LICENTIATE THESIS

Attribute-based Approaches for Secure Data Sharing in Industry

Alex Chiquito

Cyber-physical Systems

LULEÅ
UNIVERSITY
OF TECHNOLOGY

Attribute-based Approaches for Secure Data Sharing in the Industry

Alex Chiquito

Dept. of Computer Science and Electrical Engineering
Luleå University of Technology
Luleå, Sweden

Supervisors:

Ulf Bodin, Olov Schelén, and Jerker Delsing

Everything's Eventual

Abstract

The Industry 4.0 revolution relies heavily on data to generate value, innovation, new services, and to optimize current processes. Technologies such as Internet of Things (IoT), machine learning, and digital twins depend on data to bring value and innovation in discrete manufacturing and process industries. The origin of data may range from confidential sensor data to financial, user, and business critical data. In data-driven ecosystems, collaboration between different actors is often needed to provide services such as analytics, logistics, predictive maintenance, and process improvement. Data therefore cannot be considered a corporate internal asset only. Hence, data needs to be shared among organizations in a data-driven ecosystem for it to be used as a strategic resource for creating desired values, innovations, or process improvements. When sharing business critical and sensitive data, the access to the data needs to be accurately controlled to prevent leakage to unauthorized users and organizations.

Access control is an authorization mechanism to control actions of users over objects, e.g., to read, write, and delete data. This thesis studies Attribute Based Access Control (ABAC) for industrial data sharing. ABAC presents the idea of attributes to create access policies, rather than manually assigned roles or ownerships, enabling expressive fine-granular access control policies. Furthermore, this thesis presents approaches to implement ABAC into industrial IoT data sharing applications, with special focus on the manageability and granularity of the attributes and policies. The thesis studies the implications of outsourcing data storage on third party cloud servers over access control for data sharing. Finally, the thesis explores how to integrate cryptographic techniques and paradigms into data access control. In particular, the combination of ABAC and Attribute-Based Encryption (ABE) is investigated to protect privacy over not-fully trusted domains. In this, important research gaps are identified.

The main contributions of the thesis are: (1) A model to manage ABAC attributes for time series databases to enable efficient data selection, runtime adding and removal of data sources, flexible management and enforcement of access policies, and simple maintenance of policies. (2) A fine-grained ABAC model with a query rewriting step to enable the expression and enforcement of value and time constraint policies, without extensive computational overhead. (3) A state-of-the-art and gap analysis on ABE-enabled ABAC, including a proposal for how to combine these schemes into an architecture to enable the efficient use of ABAC attributes with ABE to ease administration.

Contents

Acknowledgments

It has been quite a journey, and it is still far from done. This thesis reflects a little bit more than two years of efforts, discoveries, challenges, and much more. I know this is not a Nobel Prize, and maybe these acknowledgments can be seen as too much, but these have really been my toughest moments, and my gratitude is really sincere. Of course, this work was not just from my side. First of all, this could have not been possible without the invaluable help and support from my supervisors Ulf Bodin and Olov Schelén. Their patience and guidance have helped me a lot through this process. Thanks for the opportunity to learn and grow with such amazing researchers and persons. Let's go for the second half!

I would want to thank my family. Mom and Dad, you have always been with me since the beginning, in every decision I take, even if that meant moving to the other side of the planet. You are without a doubt my strongest motivation. My eternal gratitude to my grandparents, their unconditional love is always with me. And of course, my brother, I really can't express how grateful I am that you are here, not only for the meals. What can I say? I have the best family, you all believe in me more than I probably do, and that keeps me fighting. Gracias!!

