biometrics, Hair color, Genome
- **Descriptive** Name, Place of birth, Birthday
- **Demographic** Occupations, Gender, Age
- **Geographic** Home address, Work address,
- **Psychographic** Preferences, Intrests and Hobbies

These two ways of categorisation gives a reasonably good way of defining identity attributes, which allow us to precise control who that control them, and what kind of information that attribute is carrying. In terms of storing in regards to regulations, this categorisation gives us a universal language that we can communicate what kind of attributes that is stored about an identity.

1.1.2 Identities in the digital realm

The term Digital identity has been defined by various academics, such as Pato [25] who states that *"Digital Identity is defined as a set of attributes related to an entity in a digital entity"* and Chadwick [26] states *"A set of characteristics or attributes that can uniquely identify and distinguish one entity from another in a given context"*. Last, Kim Cameron defines digital identity as: *"Digital identity is a set of claims made by one digital subject about itself or another digital subject"*[27]. All of their definitions are not completely complementary. Pato and Chadwick both talk about the characteristics of the identity, where Cameron talks about the trust which is necessary in the digital communication.

Cameron's definition is given in his paper "The Laws of Identity" where the focus is the lack of an identity layer for the internet that has led to numerous *"identity one-offs"* as he describes the identity solutions on the internet. As a solution he promotes the *"claim"* which is *"...an assertion of the truth of something, typically one which is disputed or in doubt"*[27]. In the EU project "Future of Identity in the Information Society" (Fidis) Rannenberg et al. gives a more generic term that defines the digital identity as[8]:

"Digital identity refers to the representation of the identity of a person in digital environments, in particular in terms of the representation of the characteristics (values associated to a set of attributes) of the person. The digital identity includes both the explicit representation of the person (such as name, age, email, etc.) and implicit representation of the person."

Where Cameron's definition is more focused on the assertion of attributes, it seems to be grounded in what Durand defines as a Tier 2 identity attributes and the problems that arise in digital communication. In opposite Rannedberg et al. definitions cover all Durand's tiers and fit better with the definition from the physiological science domain given by Ricoeur. However, even though these definitions are not alike, they are not conflicting. Therefore in this project, Rannenberg's definition will be used as a reference in combination with the claim based identity raised by Cameroon.

1.1.3 Usage of the digital identities

In a computer system, this digital identity enables a computer system to authenticate the user and provide access to the tailored information, being content based on profiling (Tier 3) or content based on access control (Tier 2). This could be to prove to be of a certain age, or being a specific person, or assigned a specific role. Therefore the digital identity is not a goal by itself, but it is a tool needed in order to provide this basic level of trust between computerised systems in relation to entities in the real world[7].

Whereas the concept of identification is static, identity is altogether more dynamic because it settled and underpinned by trust[28]. Trust is a construct that is defined as *"A firm belief in the truth or ability of someone or something to perform a task in a reliable matter, in a specific context."*[29] As stated this is an essential construction which is used in almost all perspectives in society. Without it, cooperation between people and enterprises could not exist, and trade and commerce would be impossible[30]. Accordingly, Identity solutions are necessary as a part of any access control procedure in computer systems. A digital identity system refers to such system which stores and manage these individual identities, throughout a life cycle[31]

It can be concluded that identity is not just a way of identifying a unique person, it is a complex construct used in many domains. A personal identity consist not just of the name they were given and a series of associated objective facts, but also what a persons think he or she is and what others think about them both the good and the bad.

1.2 Privacy

For a business to create the best possible service, products are customised and personalised based on feedback and profiling. Service experiences are tailored based on previous preferences and profit maximisation[32]. Interaction records of everything from mail and calls are kept to raise service levels. This is especially present in online services, where every interaction and transaction can easily be recorded by default. Users create accounts at the respective service provider to make use of their service. In the registration process, the user provides a vast amount of information which can be used by the service provider to create such enhanced services or tailored customer experiences. Some services might also be able to use the users interactions with the products to profile the user further. All these possibilities come at the price of privacy for the user, a price which can seem to be hard to measure and is not necessarily paid immediately. The general tendency seems to have been accelerated by the technological advance in the last decade or two. The benefit of storing massive amounts of information which later can be used to create analysis based on machine learning (ML) and artificial intelligence (AI), seems to have surpassed the cost and have created an enormous market for data-driven marketing, product development and personifications.

The result of placing detailed information about users in a single place or silo is that users tend to forget to correct and update their data. This result in deprecated information that can lead to exposure, for example when users forget to change contact details such as phone number or address. A problem that is amplified the more details is spread around at various services. Tens or hundreds of services suddenly has detailed information and a few of these joined together might be enough to reveal a good picture of a person, high chance of forgetting which information is placed where. With no rules or regulations, detailed information represents a valued that that is significant and could be sold or shared. For every service that the user utilises, it increases the chance of exposure of personal information when the users give away information about themselves and their whereabouts, purchases or usages.

The result of these resource silos and the lack of transparency into them results in poor privacy for users of the computer system using today's internet. Therefore it can be said that the user is trading their data as part of their transaction to obtain access to these customised interconnected services. Some services have chosen to monetize on this behaviour and actively make use of this as part of their business model. This behaviour is especially present in relation to Multi-sided business models in the tech industry because *"the Internet has allowed many companies to provide valuable services to consumers for free by charging businesses who also benefit from the use of the platform."*[33]. This can be said about some of the internet biggest players like google, Facebook and Spotify, which have been caugt selling information about their users as part of their business[34]. This is backed by the famous quote *"If you are not paying for it, you're not the customer; you're the product being sold"*[35] This problem further grows in relation to IoT devices, as these devices often make use of uncontrolled sensing of their environment. In the last couple of years, legislation has been put in place that establishes and strengthen the rules regarding the handling of private information for some domains within the EU. How can this problem be solved, how can IT systems deal with identifies and access control in relation to employers and customers without impairing their own security and at the same time cater for their own product.

1.3 Identities in the enterprise

Enterprises come in all sizes ranging from the start-up, large cooperates to government controlled projects. Controlling the enterprise user's access is a significant challenge with the number of applications and the diversity of IT systems in today's organisations. A trend that does not seem to stagnate provisionally. Indeed, it is not just the growing volume and complexity of many IT systems that create these problems. Many enterprises are already in a position where the complexity, dynamic and volatile structure of their organisations are so demanding that controlling not only user's access rights but also figuring out whom that control them can be a significant challenge in itself. These circumstances make it extremely challenging to prove that members of an organisation only can access just those data that they are supposed to, being confidential private or company information and data[17].

The most common problems are people having access to data, they should not have. Data being confidential might reveal secrets, being exposed might create problems with privacy regulations or similar compliance's. Larger enterprises usually deal with an extensive amount of stakeholders including employees, customers, suppliers, consultants and partners. To run the business, the underlying organisation make use of a wide range of services and applications, where data flow around in between.

All of these holds information that is relevant to one or more stakeholders to facilitate their job. These stakeholders have different positions with a certain degree of responsibility. However, these organisations structures are vibrant and change over time. Departments merge or split. People move around in the organisation enters or leaves all in dynamic patterns, which also need to be supported by the systems. Stakeholders of the enterprise continuously need access to information within the organisation, all to fulfil their jobs. Do the computer system fail and prevent access to needed information. One thing is for sure, an immediate cost of unproductivity arises for the affected group of users. For the system to provide access to some resource, the identity of the principal has to be authenticated, a crucial process that aims to prove that the claimant is whom that he says he is.
This is the first place where the element of identity meets access control, and one of the places where this report will investigate, how these work together, in order to design a solution that take privacy into account for various scenarios.

In the past, the local IT department or function has taken care of the organisations IT, this being critical enterprise applications or data storage. In today's enterprises, these duties are turning to third-party providers of cloud applications. First because the ability to quickly scale capacity to the actual demand. Second due to the ability to outsource the capital costs and staff expenses of support staff, that include application support and maintenance[36]. The conventional security model for internally hosted applications and systems is relatively straightforward with all users in the organisation authenticating against a single directory service, such as Active Directory (AD) from Microsoft or similar service. This typically means that control over identities and their attributes inside the organisation is put into the hands of the IT department. Central Identity management systems (IDM) is typically first implemented when the enterprise have reached a certain size, where organisation differentiation or size is reaching a point where not everybody knows everybody or in environments where security policy is tight. As enterprises move towards using cloud applications, data get spread around online resulting in data residing outside the local network and firewall controlling security becomes problematic, and tighter control is necessary.[36].

This change and acceptance of applications outside the network perimeter have increased revenue for many organisations by simplifying access to partners, customers and services through the internet. It is mainly the use of social media logins which laid the foundation and simplified the sign-up processed and increasing adoption rates[36]. This include services such as: Facebook Connect, Sign in with Google or Linked In. However this have also put identity and knowledge in the hands of these SoMe-companies, that are now able to profit on this knowledge by selling identity data to others without control leading to lack of privacy for both customers and employees.

Initially, Cloud-based applications have the fundamental problem not having access to an organisations data being internally or at another cloud provider. They need to make use of their proprietary directory, that takes care of user authentication and authorisation. Given the amount of cloud services typically being used within an organisation, this leads to identity fragmentation, duplication and inadequate security. That again leads to lower productivity, loss of revenue and thereby growth. Current best practises is to develop an appropriate strategy for adopting Federated identities or Single sign-on (SSO) is a crucial challenge for organisations which wants to get the most out of Cloud computing in relation to both internal and external users [36, 37].

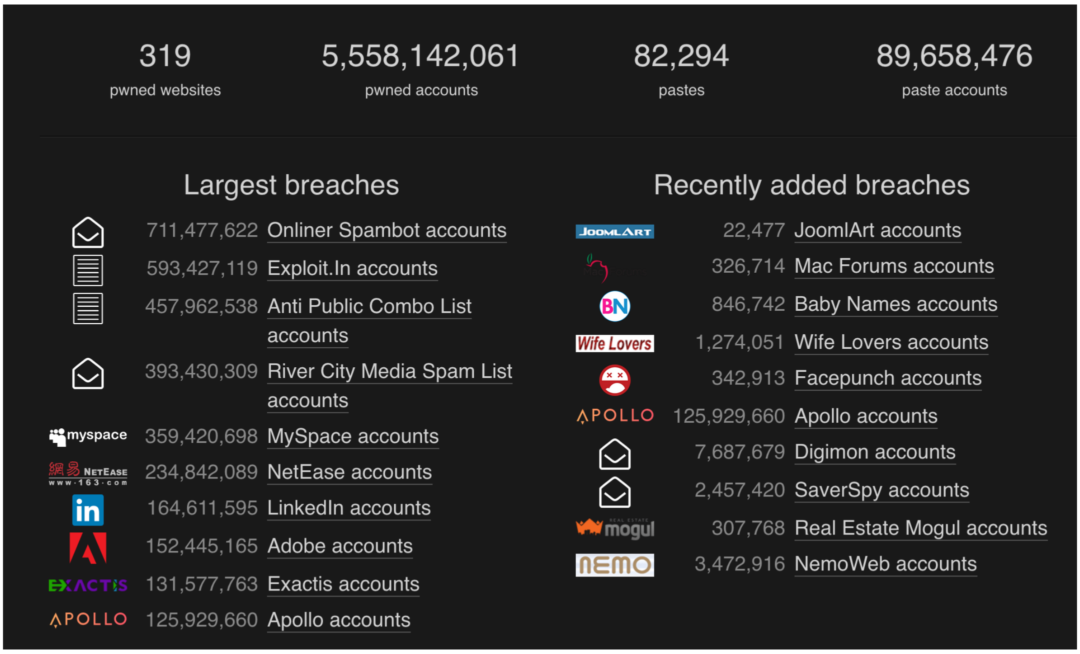
Employees in the most enterprises typically make use of a more or less static collection of applications, typically hosted within the organisation. This has to some extent changed the last 5 years, where more software has moved to the cloud[38]. Nevertheless, both models have been able to utilise federative identity technologies in combination with SSO, to hinder identity silos because preceding trust between the organisation and the service providers have been established. Identity solutions for customers face other problems. Customers or clients are often limited to access to one or a few services offered by the enterprise in order to interact and use their service. Such solutions are typically solved in large solutions know as customer Identity and access management or CIAM[39]. However, these only solve the management problem for the companies, but still leaves the customers with scattered identities around at various services and no way to see which data is made available to who. Reality is that many applications both internal and externally still hold their user database as independent silo's, which is not connected or managed by a centralised solution like IAM, CIAM or easily controlled by the IT-departments. This is not just a problem for the enterprise, but also a problem for the users as described in the privacy section.

With the introduction of collaboration and information sharing across organisations, many employees, partners, suppliers need access to resources outside of their own organisation. Conventional solutions do not support this workflow and typically results in being printed in dated revision or duplicated documents being uploaded to various storage solutions. For example, when an organisation uses cloud hosted software-as-a-service (SaaS) systems like CRM, ERP, POS or DMS solutions. Users who access these applications have enterprise employee identities. If their employment is terminated, so should their access to the SaaS solutions should be as well. In many SaaS services access is often managed in a separate identity, which is not connected with their organisation's identity management system, if they even have one in the first place. This represents a significant security challenge for the enterprise [40].

The other easy solution to sharing data or documents are just sending them over classical mail, a tool still often used in most cases.
However, the current mail infrastructure is still the same as it were 50 years ago and it has not changed much since. At that time security was not a design criterion and security was implemented by knowing everybody whom the equipment was connected to. Nevertheless, this fact that mail isn't secure, do not hinder most persons sending out confidential material. This poses another great thread to the organisation, due to the loss of confidential material, but also breach of regulations. The main reason for these problems is the external organisation managing these resources do not manage the identities of the users which have the needs[40]. State of the art security standards such as Federated identity protocols, SSO, Attribute Based Access Control (ABAC) and User managed Access (UMA) make it possible for organisations to exchange identities and authorisation decisions in a system that allows them to talk effectively with each other. It allows employees, customers, and partners to access Cloud or internal corporate applications using a single set of credentials and at the same time keeping a tight grip around access control[40].

1.4 Current problems and solutions

Lately, there have been a lot of privacy or cyber-security breaches in the media. In some of these cases persons, employees and customers have had their private information or identity stolen resulting in the loss of privacy. Many of these incidents can be related to abuse of technology, irresponsible setup or careless behaviour. Such cases include the Facebook-Cambridge Analytica scandal[41], Google Plus [42], Equifax Data Breach[43], Ashley Madison Breach [44] and many others.
This is further straightened by the fact that it can be hard or almost impossible for the implicated persons to figure out if their accounts have been compromised or not. One of the few places that can help users about this is the website haveibeenpwned.com which is powered by the security professional Troy Hunt. The website aims to create a database of accounts which have been leaked, allowing anybody to submit their mail address and get to know if this address have been in any known breaches.
In figure 1.2, a screenshot from haveibeenpwned.com showing recent and largest breaches recorded on the site. In the figure the pastes indicate the individual leaks which have been identified. In these breaches, digital identities or private data have been exposed to external stakeholders that have either been able to monetise on it or led to severe loss of privacy for the implicated. Likewise, there have been a massive amount of data-breaches from several large companies resulting in millions of login



319	5,558,142,061	82,294	89,658,476
pwned websites	pwned accounts	pastes	paste accounts

Largest breaches

711,477,622	Onliner Spambot accounts
593,427,119	Exploit.In accounts
457,962,538	Anti Public Combo List accounts
393,430,309	River City Media Spam List accounts
359,420,698	MySpace accounts
234,842,089	NetEase accounts
164,611,595	LinkedIn accounts
152,445,165	Adobe accounts
131,577,763	Exactis accounts
125,929,660	Apollo accounts

Recently added breaches

22,477	JoomlArt accounts
326,714	Mac Forums accounts
846,742	Baby Names accounts
1,274,051	Wife Lovers accounts
342,913	Facepunch accounts
125,929,660	Apollo accounts
7,687,679	Digimon accounts
2,457,420	SaverSpy accounts
307,768	Real Estate Mogul accounts
3,472,916	NemoWeb accounts

Figure 1.2: Showing status on breaches from Haveibeenpwned.com recorded as of 4/11/2018

credentials being exposed. Only some of these companies have been combining stolen credentials with re-use of such credentials on other sites or services can lead to exposing large volumes of partial identities and even stolen full identities. These challenges show there is a massive need for solutions that can help manage identities and private information in a good way and at the same time bridging the gap between technology and privacy.

1.4.1 Solutions

As mentioned in the earlier sections technology introduces solution and creates revenue by improving current solutions and creating new opportunities, however at the same time it might also add new issues. Digital identity can be seen as one of those areas. Before computer programs had to control values and confidential information concerning persons, there was no need to worry about leaked private information or exposed digital identities. The reality is that it is a growing problem with multiple solutions and no best practice. One of the technologies is user-centric identity which has been proposed by Augmented Social Network. Their work is based on *"The assumption that every individual ought to have the right to control his or her own online identity"*[8] where previous federated identity providers and technologies have put digital identities in the hands of the organisation. This initiative focuses on making identities controlled by the users so that the identity can be used inside and between organisations. The purpose is expanding the technology for a user to have more power over his identity and for trust to be decentralised. However most of today's user-centric identities are still centralised by the fact that they are stored in a central repository. This construction has shown not to be ideal in all cases. Facebooks Connect is an example of this. People can choose to use their Facebook profile to login to various services around the web, using Facebook as their identity provider. This way Facebook can track all usage of the services the user uses, and profile for their benefit.

1.4.2 The self-sovereign identity

The new kid is called self-sovereign identity (SSI) and can be traced back to Devon Loffreto in 2011 when he wrote about “Sovereign Source Authority” [45, 46]. The general principle is that individuals have to establish the right to an identity, without states and companies destroying the user’s sovereignty. This idea has been fueled by the improvement of cryptographic technologies and digital ledgers. Many others have contributed, one of them is Christopher Allen which talk about the SSI concept and its properties in. The user must be the centre of administrating their own identity. This requires not only interoperability of the identity and user consent but also fully control that allows the user to transport the identity. The SSI should also enable the user to create claims about other users, group or entities without limitations. Further, the identity should also be able to contain information that is asserted by other entities[47]. In total, the SSI concept is broad, and consist of ideas based on previous generations of identity solutions. However, there is no total consensus about its functionality and technological solutions.

1.4.3 Privacy Enhancing Technologies

Where SSI solves the problem regarding control over identities, it does not deal with consent and authorisation of personal data. It just externalises the authentication and set the user in control over their own identity. To set the users attributes free of the services and identity providers, the authorisation at each service should also be outsourced to an entity trusted by the users and the service provider. The Kantara initiative promotes this exact procedure. In their UMA standard, the design goal is to give the user a unified point of control for authorisation online. Using the UMA standard the user’s can control who and what has access to their data such as identity attributes, personal content, and access to whom may access services on their behalf. No matter where all those things are located on the internet.[48] Immediately this can be seen as a loss of control. However, this choice might provide a significant benefit for the enterprises choosing to do so. It allows for services and applications residing on the internet or intranet to protect resources without having to implement comprehensive authentication and authorisation policy infrastructure themselves, by outsourcing these entities to a trusted access control entity. This style makes applications simpler by allowing them to be more loosely coupled to their authentication methods and make more effective access decisions[49].

1.5 Problem formulation

Based on the preceding research, this report aim to answer these questions by raising the problem of how privacy aware authorisation together with state of the art identity and access control standards can be used to solve these problems in an enterprise context.

The goal is to design a privacy-aware distributed digital access control system. Such a system sees its need in both private and governmental enterprises dealing with personally identifiable information or personal data, where regulations and expectations have set a new requirement for computer systems and how they deal with personal data and privacy. In the first place the system should be able to handle traditional authentication and authorisation scenarios of the user. Thereafter enable the user to take over control of their own digital identities and data, by providing consent management for accessing personal data. Thereby letting the user manage which information that is accessible, whom they are shared with and when they can use it.

A problem statement which frame the project based on this initial investigation in this chapter is introduced based on these findings the following problem statement is established:

How can a privacy-aware identity and access management system for enterprises be designed, which enhances security for individuals and enable consent based access for services?

In order to answer this problem statement, several sub questions have been formulated which will be discussed throughout the report. Each of the sub questions deals with key areas and perspectives used to answer the general research statement and enables a better understanding and insight into the problem field that are being addressed:

- *How can identity and access control be used to increase security in an enterprise?*
- *How can a system be designed to comply with the current data protection regulation?*
- *How can identity be used in a privacy-aware manner in relation to enterprises?*
- *What architecture can enable the increased privacy for digital identities?*
- *Which technologies can enable the user to have management over their data in a privacy enhancing manner?*
- *How can partial identities be used to manage rights and access to personal information?*

1.5.1 Limitations

In order to limit the scope of the project, some limitations have been set for this project. They include:

- A full analysis of all possible organisation structures in enterprises will not be investigated.
- Closed source solutions will not be taken into consideration due to the uncertainty of how they work
- We will not be exploring a final implementation of a holistic system. However, the proposed prototype will be made with considerations to a real enterprise scenario.
- Only the European GDPR legislation will be investigated in relation to regulative work

1.5.2 Objective

The objective of this project is to develop a concept for utilising privacy aware authentication together with externalised authorisation as part of an Identity management system for a general enterprise. The requirement for the system will be a based on a combination of Theory, best practice and additional findings throughout the analysis. The project is not based on a specific company or environment but aims to meet general requirements that enterprises in a wide range deals with. The goal of the report is to describe the work and considerations which have been made during the project, together with showing understanding of the various fields covered in the report.

Chapter 2

Methodology

The method for answering the problem is outlined in this chapter. The data used to answer this is primarily based on and secondary-data, which is a combination of various forms of past-data that will be described in this chapter.

2.1 Research strategy

The details of the research methods utilised in the process of collecting data which is generates the foundation for this project is described. Furthermore, the techniques used to create the design, implementation and how this help answers the problem formulation is defined. A research strategy is an overall plan that includes clear objectives, problem formulation, data gathering sources and the different limitations affecting the project. The research strategy for this project will be specified for the various steps in this project. An illustration of this can be seen in figure 2.1.

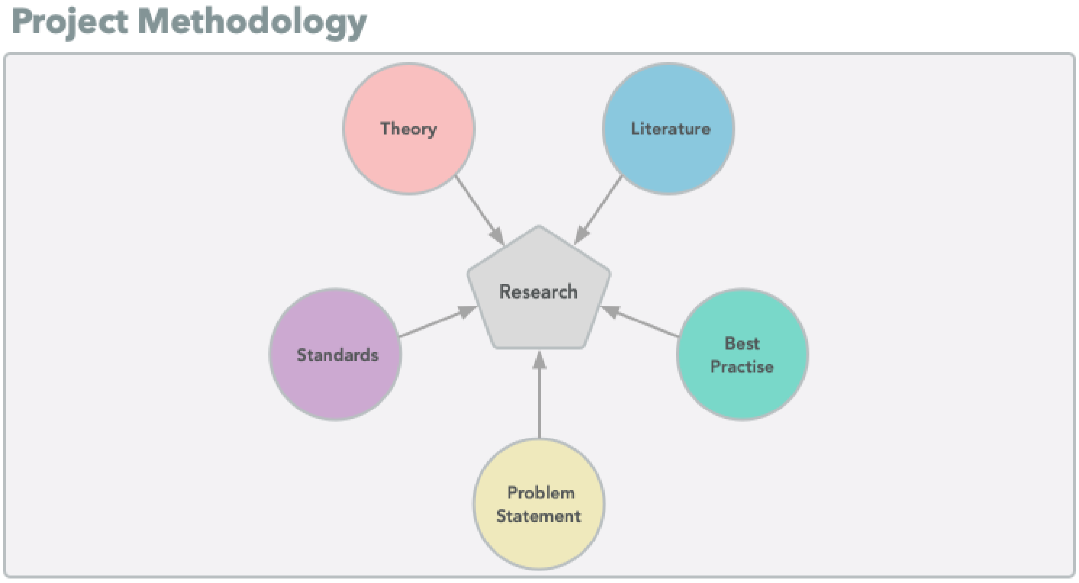


Figure 2.1: Project Methodology [2]

Figure 2.1 depicts the research strategy used in answering the problem statement. The result is achieved by joining theory with various kinds of secondary data research. I will first introduce and discuss the secondary research, which is based on a combination of technical reports, white papers, journal articles and published books. Secondly, the process which has underlined the project is discussed. The project consists of 4 phases: State of the Art, Analysis, Design and Implementation. Each of those phases uses a separate method due to the nature of the process of the specific phase. In the next sections, these three phases will be described, including the selection of a method of each of them.

The project has overall followed the standard project form used at AAU, where the initial process is to identify a problem field, and in this case, identifying a scenario where it was possible to apply some of the skills stated in the Curriculum. This includes but not limited to how it is possible to enable Identity and access management in real life settings combined with how privacy and trust can be applied to protected resources.

Some of the inspiration for the project came from the previous experience with technologies in the field of IAM, and how privacy could be respected with the current technologies. Privacy is somehow a hot topic, and newly released legislation regarding GDPR have contributed to this increase in interests and awareness. This should be seen in relation to the lack of working solution in the market in this area. There does simply not exist a standardised and widely implemented IAM technology that promote the privacy of the user. Based on this initial finding and wonder, the privacy topic was chosen. The next step in the process was to establish a case or a concept of what could be created and what could be the goal in an investigation.

2.2 Scenarios

For this project, a number of scenarios has been conceptualised and developed, with the purpose of relating the design to realistic and concrete instances of real-world use.
To identify the scenarios, an investigation of common applied problems were conducted. By researching relevant technologies and their solutions, possible scope of scenarios was chosen and used to provide the basis for the analysis. Based on this, the SOTA were conducted without specific scenarios in hand. The aim was to outline the specific research problems that could be solved by a unified design solution, and thereafter identify the scenarios could work as evaluation for the project.

For this project, two main scenarios were developed, both cases that were seen as relevant and deemed to could occur in real life in relation to an interaction between a private person and an enterprise as described in the problem field. The scenarios are presented in section 3 and we feel confident the scenarios are representative of actual real-life situations.

However, we will not claim that they are the most representative or the most significant to showcase a solution to the stated problem. Each implementation should be worked with the people in the enterprise. Due to this, it has been a priority in the process to create a general solution that can be extended and used in a variety of ways and still fulfil the problem stated in the introduction.

2.3 Literature Research

The research covers the initial investigation from the introduction which has resulted in the problem formulation at the end of that chapter. Identity management is a research discipline anchored in both business-theory as well as cybersecurity. These areas of research linked together outline the theoretical framework for this project. Most of the information regarding identity management is obtained through literature research. This is because of the problem with cybersecurity being a complicated topic with many different opinions, viewpoints and a relatively short lifespan. The field of identity management is characterised by many closed source solutions that are not rooted in models described in academic works. To understand the field, a large part of the time was spent on researching for relevant, trustworthy material, in order to understand the type of solution, the analysis and design should be based upon.

The State of the Art chapter is grounded in existing solutions and academic articles state problems and solutions for cybersecurity and privacy-related topic, which together have formed the introduction as well as state of the art (SOTA) chapter. The primary literature for the SOTA chapter has been material gathered from (System Security) Lampson[50], (Legislation) Information from the European union[51], Identity solutions Kantara Initiative[3], (Next generation identity solutions) Sovrin [52] and Access control (Kuhn)[53]. In relation to the Identity and access control in enterprises, in both SOTA and the analysis, the report [7] were used, this is a former piece of work by the same author as this report. The report includes primary research in terms of the interview and several presentations with IAM professionals from PWC Denmark. It should be noted that this work continuous been used in this report, but not made in relation to this project.

2.4 Analysis

The analysis chapter focuses on investigating the actual problem formulation and its sub-questions, as stated in Chapter 1.5. To do this, theories, standards, and best practice described in the introduction (Chapter 1) and SOTA (Chapter 4) will be applied within our specified problem field. The relvant technologies will be discussed in relation to how they fit to the problem of this project. The result of this chapter combined with the research, will lay the foundation for the design proposition of the system. Based on the analysis, a limited requirement specification for the system will be derived, that will describe the functional, as well as the non-functional behaviour of the system. Since not all requirements are able to be implemented, they will be prioritised using the "MoSCow"-Methodology. This methodology provides a framework for how we intend to prioritise and categorise the importance of functionality and features of the system design.

2.5 Design

To describe the proposed solution, a system will be designed, outlined by a variety of scenarios and use cases, which describes the semantics of the system. To do this in a standardised manner, Unified Modelling Language (UML) syntax will be utilised for the Swimlane diagram. The rest of the diagrams have been developed with similarity in mind, using the same notation.