And last but not least, many thanks to my dear friends. Everyday I look forward lunch in the Fika room and it is always a breath of fresh air and tons of fun. The band practices, skating, skiing, or just drinking and chatting. As the FUN song says: "F is for friends who do stuff together", and we have for sure done fun things that I wouldn't have done on my own, so thank you! I would need to write a whole chapter to talk about you all, but I'm pretty sure it would be the best one. Please forgive me if you don't see your name, but if you are reading this you are most likely my friend so thank you! Special thanks to Jaime and Emanuel for being amazing officemates and helping me starting up. Shai, Cristina, Konstantina, Carl, Sree, Joakim, Jacob, Gustav, György, Pedro, and many more. My friends back in México, always asking how I'm doing, old work and college friends, my D&D and Magic group, Meli, and many many more. I could have not done this without you.

Gracias a todos!

Luleå, May 2022.
Alex Chiquito

Part I

Chapter 1

Thesis Introduction

"You miss 100% of the shots you don't take.
-Wayne Gretzky"

-Michael Scott

1.1 Problem formulation

Industrial production is receiving an important paradigm shift given the rapidly emerging technologies, advance digitalization, future-oriented technologies such as smart objects, and the constant presence of the internet. This new paradigm's vision looks, among other benefits, for shorter development times, flexibility, and resource efficiency, while using the new coming technologies to improve production aspects such as the automation, digitalization, or miniaturization [1]. In this context, industries have developed an important part of their business and value processes around data. Data is often used to build data-driven models to carry out advanced analytics in tasks such as assessment and prediction of production performance, making prediction on costs, performing predictive maintenance on machinery and equipment, and much more.

A big part of the value producing operations is either done outside the organization where the data is originally produced, or require data from partners and outside parties to be effective. This opens opportunities for the creation of data-driven business ecosystems, where organizations collaborate to generate value and innovation in greater measure than what they would be able to produce on their own. In data-driven ecosystems, data is a strategic resource to create value and innovation and, as such, it is crucial to make it available to partners. Inside these ecosystems, there are a number of different types of data being produced and shared, ranging from temperature sensor readings to customer order details, and even sensitive human resource (HR) information.

Every type of data may have an important role in value producing operations for the customers, partners, and to the business itself. However, the same data that may help the business if processed properly, can be harmful should it reach the wrong hands, causing monetary losses, or even facing legal actions if agreements such as the EU's General

Data Protection Regulation (GDPR) are violated [2][3]. Therefore, business critical and sensitive data need to be protected at all times, both in transit and at rest, e.g., when stored in different types of databases.

Data protection means to safeguard important data from corruption, compromise or loss [4]. Moreover, it includes to ensure that the data is accessible for authorized purposes only, and that it is in compliance with applicable legal or regulatory requirements such as GDPR. This thesis focuses on the access control and encryption security aspects connected to the protection of data. In this context, access control is an authorization mechanism that operates on ***access policies*** that determines the who can access what data, and under what conditions. Other data security aspects include threat monitoring, authentication, and data loss prevention.

1.2 Motivation and research questions

Secure data sharing

According to the International Data Spaces Association (IDSA) [2], to achieve secure data sharing four requirements must be satisfied:

1. *Authentication and authorization*: Each participant in the data sharing process must be able to verify their identity to the other participants. Given the participant's identity, it must be possible to know its access capabilities and security features. In this, *access policies* are used to control these capabilities.

2. *Usage policies and usage enforcement*: In a data sharing ecosystem, consumers must handle data following usage policies specified by its owner. This includes restrictions for tasks such as future transfer or sharing, or aggregations with similar competitors' data.

3. *Trustworthy communication and security by design*: Applications and systems handling data must be running in a trusted software stack, and communication with external systems must be properly protected.

4. *Technical certification*: The entities participating in the ecosystem must be certified by a central trusted body. This certification may mean that their security can be trusted, or that they comply with the technical requirements to ensure interoperability.