2.6 Conventions of This Project

For this report, the following conventions were taken as basis:

- *The report uses British English spelling.*
- *All abbreviations are listed in the List of Abbreviations on page VI.*
- *All Figures are named and a complete list can be found in the List of Figures on page 4.*
- *All Tables are named and a complete list can be found in the List of Tables on page 5.*
- *The formatting of the bibliography in this report follows the guidelines laid out by the The Institute of Electrical and Electronics Engineers (IEEE).*
- *The information systems and relevant diagrams in this report are modelled using the specifications of the Unified Modelling Language (UML), as specified by the Object Management Group (OMG).*
- *Illustrations used in the report and in illustrations are created solely for this report based either on the authors ideas or based on referenced material described in their individual caption, icons or symbols used within the illustrations are based on Font Awesome.*
- *Finally, this document is formatted using the LaTeX typesetting system.*

2.7 Terms used in this thesis

Enterprise In this report, the term enterprise is used extensively. The definitions of this word are based on the enterprises used in the field of enterprise architecture, which refers to an organisation that has an ongoing project and describes an organisational unit or collection of entities that share a set of common goals and collaborate to provide specific products or services. This includes both small organisations as well as major corporates and governments.

Chapter 3

Scenarios

A series of scenarios have been developed to help analyse and design a solution for the stated problem field. The scenarios will further be used to relate the solution to real-world usage and a specific problem. The scenarios are based on the general problem field described in the introduction, not all cases described will be investigated or fit into these scenarios but they are designed to showcase how such a proposed system are able to facilitate privacy-aware computer systems for general usage in enterprises.

3.1 Scenario 1

To qualify for most professional jobs, the applicant needs to present them self with a list of qualifications. Among many, these are previous positions that the applicant have held and Academic skills. The latter includes everything from diplomas which include attended courses to obtained marks, comments awards and honours. This is high-value personal information and often sensitive for the specific student. The academic institution asserts this set of qualifications and continuously update them as the student achieve them. For the student to increase his chances of getting the best possible job both during studies and after. It is essential that the institution continuously and reliable can provide this information. For the prospective employer, they need a way to check this kind of information that is presented by the applicant. Checking information based on papers are is a cumbersome job that takes resources away from the central business. For the applicant seeking a job, gathering this information takes a lot of time, and for most job position the material is the same.

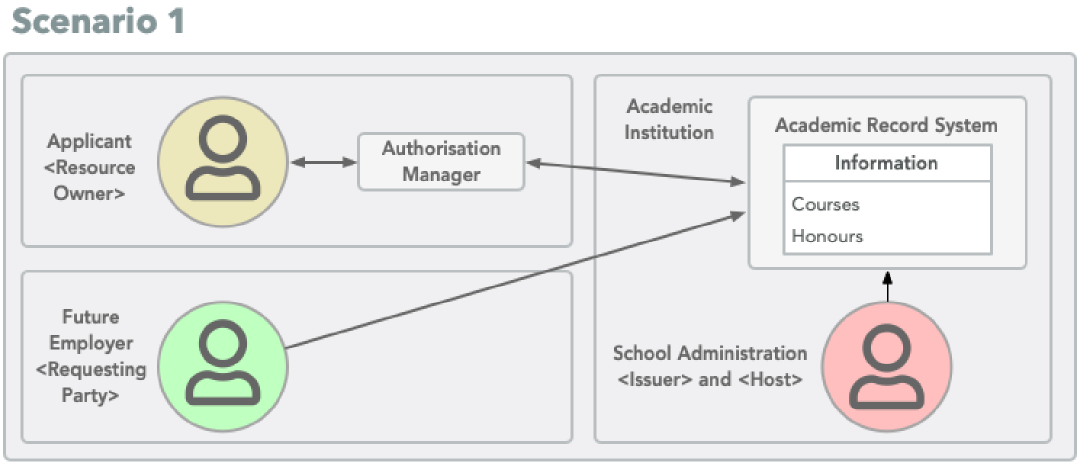

Scenario 1
Applicant <Resource Owner>
Authorisation Manager
Future Employer <Requesting Party>
Academic Institution
Academic Record System
Information
Courses
Honours
School Administration <Issuer> and <Host>

Figure 3.1: Scenario 1 - Managing access to updated qualifications where the user is not the host

3.1.1 Managing access to updated academic qualifications where the user is not the host

The Scenario and use cases described here and illustrated in figure 3.1 represents the needed interactions to showcase the usefulness of such a proposed system. The scenario is based on a thought up example, where privacy-aware systems can be used to control access to information, that the Resource owner are not the issuer of.

The Following use cases should be used to represent this:

Use case #1
The Administration entity, creates a record of qualification in the academic record system.

Use case #2
The Resource Owner create a provision profile, containing the resources he wants to share with his future employers.

Use case #3
The Resource Owner provision access to a set of resources, with use of the provision profile for the future employers that he wants to share the resource with.

Use case #4
The Future employer request access to the protected resource

Use case #5
The Resource Owner check which entities that has accessed the protected resource.

3.2 Scenario 2

Managing data in Which Employers and Employees Both Have a Stake In every company with employees there exists information where both parties have a saying concerning the specific information. Because the information is related to the reference between the two entities it is not clear who can control this. This scenario covers such problem. The employee - company relations contains information about the position that the employee holds at the company, title, union relations, salary, among others. This information might be confidential for some parties, where certain information should be available for others. Any of these parties might want to impose constraints on the sharing of the same data for others, and some of this data might be used to make authorisation decisions. The information is used both inside and outside the company and should be accessible to business partners and systems. External partners that can need access to this information include Governmental bodies such as Tax and social-authorities. Private entities include other companies such as Banks or private persons.

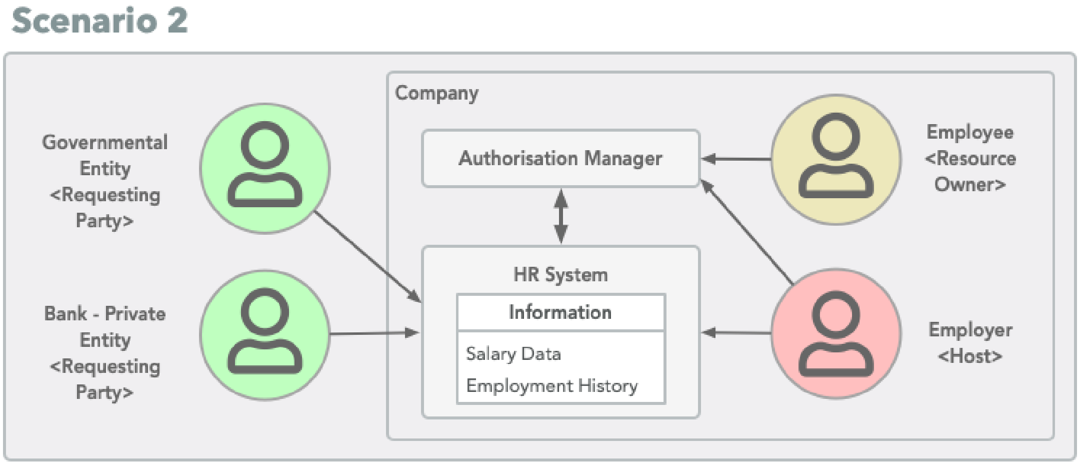

Scenario 2
Company
Governmental Entity <Requesting Party>
Bank - Private Entity <Requesting Party>
Authorisation Manager
HR System
Information
Salary Data
Employment History
Employee <Resource Owner>
Employer <Host>

Figure 3.2: Scenario 2 - Managing data in Which Employers and Employees Both Have a Stake

3.2.1 Sharing salary data with governmental body

The Scenario and use cases described here represents the needed interactions to showcase the usefulness of such a proposed system. The scenario is based on a real-life example where privacy-aware systems can increase the transparency of systems dealing with private data, even in scenarios where the resource owner full, don't have full control over his or her data.

The Following use cases should be used to represent this:

Use case #1
The employer entity, Adds Salary data to the HR system.

Use case #2
The employer provision access to the salary data for the governmental entity that needs to the resource

Use case #3
The governmental entity access the protected resource without specific permission from the employer

Use case #4
The resource Owner provision access to the salary data for his Bank which require access to the data

Use case #5
The private bank entity access the protected resource with the permission given by the employee

Use case #6
The resource Owner check which entities that has accessed the protected resource

Chapter 4

State of the Art

In this section related work and current solutions will be described and discussed in relation to the topic outlined in the introduction. The following sections will provide descriptions of different concepts related to Identity, Privacy and Access control. By evaluating the solutions in the market and relevant research, the evidence is collected will further be used in the analysis.

4.1 Security and Cyber Threats

In the last decade, cyber attacks and data breaches have moved from the IT department to the board of directors. This move is due to the combination of higher chances of getting attacked and increasing the cost of an attack or breach for most enterprises. These increasing costs come both from lost sales, compensation and fines such as described in the European GDPR[54] and indirect costs such as bad PR. The complexity of malicious attacks is on the rise. Poor security or low awareness in just a little branch of an organisation can be the Achilles heel for the entire enterprise. The last decade has seen a massive increase in cyber attacks and breaches. In 2016 there were more than 4,000 reported cyber attacks every day. That is a 300% increase from 2015 based on numbers from US FBI [55]. FRI cyber task force further state that In 2017, cyber crime is expected to cost the US economy between $57 to $109 billion. Examples of massive attacks in recent years are plenty, this includes but not limited to, giant companies such as Maersk, FedEx, HMS Scotland and Nissan. Even though they were not the target, these companies were all hit by the NotPetya as part of an attack. Nonetheless, it ended up having significant costs for all of them. Many of these companies have blamed inadequate cybersecurity policies for their losses[56]. As described in the introduction, Interconnected computer systems make use of Access control Models to govern access to systems and these ACM models base their decisions on Identity. Therefore Identity solutions are crucial in mitigating attacks on systems.
In the following chapter material about why these protection mechanisms are so important is presented. This material lays the foundation for the systems that this reports touch upon and is used to defend against malicious players and why identity plays a central role in this topic.

4.1.1 Real World Security

The overwhelming amount of cyber security incidents shows that there is no such thing as perfect security, and it does not exist in the real world. If an attacker wants to penetrate a system, he will do so. The cost might be high, but it is not impossible. The computer pioneer Butler Lampson[50] supports this claim. Even the high level of security is not enough when dealing with highly sophisticated attacks. Such attacks are known as Advanced Persistent Threat (APT), and this category of attacks is known for their ability to hit precisely and hard because they are tailored to hit a specific target. Even for well-protected systems, the history of cybersecurity shows that even highly secured systems have been broken.
By using complicated techniques such as "yet to be known" exploit also known by the name zero-day vulnerabilities. The attacker can penetrate a system, without the system owners knowledge, this fact in combination with good research and coordination can break most computer systems. Due to this knowledge of the victims systems needed for such an attack, this category of attacks is both costly and slow. Ordinary security policies have minimal chance of protecting against these specially targeted attacks. Facing this reality, there does not exist such thing as a complete security system[57]. This leads companies to the choice of balancing security and costs. To do this it requires an enhanced focus on which data is handled and who should have access to it. Therefore enterprise systems which are critical, need a high degree of protection. However, this strategy is just one out of multiple security strategies that can be used to protect their valuable resources.

Security in the physical world and the digital realm have many similarities. A significant difference is that in the physical world a competent and determined attacked would not be able to steal from 10.000 of persons in seconds. This transition to the digital world has further been accelerated by the internet that over a couple of decades have connected billions of users. Any of these connected devices have the opportunity to attack other connected devices. Advanced malware like worms makes it possible for a compromised host to attack other devices without the knowledge of their owner.
Every time a new program or an update is executed the computer could get infected or mobile devices that connect to the hostile environment and isolation is not a possibility. Complete Security is tough if not impossible, at least with the computers we use today[57]. An argument, in this case, is the PC-monoculture makes this worse due to the little diversity. Attackers have an easy job when the systems have the same flaws. In combination with the computational speed, such attacks can be created as a dangerous cocktail that can harm many people in a short amount of time.

In the article *"Computer security in the real world"* Lampson state, a goal for computer security: *"Computers are as secure as real-world systems, and people believe it"*[57] He further state that if your house ain't complete secured why should a computer be it. Real world security is not about the perfect defence, so why should cybersecurity be it. Every house can be broken into, if not through the door then through the wall. It is just a matter of time and price.
Based on this assumption Lampson creates a framework stating it is all about "value, locks and punishment". An attack in the online or offline world will happen over time if it pays off for the criminal. It can all be described by the equation: A theft will happen if the gained value is higher then the cost of punishment times the probability of getting caught[50]. In the offline world locks are no absolute security, not even in houses of banks. It is the punishment that makes the life of a criminal unattractive. However in this internet era, there exists an enforcement gap. *"Cyber criminals can operate with near impunity compared to their real-world counterparts"*[58], This is supported by the fact that less then 1% of malicious cyber attacks see an enforcement action taken against the attackers [58].

Even though modern cryptography does a good job and is close to enabling perfect security, the cost of a fully secure system is too high.
On the other hand practical security, as used today *"balances the cost of protection and the risk of loss, which is the cost of recovering from a loss times its probability."*[50]. Therefore the cost of recovering from an attack should always be examined contrary designing a secure system and proper cyber-security features. When the risk of getting attacked is high and the cost of recovery is low, costly security is not worth it. In relation to private information of users, the price of recovery is extremely high. With introduction of recent Privacy legislation companies can be fined for inadequate security measures. This fact in combination with bad publicity and loss of sales further increases the price. Therefore protection of personal or private information can be considered as extra high in relation to other information.

4.1.2 Defensive Strategies

If content or decisions of a computer system is deemed valuable, a defensive strategy should be formulated. To do this Lampson describes this model as a framework for the following defensive strategies which can be used when dealing with security in relation to the classical Access control Model (ACM) [57]:

1. **Isolate** - Keep everybody out
2. **Exclude** - Keep the bad guys out

3. **Restrict** - Keep them from doing damage
4. **Recover** - Undo the Damage
5. **Punish** - Catch the bad guys and punish them

Each of these 5 strategies can be used to defend against a malicious activity or a combination thereof. Some might fit good for some cases and others not. As described in the previous section, this should be embedded in the requirement when designing with any kind of security requirements.

Due to the problem field described in this Section, it can be concluded that any system should balance their security effort and the related costs. If the systems deal with personal information, there should be an enhanced focus on which data is handled and who that have access to it. A set of defensive strategies is given and should be used when specifying requirements for every entity in the system.

4.1.3 Identity, Theft and Phishing

Identity theft is a commonly used term that describes the action of somebody that makes use of others people's identity without their approval or knowledge [59]. In Denmark, there is no legal definition of identity theft. Legally speaking, identity theft is a misleading term because the word "theft" refers to a person who owns his identity as one owns a material thing. Since an identity or digital identity is no such thing, it does not apply to the conventional definition of theft[60]. The term is typically misused or mixed up with the term identity fraud. Which is the action of exploiting a fictional identity [59]. According to the OECD, not many countries have specific legislation on identity theft[61]. However, use of the false identity might be illegal in many countries under other legislation such as Forgery, Hacking or Defamation, hereby Denmark according to research Peter Kruize [60].

According to Dutch police the most commonly used method of acquiring digital identities or information their off is phishing: looking for personal information in the digital realm. Phishing can be done in different ways, and each time a new method is detected, and a new name will be assigned. The following kind of Phishing exists [62]:

- **Phishing** - An email with a link to a fake website where personal information can be stolen
- **Spear-phishing** - Like phishing, but where the mail specifically contains personal content

- **Pharming** - Users who contact an enterprise online and are secretly redirected to a fake webpage. This redirect is accomplished through the use of malware or a hostile setup that the computer is infected with.
- **Spy - phishing** - Phishing with use of keyloggers, that intercept keystrokes, mouse movement or screen or audio-recording.
- **Vishing** - Phishing over VoIP systems, where the phone or similar audio connection is used to lure information from.
- **SmiShing** - Phishing information with use of the SMS system

The most common kind of phishing happens due to identity theft, where the attackers target payment card information or passwords to net banking.
Research from the cybersecurity consultant Experian Information Solutions states that [63]:

- Social Security number in the US sells for about **1$**
- Driver license information **20$**
- Credit card or debit card information depending on attached information and quality **5-110$**
- Online payment service login info online banking or PayPal **20-200$**
- Medical records **1 - 1000$**
- Passport (US) **1000 - 2000$**

There exist a demand and a market for identities, making them coveted; therefore identities should be protected. The more information a possible attacker can gather, the higher the value. Therefore a system managing identities should include an appropriate level of security in comparison to the value of the identities it holds. Such an identity system should also include mechanisms that protect against known attacks such as Phishing.

4.2 Legislation

Legislation as a tool is one of the powerful forces which can influence and shape the development of technology. This has been the case the last years and with the recent introduction of the extended privacy regulation within the EU known as GDPR. The driving motivation factor for enterprises is threats about huge fines, which can be given if companies fail to protect the privacy of users. GDPR is not the first regulation to touch upon this area of legislation. GDPR builds on top of the already established EU Data Protection Directive from 1995. In addition to the GDPR, a new ePrivacy

(EPR) regulation is expected to be taken into action in the next couple of years. The regulation is meant to regulate data privacy concerning phone and internet-communication services. The scope of the EPR Regulation applies to "any business that provides any form of online communication service, uses online tracking technologies, or engages in electronic direct marketing"[64].

It is not only in the EU that regulation shapes how companies deal with user information. In the US there is no single regulation that deals with privacy of user information. The US system has a wide range of patchwork based on both federal and state level that mainly deal with user privacy in a specific domain. This includes legislation such as The HIPAA Lay in Healthcare, The GLB Act and Fair Credit Reporting Act about finance and credit ratings and The FTC Act concerning Telecommunication[65].
This array of regulations in and between countries makes it difficult for companies that want to cater for an international audience, because they have to investigate local legislation before being able to conduct business without facing the change of large fines or legal actions[66]. To limit the scope, this project will only focus on European legislation in relation to GDPR, and as of writing this, the ePR regulation is still working.

4.2.1 GDPR

As described in the introduction all companies dealing with private information have been required to comply with an updated and new set of requirements known as GDPR. GDPR reshapes the way in which organisations and enterprises manage data, as well as redefines the roles for data responsible in businesses. The regulation covers all processing and storage of personal data and introduces new rights for the user and puts restrictions on how, when and where data can be shared between controllers and processors. In regards to the problem statement, this is highly relevant and should be examined further, in order to design a system that meets these requirements.

The regulations define a set of entities that the rules in the regulation is bound to. This includes:

- *Data Subject* - A data subject entity is any person whose personal data is being collected, held or processed. By using a service anyone becomes at some point a data subject – whether they are applying for an internship, booking a hotel or just shopping online. Due to any of these cases the a person discloses some personal data to a service provider known as the controller[67].
- *Controller* - The controller entity has the principal responsibility of assuring that the processing of data is happening in compliance with the regulation. This includes storing, protecting and transferring personal data. This also covers enforcing data-protection by design

and by default. If more then one controller is present, clear and strict responsibilities should be agreed on between the controllers. The controller can let another entity process data on his behalf, this is known as a processor. However, this delegation must only happen according to the agreement between the controller and the data subject. A processor cannot delegate processing of any data to another processor without prior consent from the data subject. In most cases, this makes any service provider handling any personal data a controller in relation to the GDPR and should act as responsible of the data subjects privacy[67].

- *Processor* - The processor entity is described as someone who processes personal data on behalf of the controller. The controller shall only make use of processors which provide sufficient guarantees to implement suitable technical measures that processing of data will meet the requirements of GDPR and ensure the protection of the rights of the data subject. However, this could also be the controller which just process information for itself. In all cases the processor should follow the instructions stated in the agreement between the subject and the controller[67].

GDPR What, When and Where

The Regulation applies only to what is specified as personal data which is defined as: *"Any information relating to an identified or identifiable natural person (data subject) an identifiable natural person is one who can be identified, directly or indirectly, in particular by reference to an identifier such as a name, an identification number, location data, an online identifier or to one or more factors specific to the physical, physiological, genetic, mental, economic, cultural or social identity of that natural person;"* [67]. This means that all data or any set of it, that are able to identify a person or be linked to such a person categorise as personal data in the GDPR.

The regulations covers when either a controller or a processor is located in the EU regardless of the processing takes place within the EU or not. This means just one of these entities should be present from within the union[51].

The Regulation involves the processing of personal data of the data subjects, where the processing actions are associated with the offering of goods or services. The regulation also applies in the data subject is just being monitored within the union[67].

Consent

Consent is a description of permission that can be given between entities. Processing of data under GDPR is only legal if consent is given beforehand and under certain conditions. The GDPR forces controllers to obtain such consent from data subjects. A consent of the data subject shall be a voluntarily given, explicit and unambiguous statement of how the data subject information should be processed. This consent should also be utterly revocable to satisfy the regulation.

Rights of the data subject

Besides the introduction of consent, the regulation gives data subjects or users of a service certain lawful rights that the controller shall obey. This includes[68]:

- **The Right to information**
  This Right provides the data subject with the ability to demand a controller for the information concerning what personal data is being processed and the reason for such processing. This includes information about processors with whom his or her data is shared at any given time.

- **The Right of access**
  This right provides the data subject with the ability to demand access to his data which is stored or processed. Ir further provides the subject with the right to see their data, as well as to request copies of the data.

- **The Right to be forgotten**
  This right is also known as the right to erasure. It provides the data subject with the ability to ask for the deletion of their data. However, this is not an absolute right for the subject and depends on the retention period in line with other applicable laws which overrule this right.

- **The Right to rectification**
  This right provides the data subject with the ability to ask for adjustments to his data when the data subject believes that this personal data is not accurate.

- **The Right to withdraw consent**
  This right provides the data subject with the ability to revoke any given consent for the processing of their data. The request would then demand the company to stop the processing of the personal data.

- **The Right to object**
  This right provides the data subject with the ability to object to the processing of their data, if there are reasons that gives the processor the ability to process the data without consent.

- **The Right to object to automated processing**
  This right provides the data subject with the ability to object to the use of automated processing of the personal data. This could be the right to ask for a manual review of data instead of an automatic processing if the user thinks that he is a particular situation.

- **The Right for data portability**
  This right provides the data subject with the ability to request transfer of his data. In such a request, the data subject can both ask for transfer of is data to himself or another controller. When doing this, the data must be transferred in a machine-readable format.

Due to legislation, any company that works with personal information must take this legislation into account. Sometimes this can be done by the change of procedures or increase security in current systems. However, to mitigate risks of leakage and improve the level of compliance with legislation, systems should be designed and implemented to natively comply with the bills such as GDPR and the EPR directive. Therefore these rights should be built into the system by default. As this is closely related to the topic of this project, the solution prototype should, therefore, be built with these rights in mind. This mainly covers the **"The Right to information"**, **"Right to be forgotten"** and **"Right to withdraw consent"**.

4.3 Identity and Access Management Solutions

Making confidential and personal information available on the internet has led to the need for digital identities, a trend that has led service providers to create isolated user repositories. The result of this shift has meant that personal information is spread all over the internet without an easy way to remember, update or revoke these information [27]. Such fragmentation and distribution of user information create a challenge that can lead to loss of privacy for users.
In a business relation, this might be a thread related to loss of trade, or a security thread. Examples of this include users writing down passwords to various services, because the human brain is not capable of maintaining changing passwords, with the security requirements used in today's password policies. This is further a great example on an anti pattern in regards to the use of passwords.

One of the solutions to this problem has been known as SSO and it make use of federated identities. The term federated can be defined as *"The process of uniting smaller, localised entities in a single group"*[69]. A federation requires trust in order for the participants involved to work together. A proposed solution to this problem of fragmented digital identities is the concept of Single Sign-On (SSO). SSO was introduced by Liberty Alliance early 2000, with the goal to define a set of standards, in order to create the technology necessary to build a universal identity infrastructure. According to Liberty Alliance [70], the idea of the project is to create a federated network identity concept. This idea relates to the fact that many organisations have their own way of defining and manage the identities of users with no relationship between them. This causes problems with compatibility, considering the user's identity in one system cannot easily be directly linked to another, and thereby creating the obstacle. Using the proposed concept of federated identity creates a standardised method of sharing identities across different systems. The concept relies on separating the service from the storage of digital identities. Instead of adding an element of trust between these two entities. The provider of the service (SP) can acquire the necessary information about the user's identity from a service called the identity provider (IDP). In order for this concept to work, a previous arrangement has to be in place between the IDP and the SP. This element of trust is by Liberty Alliance called the circle of trust. This Idea can be expressed by of two circles that represent trust, one connecting the user and the IDP and another one between the IDP and the SP.[70], this is illustrated in figure 4.1 where a single user can use multiple IDP to authenticate against an array of SP's. The concept has later become known as the 2nd generation identity solution.

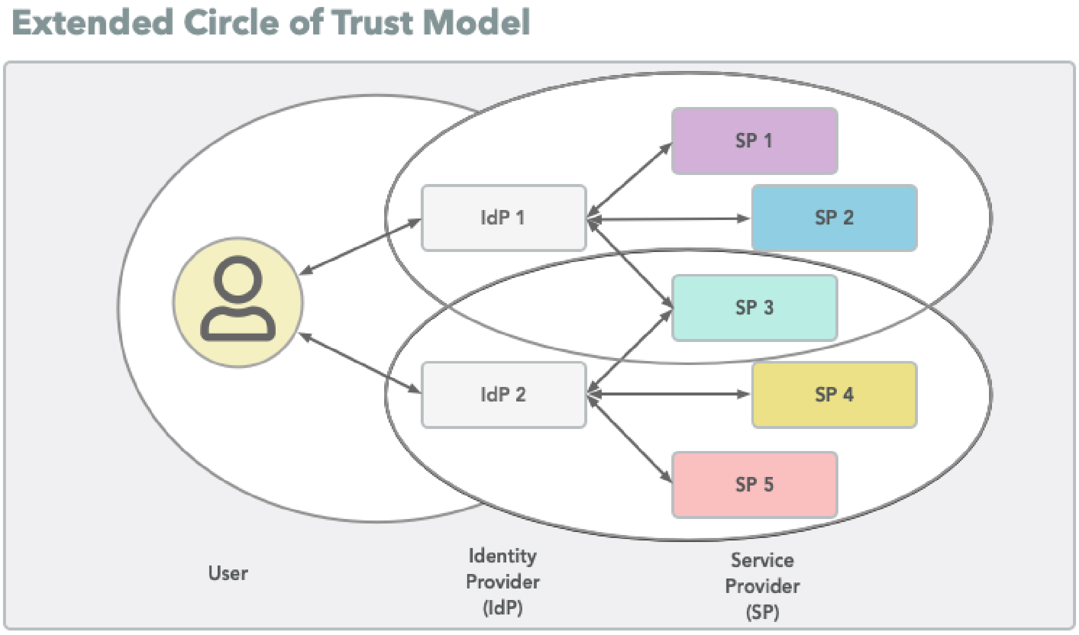


Figure 4.1: Extended version of the Liberty Alliance Circle of Trust model, with two independent Circles

This idea of digital identities and how they are dealt with, have become the best practice in modern identity systems. Microsoft's identity architect Kim Cameroon states that *"Identity should be based on claims"*. Where the claim holding a statement that a subject gives about another

entity or themselves[37]. Identity is built by presenting a claim that is issued and can be confirmed by an identity provider (IDP).
However, for some actions, a self-issued declaration might not be sufficient. For these cases, a trusted third-party authority is required to assert the claimant's identity. Therefore an assertion in federated identity is defined as a *"Confident and forceful statement of fact or belief"*[30]. Such assertions should transferred in packages with use of tokens that make use of cryptographic safe technologies, to maintain the integrity of the assertion. Confidentiality of the transferred additional information can be added with use of encryption, however it is not necessary in order for the general idea to work, and up to the specific implementation and purpose.

4.3.1 Federated Identity Technologies

In this Section a description of these technologies will be given, together with the possibilities they provide, that this project will be built on top of.

Technologies that support this separation of service have evolved over the last couple of years. This includes SAML, OAuth, OpenID Connect, Facebook Connect and others. On the operational level, identification is established by presenting a claim issued by a Service Provider, as explained in the past section. In some cases, a self-issued statement is not enough to guarantee claimants identity, therefore a trusted authority is needed to assert this identity. An assertion like this can be defined as a statement of fact or belief [30].
According to the ideas of Liberty Alliance, these assertions should be packed in secure digital tokens. This is the essence in federated identity technologies, they enable this concept of SSO by enabling a many to many relationships between IDP's and SP's.
Beside SSO, SS-Out is another feature which was introduced by Liberty Alliance, which provided users with the ability to log out on all services they are currently logged into. This feature also enhances security by improving the usability of the services. Thereby SS-OUT It defines the mechanism that can log the user out of all SPs by using a single logout function. [71].

Instead of the user having multiple scattered identities around multiple applications or services, this pattern allows the user to have only one single federated identity registered at their preferred IdP.
By using one of these federated identity technologies, it enables a secure and standardised method of dealing with identities among multiple sites, applications or organisations. When users try to access an application that supports the federated identity concept, their identity is therefore securely transmitted to the application, without the need of transferring all of the user's credentials [69]. This use-case will typically increase the usability of the system together with enabling aspects of better security and

privacy. This concept of IDP and SP separation has also been described by Kim Cameroon et al. which state that *"Identity should be based on claims"*. A claim being a statement an service provider makes about themselves or another subject [27]. However, the idea of federation does not solve the privacy problem from the users perspective and an amplification of this issue might even occur in some contexts. An example of this is when the IdP is used to verify the identity, it gets knowledge about the place that the identity is used and thereby eliminate the user privacy.

Another problem with this concept is that the User doesn't exactly know which information is being shared between the IDP and SP. Maybe the user only wants to expose a specific part of his or her identity. This problem will be further investigated to find a solution that respects the privacy of the user at the same time providing the advantage and possibilities as federated identities have given.

4.3.2 SAML

SAML or Security Assertion Markup Language is a standard proposed and maintained by OASIS. The protocol can be used to provide authentication and authorisation by letting a user transmit assertions between entities in a computer system[72],[73]. The standard is based on an XML-based framework which allows the entities in SAML to interact in a standardised and non-proprietary way at the same time supporting customisation which allows for different types of data to be sent to the external service provider [74]. This possibility is widely used both on a local scale but also on national and is for example used in the Danish National Identity and Access solution MitID[75]. Due to SAML provides the mechanism for exchanging information about a verified user without ever sharing the user's credentials with the third-party or the service provider[72].

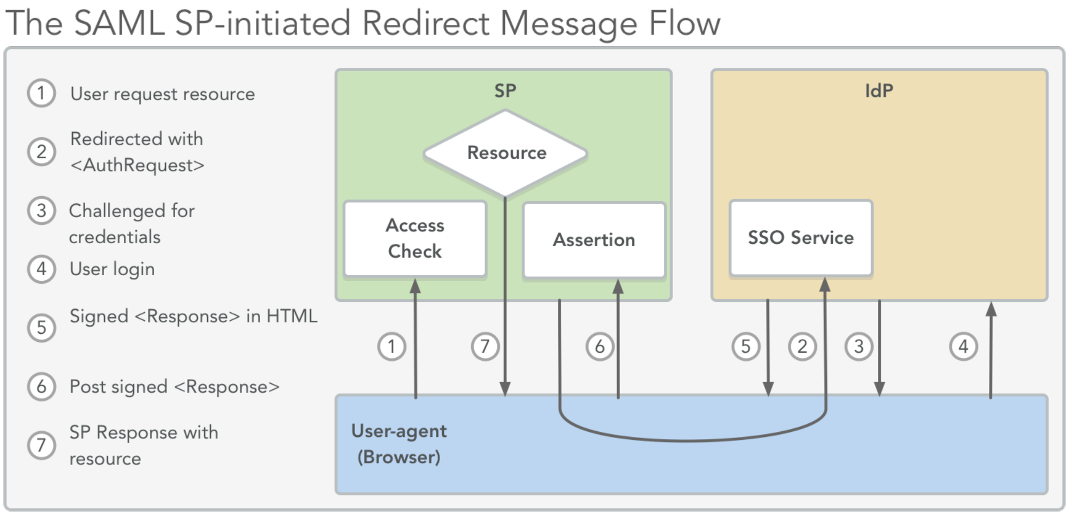

The SAML SP-initiated Redirect Message Flow
1 User request resource
2 Redirected with <AuthRequest>
3 Challenged for credentials
4 User login
5 Signed <Response> in HTML
6 Post signed <Response>
7 SP Response with resource
SP
Resource
Access Check
Assertion
IdP
SSO Service
1
7
6
5
2
3
4
User-agent (Browser)

Figure 4.2: One of possible SAML Flow specified in the Standard

One of the possible SAML interaction is illustrated in figure 4.2: The subject needs to be authenticated and thereby granted access to a set of restricted resources, protected by Access Control (AC). The process is bootstrapped by the subject uses a user agent to request access to the protected resources. The asserting entity or Identity Provider (IDP) generates a SAML assertion concerning the subject which is then sent to the SP which hosts the resources. One way of doing this with web technologies is by storing it as information on the client side in a cookie, or similar storage. The subject is either given or denied consent to access resources by the ACM based on the information carried over by the assertions. In order for the relying party to trust the assertion statements received from the asserting party, there has to be pre-defined trust relationship. This is enabled by the party using a cryptographic protocol to ensure integrity on the assertion.

4.3.3 OAuth

OAuth is one of the central identity and access management technologies available on the market. The standard has evolved a lot during the last 10 years and has reached version 2.0. The standard is described in RFC 6749 and maintained by IETF. The Protocol does unlike SAML not deal with authentication but is used for constrained delegation to applications and services. The technology provide the user with control over which attributes are shared with other systems. The usual scenario where Oauth is used is to enable third-party applications to obtain access to the API on behalf of a particular resource owner.

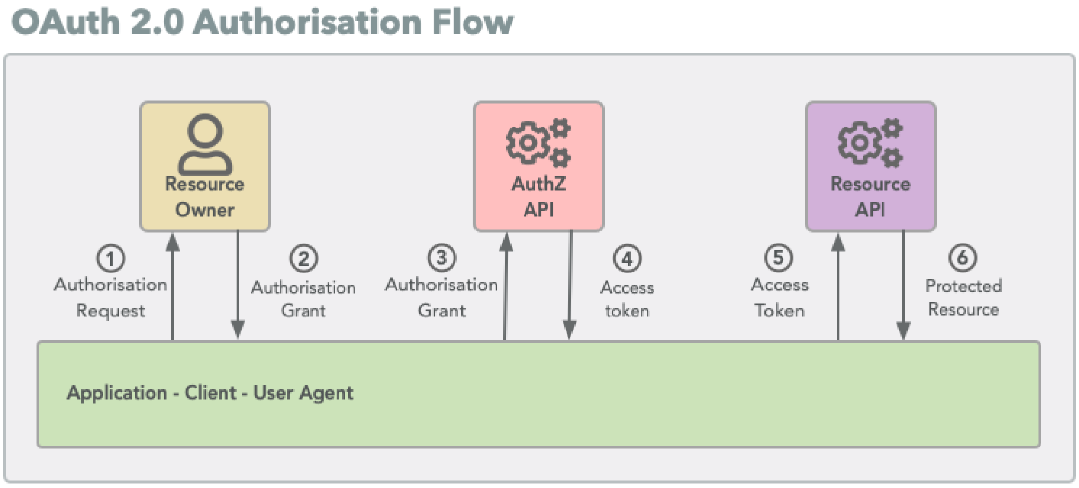


Figure 4.3: The OAuth 2.0 Implicit Authorisation Flow

In figure4.3 The OAuth flow is illustrated. The OAuth standard allows multiple flows depending on the scenario, and all these can be found in the RFC and shown in appendix A.4. The flows involve the following entities[76]:

- **Resource owner (RO) - The User**
  The entity capable of granting access to a protected resource.

- **Resource server (RS) - The requested API**
  The entity hosting the protected resources, which are able to consume and respond to requests using access tokens.

- **Client - The application)**
  An application such as a browser making requests on behalf of the resource owner and its authorisation.

- **Authorisation Server (AS)**
  The entity issuing access tokens to the client. This entity also responsible for authenticating the resource owner and obtaining authorisation.

As shown in figure 4.3, the user authorises the client or application to the authorisation entity and get an authorisation code. This code can then be sent to the client, which exchange it for an access token at the Authorisation server. The Access token can then be used to access the desired protected resource at the resource server. At the resource server, the AC check the validity of the token and permit the related access. The main benefit of this flow is the client do not need to store any of the resource owners credentials and thereby reducing a security threat.
Besides this critical fact, like SAML this flow also enables a user to keep his or her attributes in a single repository, and thereby making it easier to keep them updated. Nevertheless, SAML has only support transferring user attributes between the Asserting entity or IDP and the SP.
In Case of OAuth, because of the added entity, the protocol permits to access the attributes in an API which might not be directly related. However in order for this to work initial Trust is still needed in order for this setup to work. In this report, the term OAuth will be used to refer to OAuth 2.0, as this is the newest version and most used in the market.

4.3.4 OpenID Connect

As described in the previous section about the OAuth standard is meant to provide constrained delegation for conveying authorisation decisions. Therefore the protocol does not solve the identity problem specified in the introduction. Over the past years, the standard has become popular in relation to use with mobile and web-applications and is supported by most of the large tech companies.
OpenID Connect is the third-generation of the OpenID identity standard, unlike some of its predecessors, OpenID connect built on top of the existing OAuth flow, aimed to provide such a missing layer of authentication. In identity forums some experts have even called it an identity layer on top of the OAuth protocol.
In addition to the OAuth access token, OpenID Connect introduces an id token. The id token contains attributes about the authenticated user. The

token is signed by an identity provider entity or IDP and can be verified afterwards. This is currently one of the best proposals for an open standard, which support both authentication of the user and delegation of authorisation for the SP with respect to user identities and privacy. For the rest of this project, the term OpenID will be used to refer to OpenID Connect, as this is the newest version and most used in the market.

4.3.5 User Managed Access

User-Managed Access or UMA, is an OAuth-based protocol created to give web users control over their data through unified control. The first work started in 2009 and the standard reached version 1.0 in 2015.
In 2018 the second and updated version of UMA called UMA 2.0 was released. The general concept about UMA is, it allows the users to manage and authorise who and which of their online resources other entities can get access to, thereby providing proper delegation authorisation to user identities and data belonging to the user. This includes both online personal data (such as identity attributes), content such as images, text and videos, and services, no matter where all those reside on the internet.
UMA is built on top of OAuth 2.0, which UMA inherit and reuse the data flow architecture and technology laid by the OAuth standard[3].
Thereby can UMA achieve fine-grained authorisation control that involves implicit user permission in the form of policies or explicit by incorporating run-time access grant.

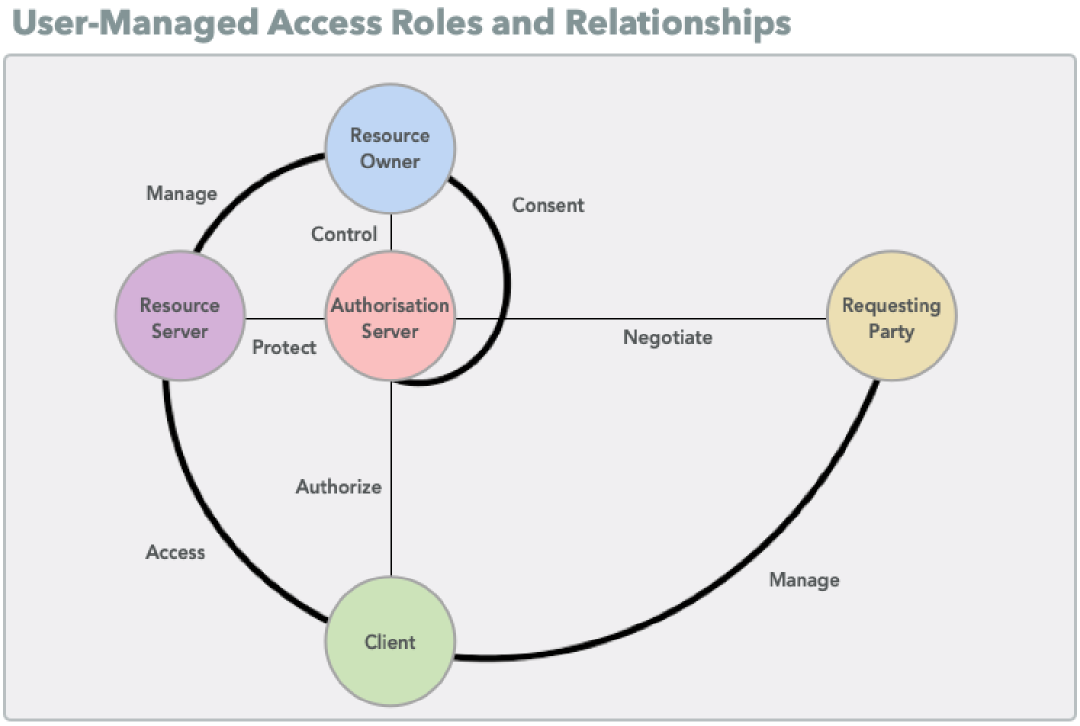


Figure 4.4: Entities in UMA and their relationships inspired by the UMA interaction diagram[3]

The UMA standard built on top of the entities described in OAuth[3]. An illustration of the relations between the entities in UMA can be seen in figure 4.4.
The design of each entity is different because UMA introduces different roles for the entities compared to OAuth by introducing new definitions, scopes and endpoints for the entities.
The protection API is a standardised endpoint as they are called in UMA is available at the Authorisation Server (AS) which is used to communicate with the Resource server (RS) This interaction enables multiple RS to be used in a scenario, and thereby accommodating a many-to-many relationship in UMA. the result of this is that a user can control, monitor and manage consent for his or her resources in a single place instead of at each SP.
The Security is inherited from OAuth and allows formal trust between the between the entities. Another addition to UMA which differs from the OAuth specification is the notation of the requesting party (RQP). The RQP can be separated from the Resource Owner (RO) in UMA, this construct enable a delegation of access for individual resources and party to party sharing. UMA support claim based identity from the RQP which result in trust elevation based on tokens, which can be associated with the authorisation details. In this report, the term UMA will be used to refer to UMA 2.0, as this is the newest version and most used in the market.

4.4 Next generation Identity solutions

The following sections will introduce the current and upcoming solutions in regards to identity technology. First, a little theory will be described, then the current solution regarding PKI, lastly distributed identities will be described.

4.4.1 Verifiable Claims

For the large majority of the industrialised world, the digital signature is already in place and can be used as valid legally binding signatures. Many of these solutions make use of public key cryptography where the private key is used for signing, and the corresponding public key can be used for validation. With the use of this tool, it is possible to determine and ensure the integrity of both documents and signatures. This technology has been an enabler for most of the systems that we rely on today, when conducting business online. However, a central problem still exists concerning the use of this technology. The main challenge is verifying, that the public key is correct and belong to the right issuer. For the widespread adoption of digital credentials, a solution to this problem is needed. For a long time, the answer to this problem has been Public key Infrastructure (PKI).

The technique provides the possibility to define, exchange and verify digital credentials over the internet. The strength of Verifiable claims lies in the level of trust the verifier has in the issuer. This is presented in Figure 4.5. For example: if a university issues a claim stating that a person has taken this degree. A miring manager can rely on the claim if he has a high degree of trust in the university.

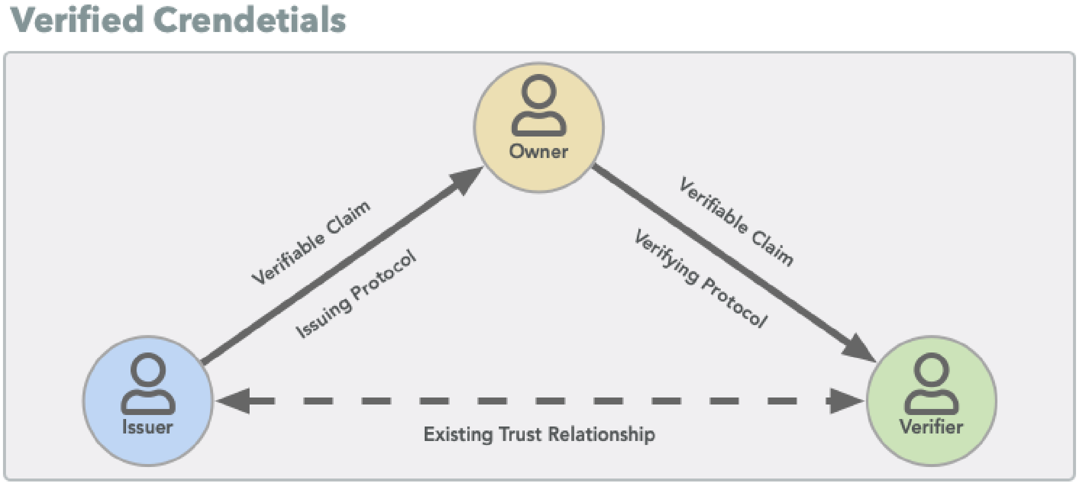


Figure 4.5: Flow of verified credentials between owners, verifies and issuers

4.4.2 Public Key Infrastructure

PKI is the technology used in modern browsers and other information systems, to verify the relation between keys and identities. The system architecture of PKI is centralised and relies on a few hundred trusted certificate authorities also known as CA's. The number of these fixed authorities are small enough for the browser to control. In Figure 4.6 the relation between entities is mapped.

The PKI system is the primary solution for verifying certificates. Almost any transaction uses this infrastructure to verify the identities of web pages and web applications. However, the PKI is not limited to use on the key, due to the possibility to verify keys without contact to the issuer of the certificate. For all instances, where a web page is accessed through the browser ny HTTPS, the browser checks whether the site was responsible for the request, and whether it was placed from a particular domain. This is done, to ensure that malicious pages and entities are not redirecting a user. Similarly, when a new piece of software is installed, the OS checks whether the installed software is from a reliable vendor.

- User - Entity verifying the certificate .
- Subject - The entity requesting a certificate.
- Certificate Authority (CA) - The entity that issues the digital certificate binding subject's identity with subject's public key.

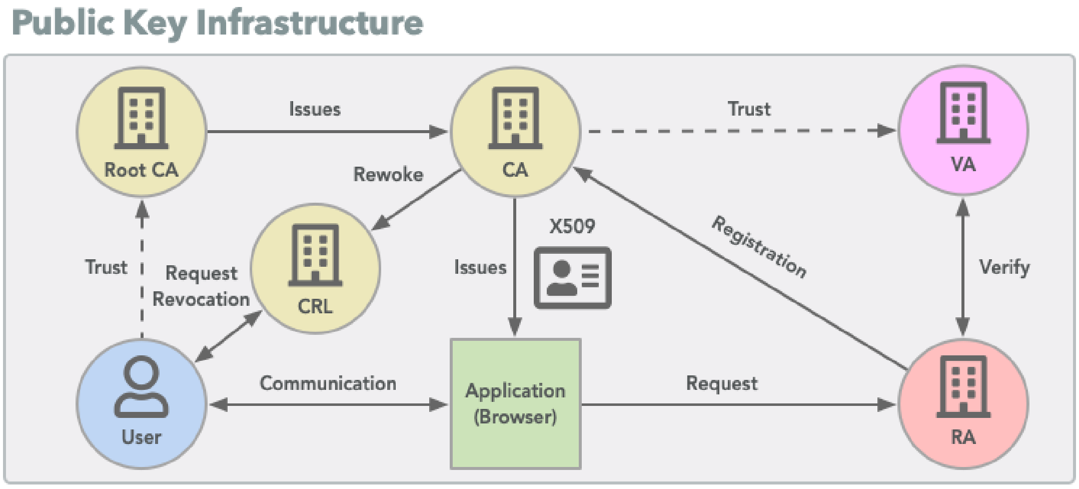


Figure 4.6: Entities used in Public Key Infrastructure (PKI)

- Root Certificate Authority (Root CA) - The entity that issues the certificates to the CA's.
- Registration Authority (RA) - The entity that verifies the identity of the subject.
- Validation Author (VA) -The entity that verifies the credentials of a subject on behalf of an RA.
- certificate Revocation list (CRL) - Holds a register of revoked certificates.

The root certificate (Root CA), also defined as the trusted root, is at the centre of the PKI trust model which underpins the HTTPS (SSL/TLS) protocol. Each application which utilises encrypted HTTP traffic, includes a root store. Some applications run their own, like most web browsers, whilst others make use of a third-party, as most Operating systems (OS) have. The root store contains a group of pre-downloaded root certificates which is stored on the devices. A root certificate is invaluable because the browsers will automatically trust any certificate that is signed with its private key. The organisations that operate Root CAs and CA, validate and issue SSL certificates by validating credentials from the subject against public records [77].

In modern PKI infrastructures, CA does not issue server certificates directly off-of the Root CA. This could cause unnecessary risks, due to potential mistakes, and would require the root to be revoked. This would result in distrust for all certificates that was signed using the root. To avoid this, an intermediate root with its own private key is placed between the Root CA, and the entity applying for certificates. The CA's certificate is signed by the Root CA, and can with its own private key sign either underlying intermediate CA or individual SSL certificates [77]. This infrastructure has shown, that it functions and scales well. However, PKI has a number of potential issues. Initially, certificates from CA, that are respected by hardware and software vendors, takes time and effort to obtain. Secondly, validation of subjects seems to be somewhat superficial, and there have been many cases where it has been possible to obtain certificates without having the required identity [78]. It has occurred that CA was hacked, and allowed for malicious people to issue false certificates to a number of large sites [79]. A business, running a respected CA, can in some manner be described as a business that has a license to print it's own money [52], in the sense that running a CA has a low cost while customers are continuously charged for the service. As soon as these large CA certificates are built into the software, they typically remain there. It is too hard for most people to deal with managing the trust of CA by themselves. This centralisation around CA's can lead to censorship and in worst cases, a single point of failure [52].

4.4.3 Distribution of Identities

Traditional IdM systems are built on centralised authorities such as certificate authorities (CA) or directory services (DS). These services are based on the concept of trust verification and each of them are centralised authorities, which serve as their own root of trust or a hierarchy thereof. To make the identity management work across the organisation or between enterprises, systems require implementing federated identity management, as described in Section 4.3.1. This paradigm of identity technology is sometimes referred to as the second generation [80]. Various researchers in identity management have stated that Blockchain, or distributed Ledger technology (DLT) [80, 81], forms a 3rd generation identity technology. This new generation will provide the possibility of making purely decentralised identity management [82]. One of the proposals for such a system, is The W3C CCG's draft for a decentralised identity system (DPKI) The goal for decentralised IdM systems, is to create "self-sovereign" digital identities, which do not have the same limitations as 1st or 2nd generation identify technologies [82].

Global DLT accommodates the tools for managing a root of trust with no centralised authority or any single point of failure [82]. In combination, DLTs and decentralised IdM systems have the power to enable anybody to create and manage identities, based on any independent roots of trust [82]. This design eliminates dependence on centralised registries for identifiers as well as centralised certificate authorities for key management. Another benefit of this system is, that it allows for building systems incorporating identities, that are made with "Privacy by Design" principles in mind [80].

4.4.4 Distributed Identities

Entities in the identity are based on the Decentralised identifiers or (DID). A DID points to a DID Document, which describe service endpoints that can be used to interact with the entity. There is no limit to the amount of DID that an entity can have, so that a person-entity could point to persona like "Employee", "Dad" or "Shareholder"-identities. One of the core elements of implementing DID is the definition of DID methods. These methods specify the set of rules for how a DID is registered, resolved, updated, and revoked on that specific ledger or network. Another benefit of this solution described in the W3C document, is that DID methods might also be able to target identifiers registered in federated or centralised identity management systems. This produces an interoperability link between the systems of centralised, federated, and decentralised identifiers and their systems [82].

Because each DID has an associated public-private key pair, everybody with a DID is able to issue and sign verifiable claims and other digital documents. The only requirement is, that the verifier needs the DID of the issuer, which is a simple matter to look up the issuer's public key and verify the signature of the claims. The introduction of DID means that nobody should need to create identity federation in order to verify identities over the internet. This structure thereby breaks down the enterprise silos of identity, and are able to decentralise the CA role as current solutions make use of. This model has been known under the term Web of trust, which was initially introduced by the founder of the Pretty good privacy (PGP) protocol. The requirements for such a solution have been set by W3C and can be read in Appendix A.7.

4.4.5 Sovrin

Sovrin is a self-sovereign identity solution which is built on top of W3C specifications in combination with digital ledger technology, also known as Blockchain. Sovrin is designed, based on the following requirements [52]:

- **Governance** - The network can be trusted by all stakeholders.
- **Performance** - The network can provide self-sovereign identity at Internet scale.
- **Accessibility** - The network can ensure that identity is available to all.
- **Privacy** - The network can meet the strongest privacy standards in the world.

These requirements are met with the use of the DID technology. In Sovrin, for all relations, a DID is generated. This means that if the DID is exposed, it doesn't affect the users' other relations, which is a solution known as the Pairwise-pseudonymous identifiers. In order to comprehend this, the Blockchain must allow a massive scale of DID, that would enable every entity using the system to have a DID for all their relations that each contains a public-private key pair. As shown in Figure 4.7, a Sovrin identity ledger is illustrates with the individuals Victor and Peggy and their wallets containing the keys. With the use of Sovrin on top of a blockchain, they can verify each others claims with use of DID because the system allow for zero-knowledge proofs.

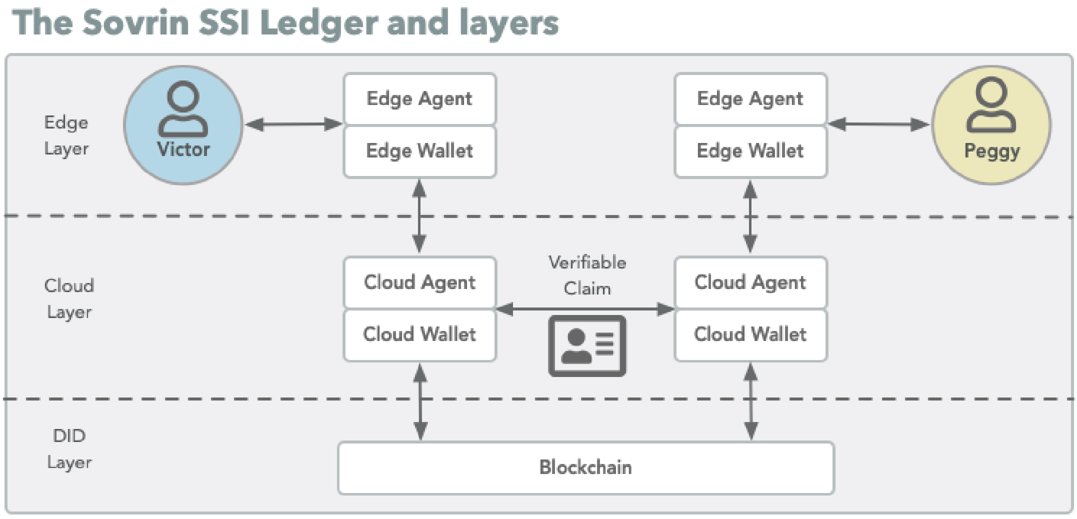


Figure 4.7: The Sovrin Identity ledger architecture

4.4.6 Solid

Solid was presented in 2018 as a privacy-aware architecture for web applications. Solid a set of web specifications that, used together, enables a Solid framework that can help to build privacy-aware web applications. Solid is co-created and supported by Tim Bernes Lee, known as "inventor of the internet". T. B. Lee was the initial proposer of combing hypertext and the TCP network protocol at CERN in the eighties. Solid build on top of traditional internet technologies and principles such as REST, HTML and Javascript. Solid was also build on top of one of Tim's other inventions called the "semantic web", which enables data to be shared and reused across the interconnected host on the internet. The concrete standards behind the idea, are mainly driven by the Resource Description Framework (RDF) that enable information to be statements about other resources in a subject-predicate-object model, known as triples. Using this in combination with the URI and URL specification form the concept of Linked Data, it enables data within a document to be linked together. Solid introduces an identity layer, that is based on the WebID standard by W3C. Even though the technology provides technical solutions to one of the highlighted problems of the internet, the solutions is not widely used.

This could change over time, but current support of the solutions seems to be limited.

4.5 Access control

The term access control refers to the process of controlling access to a resource, such as a service, a room, a system or a document within a system to authorised entity. Physical access control can e.g. be controlled by a manager, a bouncer, or receptionist. Such an entity is known as the guard of the object. In the information security context, this is also referred to as access management. Wood [83] defines this as the process of controlling access to resources by providing a policy-based control of who can access these specific resources based on individual permissions. Services enforce access to resources by requiring the subject to provide credentials that prove their identity. Credentials such as username and password or physical tokens are commonly used as authentication method in order to gain access to resources. However a movement towards multi-factor authentication becomes more normal in recent years [84]. In inter connected systems, such as the ones connected to the internet, authentication could also be externalised to a trusted party, this model is known as federation, and described in section 4.3.1.

4.5.1 History of Access Control Models

The first models of access control seen in the field of software were introduced with the Discretionary Access Control (DAC) and Mandatory Access Control (MAC)-model in the 70's. These models originated from research conducted at Bell Labs, on the support from the U.S. Defence science Board. The Department Of Defence (DoD) knew there existed a potential in this area due to previous studies which had unveiled that governmental systems were without proper security and at high risk and vulnerable [53]. This work ended in a publication which described guidelines for the use of access control models in computerised systems. Today this publication is known as the "Orange Book". The models introduced in the book were already proven models by researchers in the field of access control and covered mathematical proofs for ensuring data integrity and confidentiality in systems with either the Biba Model or the Bell–LaPadula model. The publication was not only targeted for the military systems but could also assist enterprises and private as wells as governmental systems.
Despite the promotion of these models, the result turned out to be slow implementation, where it took a decade or more before proper access control models were implemented in many systems [53, 85].

4.5.2 Modern Access Control Models

A modern access control model (ACM) as the one described by Lampson in [57], refers to what he calls the Gold plated standard, due to the Authentication, authorisation and auditing all starting with the letters "AU". This model has become the de-facto security model used in information technology today. The guard authenticates the subject by verifying the presented identity and subsequently authorises the subject to access the resource based on a specified policy. This model is fairly abstract, and do not dictate implementation. However due to its construct it can be used to describe a wide range of scenarios.
In the next sections, the process of authenticating and authorise will be described further to set a baseline for how the proposed solution should be understood.