Requirements 4 and the security by design aspect of requirement 3 presented by IDSA rely heavily on the platform on which the ecosystem is based. In contrast, requirements 1 and 2 demand higher level mechanisms to satisfy and enforce, such as application level enforcement or a central authorization system. The work presented in this thesis is motivated by the authorization part of requirement 1, aiming at providing efficient mechanisms to enforce authorization using access-policy control to shared data in industrial settings. Moreover, this thesis studies the trustworthy communications aspects of requirement 3 by exploring encryption mechanisms for data in transfer. Requirements 3 and 4 are considered mainly for future work in the context of the industrial Service-Oriented Architecture (SOA) based framework Eclispe Arrowhead [5, 6]

Given the risks of leaking data that may be sensitive, it could be tempting to impose strict access restrictions on all produced data. However, in that case, the innovation and value processes from the ecosystem would be hindered by slowing down the data sharing. There is a need inside organizations to find balance between data privacy and security, and agility in the data sharing process [2]. Not all data produced by an industry demands the same level of protection. Hence, the work on this thesis aims on providing efficient, flexible, and fine-granular mechanisms to express and enforce access control over data.

Research questions

This thesis aims to answer two different research questions:

Q1 *How can data protection be provided in industrial settings and what are the essential properties in achieving such protection?*

For this question, the boundaries and properties of an industrial setting need to be defined. Industrial environments are dynamic in nature, due to constant staff rotation, sensor updates, and even machinery upgrades. Furthermore, the type of data in an industrial environment is heterogeneous, including types of data such as time series sensor readings, machinery logs, financial data, and even personal information.

The sharing of data can happen between teams and departments inside an organization. However, industrial data sharing is not always bound to the trusted domain of the data owner. Data driven ecosystems depend on data from various sources to be shared with a number of partners and 3rd party organizations, adding a whole new level of complexity and challenges requiring detailed analysis.

Finally, industrial settings demand quick responses to issues and management of assets and users. Industries usually invest manpower and time into roles targeted to solve those issues and avoid stopping or delaying production. This challenge brings attention to specific properties that are considered crucial for industrial data sharing.

When considering data protection in industrial settings, this thesis focus on exploring approaches and mechanisms to cover the authorization and trustworthy communication aspects of the above-mentioned IDS requirements.

Q2 *How can fine granular access control policy creation and maintenance be automated and how can those policies be efficiently enforced for secure IoT data sharing?*

As mentioned above, industrial environments are dynamic in nature, which challenges the access policy control and maintenance. Consequently, one of the main deterrents to implementing a secure fine-granular access control model for industry applications is the administrative task of creating secure policies, roles, and data ownership definitions. Furthermore, the manual creation and management of policies is error prone where mistakes could potentially compromise the privacy of sensitive data and harm the data owner.

Automating the creation and maintenance of policies could facilitate the adoption of modern and secure access control schemes such as attribute-based approaches and is thus important.

Industrial data is often characterized by its big number of data elements. Therefore, the enforcement of fine-granular policies over industrial IoT data presents some challenges. The traditional approaches of enforcing the access control policies to each data element at the desired granularity level are often not efficient in these types of environments.

1.3 Research methodology

For this thesis, the Experimental computer science and engineering (ECSE) methodology was used. This methodology is based on the creation and experimentation of computational artifacts to better understand the computational processes and algorithms involved in the computational phenomenon. This methodology defends that when computational processes are implemented as an artifact, the behavior and interaction of its components can be seen in action[7].

The work was developed in two sequential stages, where the first stage addressed a knowledge gap identified by an industry partner through a use-case in the Arrowhead Tools project. Through an iterative process and interaction with the industry partner, a set of requirements was defined. These requirements motivated the creation of the first artifact used to explore ideas looking for answers to **Q2**, which lead to Paper A and Paper B. Moreover, thanks to the collaborations with additional partners, like the construction industry, additional requirements were explored leading to Paper C.

The analysis of gaps between the proposed solutions and the initial requirements inspired a second stage of the research, the literature review presented in Paper D. The goals of the literature review were threefold: further exploration of the gaps found in the first stage related to **Q1**, a proposal and analysis of a conceptual solution, and lastly the identification of open knowledge gaps in the context of the proposed solution.