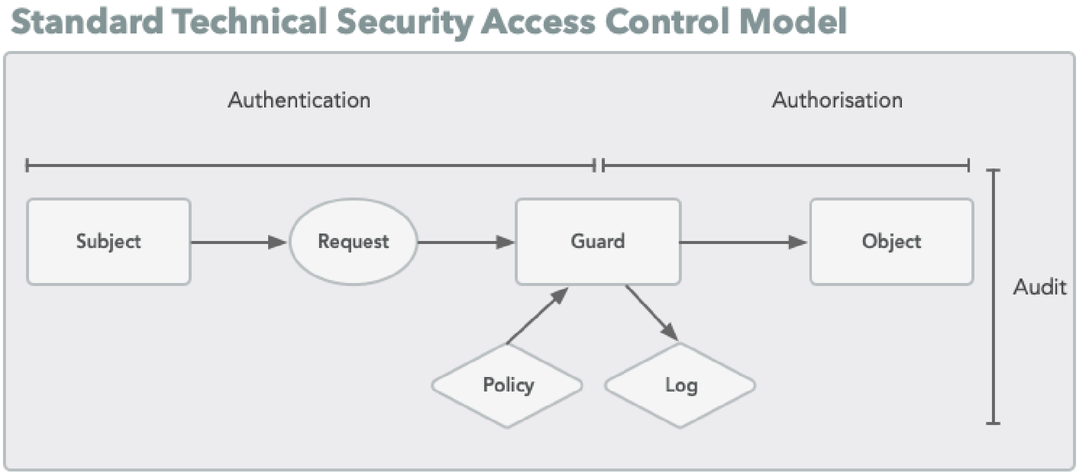


Figure 4.8: Illustration of the modern Access Control Model inspired by Lampson

As illustrated in 4.8, the Lampson's described ACM also known as the Gold plated standard.

4.5.3 Authentication

In the ACM, for a user to get access to a system, their identity has to be verified to be who they claim to be. This process is known as authentication. In this section, authentication in general and how the method applies to the modern computer systems is described previously. Newer models have been proposed to deal with authentication in various contexts. One of these are the NIST which describes authentication through networked systems as "E-authentication model"

NIST refers to E-authentication as: *"The method of establishing confidence in user identities that are electronically presented to an information system"* [4]. When a user is requesting access to a resource, The authentication process establishes confidence in the user's identity and might endorse specific attributes [4]. Authentication cannot determine the requester rights. The guard of the services rely on the authentication process and is responsible for using the verified identity gained in the authentication step to make access control decisions [57]. The model is separated into two phases, the Registration Phase (in figure 4.9 to the left) and the Transaction Phase (in figure 4.9 to the right) The entities used in this model are [4]:

- **Registration Authority / RA** - *The responsible entity for establishing and vouching the identity and attributes of a claimant. The RA can also be deployed together with the CSP.*
- **Claimant / Subscriber** - *The entity whose identity is going to be verified using some authentication protocol.*
- **Relying Party / RP** - *The body that relies on the token or credentials provided by the Claimant, or the assertion supplied by a Verifier, Entity responsible for processing a transaction, represented by the guard.*
- **Verifier** - *The accountable entity for verifying the token that is owned by the Claimant. However, the verifier may validate the token for the CSP.*
- **Credential Service Provider / CSP** *The entity that holds the register and issues tokens for subscribers.*

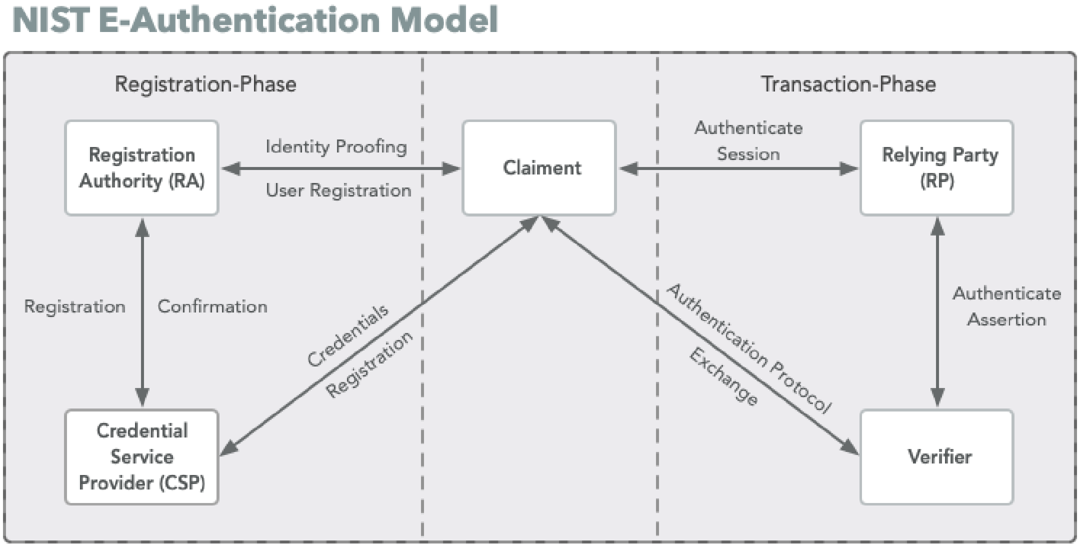


Figure 4.9: E-Authentication Model by NIST [4]

The Registration Step

As shown in figure 4.9. The registration phase bootstraps the model by provisioning of the users. The following interactions between parties in this phase are as follows[4]:

1. A subject applies for registration with a RA. The RA might require to authenticate the Claimant.

2. The identity of the Applicant(Claimant)
   The RA verifies the identity of the applicant, either by authenticating it against some trusted source or confirm the identity and take the role of being a root identity source.

3. The RA sends a notification to the CSP with a registration confirmation message

4. The CSP issues credentials to the Claimant.
   A token is created that corresponds to the credentials either by the CSP or generated directly by the Subscriber.

5. The claimant retains the token
   The CSP maintains the identity with the associated credentials and their lifetime while the Claimant retains the token. The CSP has the registration of the Claimant. The claimant can now make use of their distributed token to proceed with the next step of the authentication procedure.

The Transaction Step

When the registration phase is finished, the claimant can start to make requests to the RP and continue to the authorisation phase. The transaction phase includes the following steps [4]:

1. The Claimant shows the Verifier that he holds the token

2. Verifier validates this token at the CSP, and make sure that the credential binds the Subscriber's identity to the token

3. If the Verifier and the RP are two distinct entities, the Verifier will give an assertion about the Subscriber (later claimant) to the RP. The RP can then use this information to make authorisation decisions.

4. When all the preceding steps are finished, the authenticated session is established.

Even though the authentication phase was successful, the claimant might not be allowed access to any resources. This is due to missing authorisation which determines if a subject have access to their requested resources.

Credentials

In the authentication model, the CSP maintains and issues credentials. They are used to authenticate the Claimant's identity by the Verifier. The idea of the credentials is quite similar to the use of traditional physical

credentials like National ID-cards and passports [4], as described in the previous section. In the E-authentication model, the token possessed by the subscriber is bound to identity by the credentials.
In the physical world showing a credential might be enough in most scenarios. In the digital world, the authentication process are increasingly tedious due to the uncertainty in electronic communication. In this case, the Claimant has a token and presents this to a token authenticator, but the claimant does not necessarily possess electronic credentials [4].

Assertions

The assertion is defined as a confident and forceful statement of fact or belief and the Verifier gives assertions to provide the outcome of the authentication to the RP. The RP and the Verifier entity can be implemented together in which case the assertion is implicit. If this is not the case, the Verifier makes use of the assertion to send information of the desired kind about the Claimant to the RP [4]. To send the assertion, the RP and Verifier can communicate directly together, or use the Claimant as a messenger. Since the authentication decision at the RP is computed with the use of the assertion details, it can contain sensitive information about the Claimant. Therefore the Verifier must ensure a high level of integrity and confidentiality[4].

Authentication Factors, Tokens and Types

Authentication factors can be divided down into three types. For the process to be successful, at least one factor of at least one type has to be used. A higher assurance level can be given if the claimant can provide more than one of these factors. This is known as multi-factor authentication. Systems can utilise this at the access control level, making the guard require a minimum level of multi-factor authentication or more for operations[4] The following is categories represents the factors which can be used for authentication:

- *Something you know* (eg. Persistent password, One time password(OTP))
- *Something you have* (eg. ID badge, crypto keys, hardware token)
- *Something you are* (eg. Biometric data (Fingerprint, Iris))

However, in e-authentication the process is slightly different. In order to prove their identity, the Claimant presents a token which is registered with the CSP and contains a secret. The token can take many forms. The token provides an output that is used to confirm the Claimant's identity[4].

Assurance Levels

An assurance level can be determined based on the authentication of a user's Identity. Assurance in the NIST e-authentication model is divided into four levels. Each assurance level specifies a level of confidence about the credentials provided by a subject that points to an identity. The goal of assurance for this context is to establish certainty in that the subject of the credential is the one to whom the credential was issued initially [86]. The required level of assurance for a transaction is determined by the type of activity and the sensitivity of the information required in the transaction. The four levels are described as:

- **Level 1** No confidence or little in the asserted identity's validity
- **Level 2** Confidence to some degree in the asserted identity's validity
- **Level 3** A high degree of confidence in the asserted identity's validity
- **Level 4** A Very high degree of confidence in the asserted identity's validity

These levels can be used to communicate how certain an entity is about a specific assertion. For example, an assurance level is used in a system to guard specific confidential information, so that the data is only accessible if the system has a high enough level of assurance of the asserted identity's validity.

4.5.4 Authorisation

The next step in the ACM is Authorisation, which is the element that determine which actions that can be performed based on the known identity. In today's IT-systems this concept of access control is embedded in a wide range of products, spanning across almost any industry and system. However, industries associated with a high level of confidentiality usually see the biggest need of access control. It is these types of industries which have shaped the popular access control models, which we use today.
In organisations it is the danger associated with an unauthorised disclosure of information, which has the ability to disrupt the organisation's operations that is feared. Such a disclosure could in worst case cause harm and create serious implications in the areas of competition, financial, legal or human safety [87]. Several examples exists, where confidential information have been compromised, leading to lost trust from the customers. It is such undesirable events which lay the foundation for the origin of access control, and making it highly relevant in relation to privacy in computer systems.

It is not only businesses that wants to protect their inventions and finances where access control is relevant. Businesses dealing with Personal information of any kind, have a responsibility that are governed by regulations which stipulate that they need to comply with higher levels of integrity, confidentiality and availability or a mix of these. In order to run their business these companies are regulated by law and it is here that access control comes into play. Access control takes care of the authorisation and is critical to preserving the confidentiality and integrity of information. [53]. Therefore it can be concluded that for any computer system to preserve privacy it must include an access control model which is comply with privacy concerns.

4.5.5 Access control in Enterprises

Today most enterprises today make use of Role Based Access Control (RBAC) to specify access to the applications and systems based on functional working role. This has proven to be a time saver in assigning rights across organisations. However, if an employee changes roles or leaves the company, a person from the IT department or similar must manually change these access rights across all relevant systems.

When organisations grow, partner with external businesses or systems and change systems, the process of managing access control becomes increasingly cumbersome, due to disadvantages related to RBAC[88]. Research from Gartner has recently predicted that *"by 2020, 70% of enterprises will use attribute-based access control...as the dominant mechanism to protect critical assets, up from less than 5% today."*[89]
Attribute-Based Access Control (ABAC) is an access control technology, which grants access rights to an organisation's systems based on attributes of the subject such as the person's attributes, or the objects that the person tries to access.
This new paradigm can not only increase security but it also enables stronger relationships for enterprises.

4.5.6 Role-based Access Control

Role Based Access Control or (RBAC) is rooted in the belief that the general structure of an organisation does not change much over time.
However people come and go, advance or change position more commonly than before. To comply with this constant changing set of rights, RBAC was introduced[90]. It has become one of the favoured access control models in business software. Relative to the simpler ACM's, where individual users needed promotion frequently. RBAC creates a layer of abstraction which consist of roles. These Roles are modelled after the organisation and job descriptions and address the business needs instead of technicalities. When a user needs privileges for the business process, they can be assigned to the associated role[7].

Nonetheless, this new abstraction layer generates a demand for proper role-definitions within the organisation, that needs continuous updates in order to fit the business processes and organisation hierarchy. This area of analysing and designing roles and their associated properties are known as role-engineering and is crucial to the outcome of implementing

an RBAC into the enterprise [91].
Since the invention in the 90's It has been developed and many researchers have contributed to the area. Today a standard exists for RBAC in NIST (INCITS 359-2004) [92]. Research by [53] Ferraiolo has revealed that the characteristics of RBAC fit very well with many areas of business. Which makes it fit for commercial applications such as identity management systems due to the potential to decrease administration processes.
In Figure 4.10 an illustration of RBAC is given[7] that show the named relations, User assignment (UA) and Permission assignment (PA) in relation to the entities of RBAC.

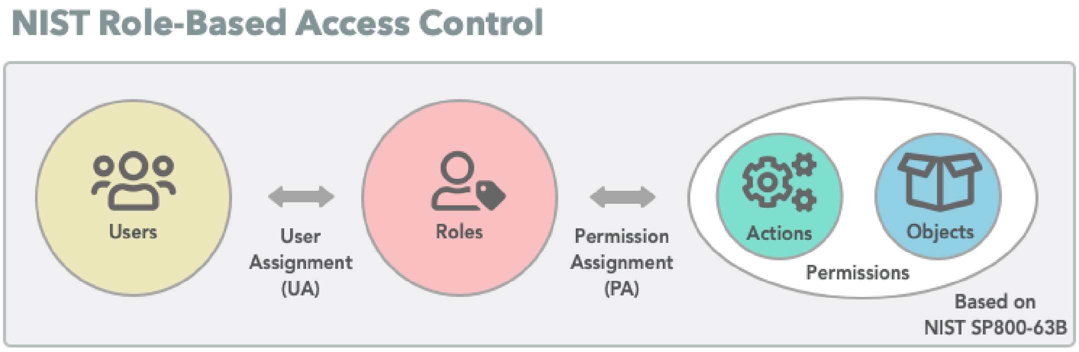


Figure 4.10: RBAC role abstraction hierarchy

4.5.7 Role Engineering

Due to the widespread use of RBAC in many software products and a considerable emphasis on its layer of roles, a need standardised method for creating and maintaining roles have been developed in the ANSI INCITS 359-2004 standard [93]. The standard contributes with an RBAC reference model, made for large projects using the RBAC access control model. The model describes a method to identify the smallest working set of features in Role-based systems and their hierarchies. Furthermore, it presents a semantic language for describing elements and functions. The standard introduces two variations, a core RBAC and a hierarchical RBAC. These models are recommended for very complex implementations of RBAC in larger organisations[94].

1. Design use cases for the system in a straightforward language.
2. Based on the use cases, recognise the roles that are needed. If junior roles are identified in the process, identify the hierarchical structure
3. Based on the use cases, identify protected resources and the related operations which the system can perform
4. Create a list of assigned pairs, listing resources and operations concerning the found roles, thus creating permissions.
5. Assign roles to users in the system according to findings

4.5.8 Attribute-Based Access Control

In contradiction to RBAC, Attribute-Based Access Control (ABAC) was created with the classical subject-object model in mind, rather than domain specific roles and identifiers[95]. Because RBAC entirely relies on roles, is not sufficient to describe complex and fine-grained policies which might be a requirement for specific operations. ABAC Opposite RBAC uses resource attributes linked to entities in the access control model. This includes Subject, objects and actions or environment, which can be described in policies. A combination of policies can, therefore, be used to decide if proper access can be given for a subject to a specific object. These properties make ABAC a potent tool to tailor specific solutions and at the same time being more straightforward from a design perspective. In contradiction to RBAC append an extra abstraction layer of roles, which might decrease transparency [95]. As an outcome, RBAC does not take into account the fact that a user may have special permissions for individual resources. The paper [95] also describe that in RBAC the properties of the resources are not taken into thought, other than their identifiers[7] ABAC do not specify a static set of attributes that need to be available for an access control decision to be made, in NIST ABAC recommendation in SP800-162, four different areas are defined, these can be seen in 4.11 in relation to the Access Control Model.

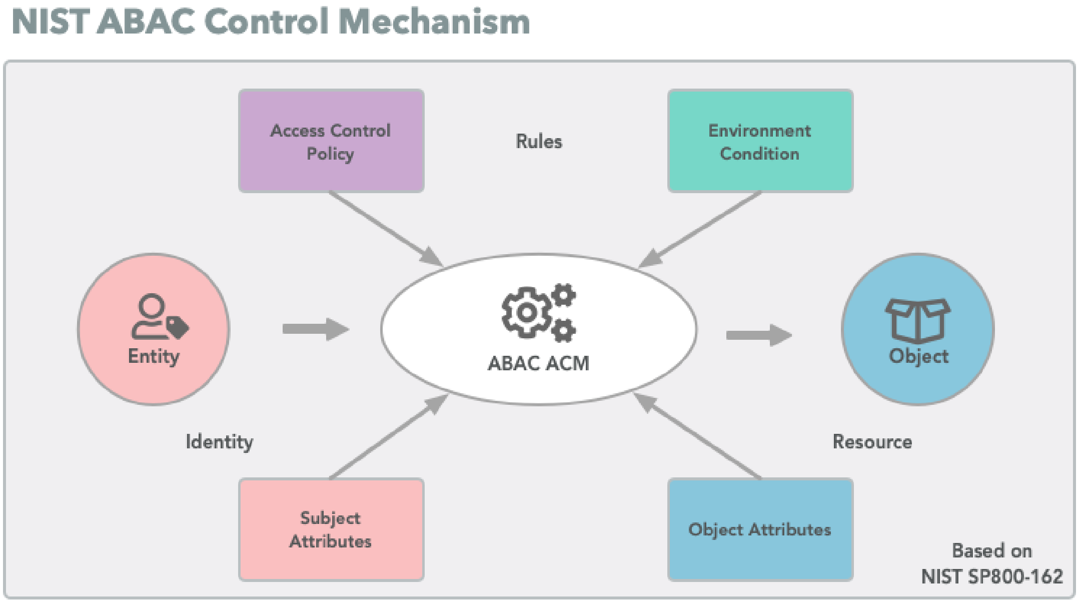


Figure 4.11: NIST ABAC Control Mechanism

As shown in the figure, the guard in charge of the decision utilises a combination of Policies, Environments condition, subject attributes and object attribute to make an access control decision. In some frameworks or standard, this guard is also known as the PEP, taken decisions from a PDP which calculated the decision based on the inputs. This would be the case if the ABAC model were implemented with the use of the XACML technology.

4.5.9 eXtensible Access Control Markup Language

XACML v3.0 by OASIS is a standard which allows access control requirements to be implemented in a standardised fashion. XACML specifies an architecture as well as a language these access control policies should be modelled into. These two can in combination be used to create access control decisions within a system. XACML is built around the XML syntax which is used to define action-rules for subjects and their targets. In most cases, actions in XACML corresponds to requests and related responses, while subjects and their targets are mapped to users their requested resources.
Further, XACML is a standard which defines both the access control policy language and a request/response based policy decision language. These are both used in the XACML processing model [96]. The policy language referes is a syntax for how to describe the access control requirements and the request/response language is a syntax used to describe if an access request is permitted[97].

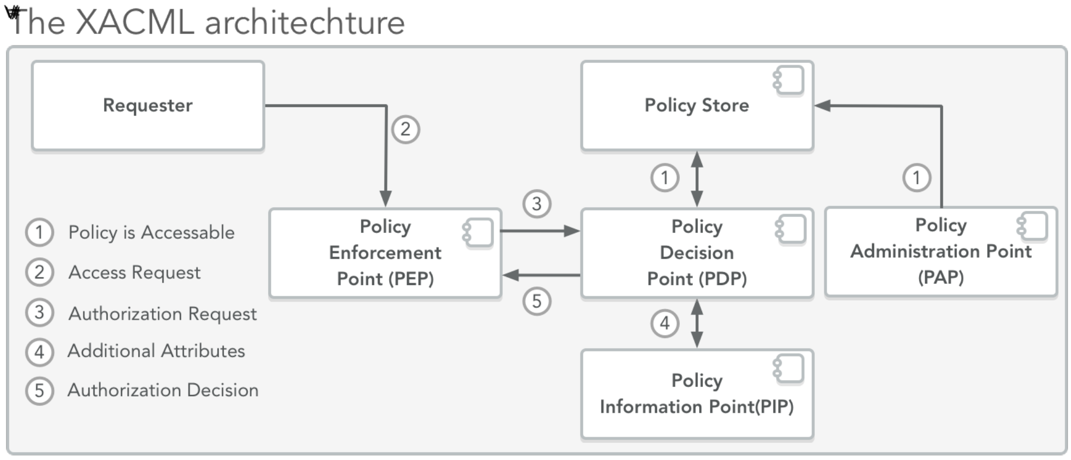


Figure 4.12: The XACML Process model, inspired by the illustration by OASIS[5]

The following entities are introduced in the XACML processing model; these are further illustrated in figure 4.12. The functions of the entities in the illustration are described below [96]:

[noitemsep]**Policy Decision Point (PDP)** - The entity in XACML that provides the authorisation decision after assessing and evaluating the applicable policies. **Policy Enforcement Point (PEP)** - The entity in XACML that enforces authorisation decisions and thereby performs access control, in relation to the theory described in relation to the ACM in Section 4.5.2, this entity represent the guard in the system. **Policy Information Point (PIP)**

- The entity in XACML which acts as a source for attributes required in order for decisions to be taken. This might be a combination of attribute sources or an aggregation unit. **Policy Administration Point (PAP)** - The entity where policies are administrated. This includes creation, modification and removal. Managers use this to provision policies. As described in the Standard [98] and illustrated in figure 4.12, the flow of data in XACML[1] is as follows[7, 98]:[noitemsep]

- 1. The policies are built in the policy administration point. The policy administration point makes these specifications available to the policy decision point.
  2. The policy enforcement point gets an access request about an object from the requester (subject)
  3. The policy enforcement point the creates an XACML request based on the received request and sends it to the policy decision point. The request may include the attributes of the subjects, resource, action or environment
  4. If needed, the policy decision point can request any additional subject, resource, action, environment and other categories of attributes from the policy information point. The policy information point obtains the attributes and returns them to the policy decision point.
  5. The policy decision point assesses the policies and returns the decision to the policy enforcement point.
  6. The Policy enforcement point enforces the decision by allowing or denying passage to the requested resource.

Based on this flow, a system implementing XACML can address a wide range of access control requirement by implementing them as policies in the XACML. Further, this can be used to combine requirement from various entities, because the XACML language is standardised and can be merged into a resulting policy.

4.5.10 Conclusion

In this section, two access control models have been described. RBAC which is one of the most used models used in many places from OS's to physical access control. Another model ABAC were also described, one which is can be can be used in fine-grained access control. These models will be used as reference models in the further works of this report. With use of a Access control policy decision language like XACML security requirements can be implemented. To meet Business logic and organisation structure model like RBAC and ABAC can be used. These solutions will further be used to design and implement the proposed solution.

[1] The data flow is simplified compared to the standard in order to fulfil the needs of description and to fit of problem field

Chapter 5

Analysis

Not so long ago, IT security was about protecting the organisation's perimeter. with recent extension and advancement in service and deployment models the traditional perimeter might not exist anymore. If it does, its primary objective is to protect the remaining legacy systems that the business still rely upon on [99]. The move from proprietary protocols and technologies to standardised API in combination with "Everything over IP" and *-cloud solutions are changing the landscape of private networks and dedicated connections. Because of this new distribution of Data, alternative ways of dealing with the flow of data and information security need to be addressed in other ways than just a decade ago. Diagrams over racks with network equipment can in some instances be able to deem obsolete, do to major parts of it being virtual. Applications might not reside inside the perimeter and in today's IT landscape Privacy is a larger concern. IT systems need to open up against all their users and break down the isolated silos of identities and attributes.

Various solutions and technologies that solve these challenges are presented in the previous chapter 4 SOTA. Current federated identities technologies let systems authenticate by assertions from trusted third-parties locally or over the internet. This 2nd generation identity solution has proven its worth. However, these solution does not support delegation of authorisation or support for consent from an enterprise perspective either inside or outside the border of the enterprise. As stated in the introduction this is a requirement for modern IT-systems to stay relevant when dealing with private information. Newer IAM technologies such as UMA and Sovrin contribute which these missing building blocks, however, none or few implemented solutions exist today. This chapter will analyse how these technologies and concepts can be used in an enterprise context to create a privacy-aware systems with use of a privacy aware ACM. One thing is to build a stand alone system with a certain amount of requirement. Another is to implement a solution in the real enterprise.

To understand the environment that such a system is going to be incorporated in to, an analysis of current enterprise practice is needed, how identities are used in enterprises and what current practise is. In the following sections, this will be analysed. Further an analysis of the application and services in relations to how they use identities will also be covered. this involve how identity and consent management can be incorporated into an enterprise context.

5.1 Identity Management

The sequence in the identification process highlights a traditional problem field to understand the concept of identity. For example, when a person interacts with an enterprise from a remote location using the phone or internet. In such scenario, identification of the parties is necessary to establish needed trust around the agreement. Such an identification process relies on a good identity store. In the analogue world, this can be a list of clients, partners or customer. With such interactions and communications moving from the physical world to the digital, centralised identifiers introduces a higher privacy problem due to the nature of computers and data storage stated in the introduction. By reusing ids for people poses a critical point because of citizens risk being stripped of their privacy, such as personal security numbers, TAX numbers or mail addresses. A solution to this problem is providing all parties with partial digital identities that can be used as they prefer, no matter the context. An example of this could be, one partial identity with an associated identifier for each system, subsystem or action. Thereby only the system can connect the related person or entity to their legal entity, and exposure of identifiers won't harm the privacy of the user, as it not used anywhere else. The problem is amplified by the skyrocketing amounts of systems incorporating an element of identity, and the trends do not seem to slow down anytime soon according to IAM consultants KPMG [100].

Besides the privacy problem, It becomes quickly evident that a system where an IT-administrator needs to provision users together with their permissions, one by one, is very time consuming and expensive. This might also introduce security problems by old users not being deprovisioned or vice versa. Further, it is easy to see that the structure of these roles in the organisation is fixed and is not changing on a regular basis. Therefore there exists a market for handling identity, and this is what Identity Management is all about. Much work and research have been conducted in this field, and there exist frameworks today that respects these ideas of full as well as partial identities. The general goal is to tackles this problem and aims to guarantee a secure, reliable and privacy-respecting management of identity information, with both the needs of individuals and organisations in mind [47]. As described previously, Identity Management (IdM) is the process, and Identity Management Systems (IdMS) is the system that takes care of these processes. A complete definition is given by Windley [11] as:

- **Identity Management (IdM)** - The organisation and IT-related processes for handling identities (partial or full) and their transformations, taking into consideration an identity lifecycle and the context an identity is acting in (e.g., governmental, enterprise, or private)
- **Identity Management Systems (IdMS)** - The computer systems that support handling identities (partial or full) and their transformations, taking into consideration an identity lifecycle and the context and control of transferred (personal) data

However, Royer [17] further states that IdMS also takes care of:

- Linking identities to physical persons.
- Fulfilment of Authentification, Authorisation, Administration and Audit (AAAA).
- Anonymisation, Pseudonymisation or Identies.

Most of the research found during this project in the area of identity management utilise the title "IDM" and "IDMS". However, it seems that over the last couple of years there has been a development which has introduced the term Identity and Access Management as a substitute for IDM (IAM). According to definitions of IAM, IDM is a central subset of that term not dealing with access and governance in any certain degree[38]. Nonetheless, it seems that most of the time these names are used interchangeably in most literature, and the only difference is that the term IAM is used primary today, where IDM was more used 5 to 10 years ago disregarding it is the similar definitions they hold. In this project, the term IAM will be used when the elements of access control are integrated into the system. Moreover, IDM will be used for systems that solely deals with identity management.

This identity and access management systems do not only control the identities of the employers in the enterprise, features including access governance and process management have also become standard in these solutions. This change is sustained by a regulatory requirement which imposes requirements for higher information security standards such as ISO2700X[38], ISF [101], NIST Cybersecurity guidelines [102] and CoBit [103] according to [7]. The governance of these standards is even more critical when the enterprises must deal with legislation concerning the area of business they are operating within. A general example of this is GDPR regulation, which came into force in may 2018. In such case, some of these frameworks can help provide the general baseline for safely managing data. However, it is deemed out of scope for this project to investigate these frameworks.

5.2 IAM systems in Enterprises

Identity Management Systems (IdMS) or just IDM aim to address this area of problems with automated processes and easy administration of identities in regards to user identities. Nevertheless, IDM's do not only deal with the technical challenge, but there is also no single solution to the identity problem described in this introduction, that will work for all enterprises. Therefore the creation of such a system raises questions about who in the organisation that determines the security policy and rights and for which data[17]. Denis Royer further describes in [17] the implementation of IDM's systems as they touch the power balance and culture of many organisations "therefore, currently no simple way to decide whether and how to introduce or change IDM's" in the organisations. This fact further emphasises the topic for this thesis, that there is no golden route to introducing identity and access control system in enterprises, especially regarding the extensive requirement set by users and regulations as described in the introduction chapter.

5.2.1 Why enterprises introduce identity

In this section, details on identity management will be studied. The IDM system is not a goal by itself, and they are incorporated in order to solve other problems that the enterprise faces. In 2009 KPMG did a study which showed that the main reasons for introducing IDM/IAM into the enterprises was to *improve compliance, reduce risk* and *increase business values* [100]. An alternative definition was provided by Rannenberg [8] who provided the following description:

- ***Primary goals:***
  - *Business Related goals as Efficiency, Automation And Cost reduction*
  - *Compliance goals*
- ***Secondary goals:***
  - *Risk management and IT security goals*
  - *Enabler for future business opportunities*

Rannenberg further states that there can be overlaps between these goals, and a combination of these goals is the actual driver for implementation of such systems. The point that improving compliance is the central driver for IDM projects, indicates that IDM is seen as a solution to comply with the increasingly severe regulatory requirements imposed by law such as GDPR and Basel II[100]. Many companies anticipate a synergy effect when investing in an IDM, due to clean up and streamlining of the process when incorporating such a system.

5.2.2 Identity Lifecycle

The leading roles of any IAM system for enterprises are related to lifecycle management of identities. Looking at identities in an enterprise, the associated entitlements of users such as roles, attributes or access rights change over time, due to personal events or from the organisation that generates the need for creation or change of accounts[104]. Identity Management, therefore, refers to the process of managing these digital identities during their lifecycle from conception to termination[105]. This also includes the process of updating, maintaining, and auditing these identities. Meints and Royer [106] introduces the identity lifecycle, to consists of the following elements:

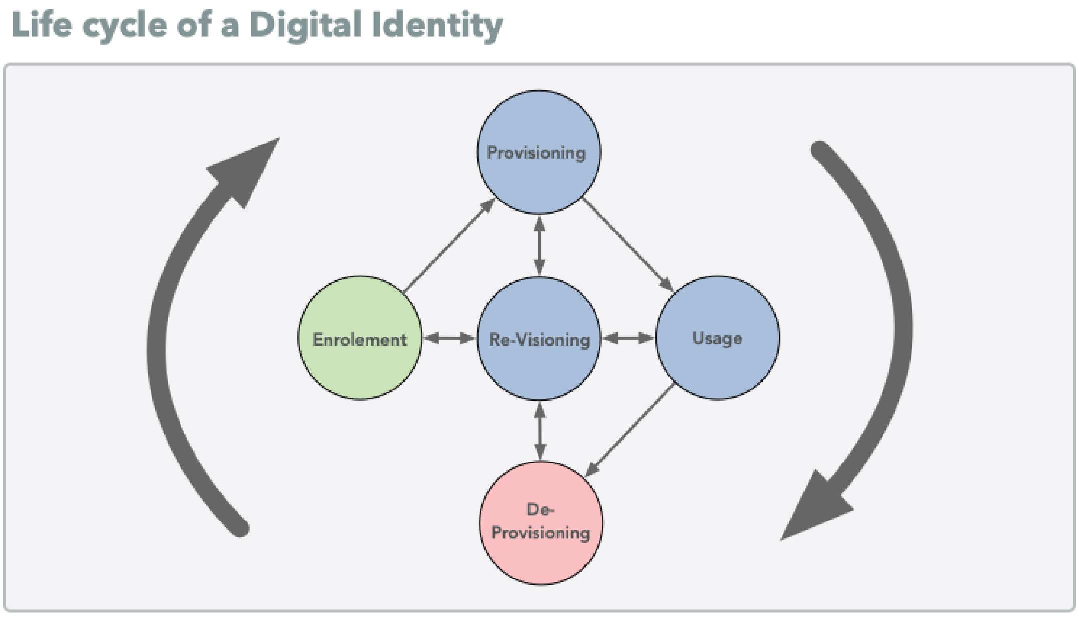


Figure 5.1: Identity life cycle

- **Enrolment** Creation of the account and mapping of the identifier to a physical entity such as a person, device or object.
- **Provisioning** Granting access permissions and entitlements to the account
- **Usage** Usage of the assigned account for accessing computer systems
- **Re-Visioning** Monitoring and auditing of the account and identity management process
- **De-provisioning** Deactivation of the account, might also include anonymisation or deletion of associated identity

Royer [17] states that planning of the lifecycle model is essential when designing identity-aware infrastructure since faults in this can lead to security issues, where de-provisioned account still can be used and thereby introduce severe security problems. In figure 5.1 the identity a lifecycle is illustrated. As described, the identity lifecycle is an essential part of administrating identities in enterprises. Lack of control can introduce security problems. Therefore IAM systems are needed and thus, the number of systems incorporating identity increases this demand exponential.

This problem further highlights the scope of this project and calls for federated identity solutions within enterprises. Thereby Lifecycle management is limited to a single point, and advanced identity governance solutions can be easily applied to all of the enterprise and its IT operations. Such centralisation might also make it easier for systems to become GDPR compliant, because identities are linked, and records of use will be linked to a central identity instead of being spread around. This should make it easier to comply with *The right to be forgotten* and *The right of access* as introduced in Section 4.2.1.

5.2.3 Stakeholder in IDM

In order to implement an IDM system into an enterprise, it is crucial for the adoption process that the system can fit the organisation and its stakeholders. Therefore an investigation into the stakeholders for enterprise IDM is necessary in order to create such a system, as stakeholder will both be represented in the design and implementation phase. The stakeholders also need to be identified in order to figure out where relevant data are stored as well as actors in the system [17]. To establish such a list, Royer [107] conducted a series of interviews with suppliers and buyers of IDM systems. This were conducted by analysing these interviews and he were able to conclude which stakeholders that are important from a general perspective. This list is presented in appendix table A.5. Moll [108] states that it is essential to have full acceptance and involvement of senior management functions, as this will work as a driver in the implementation process.

5.2.4 Organisation structure in relation to IAM

To identify how the appropriate identities and access control can be developed for the enterprise, a study of how organisations are structured is needed. Initially, an organisational structure can be defined as “a system used to define a hierarchy within an organisation. It identifies each job, its function and where it reports to within the organisation"[109]. Various structures exist, which emphasises different strategies or goals in terms of growth. According to Pearson [6], an organisation of almost any size can be seen as a group of the following building blocks, as illustrated in Figure 5.2.

Small organisations might not make use of divisions or departments, so they only have one. But as they grow, building organisations around this

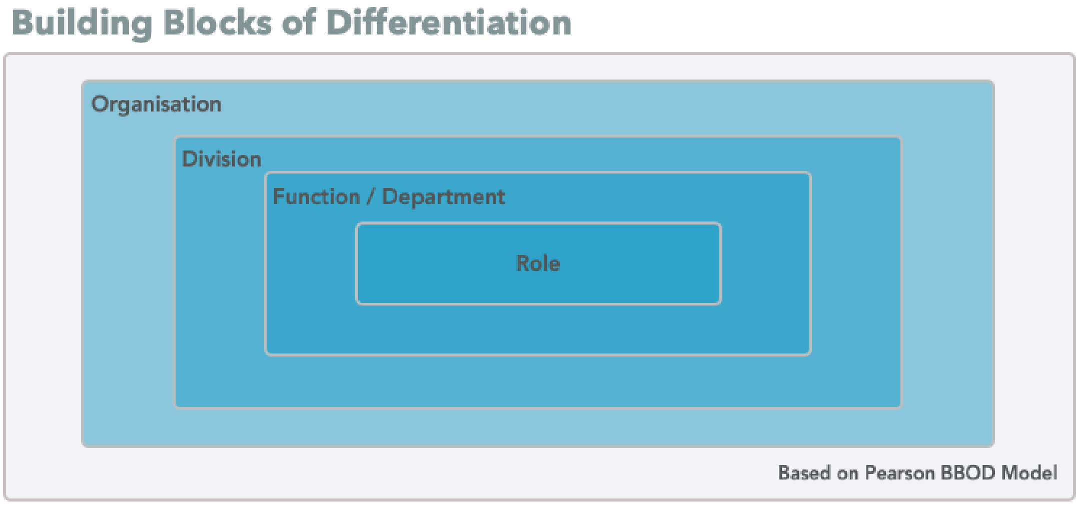

Building Blocks of Differentiation
Organisation
Division
Function / Department
Role
Based on Pearson BBOD Model

Figure 5.2: Differentiation in the organisation, inspired by the Pearson BBOD Model [6] page 95

principle will enhance communication and productivity within the organisation. This is due to the general principles in managerial economics, that organisations exist because they are able to: *Increase the specialisation and division of labour, Use-large scale technology, Manage the external environment, Economise on transaction costs, Exert power and control.* The organisation's role is a set of task-related behaviours that are required to be fulfilled by an individual with a specific position within said organisation. An example of this can be seen in Figure 5.3 where the modelled company delivers an ERP system on-premise or cloud-based for small companies. The goal is to provide customers with software that enhance and improve their workflow. The role of R&D Personal is to create software solutions with the best functionality and high stability, and the role of the production personal is to host, maintain and support the current software. As the division of labour increases in the organisation, all personal specialise in roles and new people are hired into specific areas. This specialisation allows people to develop their competencies and knowledge, which together benefit the organisation's goal.

Based on this knowledge and the differentiation between entities within the organisation, we can design an partial identity that can take roles, functions and divisions into account. An example of this can be seen in 5.4, where a concept of a (partial) identity is presented. In the illustration, John Johnsons partial digital corporate identity is showcased. E100123 is his corporate identifier and employee number which is used as a primary key in the IT systems because multiple employees might be named John Johnsen. His first and last name is also used by the computer system, however to for accounting, but to increase the User friendliness, as his close co-workers and department known him to be that name. The Role, Function, division, organisation and location is used to determine files, and document john has access (Access control). Datetime and signature is information used by the system to determine when and by whom this

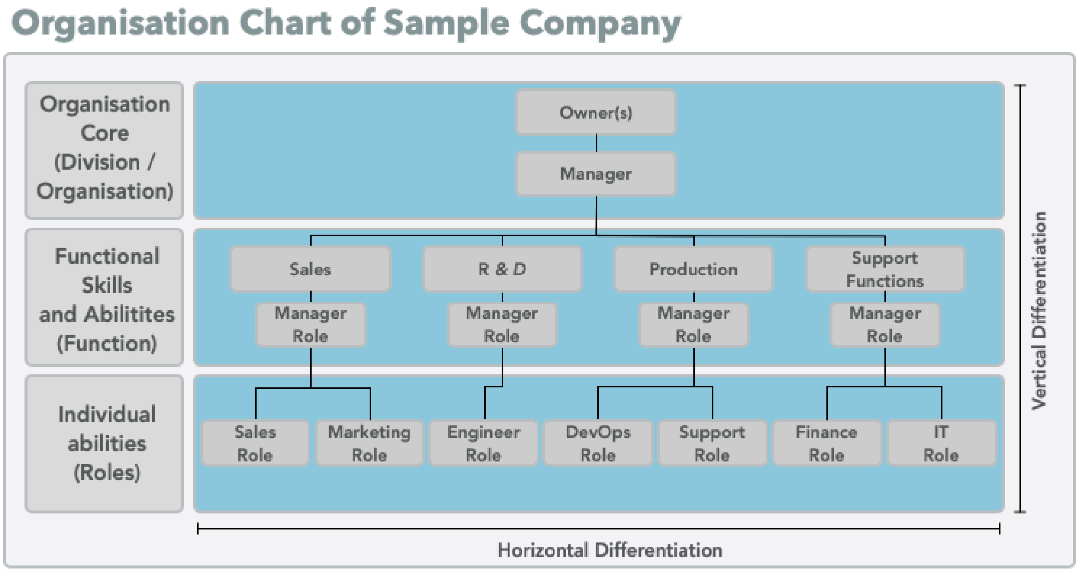
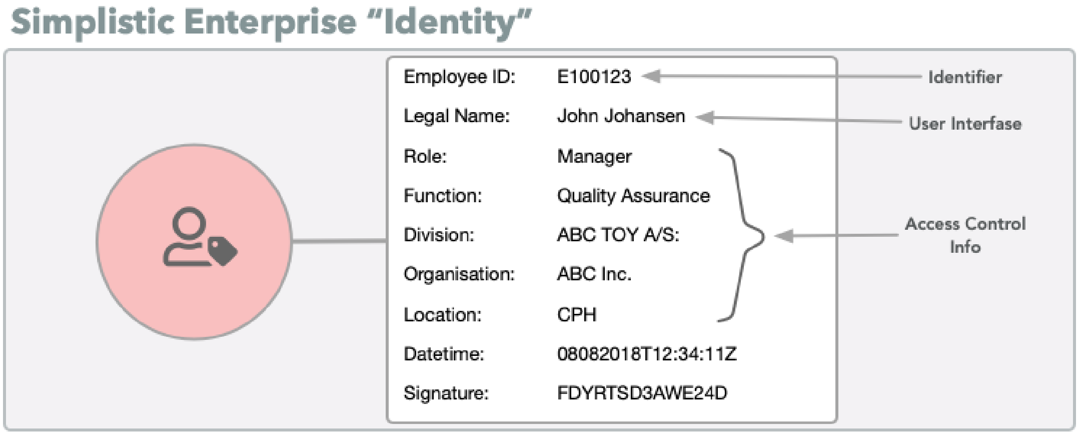


Figure 5.3: Organisational chart, based on example IT-service company

identity was verified by.



Employee ID:	E100123
Legal Name:	John Johansen
Role:	Manager
Function:	Quality Assurance
Division:	ABC TOY A/S:
Organisation:	ABC Inc.
Location:	CPH
Datetime:	08082018T12:34:11Z
Signature:	FDYRTSD3AWE24D

Figure 5.4: Example of an enterprise Digital identity

5.3 Integration with Existing Systems

None or very few enterprises, of considerable size, have all of their business operations based in a single application. Usually, these operations and processes are scattered over tens if not hundreds of services and applications depending on the size of the enterprise. Recent studies show that these numbers varies from each industry, but an average business uses between 25 to 50 systems at any time [7]. This finding fits well with the general trend, and the numbers do not seem to diminish over time, according to leading vendors of IAM systems. Most of these systems incorporate some identity component, to provide proper authentication and authorisation[8], and provide necessary levels of security and privacy. This means that such a proposed system should not be centred around a single application because it will not be viable for an enterprise which will work across multiple applications or systems.

Most larger enterprises have complicated IT Environments, which include a comprehensive combination of newer and legacy systems, that deal with the core of business operations. Modern applications are built with a high degree of different languages, architectures and frameworks. As an example, new SaaS applications hide the underlying complexity of the system and expose only end-user functionality, through the internet. This new reality enables the business to eliminate support functions and focus on core business operations. However, it comes at the cost of having complete control over the essential data and process layer[99]. This change in IT operations calls for a higher need of governance and control in terms of security. 2nd generation identity standards have introduced authentication beyond the enterprise perimeter by utilising standardised communication protocols and flows. This in combination allows authentication from Cloud-based SaaS against enterprise identity stores. UMA utilises the idea behind these flows and expands them to provide delegation of consent. This means that enterprises can make use of SaaS solutions and simultaneously gain the advantage of UMA, regardless of where the resource is located, as long as it is connected to the internet. The problem is that no company of considerable size will be able to change all of their systems in order to comply with such a new proposed Privacy-aware ACM.

In [7], an investigation into how enterprise application are connected to individual or common identity stores are presented, the findings of this report is illustrated in 5.5. As application purpose and functionality differ, applications incorporate various identity solutions. Legacy applications typically make use of proprietary identity stores, that either has their own life cycle or can be imported from other sources. Newer applications usually have the possibility to make use of federated identity solutions such as SAML or OAuth. To limit unnecessary work, multiple solutions is presented that allows a central IAM system to control and distribute identity and access to the legacy applications depending on their technology.

In Figure 5.5. The legacy applications are shown in the grey area. The report [7] states that many of such applications either have their own identity database or can be connected to an enterprise directory service with the use of LDAP or other DS interfaces. By using these technologies, it is possible for the IAM system to control which users that have access to each individual application. However, authentication is still only present directly in the application and cannot be distributed. In the figure, each application, as highlighted with the colour blue, resembles a scenario for how these local identity stores can be controlled from a central point,

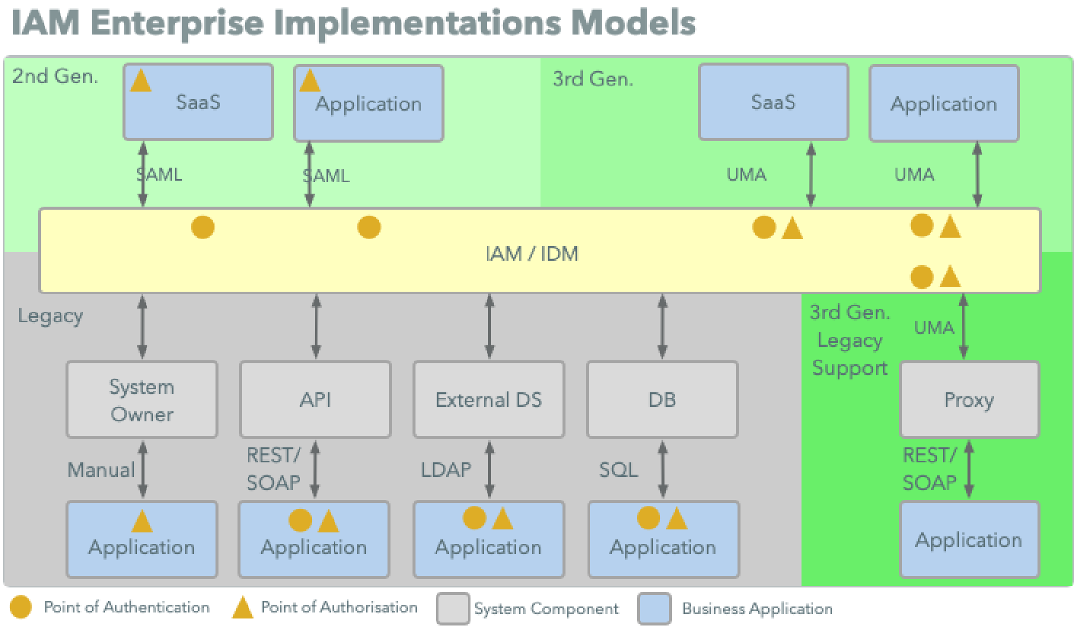


Figure 5.5: Implementation models with various generations of identity technologies and proposed system, inspired by [7]

an IAM system. For each scenario, the point where authentication and authorisation is conducted, is illustrated with respectively a yellow circle (Authorisation) or triangle(Authentication).

Modern applications and IT-services used by enterprises today, are distributed beyond the perimeter of the IT-department. 2nd generation identity solutions enables this architecture. With use of the separation of identity into an identity provider, application and services can authenticate seamlessly no matter if they are inside or outside the company, as long as underlying trust as established. The proposed 3rd generation identity and access control solution with use of UMA make use of the same principle as the 2nd generation and thereby has the same disadvantages in relation to use with legacy applications.

The problem related to former standards, both legacy applications and federation-enabled applications, is that none of these allows for both authentication and authorisation to be conducted outside the application. Due to this constraint, applications need access to the actual information about each object and subject for each authorisation decision, making it hard, if not impossible, to outsource or distribute the application anywhere due to privacy concerns. As illustrated in the figure, it should be possible to implement a web proxy if the applications are web-based and thereby lettings the web proxy act as an intermediate client which can be designed to be UMA compliant.

In the proposed 3rd generation identity solution, An Access control solution should be able to provide authorisation from internal as well as external systems as well, however with few exceptions it is not possible to implement such concept as UMA in existing software as shown in the figure. Therefore it can be concluded that it will be hard for enterprises to implement such privacy-aware technologies as this project described, because authorisation is a central and integrated part of most application into their existing infrastructure of applications. Applications should be made with consent and distribution of ACM in mind for them to work with privacy and consent as proposed with UMA.

5.4 Implementation in in a actual Company

To implement such an IAM system and enterprise, Royer has developed a framework for how IAM systems should be integrated and which process should be used to design such systems. This framework requires a thorough analysis of the enterprise which the IAM system is going to be implemented in to. Since this project is not based on a specific enterprise, this process and framework cannot be followed, therefore this related academic work is out of the scope of this project. However, The process is illustrated in appendix A.6 for reference.

5.5 Identification

As described in the introduction the process of identification is essential to the concept of identity, therefore this concept will be analysed to determine the circumstances of identification for such a project. In this section, a thorough explanation of the concept with examples will be given.

"Human individuals have continuity of personal existence: you are today the same person you were yesterday, and indeed you remain all your life the same person you were on the day of your birth, despite the many changes that have occurred in you since that day."[110]. This fact allows useful constructs like names to identify a person. Because names are used to point to specific identities, they enable people to distinguish and share relations between people. However, most people do not just use a single identifier. In a familiar relation, some would be known by "Mom" or "Boss", where friendly and professional relations will differentiate will use first name or full name. However, these construct works excellent in smaller communities, the opposite is present in larger ones. Everybody has multiple identities or personas, and sometimes this is relevant for a computer systems to make decisions based on those[110].

Problems first arise in relation to identities when the primary identification attribute identify more then one person. This problem appear when two different persons make use of the same name, This could be "Stefan Jørgensen", there is simply too many persons with this name. Neither can it be assumed that an individual has a single unique name in a larger population. The same person might be known by "Son", "Stefan", "Developer" or "Stefan Reinholdt Jørgensen". All of these might be the primary name, but all in various contexts and some of them might only work in specific contexts. The name might be unique, but no certainty can be established. A naming collision between two persons, work fine in social settings, but computer systems in contrary to humans, do not automatically adjust and use a better name that works for this particular context. A collision of names might break an identification component or at least makes it more complicated to distinguish one individual from another. Depending on the system, this could impose a direct security vulnerability, crash a computer system or maybe be an inconvenient from a user experience point of view. On the other hand could it also mean that a person could not be uniquely identified. This fact increases the requirements for systems dealing with identities. A common result of this construct is to assign an ID to a person, like a social security number or a persons mail address.

5.5.1 Example of Identification

A person born in Denmark has an entry in the states register of births. The system holds records of the date of birth, place of birth, official addresses over time and reference to known parents, This information is directly available for the governmental sector and enables to identify and keep track of people living in Denmark. This includes an electronic patient journal (EPJ), criminal record, information in regards to income and tax and many others. In all of these cases does the national identification number work as a universal identifier across systems[111].

Information about a person’s current name, address and job position can be accessed by companies or private persons who can provide information that a relationship worthy of recognition exists between them the person they are obtaining information of[112]. This information can be accessed easily by providing one of the following identifiers[111]:

- Person number (CPR)
- Combination of date of birth and name
- Combination of current name (or earlier) and address (current or earlier)

One of the use cases of the system is the enterprise’s ability to identify the entities of the persons they want to establish a relation to, such as employers, customers or contractors. Thereby ensuring that the individuals that an enterprise wants to do business with exists. Where the CPR-register holds information about legal person identities, there likewise exists a national register of companies and foundations in Denmark (CVR). This register contains information about legal entities and their owner relations. By using these two registers, enterprises can identify that a certain legal identity exists that they are presented for. However, the system does not allow the enterprise to authenticate the identity of the claimants. Yet, many enterprises consider the facts as evidence that the person who identifies with a certain CPR number is whom they claim to be[110] Or said in a different way they misuse an identifier as an authenticator.

In appendix A.3 an example of the danish CPR number is given based on this example of a national identifier shows that even though there may be excellent ideas and relatable requirements for a system, external changes will be most likely set new requirements for a system over time. Solutions can be either a complete re-engineering or specific fixes that aim to solve that single problem and let the rest run as previously. Such changes show that interfaces between systems should be adaptable for such changes. In many companies, CPR-numbers could be used to relate the employee identity to the government identity. Further, many companies make use of the country's identification number. This can be a strength. However, it also poses several weaknesses. Therefore systems using this number of interactions in both digital and analogue communication must be aware of its shortcomings and design for it.

5.5.2 The privacy aware solution for identifiers

As described within this section, the traditional approach to understanding the concept of identity, highlight the sequence required in the identification process. An example, when a person interacts with an enterprise, over the internet only that person should be responsible for his action. Such as filing required information or applying for support or other personal tasks or actions. With such interactions and communications moving from the physical world to the digital, centralised identifiers introduces a privacy problem. By reusing unique identifiers for people poses a critical point because of citizens risk being stripped of their privacy. Therefore should all persons be able to have partial digital identities that can be used as they prefer, no matter the context. An example of this could be, one partial identity with associated identifier for each system, subsystem or action. This design choice, will eliminate foundations problems in regards to privacy for users.

5.5.3 Personas and Partial identities

Common names like 'Jens Jensen' might be sufficient to identify a person in a single department, but the chance of the name being unique in the entire company are not that certain. The chance of all individuals having unique names in the company is even less likely. 'Jens Jensen' might have kids that don't use which use alternative identifiers such as 'dad', daddy or other nicknames. Any of these work, but might be limited to specific contexts. This mix of names concludes that names are only valid and work as identifiers in specific contexts. Like the physical world make use of identities, a computer system should be able to facilitate this construct. Examples of this could be a work identity, a student identity and a family identity. A one-to-many relation can represent this construct. Personal information associated with the identities can reveal significant information about a person. The information which a person wants to share concerning his family identity might not be the same as the information that wants to be shared in a work relation and Vice versa. In the enterprise context a person might have multiple affiliations to the enterprise, this could be employee, customer or board-member, based on the setting the employee does a specific action various rights or level of trust should be given. This fact created an essential requirement for systems dealing with identity in the enterprise.

Persona originates from the Greek ‘persona’ and is often used in the field of law. Where the persona was the mask used in Greek theatres to cover the face of the performer it provided the actor with the possibility to take a physical cover of other attributes or another identity [8]. All subsets of the attributes and properties of the identity are a description of that specific person, is the partial identity or persona. The partial identity is thereby a subset of attributes of the complete identity, and The sum of all these personas or partial identities makes up the identity. Each of these partial identities represents the person in some context[17]. An example of this is depicted in figure 5.6 where six partial identities of a full identity are illustrated. Some attributes are shared, some are not and as stated before in the identity section these change over time in relative to the person as the identity points at. This image could reflect the digital identities stored at each of the companies. In the FIDIS project, these personas have a close relation to the concept of virtual identities, that represent a partial identity with a set of predefined attributes, that can be presented to a certain service[8].

5.5.4 Conclusion

Based on the information specified in this section, systems dealing with identities should be aware of the problems highlighted in here. If a user uses a system in multiple contexts, partial identities should be supported in order to preserve privacy and increase security. Further all identities and partial identities should have unique identification numbers for each context to preserve privacy and hinder unexpected interconnection between identities.

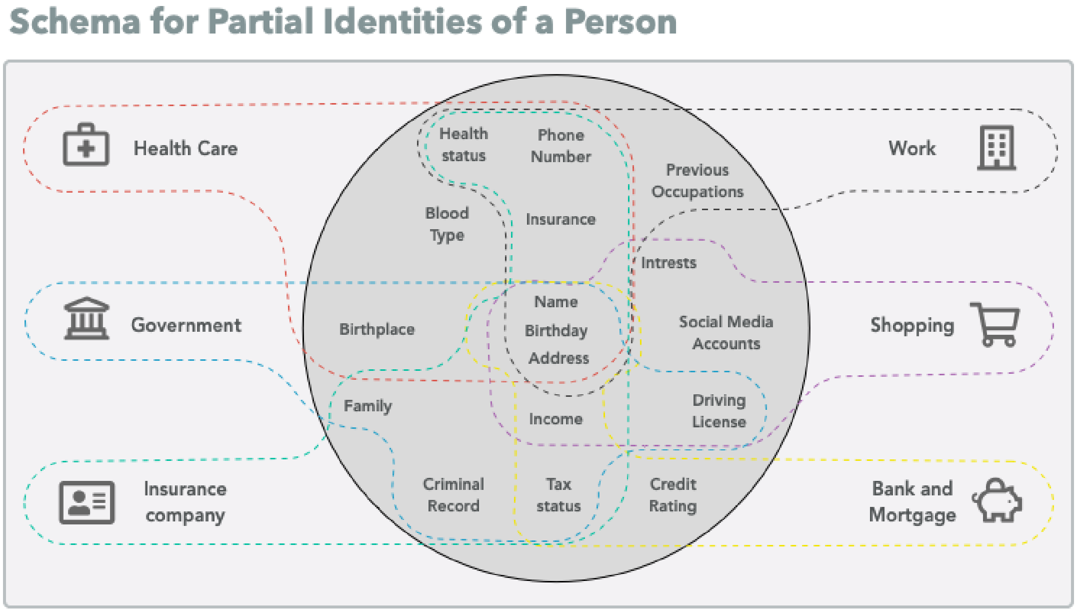


Figure 5.6: Several partial identities in relation to external enterprises, inspired by illustration in [8] section about identity

5.6 Consent Management with UMA

As described in the SOTA chapter 4, the UMA specification enables parties to operate and use software entities and devices to distribute rights and obligations fairly in access federation trust frameworks. This framework can then be used to design and configure a privacy-aware system where each entity can manage their information. As stated in the UMA working group UMA's primary use case is centred on individual people. It is the "users" who control the access to their online resources[113]. The UMA concept of "Authorisation as a service" is also relevant to modern enterprises that must secure their IT and thereby the APIs and other web resources in a natural way.