The gap analysis of the state-of-the-art solutions and our own proposed idea lead to the requirement definition for various computational artifacts to iterate in the methodology. Thanks to collaboration with PhD colleges, students, and project groups, some of these artifacts have been built with the aim of building more complex proof-of-concepts.

1.4 Thesis outline

This compilation thesis is structured in two parts. The aim of Part I is to set the context and motivation for the work presented in the appended papers, this includes setting the background and presenting a compiled summary. Part I comprises four chapters: Chapter 1 sets the motivation and problem formulation for the work as well as defining the guiding research questions, Chapter 2 presents the theory and background used to solve the research questions, Chapter 3 describes the contributions made in each of the appended papers, and Chapter 4 concludes this thesis work and presents opportunities for future work.

Chapter 2

Research Background

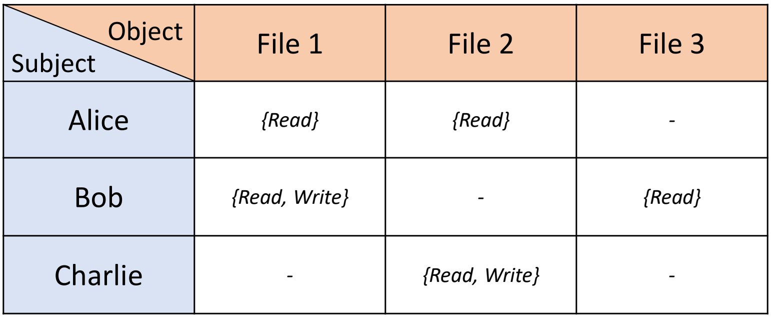
2.1 System of systems (SoS) and Service-Oriented Architecture (SOA)

As part of the advancements in industrial processes, the concept of systems composed of another individual and autonomous systems collaborating towards a mutual goal is emerging. This concept is called System of Systems (SoS) [8]. The systems inside a SoS can range from a sensor device in the manufacturing plant, to an Enterprise Resource Planning (ERP) node.

The interaction of systems inside a SoS is made easier thanks to frameworks such as the Reference Architectural Model for Industry (RAMI) 4.0, following a Service Oriented Architecture (SOA) style. In SOA, systems have their functionalities and data endpoints offered as a service through an interface. Moreover, they have loose coupling properties in that there is no need for services to know about other services at design time, they rather learn about the needed services with the aid of some form of orchestration system [9]. One of such frameworks is the Eclipse Arrowhead Framework. This framework allows systems to consume and provide services providing real time runtime, security and interoperability features with the support of three core systems: Authorization, Orchestration, and Authentication [10]. It is important to note that the authorization features of this framework are limited to service consumption authorization.

2.2 Access Control

Access control is a paradigm existing almost since the beginning of computer history. Its purpose is to limit the actions that users are allowed to perform in a computer system. It concerns actions such as executing certain programs, accessing folders, reading files, or even accessing memory addresses inside a system. In general, the goal of access control is to prevent any activity that may lead to security breaches. In the early days of access



Object Subject	File 1	File 2	File 3
Alice	*{Read}*	*{Read}*	-
Bob	*{Read, Write}*	-	*{Read}*
Charlie	-	*{Read, Write}*	-

Figure 2.1: Example access matrix for a file sharing system

control, this protection usually meant limiting the memory address that a given user or process could interact with, often in an all-or-nothing fashion. These early approaches provide full isolation, but fine-granular sharing of memory is not possible [11].

With the aim of providing more fine-granular methods to achieve access control, the reference monitor concept was proposed in 1972. The goal of the monitor reference is to validate all references by any program to any program, data, or service to a list of authorized types of reference based on the user or program function. The monitor reference model is known as a subject-object model, where subjects are entities trying to get access to objects. To achieve this, an authorization list for each subject is needed, describing the authorization capabilities to each object. For each subject-object pair, an authorized set of permissions can be defined, e.g. read, write, delete. This authorization list is modeled as an *Access Matrix*, with subjects as rows, objects as columns, and access permissions in each cell as shown in Figure 2.1. The idea of the reference monitor was to intercept access attempts a to a monitored object o from any user u, and evaluate the entry (u,a,o) in the access matrix, allowing the access only if authorized [11].