Standardised OAuth (RFC6749) is just a framework that does not specify the implementation details of the specific scenario or use case. OAuth only defines a set of standardised flows, for authorisation between person and clients. The standard document itself does not even specify which format of tokens should be used. This is placed in a different standard (RFC 6750). Due to these characteristics, OAuth can be reused for other purposes as described with OpenID Connect and UMA. OpenID which define how persons can be authenticated, discovered and registered and UMA that enables an enforcement point to authorise a request from an authorisation server[114].

5.6.1 OpenID Connect on top of OAuth

In an authentication process, the identity does not necessarily have to be revealed, only that the identity is verified, by some party. This concept has been described in relation to authentication and assertions in section 4.5.3 and 4.5.3. The Internet is in nature very faceless and does not provide any means of authenticating. OpenID Connect is a technology which aims to do just that. As previously described it is built on top of OAuth like UMA and adds an id token. The id token is a JSON Web Token (JWT) and contains information about the authenticated user. The token is signed by the IDP and is verifiable. Beside this OpenID specifies several aspects that is that are optional in OAuth 2.0 like scope and endpoint discovery among others[22].OpenID is further being supported by notable prominent companies like Microsoft, Google and Salesforce among others.

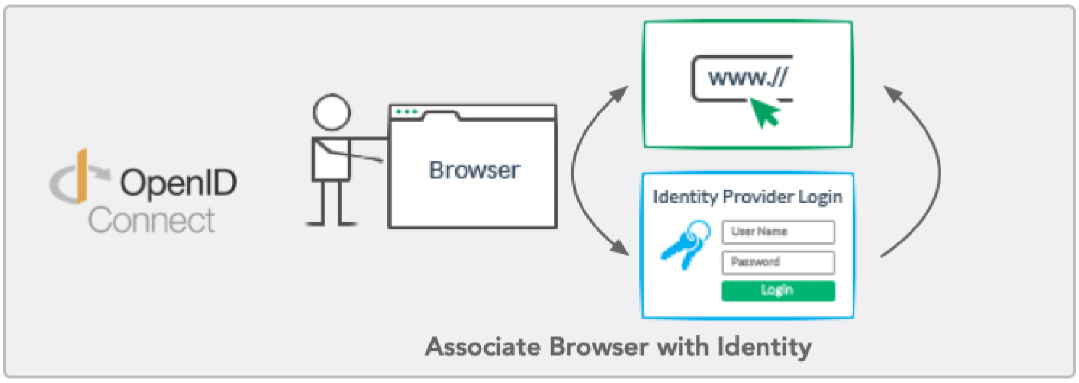


Figure 5.7: Illustration showing the concept of OpenID Connect, inspired by [9]

5.6.2 UMA on top of OAuth

The UMA standard is separated in two specifications: The UMA 2 Grant Spec and The UMA 2 federated Authorisation Spec. The first adds some changes to the OAuth protocol and the second adds on top of the first. This way Essential changes to OAuth to work as intended and specific UMA functionality is added after that in the second specification. The main changes that the UMA 2 grant makes is:

1. Functionality for the RO to authorise access to clients used by the RP. Thereby UMA enables party-to-party authorisation instead of authorisation of application access as in OAuth.

2. States how the communication between the AS and RS can happen in an asynchronous manner about the RO interactions.

3. How the RO can configure the policy conditions at the AS instead of authorising access tokens as specified in OAuth

The UMA 2 federated Authorisation Specify:

1. How multiple RS can operate in separate domains and still communicate with a single AS in yet another domain on behalf of the RO
2. How the RO monitor and control the authorisation grant rules through the AS.
3. How the Authorisation grant can increase and decrease for the individual resource and scopes.

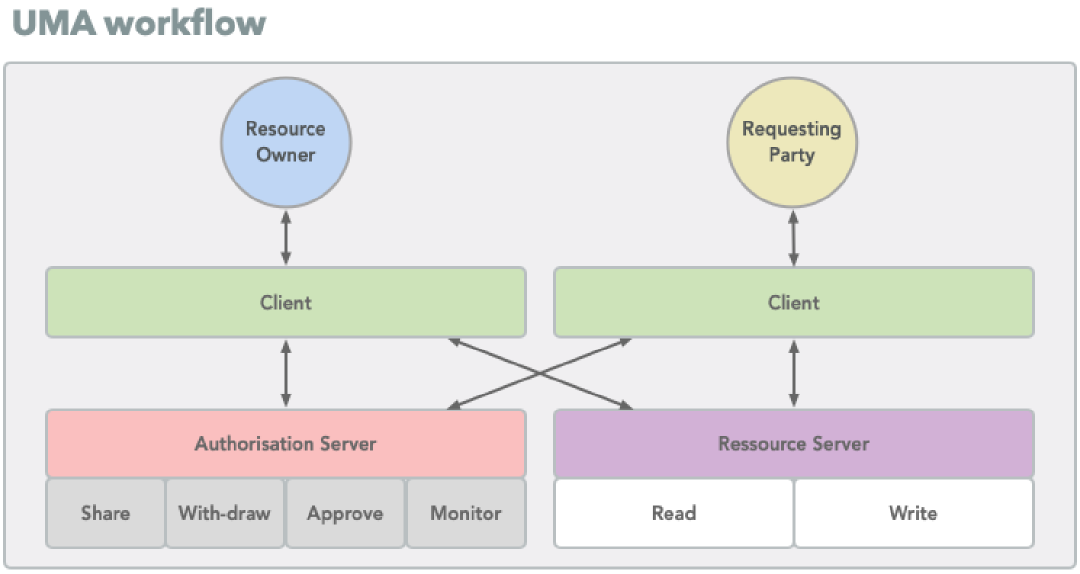


Figure 5.8: UMA workflow

UMA defines the interfaces between AS and RS that, enables centralised policy decisions. This characteristic can enables enterprises to distribute access control decisions beyond business applications and enables more straightforward applications, improved auditing and easier policy administration, even in a loosely coupled "public API" environment. APIs are by default are groups of functionality available at a domain. These maps excellent to arbitrarily fine-grained policies, for example, at the method and parameter level. However, outside the use of XACML, authorisation policy granularity is coarse at scale. Further, it is often ineffective to perform according to policies from multiple authoritative sources[113]. Therefore UMA defines a RESTful, JSON-based, standardised flows for the protection of any API or web resource in a way that will be easy to any project already acquainted with OAuth. UMA does not standardise a policy expression language such as XACML or OPA, thereby letting the enterprise or vendor choose the right policy technology for their purpose. UMA's concept of standardised resource set and scope descriptions makes an access control mechanism that allows management of particular API scopes, not just domains possible. Which were established as one of the requirement to achieve increased possibility for privacy for the user.

With the use of UMA, client app developers can manage authorisation tasks by requesting simple REST-based JSON endpoints. Due to the conveniences, UMA is a good fit for enterprise use, compared to current IAM technologies like SAML. UMA bring possibilities regarding Consent, management and monitoring at the same time simplifying the process of authorisation for current services.

As presented in the last chapter, UMA introduces some significant possibilities by building on top of open, standardised, free and popular technologies such as OAuth 2.0, OpenID and JSON. Using UMA enable Authorisation to be delegated to a trusted entity called the Authorisation server, which can be used in combination with many resources servers, no matter the framework or technology stack used. This characteristic can provide improved privacy to the web and the scenarios used for this report. By using an outsourced authorisation server, the architecture of UMA enables sharing beyond the known identities of the resource server. Most users have a lot of various resources spread around on platforms, hosts and security domains. This applies both to private and enterprise scenarios and thereby enables use cases in between the private and enterprise realm, for example, everyday transactions where persons deal with companies, wherever the company need some access to a particular resource. By using UMA, access can be granted under the precautions stated by the resource owner. Because resource Owners in UMA not are restricted to be individual persons, this can also be used in an enterprise context. These possibilities have been explored and used before by various vendors, but UMA brings a standardised mechanics that allow sharing of resources between any vendor ecosystem as long as they support the UMA workflow.

5.6.3 Combining UMA, OAuth and OpenID Connect

By using these standard together, a synergy effect could be created due to the same foundation of OAuth together with the non-overlaying purpose. Several advantages for this combination of technologies can found. First, they rely on the same platform, which could result in simpler implementations and fewer dependencies. Second, the OAuth protocol is built to be platforms independent, due to the web and mobile design requirements for the standard. Third, the standards do not overlay in their purpose or objective. In figure 5.9 a Venn diagram showing possible goals when combining the mentioned technologies can be seen. It can, therefore, be concluded that it is the combination of technologies which fulfils the problem.

It should be added that in order to create a full IAM solution many other technologies should be added. For example Sovrin for identification, FIDO for multi-factor authentication, SCIM for provisioning and XACML as Policy language. However, this is out of scope except for Sovrin.

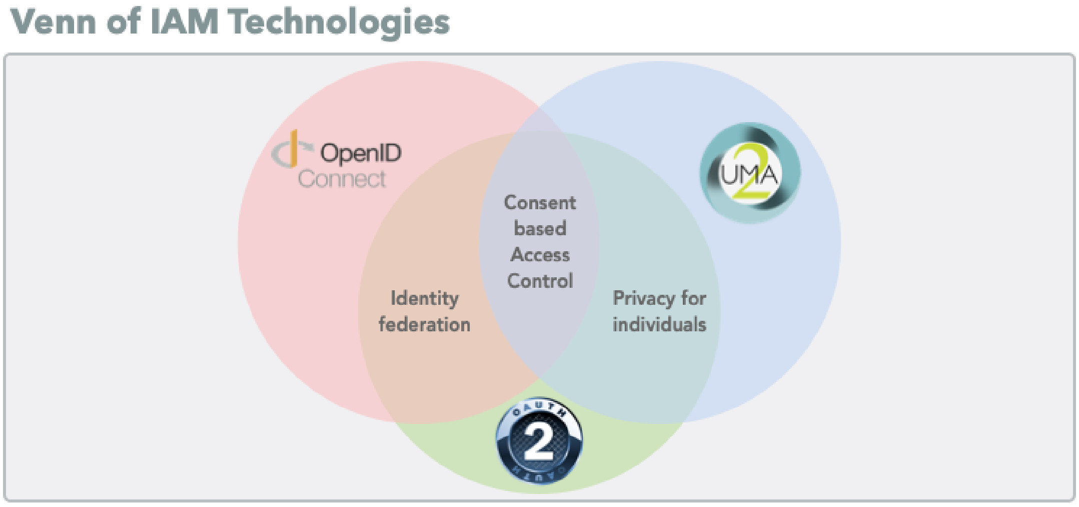


Figure 5.9: Venn diagram of the three technologies and their complementary effects, inspired by [10]

5.6.4 Conclusion

This section summarises the need for consent management and how it can be achieved with the use of UMA. This further demonstrates how UMA work together with other technologies to fit into an enterprise scenario. The properties of UMA allows for consent management, which can be used to increase privacy for users concerned about their privacy in relation tot he information they share with various entities.

5.7 Identity Management With Sovrin

As described in the SOTA chapter/ref(), a distributed identity technology in the form of Sovrin is presented with its related pro and cons. The foundation for this solution relies on the highly discussed concept of distributed ledger technology. As described in the development of this technology is still in an early stage, and an only early prototype implementation is active. Concerning this project, such identity solution would be beneficial for solutions where multiple legal entities deal with a system that has the purpose of doing transactions between the legal entities so that the control of the identities is not only located inside the IT systems of the enterprises. Concerning the scope of this projects, the focus is on the interaction of such two entities. However, the extent of this project is put on the interaction between a person and an enterprise entity as exemplified with the scenarios in Chapter 3. Therefore a technology like Sovrin can work as an enabler for an IAM system like the one proposed in this project with a focus on privacy as the one proposed in this project. Sovrin works by the person, and the enterprise can use the distributed ledger to find each other and negotiate a private channel that is unique to that particular relationship and has no intermediaries. The parties can then share verifiable credentials on this secure channel to establish the essential level of trust to perform business. If anybody were to tap in on the line and listen, he or she would not be able to understand a thing. Such a solution should utilise the power of the distributed identifiers. Thereby the Identity Owners (IO) can make use of identities and the verifiable claims in any system connected to the ledger. This makes it independent of other entities in the interaction and provides control over the distribution of the identity. From each party establishes a private connection with the use of the DID's that they share with the each other, all communication takes place off the ledger. This pattern ensures both better privacy and stronger security as none of the parties is reusing the public key that they have used for any other interactions. However, it is not the pairwise pseudonymous DIDs that is used to establish the overall trust for this new connection. The established cryptographic trust is based on the assurance that each party has a secure private channel with the other. This construct is also more scalable due to the ledger is only involved with the initial establishment of the relationship. This should be seen as an alternative to the PKI model described in the SOTA chapter, Section 4.4.2.

To foster the exchange of verifiable credentials, the ledger stores two kinds of objects schema definitions and credential definitions. A schema definition is a set of attribute data types and formats that can be used to specify the claims on the credentials for example with the grades scenario. The schema for creating academic qualifications credentials could include a definition of attributes such as given name, family name, date of birth, issuer name, issuer course, issuer mark and issuer note etc. A schema definition can be used by many credential issuers and is a way of achieving standardisation across issuers as academic qualifications.

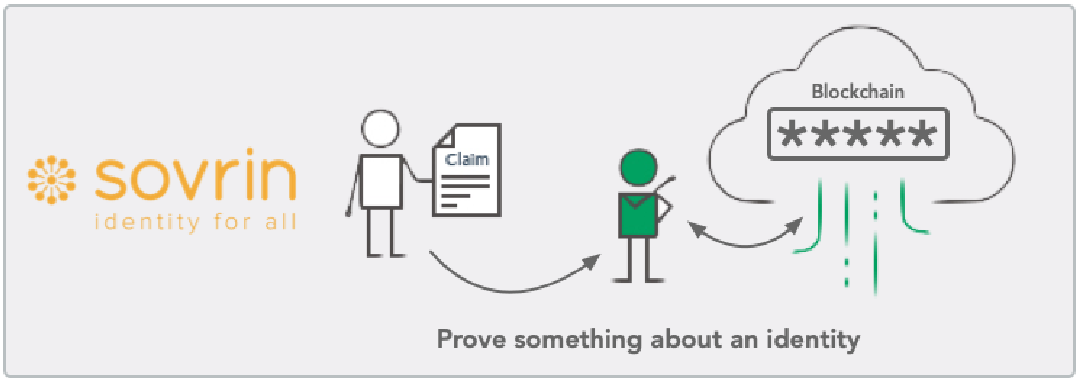


Figure 5.10: Illustration showing the concept of Sovrin, inspired by [9]

The credentials information is the structure of the data which is an instance of the schema plus the attribute-specific public verification keys. This construct allows the issuer to reuse existing schemas. It further enables the verifier to receive a proof for the contained data from the issuer, by looking up the issuers own credentials information on the ledger. By doing this, the verifier gets the ability to verify the origin and integrity of the specific credential information. Sovrin makes it possible for anybody to publish their own schemas and credential definitions on the ledger. If Sovrin were designed to place credentials (or credential hashes) directly on the immutable blockchain was it impossible to revocate them, But the Sovrin ledger does not store the credentials even hashed or encrypted to the ledger. They are only issued and exchanged off-ledger between Sovrin agents and their respective wallets. This is a significant advantage because there is no central authority and everybody can define and issue their own types of credentials making the exchange of verifiable credentials decentralised. This construct puts the control over identities in the hand of the IO and thereby improving privacy in relation to current standards.

5.7.1 Conclusion and limitation of self-sovereign identity

Based on these possibilities, Sovrin fits the scope of the project outlined in the introduction chapter 1 due to sharing of the objective for the identity control. Therefore it can be concluded that storing identifiers in Sovrin can be used as the foundation for such a system. Although the analysis of this technology shows that self-sovereign and specific Sovrin will contribute to the solution of this problem, further analysis and design of such a solution will is out of the scope of this project, due to the limitation of the scope.

5.8 Requirement

The conceptual requirements and motivation were described in the introduction chapter, resulting in the problem statement in section **??**. Related theories were introduced in the state of the art chapter, where core concepts such as the general ACM and its related concepts such as authentication and authorisation were described. Possible Technologies which can solve these problems and their objective were also described. Based on these candidate technologies an analysis of their capabilities was made in this analysis chapter together with an additional analysis of how the component could work together with the other proposed solutions to create a privacy-aware IAM solution. In this section, the requirements for the system will be introduced based on the problem statement and related research. The requirements will be introduced in categorised of functional and non-functional requirements. The functional requirements are denoted by "FR" and the non-functional are denoted by "NFR" abbreviation.

In order to limit the scope for the system a requirement elicitation is first conducted and a prioritisation is done afterwards. In this Section, a set of requirements will be specified. These requirements are formulated based on problem statement and about the scenarios described in chapter 3 and the further analysis in in this chapter. The list should be considered as a limited list of requirements for the proposed Privacy-aware IAM system. should the idea of this report be used in an enterprise setting further analysis of the setting, enterprise and external requirements be analysed. For the time being, these requirements have been identified based on the findings in the analysis chapter.

The key words "MUST", "MUST NOT", "REQUIRED", "SHALL", "SHALL NOT", "SHOULD", "SHOULD NOT", "RECOMMENDED", "NOT RECOMMENDED", "MAY", and "OPTIONAL" in this section are to be interpreted as described in the following RFC [115].

5.8.1 Requirement Elicitation and Prioritisation

In appendix A.2 the full list of requirements is placed. The list contains 23 requirements isolated from the analysis. The list of requirements are based on the basis of the presented scenarios and their use cases in the scenario chapter 3. The requirements are prioritised, following the method described in the methods chapter 2. However, there is an overwhelming amount of the requirements with the priority of "Must", which is due to the system being a prototype, where only essential requirements are taken into consideration. Had the system been anything else, then the list would have contained many more requirements but also with a larger distribution of importance. It could be argued that even a prototype should have a longer list of requirements so that more functions could be tested, but due to a limit of resources only essential requirements were described here.

Chapter 6

Design

In this chapter, the design of the proposed system will be described. The design is based on the analysis in the last chapter which resulted in a set of requirements. It will be described how the overall architecture is made and documented with use of illustrations of the system. The goal of the design realises a solution based on the conclusions from the analysis chapter, with use of the technologies presented in the SOTA chapter 4 Using this material as input, it should be possible to design a system which addresses the general problem statement of the project.

6.1 Architecture

A high-level architecture of the system will be presented in this Section, based on the analysis from the previous chapter. Including elements from the relevant technologies; OAuth, UMA, OpenID Connect and XACML.

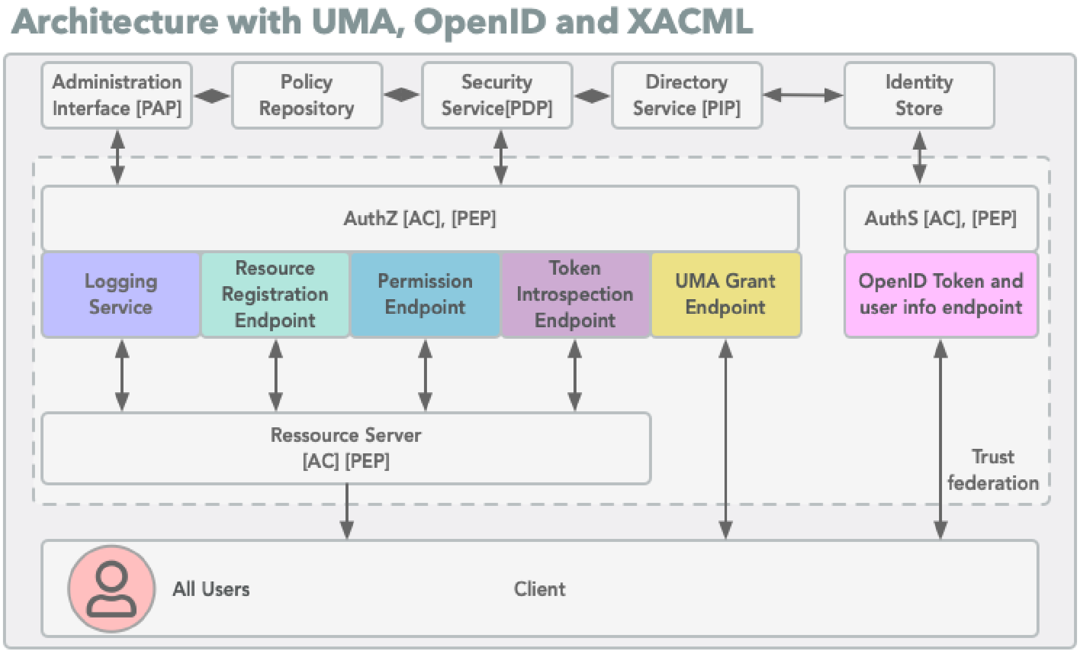


Figure 6.1: Architecture with UMA, OpenID and XACML

How the various entities are linked are shown in figure 6.1. All users interact with the system with the use of a client. Client devices are authorised with the use of OAuth as described in the SOTA chapter 4. Authentication is done with the use of OpenID, and this is formed as an interaction between the Authentication server and the Client (User). Thereby can the client be authenticated to the system. To increase the security around the authentication process of the user, FIDO can be implemented, FIDO is out of the scope of this project but is a relevant technology which can fit into this purpose. When the identity of the client is known, he can request a resource at the resource server, UMA takes care of authorisation of resources for users, this interaction takes place between the Authorisation server, Resource server and the Client. As described in the SOTA chapter, UMA does not specify a specific way of applying access control, here XACML can fill the spot by specifying a policy language and a process model explained in the analysis chapter 5. To determine if the user has the proper rights, a combination of UMA consent management and the XACML policies are calculated. The XACML decision is taken in the PDP, which is a service that is accessible at the external security Service (PDP). The PDP calculate access rights based on policies specified at the Policy repository, which is administrated from the Administration interface (PAP). If additional information is needed, the PIP can be contacted to aggregate this information might comes from the identity store (DS). The PDP calculate a simple boolean answer based on the policies and the aggregated information and returns it to the PEP, which can process the original request after that. Based on this the requested resource or answer is returned to the client.

The separation of entities in this diagram represent the possibilities with this architecture. The Resources server, Authorisation server and Authentication server are all loosely coupled with the use of standardised interfaces. Components in the architecture could be placed differently, however for the sake of storytelling and showcasing the wide possibilities with these technologies, they ended up this way. Nevertheless, it should be noted that in a real-life scenario some of these will probably end up being hosted the same place, but it shows the strength and adeptness of this stack of technologies.

In figure 6.2 and 6.4, a detailed look into the authorisation and authentication servers is illustrated. A special remark should be put on the rotated rectangle representing trust. This is one of the most important constructs of the architecture and the reason why the entities in OpenID, UMA and OAuth can be separated. The trust is established with the use of cryptographic technologies and preceding exchange of identities.

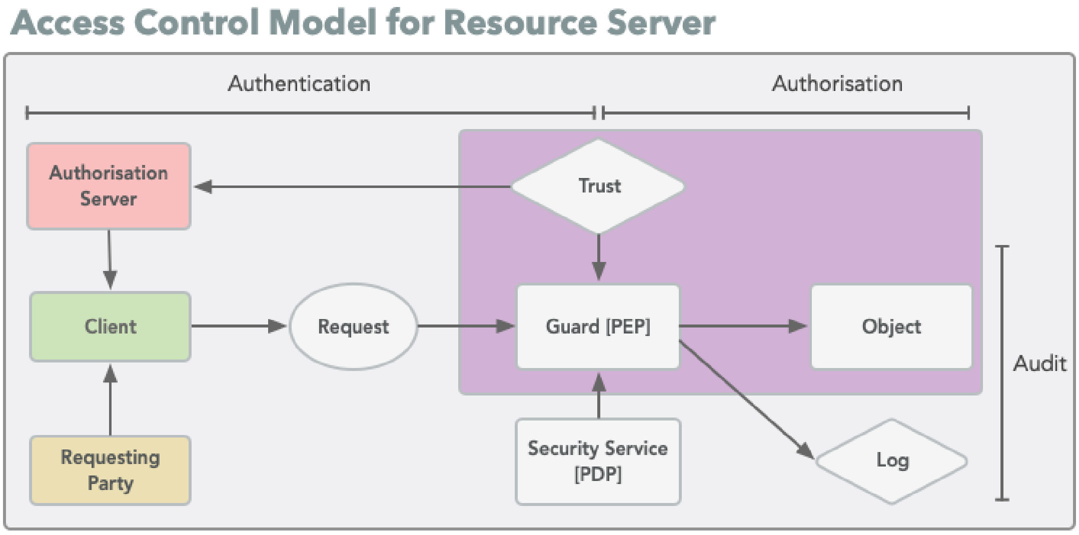


Figure 6.2: Access Control Model for Resource Server

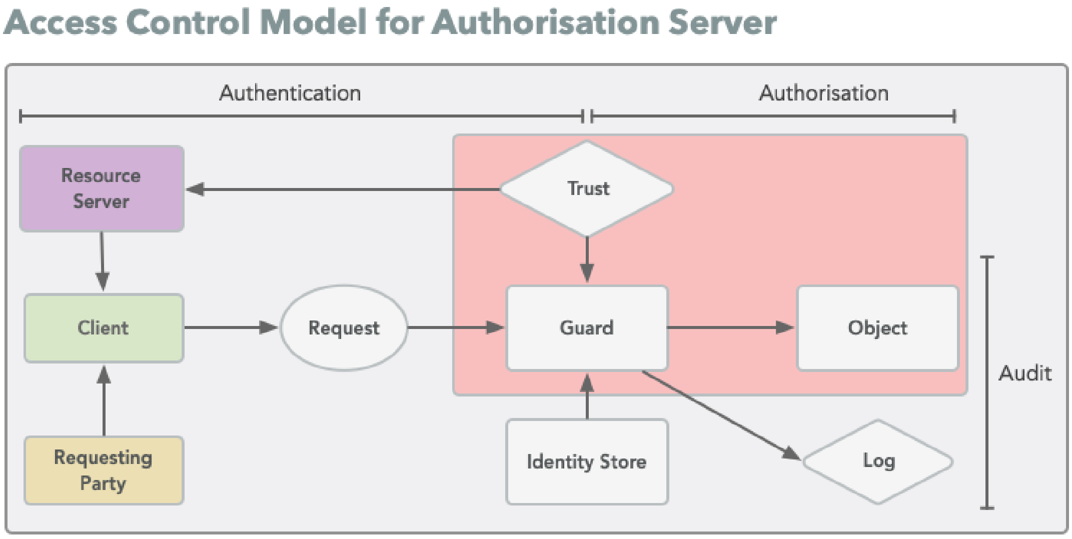


Figure 6.3: Access Control Model for Authorisation Server

6.2 Platform

The Web is, by far, the most powerful information aggregation and distribution tool currently available to mankind[116]. No longer must anybody search through countless libraries and catalogues to discover what they are looking for. Instead, hyperlinks enable users to explore rich interconnected structures of information with ease. With just a few clicks, it is possible to navigate through a vast space of information comprising documents on virtually any subject. Improvement of browser compatible client-side programming languages and API's have enabled complex applications to be made with the use of standardised and straightforward building blocks such as HTML, CSS and JS. This development had meant that many applications that before in time were made with native desktop application technologies have moved to the web. This move hasn't been without problems and comes with specific pro and cons as any system owner should be aware of before designing a web application. However, this discussion is out of scope for this report but based on the trends regarding web applications, distribution and easy prototyping capabilities the prototype developed in relation to this report will be built using web technologies.

The client main challenge is to present information and returned data to the user. In these clients could take the form of any platform, however, most likely a webpage or mobile app would fit perfect with these technologies. However, a thicker client implementing its own business logic could as well make use of these technologies because all traffic is handled over a standard HTTPS channel with use of traditional and popular REST / JSON syntax. No matter the purpose the client will interact in the mentioned information flow, retrieve tokens and use them to request the actual resource. For implementation purposes, simple API with sample data will be made.

6.3 Flow

The following flow as shown in figure 6.4 were specified for use within the system. The flow shows the way that the components should interact in order to gain access to a protected resource

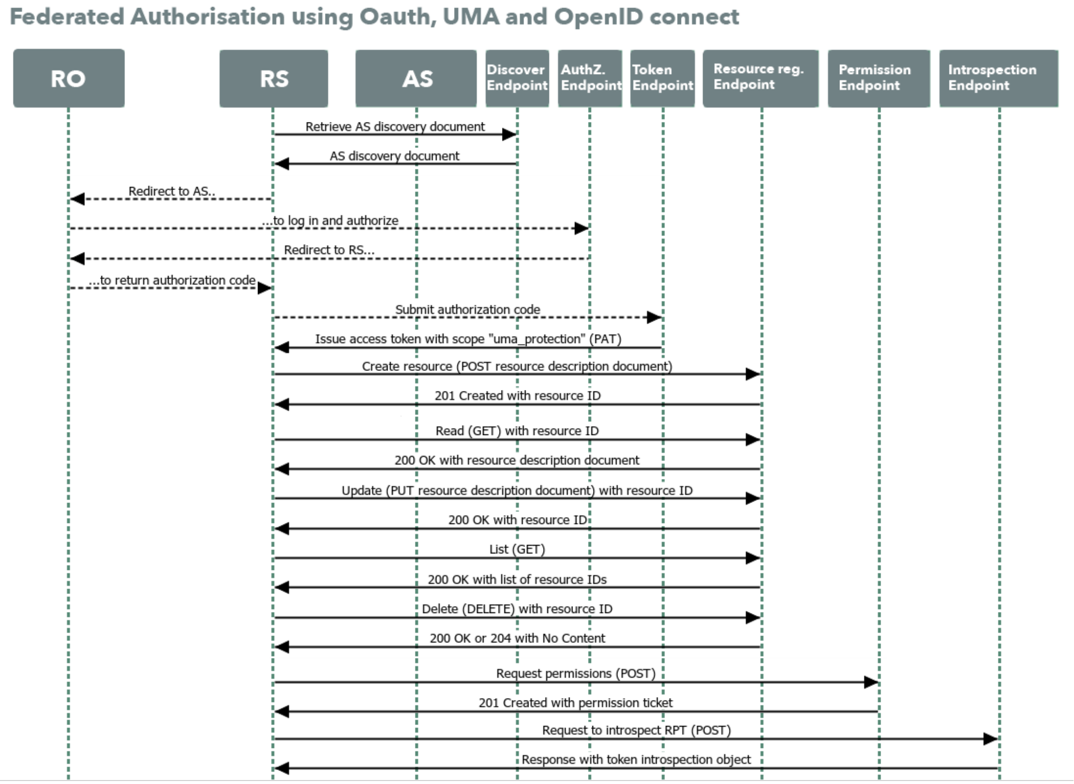

Federated Authorisation using Oauth, UMA and OpenID connect
RO
RS
AS
Discover Endpoint
AuthZ. Endpoint
Token Endpoint
Resource reg. Endpoint
Permission Endpoint
Introspection Endpoint
Retrieve AS discovery document
AS discovery document
Redirect to AS..
...to log in and authorize
Redirect to RS...
...to return authorization code
Submit authorization code
Issue access token with scope "uma_protection" (PAT)
Create resource (POST resource description document)
201 Created with resource ID
Read (GET) with resource ID
200 OK with resource description document
Update (PUT resource description document) with resource ID
200 OK with resource ID
List (GET)
200 OK with list of resource IDs
Delete (DELETE) with resource ID
200 OK or 204 with No Content
Request permissions (POST)
201 Created with permission ticket
Request to introspect RPT (POST)
Response with token introspection object

Figure 6.4: Swinlane diagram over interactions for UMA. Inspired by swimlane in the UMA Specification [3]

6.4 Implementation

For the implementation of this prototype, a series of open source projects implementing the various technologies were chosen. A described in the methodology chapter 2, only projects that had an open source were taken into consideration. Based on the overall architecture specified in this chapter, software that implements each component in the project needed to be chosen. As Highlighted before in this chapter, the following standards were choosen to be used for this project: UMA 2, OpenID Connect, OAuth 2 and XACML. This together with the architectural element stated the requirement for the implementation. A short investigation into the possible implementation of the technologies was made. The primary goal was to determine possible project implementing these technologies and being able to use in the project as described, being open and used in the real world. The following candidates were found: - For UMA, Open ID Connect and OAuth: OpenUMA (Forgerock), Gluu (Glue foundation), Keycloak (Red Hat), Privacy services (Telia), NuveAM (synergetics) and identity Server (WSO2) for XACML: API Manager (Wso2), Access Manager (OpenIAM), SUN XACML (SUN) OpenAM (ForgeRock).

This list does not contain all implementations available, but candidates evaluated for this prototype. A structured evaluation was not chosen, due to resource constraints, but should be made, if the project goes further then a prototype stage.

Based on a quick assessment the Gluu product were chosen, because it implemented all targeted technologies chosen, it has an active community, being Open Source, and uses the standards as envisioned for this project.

6.5 Sub conclusion

The purpose of this chapter is to lay down the basic structure needed to implement the prototype of the proposed system and enable to initiate a development process of the IAM system. The design process presented in the chapter is used to outline the overall architecture with the different identified entities and their placement. This has been made to conform with both the UMA, OpenID, OAuth and XACML processes and architectures. The flow of the UMA solution was identified and presented.

Chapter 7

Discussion

With the use of these privacy enhancing technologies, virtually any resources hosted on a web server can be protected by applying the UMA flow in combination with an element of Access control. With the increase of web services, that handles everything from IoT to delivering packages to the doorstep, the future internet has the possibilities to move closer to a reality where user privacy is the default settings. The legislation highlighted in the report might amplify this. The general idea described in the introduction was to prove how such new IAM technologies as UMA, OpenID and OAuth could play together to increase privacy in interactions between the users and one or more enterprises. This is a typical pattern in real life, but most UMA use cases are built on end-user to end user interactions. Overall, the suggested design can be used as the underlying building blocks for this type of use cases and thereby used for many kinds of services.

It should be noted that many other solutions could be chosen. If this study were to be remade an increased focus on the essential identification technologies would be chosen. This is due to the privacy-enhancing possibilities that such service enables. It should be noted that many other solutions could be chosen. If this study were to be remade an increased focus on the essential identification technologies would be chosen. This is due to the privacy-enhancing possibilities that such service enables. Considering the problem statement defined, other scenarios could also have been derived where different actions and user interactions could be considered. In this report, only a minimal set of scenarios were chosen, but it should be noted that many other conceptual scenarios should be made in order for such to prove relevant in a practical sense.

Chapter 8

Conclusion and Future Work

In this report, a privacy problem were introduced in relation to personal identities and privacy, the idea of having siloed identity stores placed at shady service providers takes control of people's identities and they end up being a product of some internet juggernaut of a company. The next best case is putting the identities in the hands of service providers which might get hacked, and personal information exposed. In both cases harm the end user because privacy aware identity solutions not are built into the foundation of the internet. A combination of modern IAM technologies has the possibility to change this. This proposed idea envisions that users take control over their own identities and control over their resources by using services that support such privacy aware technologies. This is not just a win for the user but has the ability to simplify the business logic and general responsibility at the service providers. The system relies on the entities establish a trust relationship so that authentication and authorisation decisions can be offloaded to external services. However, this concept increases needed interactions between entities, but does it in a standardised way, that most services would be able to integrate it. this fact set the scope for the project and the main goal is to create a working prototype with the use of this concept.

8.0.1 Sub Questions

In the problem statement given in the introduction, the problem statement were followed by a series of sub questions, which were deemed to be answered first in order to answer the general problem statement.

How can identity and access control be used to increase security in an enterprise? - In the SOTA chapter4, various digital identity technologies were described, it was also found that with the use of such technologies the users and enterprises can take control over the digital identities used within and outside the enterprise, this concept strength security and benefit both the user dealing with the enterprise and enterprise itself. So with the use of federative identities technologies security can be increased
How can a system be designed to comply with the current data protection regulation? - As investigated in Section 4.2.1 the core concept of such a data protection regulation was investigated. The investigation

further introduced the Rights, which system like the one proposed in this project must comply with in order to be used. Therefore by following the legislation and implementing the rights from the regulation into the requirement specification, as a necessity. A system should be able to comply with privacy regulations like the European GDPR. Further, it can be made easier to adopt such regulation by utilising technologies where such rights as introduced by the regulation are included as a base requirement. An example of such technology is UMA as part of the authorisation process, where consent management is a core concept.

How can identity be used in a privacy aware manner in relation to enterprises? Enterprises can utilise federated authentication technologies when dealing with external stakeholders. In the report, it was found that by using such technology can minimise administration work in term of maintaining an identity lifecycle. By using technologies like that utilise self-sovereign identities, a higher level of privacy for the users can be archived. However, this is not the only solution, 2nd generation identity solutions such as SAML and OpenID connect can contribute with separating the authentication process from the business logic. Thereby authentication can be placed remotely, where the users can take control over their identities.

What architecture can enable the increased privacy for digital identities? - As described in section 6.1, an architecture for the proposed prototype is made, this architecture has the possibility to increase the privacy of the users by separating the authentication and authorisation servers from the main resource server. In this project, two scenarios were introduced as an example to prove this. This distribution of entities enables the service to focus on its primary business and outsource the consent to a centralised place, where the end user can take control. In scenario 2, a solution where two parties have a saying in relation to access to the resource were given. This possibility makes the solution should further compatible with more business cases.

Which technologies can enable the user to have management over their data in a privacy-enhancing manner? - As shown in this project, technologies like UMA, OpenID Connect, OAuth and Sovrin can enable this increased level of control and thereby giving the user a higher level of privacy. However, this can only be archived by having the right architecture, by default the technologies just specify a flow of information. Many other components are required in order to facilitate such higher levels of privacy. As specified in the security strategies, if a system doesn't require access to personal attributes, then it should not be included, because it makes the cost of such a system higher.

How can partial identities be used to manage rights and access to personal information? - As briefly described in section 5.1 utilising partial identities can increase security because computer systems will only have access to a limited scope of the identity and will not have the possibility til couple to the actual person. This construction could lead to increased privacy for users. However, this was not utilised in this project but should be included in the further works. Sovrin is a technology which enables this by making it possible for the user to assign an identifier to each interaction.

By answering these subquestions, we believe that it is feasible to conclude that a privacy-aware identity and access management system for enterprises can be designed with the use of the proposed technologies, distribution, architecture and right setup, can enhance the security for individuals dealing with enterprises both on the inside and outside.

8.0.2 Further Work

As this report touches upon, many possibilities still exist and should be explored that increase privacy and strength the security for enterprises dealing with individuals. If further work should be done, the self-sovereign in terms of Sovrin should be explored and added to the solution. This would demonstrate how identities truly could be distributed so that it is in the hand of end users. Further work could also be done in relation to implementing specific policies that utilise the guidelines set by the GDPR regulation. Further, an investigation into how the upcoming ePrivacy regulation deals with the privacy of information between entities should be carried out.

Bibliography

[1] M Hansen, H Krasemann, C Krause, M Rost, and R Genghini. Identity management systems (ims): Identification and comparison. Technical report, Technical report, Independent Centre for Privacy Protection (ICPP), Kiel (Germany), 2003.

[2] Ib Andersen. *Den skinbarlige virkelighed (Studieteknik)*. Samfundslitteratur, 2013.

[3] User-managed access (uma) 2.0 grant for oauth 2.0 authorization. `https://docs.kantarainitiative.org/uma/wg/oauth-uma-grant-2.0-05.html`, 2018. Accessed: 2018-10-11.

[4] NIST Electronic Authentication Guideline. Nist special publication 800-63b version 1.0. 2. Technical report, NIST, December 2017.

[5] Bill Parducci and Hal Lockhart. eXtensible Access Control Markup Language (XACML) Version 3.0. *OASIS Standard*, (January):1–154, 2013. URL `https://www.oasis-open.org/committees/download.php/4412/oasis-xacml-2{_}0-core-spec-wd-01.pdf`.

[6] G.R. Jones. *Organizational Theory, Design, and Change*. Prentice Hall, 2010. ISBN 9780136087311. URL `https://books.google.dk/books?id=aZggAQAAMAAJ`.

[7] Zaharia. S Jørgensen. S, Gitau. N. Identity and access management system in a university context. Technical report, Aalborg university Copenhagen, 2017.

[8] Kai Rannenberg, Denis Royer, and André Deuker. *The future of identity in the information society: Challenges and opportunities*. Springer Science & Business Media, 2009.

[9] Know identity conference wrap-up and decentralized identity panel notes. `https://www.gluu.org/blog/know-2018/`, 2017. Accessed: 2018-22-11.

[10] An introduction to user-managed access (uma). `https://openid.net/wordpress-content/uploads/2015/02/UMA-for-HEART-2015-02-09.pdf`, 2015. Accessed: 2018-22-11.

[11] PJ Windley. *Digital Identity*. O'Reilly, 2005. ISBN 0596008783. URL `https://books.google.dk/books?id=WTmbAgAAQBAJ&printsec=frontcover&dq=isbn:0596008783&hl=&as_brr=3&cd=1&source=gbs_api#v=onepage&q&f=false`.

[12] coop finder mønstre i 1,5 millioner danskeres indkøb. https://kommunikationogsprog.dk/files/KOMmagasinet/KOM2015/2015-06-11_KOM_88.pdf, 2017. Accessed: 2018-08-01.

[13] Nu i danmark: Lad supermarkedet registrere dine indkøb med dankort - og få rabat. https://www.version2.dk/artikel/supermarked-laver-koebsprofiler-paa-kunderne-direkte-forbundet-med-dankor 2017. Accessed: 2018-08-01.

[14] Big data business academy har fundet de første 20 foregangsvirksomheder. https://alexandra.dk/dk/aktuelt/nyheder/2016/big-data-academy-20-foregangsvirksomheder, 2016. Accessed: 2018-12-01.

[15] The invention of the internet. https://www.history.com/topics/inventions/invention-of-the-internet, 2018. publised: 2010-07-09.

[16] Individuals using the internet (https://data.worldbank.org/indicator/IT.NET.USER.ZS?year_high_desc=true, 2018. Accessed: 2018-10-27.

[17] Denis Royer. Enterprise identity management-what's in it for organisations?. In *FIDIS*, pages 433–446, 2007.

[18] Paul Ricoeur. *Oneself as another*. University of Chicago Press, 1992.

[19] Daniel Candel Bormann. If you want to know more: Ricoeur's idem and ipse, 2018. URL https://kpmg.de/media/20080101_2008_European_Identity_Access_Management_Survey.pdf.

[20] Navne loven. https://www.retsinformation.dk/Forms/R0710.aspx?id=173271, 2018. Accessed: 2018-10-27.

[21] Bekendtgørelse af lov om det centrale personregister. https://www.retsinformation.dk/Forms/R0710.aspx?id=191719, 2018. Accessed: 2018-10-27.

[22] Andre Durand. Three phases of identity infrastructure adoption, 2018. URL http://www.literarycrit.com/addendamodule3individualityipseidemb.html.

[23] HubSpot. The definition of a buyer persona [in under 100 words], 2017. URL https://blog.hubspot.com/marketing/buyer-persona-definition-under-100-sr.

[24] Christine Satchell, Graeme Shanks, Steve Howard, and John Murphy. Knowing me–knowing you. end user perceptions of digital identity management systems. In *Proceedings of the 14th European Conference on Information System*, 2006.

[25] Joseph Pato. Identity Management: setting context. *Encyclopedia of Information Security*, pages 1–5, 2003.

[26] David W. Chadwick. Federated identity management. In *Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)*, volume 5705 LNCS, pages 96–120, 2009. ISBN 364203828X. doi: 10.1007/978-3-642-03829-7{_}3.

[27] Kim Cameron. The Laws of Identity. *Microsoft Corp*, 5(May): 8–11, 2005. ISSN 0036-8075. doi: 10.1126/science.22.555.206-a. URL `http://www.identityblog.com/stories/2005/05/13/TheLawsOfIdentity.pdf`.

[28] Roussos, George Roussos, Don Peterson, and Uma Patel. Mobile identity management: An enacted view. *International Journal of Electronic Commerce*, 8:81–100, 2003. ISSN 1086-4415. doi: 10.1080/10864415.2003.11044287. URL `https://www-tandfonline-com.zorac.aub.aau.dk/doi/abs/10.1080/10864415.2003.11044287`.

[29] What is trust, why is it important and how can it be measured? `https://www.google.com/url?sa=t&rct=j&q=&esrc=s&source=web&cd=14&ved=2ahUKEwi9nNXP5breAhWSxIsKHcpoD5Y4ChAWMAN6BAgFEAI&url=https%3A%2F%2Fwww.institutelm.com%2Fasset%2F72B22B4F-52DC-45F8-A0AB6EDAB16F56D5%2F&usg=AOvVaw08zkxY7RwIrtUs2UIPW1rB`, 2018. Accessed: 2018-10-01.

[30] 3g. Identity Management. *3g*, 2009. URL `https://observatorio.iti.upv.es/media/managed{_}files/2009/01/12/3GAmericas{_}Unified{_}Identity{_}Management{_}Jan2009.pdf`.

[31] Julia Clark, Dahan Meriana, and Desai Vyjayanti. Digital Identity: Towards Shared Principles for Public and Private Sector Cooperation. Working Paper, World Bank, August 2016. URL `https://secureidentityalliance.org/public-resources/4-july-2016-report-digital-identity/file`.

[32] The amazon recommendations secret to selling more online. `http://rejoiner.com/resources/amazon-recommendations-secret-selling-online/`, 2016. Accessed: 2018-12-01.

[33] Antitrust in 60 seconds: Multi-sided business models. `http://www.project-disco.org/competition/092518-antitrust-in-60-seconds-multi-sided-business-models/#.W-be1vzZBTY`, 2018. Accessed: 2018-10-06.

[34] Who's selling private data about us? more importantly, who's buying it? https://sociable.co/technology/selling-private-data/, 2018. Accessed: 2018-10-08.

[35] Who's selling private data about us? more importantly, who's buying it? https://www.metafilter.com/95152/Userdriven-discontent#32560467, 2018. Accessed: 2018-10-08.

[36] EmpowerID. The top 5 federated single sign-on scenarios. Technical report, EmpowerID, 2017.

[37] Arun Nanda, Andre Durand, Bill Barnes, Carl Ellison, Caspar Bowden, Craig Burton, James Governor, Jamie Lewis, John Shewchuk, Luke Razzell, Marc Canter, Mark Wahl, Mike Jones, Phil Becker, Radovan Janocek, Ravi Pandya, Robert Scoble, and Scott C Lem. The Laws of Identity. *none*, page 13, 2005.

[38] Director of Security PWC, Claus Nørklit Roed and Technology in Denmark. Presentation: Iam in the enterprise. Technical report, PWC, 2017.

[39] Janrain. How customer identity & access management (ciam) is different from traditional identity & access management (iam). Technical report, JanRain, 2018.

[40] Ping Identity. The primer: Nuts and bolts of federated identity management. Technical report, Ping Identity, 2017.

[41] Statement by the acting director of ftc's bureau of consumer protection regarding reported concerns about facebook privacy practices. https://www.ftc.gov/news-events/press-releases/2018/03/statement-acting-director-ftcs-bureau-consumer-protection, 2018. Accessed: 2018-11-01.

[42] Google exposed user data, feared repercussions of disclosing to public. https://www.wsj.com/articles/google-exposed-user-data-feared-repercussions-of-disclosing-to-public-153
2018. Accessed: 2018-10-01.

[43] The equifax data breach. https://www.ftc.gov/equifax-data-breach, 2018. Accessed: 2018-10-28.

[44] A dating site and corporate cyber-security lessons to be learned. https://www.pandasecurity.com/mediacenter/security/lessons-ashley-madison-data-breach/, 2018. Accessed: 2018-10-01.

[45] Christopher allen misattributes. https://medium.com/@sheldrake/see-also-http-www-lifewithalacrity-com-2016-04-the-path-to-sel
2018. Accessed: 2018-10-01.

[46] The path to self-sovereign identity. http://www.lifewithalacrity.com/2016/04/the-path-to-self-soverereign-identity.html, 2018. Accessed: 2018-10-01.

[47] Kai Rannenberg and José Fernando Carvajal Vión. "online identity management - make it respect our privacy". *ISO - Focus*, April 2013. URL http://www.iso.org/iso/isofocusplus_2013-04.

[48] Uma 2.0 - wg - user managed access. https://kantarainitiative.org/confluence/display/uma/UMA+Scenarios+and+Use+Cases#UMAScenariosandUseCases-employer_scenarioScenario:ManagingInformationinWhichEmployersandEmployeesBothHaveaStake(Pending), 2018. Accessed: 2018-10-01.

[49] Justin Richer. UMA 2.0. Technical report, Medium, 2017. URL https://medium.com/@justinsecurity/uma-2-0-437c293c3283.

[50] Butler Lampson. Privacy and security usable security: how to get it. *Communications of the ACM*, 52(11):25–27, 2009.

[51] Who does the data protection law apply to? https://ec.europa.eu/info/law/law-topic/data-protection/reform/rules-business-and-organisations/application-regulation/who-does-data-protection-law-apply_en, 2018. Accessed: 2018-10-08.

[52] Sovrin Foundation. Sovrin: A protocol and token for self-sovereign identity and decentralized trust. Technical report, Sovrin Foundation, 2018.

[53] David F Ferraiolo, D Richard Kuhn, and Ramaswamy Chandramouli. Role-Based Access Control. *Components*, 2002(10): 338, 2003. ISSN 00200190. doi: 10.1016/S1361-3723(02)01211-3. URL http://books.google.com/books?hl=en{&}lr={&}id=48AeIhQLWckC{&}pgis=1.

[54] Fines and penalties. https://www.gdpreu.org/compliance/fines-and-penalties/, 2018. Accessed: 2018-08-03.

[55] US Department of Justice. How to protect you network from Ransomware, 2017. URL https://www.justice.gov/criminal-ccips/file/872771/download. Accessed: 2018-05-21.

[56] World Economic Forum. Securing a common future in cyberspace, 2018. URL https://www.youtube.com/watch?v=Tqe3K3D7TnI.

[57] Butler W. Lampson. Computer security in the real world. *Communications of the ACM*, 37(6):37–46, 2004.

[58] Ishan Mehta Photo of Ishan Mehta Ishan Mehta Mieke Eoyang, Allison Peters. "to catch a hacker: Toward a comprehensive strategy to identify, pursue, and punish malicious cyber actors". *Thirdway*, October 2018. URL https://www.thirdway.org/report/to-catch-a-hacker-toward-a-comprehensive-strategy-to-identify-pursue-and-

[59] Megan M McNally. Charting the conceptual landscape of identity theft. *CRIME PREVENTION STUDIES*, 23:33, 2008.

[60] Peter Kruize. Identitetstyveri. *Forskningspuljen*, pages 1 – 46, 2009. URL http://www.justitsministeriet.dk/sites/default/files/media/Arbejdsomraader/Forskning/Forskningspuljen/2011/2009/Rapport_identitetstyveri.pdf.

[61] OECD. Online identity theft. *None*, pages 1 – 137, 2009. URL http://www.oecd.org/sti/consumer/onlineidentitytheft.htm.

[62] KLPD Dienst IPOL. National Threat Image for the netherlands. Working Paper, Nationale Dreigingsbeeld, NDB, 2008. URL https://anzdoc.com/klpd-dienst-ipol-nationaal-dreigingsbeeld-georganiseerde-cri.html.

[63] Experian Information Solutions. Here's How Much Your Personal Information Is Selling for on the Dark Web. White Paper, Experian, 2018. URL https://www.experian.com/blogs/ask-experian/heres-how-much-your-personal-information-is-selling-for-on-the-dark-web/.

[64] Proposal for an eprivacy regulation. https://ec.europa.eu/digital-single-market/en/proposal-eprivacy-regulation, 2018. Accessed: 2018-10-11.

[65] Data protection in the united states: overview. https://content.next.westlaw.com/Document/I02064fbd1cb611e38578f7ccc38dcbee/View/FullText.html?contextData=(sc.Default)&transitionType=Default&firstPage=true&bhcp=1, 2018. Accessed: 2018-10-08.

[66] Is hipaa compliant with the gdpr? https://blog.ipswitch.com/is-hipaa-compliant-with-the-gdpr, 2018. Accessed: 2018-10-08.

[67] General data protection regulation (gdpr). https://gdpr-info.eu/chapter-4/, 2018. Accessed: 2018-10-11.

[68] 8 data subject rights according to gdpr. https://advisera.com/eugdpracademy/knowledgebase/8-data-subject-rights-according-to-gdpr/, 2018. Accessed: 2018-10-11.

[69] Ping Identity Corp. Saml101. Technical report, Ping Identity, 2016.

[70] Aries Fajar Dwiputera and IS Ruppa. Single sign-on architecture in public networks (liberty alliance). In *INFOTECH seminar on advanced communication Services (ACS)*, 2005.

[71] K. D. L. a. E. Lewis. Web Single Sign-On Authentication using SAML. *ArXiv e-prints*, September 2009.

[72] Vika Felmetsger. Security Assertion Markup Language (SAML). URL `https://www.cs.ucsb.edu/{~}bultan/courses/595-W06/SAML.pdf`.

[73] Introduction to SAML | Download your free guide on SAML. URL `http://info.forumsys.com/introduction-to-saml-white-paper`.

[74] Kelly D Lewis. Web single sign-on authentication using saml. *arXiv preprint arXiv:0909.2368*, 2009.

[75] Digitaliseringstyrelsen. Fælles offentlig referencearkitektur for brugerstyring. Technical report, Digitaliseringstyrelsen, 2017.

[76] A guide to oauth 2.0 grants. `https://alexbilbie.com/guide-to-oauth-2-grants/`, 2018. Accessed: 2018-10-11.

[77] The difference between root certificates and intermediate certificates. `https://www.thesslstore.com/blog/root-certificates-intermediate/`, 2018. Accessed: 2018-10-11.

[78] Ssl certificate for mozilla.com issued without validation. `https://www.sslshopper.com/article-ssl-certificate-for-mozilla.com-issued-without-validation.html`, 2018. Accessed: 2018-10-11.

[79] How a 2011 hack you've never heard of changed the internet's infrastructure. `https://slate.com/technology/2016/12/how-the-2011-hack-of-diginotar-changed-the-internets-infrastructure.html`, 2018. Accessed: 2018-10-11.

[80] Is self-sovereign identity the ultimate gdpr compliance tool? `https://medium.com/evernym/is-self-sovereign-identity-ssi-the-ultimate-gdpr-compliance-tool-9d811075`
2018. Accessed: 2018-08-03.

[81] The laws of identity on the blockchain. `https://www.identityblog.com/?p=1658`, 2018. Accessed: 2018-08-03.

[82] W3C Draft Community Group. Decentralized Identifiers (DIDs) v0.11. Draft Community Group Report, W3C, 2018. URL `https://w3c-ccg.github.io/did-spec/`.

[83] Peter Wood. Implementing identity management security - An ethical hacker's view. *Network Security*, 2005(9):12–15, 2005. ISSN 13534858. doi: 10.1016/S1353-4858(05)70282-8.

[84] Andreas Pashalidis and Chris J. Mitchell. A taxonomy of single sign-on systems. In *Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)*, volume 2727 LNCS, pages 249–264, 2003. ISBN 9783540405153. doi: 10.1007/3-540-45067-X_22.

[85] S. R. Jorgensen C. N. Gitau, S. Zaharia. Identity and access management system in a university context. Technical report, Student projects, AAU, ICTE, 2017.

[86] Joshua B Bolten. E-authentication guidance for federal agencies–memorandum to the heads of all departments and agencies (omb m-04-04). Technical report, Technical report, Executive Office of The President, Office of Management and Budget, Washington, DC 20503, 2004.

[87] David Ferraiolo, John Barkley, and Richard Kuhn. Role-Based Access Controls. *ACM Transactions on Information and System Security*, 2(1):34–64, 1992. ISSN 10949224. doi: 10.1145/300830.300834. URL http://portal.acm.org/citation.cfm?doid=300830.300834.

[88] Ping Identity. Ping identity and nist collaborate on guide to implementing attribute based access control. Technical report, Ping Identity, 2017. URL https://www.pingidentity.com/en/company/blog/posts/2015/ping-identity-and-nist-collaborate-on-guide-to-implementing-attribute-bas
html.

[89] NIST. Attribute based access control. Technical report, NCCoE, 2017. URL https://www.nccoe.nist.gov/sites/default/files/library/sp1800/abac-nist-sp1800-3a-draft.pdf.

[90] D. Richard Kuhn, Edward J. Coyne, and Timothy R. Weil. Adding attributes to role-based access control. *Computer*, 43(6):79–81, 2010. ISSN 00189162. doi: 10.1109/MC.2010.155.

[91] Michael P. Gallaher and Alan C. O'Connor and Brian Kropp. The Economic Impact of Role-Based Access Control, 2002. URL http://citeseerx.ist.psu.edu/viewdoc/download?doi=10.1.1.134.7438{&}rep=rep1{&}type=pdf.

[92] Ravi Sandhu, David Ferraiolo, and Richard Kuhn. The NIST Model for Role-Based Access Control: Towards a Unified Standard. In *Proceedings of the Fifth ACM Workshop on Role-Based Access Control*, pages 47–63, 2000. ISBN 9788578110796. doi: 10.1017/CBO9781107415324.004.

[93] ANSI INCITS. Incits 359-2004. *Role based access control*, 2004.

[94] NIST. Role-based access control (rbac) role engineering process, 2004. URL `http://csrc.nist.gov/groups/SNS/rbac/documents/HealthcareRBACTFRoleEngineeringProcessv3.0.pdf`.

[95] Eric Yuan and Jin Tong. Attributed based access control (abac) for web services. In *Web Services, 2005. ICWS 2005. Proceedings. 2005 IEEE International Conference on*. IEEE, 2005.

[96] 3G Americas. Identity management – overview of standards technologies for mobile and fixed internet. *Whitepaper*, pages 52–60, 2009.

[97] Diala Abi Haidar, Nora Cuppens-Boulahia, Frederic Cuppens, and Herve Debar. An extended rbac profile of xacml. In *Proceedings of the 3rd ACM Workshop on Secure Web Services*, SWS '06, pages 13–22, New York, NY, USA, 2006. ACM. ISBN 1-59593-546-0. doi: 10.1145/1180367.1180372. URL `http://doi.acm.org/10.1145/1180367.1180372`.

[98] OASIS XACML Technical Committee et al. extensible access control markup language (xacml) version 3.0. *Oasis standard, OASIS*, 2013.

[99] Richard Spires. CIO PerspectivemThe enterprise risk management approach to cybersecurity, 2017. URL `https://fcw.com/articles/2017/07/25/cio-perspective-erm-cyber-spires.aspx`. Accessed: 2018-05-23.

[100] KPMG Netherlands. Kpmg's 2008 european identity and access management survey. Technical report, KPMG Netherlands, 2008. URL `https://kpmg.de/media/20080101_2008_European_Identity_Access_Management_Survey.pdf`.

[101] ISF. Ifs the standard of good practise, 2014. URL `https://www.securityforum.org/tool/the-isf-standardrmation-security/`.

[102] NIST. Nist cybersecurity quidelines, 2014. URL `https://www.nist.gov/sites/default/files/documents/cyberframework/cybersecurity-framework-021214.pdf`.

[103] ISACA. Cobit 5 assessment programme, 2015. URL `http://www.isaca.org/COBIT/Pages/COBIT-Assessment-Programme.aspx`.

[104] & Takahashi K. Bertino, E. *Identity management : concepts, technologies, and systems*. ProQuest Ebook Central, 2011.

[105] E. Damiani, S. De Capitani di Vimercati, and P. Samarati. Managing multiple and dependable identities. *IEEE Internet Computing*, 7 (6):29–37, Nov 2003. ISSN 1089-7801. doi: 10.1109/MIC.2003.1250581.

[106] & Royer D Meints, M. Datenschutz und datensicher-heit (dud). In *Der Lebenszyklus von Identitaten*, 2008.

[107] M. Turpin & J. P. van Deventer Royer d., P. M. Alexander. Supporting decision making for enterprise identity management – an explanatory model for describing the relevant impacts. *Informatik-Spektrum*, 2010.

[108] Karl-Rudolf Moll, Manfred Broy, Markus Pizka, Tilman Seifert, Klaus Bergner, and Andreas Rausch. Erfolgreiches management von software-projekten. *Informatik-Spektrum*, 27(5):419–432, 2004. ISSN 0170-6012.

[109] 4 types of organizational structures. https://online.pointpark.edu/business/types-of-organizational-structures/, 2018. Accessed: 2018-08-21.

[110] Nicholas Bohm and Stephen Mason. Identity and its verification. *Computer Law & Security Review*, 26(1):43 – 51, 2010. ISSN 0267-3649. doi: https://doi.org/10.1016/j.clsr.2009.11.003. URL http://www.sciencedirect.com/science/article/pii/S0267364909001885.

[111] Personnummeret i cpr-systemet. https://www.cpr.dk/media/17534/personnummeret-i-cpr.pdf, 2018. Accessed: 2018-08-01.

[112] Bekendtgørelse af lov om det centrale personregister. https://www.retsinformation.dk/Forms/R0710.aspx?id=144955#Kap12, 2018. Accessed: 2018-08-02.

[113] Introducing the dweb. https://kantarainitiative.org/confluence/display/uma/Case+Study%3A+Access+Management+2.0+for+the+Enterprise, 2018. Accessed: 2018-03-12.

[114] 5 reasons you need openid connect and uma in your iam stack. https://www.gluu.org/blog/5-reasons-you-need-openid-connect-uma-identity-access-management/, 2015. Accessed: 2018-22-11.

[115] Key words for use in rfcs to indicate requirement levels. http://www.rfc-editor.org/rfc/rfc2119.txt, 1997. Accessed: 2018-12-02.

[116] Introducing the dweb. https://hacks.mozilla.org/2018/07/introducing-the-d-web/, 2018. Accessed: 2018-10-11.

[117] Professor: Cpr-numre er rystende usikre. https://www.denoffentlige.dk/professor-cpr-numre-er-rystende-usikre, 2018. Accessed: 2018-08-01.

[118] Vidensbanken om kønsidentitet. http://www.thranesen.dk/artikler/Personnumre_ID-kort_20130201.pdf, 2018. Accessed: 2018-08-01.

[119] CPR-Kontoret. CPR-KONTORET Udviklingen på CPR-området i de seneste 20-25 år frem til 2009. Technical report, "CPR-Kontoret", 2017. URL https://www.cpr.dk/cpr-systemet/historie/.

[120] Aktivister havde tre muligheder for at validere cpr-numre via tinglysningen. https://www.version2.dk/artikel/gaetteleg-aktivister-havde-tre-muligheder-validere-cpr-numre-tinglysninge
2018. Accessed: 2018-08-01.

Appendix A

A.1 Account chooser for partial identities

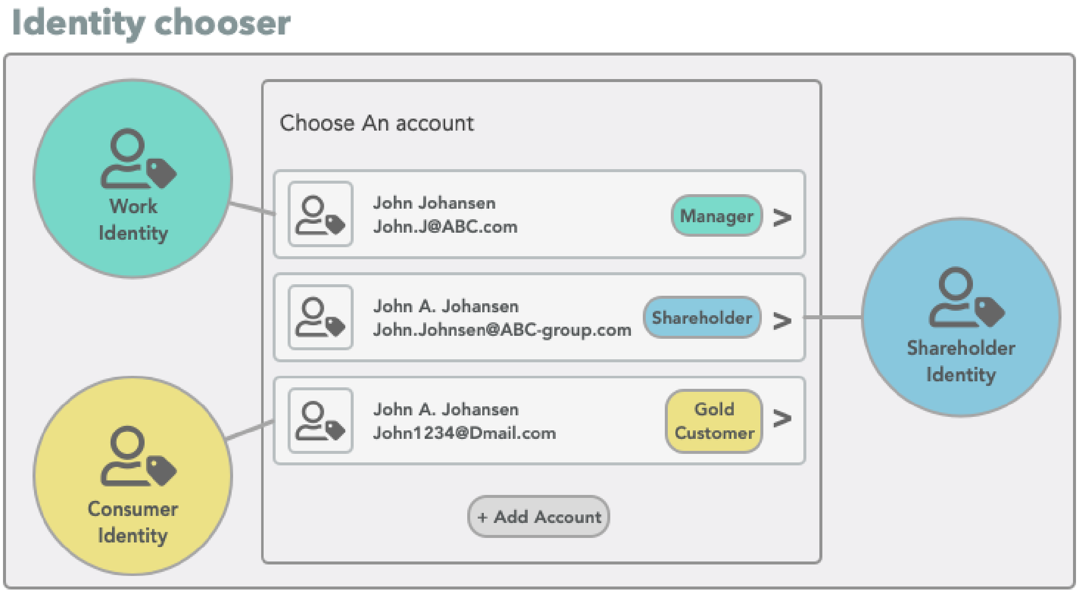
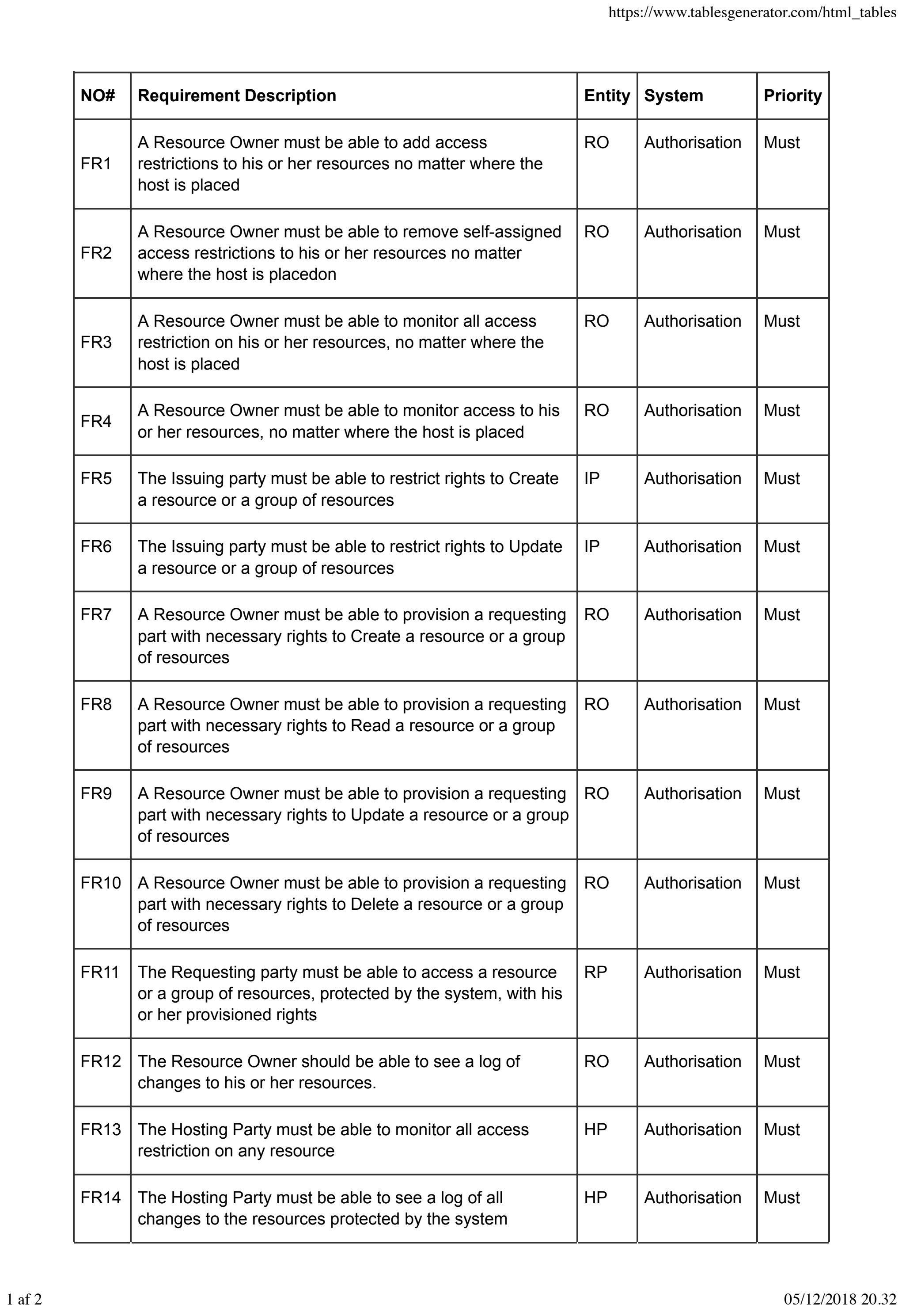


Figure A.1: Flow of choosing an account to signify an identify

A.2 Requirements

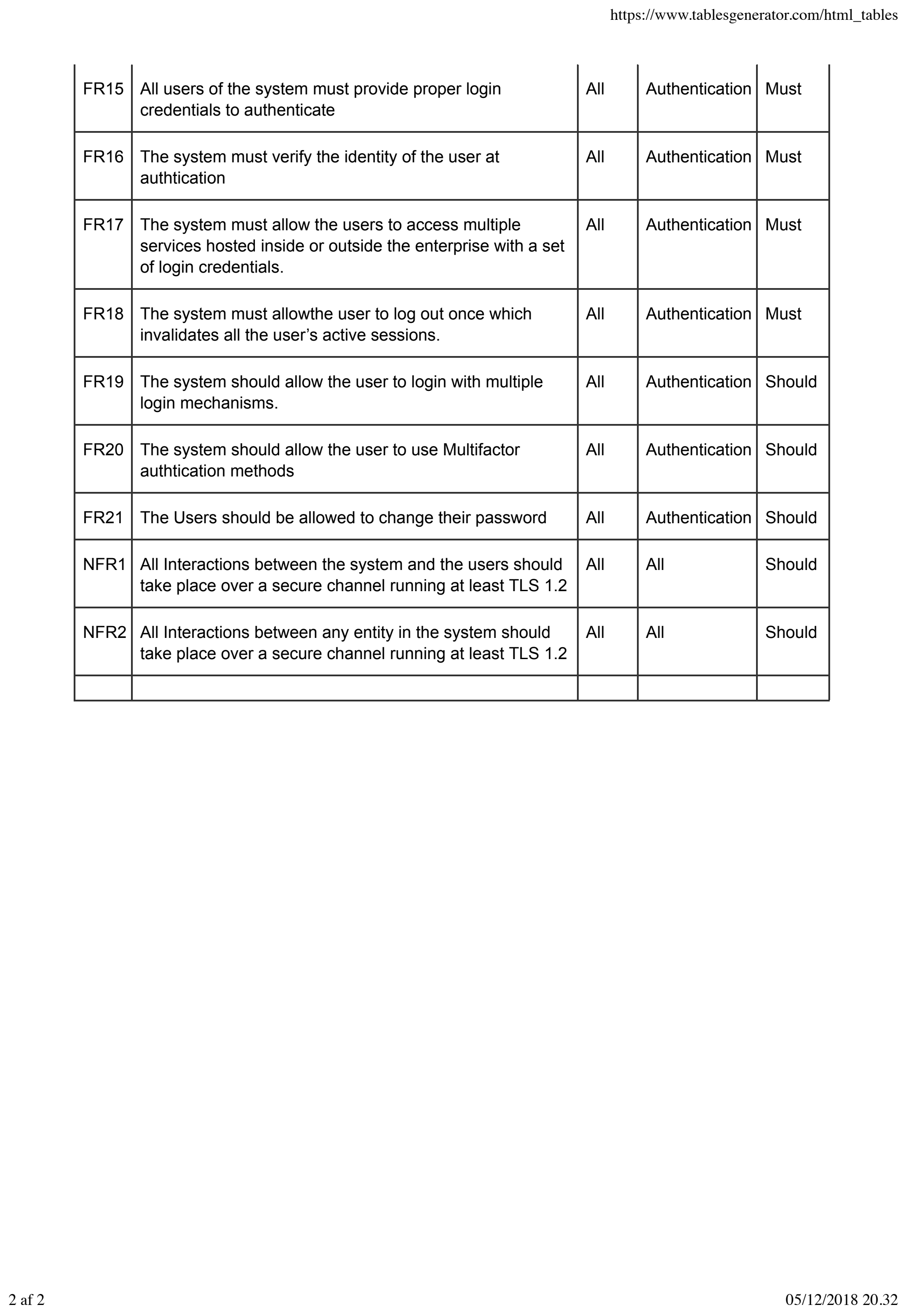


https://www.tablesgenerator.com/html_tables

NO#	Requirement Description	Entity	System	Priority
FR1	A Resource Owner must be able to add access restrictions to his or her resources no matter where the host is placed	RO	Authorisation	Must
FR2	A Resource Owner must be able to remove self-assigned access restrictions to his or her resources no matter where the host is placedon	RO	Authorisation	Must
FR3	A Resource Owner must be able to monitor all access restriction on his or her resources, no matter where the host is placed	RO	Authorisation	Must
FR4	A Resource Owner must be able to monitor access to his or her resources, no matter where the host is placed	RO	Authorisation	Must
FR5	The Issuing party must be able to restrict rights to Create a resource or a group of resources	IP	Authorisation	Must
FR6	The Issuing party must be able to restrict rights to Update a resource or a group of resources	IP	Authorisation	Must
FR7	A Resource Owner must be able to provision a requesting part with necessary rights to Create a resource or a group of resources	RO	Authorisation	Must
FR8	A Resource Owner must be able to provision a requesting part with necessary rights to Read a resource or a group of resources	RO	Authorisation	Must
FR9	A Resource Owner must be able to provision a requesting part with necessary rights to Update a resource or a group of resources	RO	Authorisation	Must
FR10	A Resource Owner must be able to provision a requesting part with necessary rights to Delete a resource or a group of resources	RO	Authorisation	Must
FR11	The Requesting party must be able to access a resource or a group of resources, protected by the system, with his or her provisioned rights	RP	Authorisation	Must
FR12	The Resource Owner should be able to see a log of changes to his or her resources.	RO	Authorisation	Must
FR13	The Hosting Party must be able to monitor all access restriction on any resource	HP	Authorisation	Must
FR14	The Hosting Party must be able to see a log of all changes to the resources protected by the system	HP	Authorisation	Must

1 af 2

05/12/2018 20.32



https://www.tablesgenerator.com/html_tables

FR15	All users of the system must provide proper login credentials to authenticate	All	Authentication	Must
FR16	The system must verify the identity of the user at authtication	All	Authentication	Must
FR17	The system must allow the users to access multiple services hosted inside or outside the enterprise with a set of login credentials.	All	Authentication	Must
FR18	The system must allowthe user to log out once which invalidates all the user’s active sessions.	All	Authentication	Must
FR19	The system should allow the user to login with multiple login mechanisms.	All	Authentication	Should
FR20	The system should allow the user to use Multifactor authtication methods	All	Authentication	Should
FR21	The Users should be allowed to change their password	All	Authentication	Should
NFR1	All Interactions between the system and the users should take place over a secure channel running at least TLS 1.2	All	All	Should
NFR2	All Interactions between any entity in the system should take place over a secure channel running at least TLS 1.2	All	All	Should

2 af 2 05/12/2018 20.32

A.3 The CPR identifier

Problems arises when a construct like the CPR number is used as an authenticator as described above. A malicious person would be able to quickly gather sufficient enough information about a person, to guess the CPR number. Then validate it and use it as part of identity theft or gather confidential information about its rightful owner. This is because the certain enterprises trust the CPR number for being a sufficient authenticator, while it originally only was meant as an identifier[117]. Therefore CPR numbers like other identifiers are good at identifying people, but should never be used to authenticate. The current CPR and CVR solutions are registers which makes it possible to identify persons and companies uniquely. The solutions do not make it possible to authenticate identities, just make sure they exist [60]. This conclusion does not only apply to a national identifier as the CPR number, but also for identification numbers in companies such as employee, customer-number etc.

The unique Person number is a semi-random identifier which takes base in date of birth and gender. It has the syntax "DDMMYY-XXXX" where "DDMMYY" is the person date of birth. The last part "XXXY" is an incrementing number, where the last digit (Y) is a modulus-11 control-digit, which also indicates the gender of the person (even for women, odd for men). The incremental number follow another static schema which makes up for people born in different centuries. These properties of the identifier provide systems to check for certain types of input errors[118]. This numbering system allowed for up to 270 people being born the same day (2000-2036). Due to constraints in the system, the authority of the system[1] changed the last digit from a control number to part of the incremental numbers in 2002, to extend the capabilities of the system regarding handling more persons born the same day. This change was initiated because there had been occasions where the current system was not sufficient due to many registered with the same date of birth. However, this limit was not met due to a sudden increase in birthrate or missing requirement for the property for the original system.

An investigation of the problem deemed that it was the increase in general immigration that leads to missing numbers at the start of each month, especially the date first of January. The heart of the problem was identified as foreigners with unknown date of birth were given a fictional date of birth, and thereby a CPR number there did not reflect their true date of birth as were one of the ideas behind the system. These CPR-numbers were given by city officials that had the responsibility to register new people, coming into Denmark but if their real date of birth was unknown, they were just randomly assigned. However, even though procedures said that these should be spread out uniformly over the year, specific dates in the

[1] CPR Kontoret

system became unavailable already in 1993, and the system had to be re-engineered. The actual fix for this problem was to make use of the first digit number and let the numbers begin with the number 9 for a person born on January 1, 1950, for people coming into the country[119].

This change and a couple of others changes made between 1993-2000 to the numbering procedure made it possible to assign at least 2000 persons per gender for each day, without breaking the system completely[111]. Before this change, there was a direct relation between the date of birth and the last four digits. This construct made it easy to guess a persons CPR number by calculate the various possibilities and check them against public systems, which required no or low levels of authentication[120].

A.4 OAuth Flows

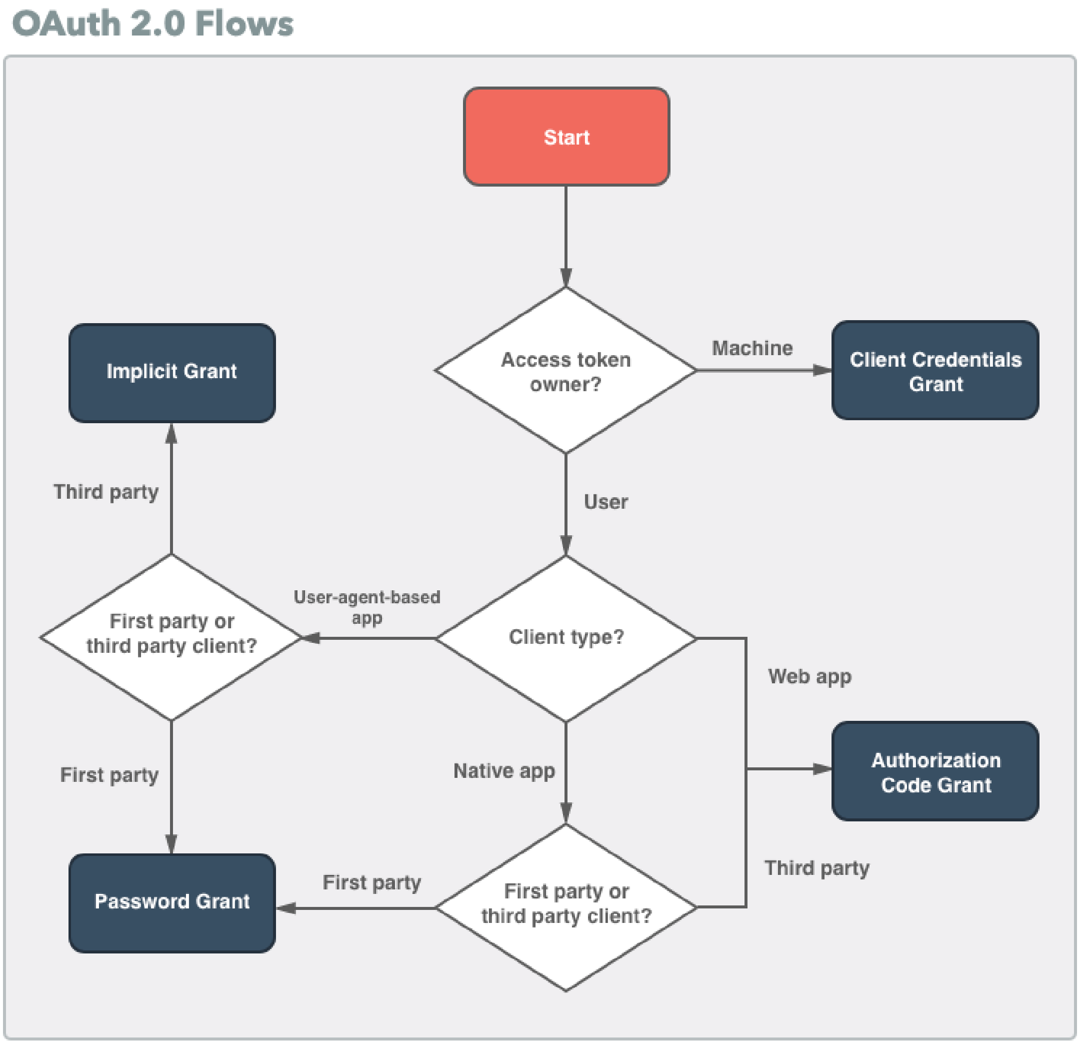


Figure A.2: Various OAuth grant types, and which one should be chosen based on characteristics of the scenario

A.5 IAM Stakeholder

Rank	**Category**	**Related comments and explanation**
1	IDM users	Operating departments and application administrators which is directly affected by the change. They also have the knowledge about structure and tasks, which is necessary in order for IDM to work
2	IT department	The executing stakeholder, which is implementing the IDM solution, typically also work as the project owner for the system. External SP has to integrate as well as internal services.
3	Management	Decisionmakers should be responsible for the execution of the IDM project, has the leadership role.
4	HR department	The entry point for employees in the organisation, typically main data supplier and is needed in order for IDM to be properly integrated into the organisation. A stakeholder which typically has an actor role within the system.
5	Works Council	Involvement due to interest in protecting employees interests(e.g. privacy issues, performance tracking, personal data). Can be a supporting role when aggregating requirements
6	Auditors	Both internal and external auditor responsible for checking compliance concerning both security and processes.
7	DPO’s	Personal with responsibilities for data protection and information security within the organisation. Also responsible for regulatory requirements such as GDPR, for both employees and customers.
8	Organ. Mgmt.	Operational Management for the organisation, responsible for account assignment in an organisation.

A.6 Decision support system framework

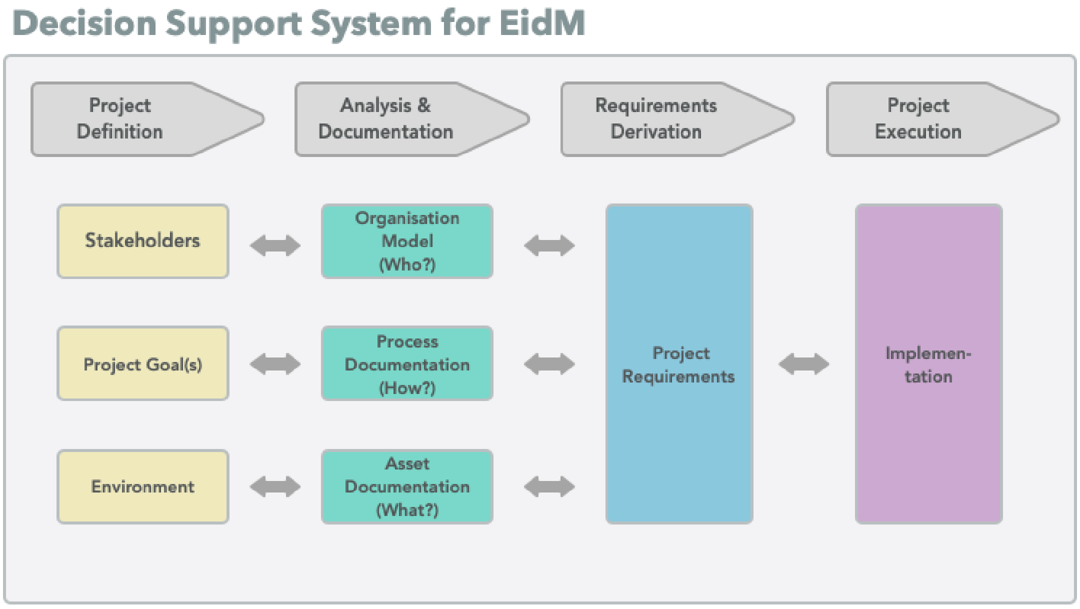


Figure A.3: Decision support system for IDM in enterprises

A.7 W3C's Requirements for DID

Goal	**Definition**
Decentralization	The architecture should eliminate the requirement for centralised authorities or single points of failure in identifier management, including the registration of globally unique identifiers, public verification keys, service endpoints, and other metadata.
Self-Sovereignty	The architecture should give entities, both human and non-human, the power to directly own and control their digital identifiers without the need to rely on external authorities.
Privacy	The architecture should enable entities to control the privacy of their information, including minimal, selective, and progressive disclosure of attributes or other data.
Security	The architecture should enable sufficient security for relying parties to depend on DID Documents for their required level of assurance.
Proof-based	The architecture should enable an entity to provide cryptographic proof of authentication and proof of the authorisation rights.
Discoverability	The architecture should make it possible for entities to discover DIDs for other entities to learn more about or interact with those entities.
Interoperability	The architecture should use interoperable standards so DID infrastructure can make use of existing tools and software libraries designed for interoperability.
Portability	The architecture should be system and network-independent and enable entities to use their digital identifiers with any system that supports DIDs and DID Methods.
Simplicity	To meet these design goals, as simple as possible but no simpler.
Extensibility	The architecture should enable extensibility as long as it does not significantly hinder the other principles of interoperability, portability or simplicity.