In a Mandatory Access Control (MAC) system, it is the administrator of the system who sets the confidentiality level of objects and the clearance level of users. This means that in a MAC system, the data owners do not have direct control over the policies regarding their data. This is useful for organization-wide policies, like in government or military. However, this hinders the capability of data owners to dynamically set policies for their data [12]. Discretionary Access Control (DAC) differs from MAC in the sense that it is the data owner of a resource who has control under those resource's policies. In other words, this model leaves at the discretion of the data owner which users have access to data and can pass those permissions. This model makes use of the access matrix described previously to define the policies [13].

Role-based Access Control (RBAC) is proposed as a way to enforce both MAC and DAC policies, using the notion of roles. Roles are inspired by the fact that in many organizations the access control decisions are taken based on the role of the user following an organizational chart. These roles can be used as groups of users who perform similar jobs within the organization and hence require similar access rights. Assigning access

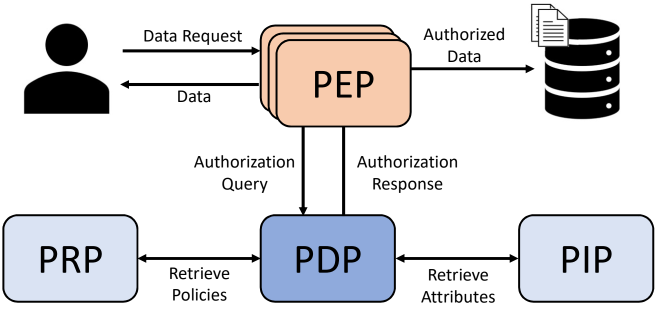


Figure 2.2: ABAC basic architecture

rights to a role rather than to a user provides multiple management advantages. One of the main advantages is that the responsibilities associated with a specific role are often static, whereas the user's membership to a role tends to change over time. Moreover, the same role hierarchy existing in an organizational chart can be translated into role hierarchy inside a RBAC model. RBAC role hierarchy allows for higher authority roles (senior) to inherit access rights from those below it (junior) [12].

These advantages make RBAC a good approach for managing access control in a moderate-size organization with static policies. However, managing a RBAC scheme in a large heterogeneous setting can be a challenging task. Role explosion issues are common in large organizations whose users perform several specific tasks, making the amount of roles and users unmanageable. Such disadvantages motivate more flexible and manageable access control mechanisms that can handle fine-granular policies and can provide dynamic properties desired in previously described industrial settings [14].

Attribute-based access control (ABAC) emerges as a more flexible alternative to RBAC, where access control policies are now defined based on attributes rather than identity or role. Attribute are characteristics or properties of the elements involved in the access process. In other words, attributes describe the users, objects and context at the time an access event is given. These can, for example, describe the user's age, department, or job title; or describe the resource type, format, or date of creation. Moreover, attributes describing the context in which the access is requested can be used in policies, such as time of the day, day of the week, or location. The flexibility of attributes allows for the expression of MAC and DAC policies similarly to RBAC, but also provides additional levels of abstraction improving the expressiveness and granularity of policies [15]. An in-depth description of the ABAC model is presented in Paper D.

The discussed access control models work by intercepting the access request from the user, evaluating the relevant policies, and performing the correspondent action depending on the authorization decision. This process is illustrated in Figure 1. The intermediary element between the users and the data is called Policy Enforcement Point (PEP). There can be multiple PEPs connected to the same Policy Decision Point (PDP) in a system, this may be advantageous for reasons such as the need for application-specific PEPs, or

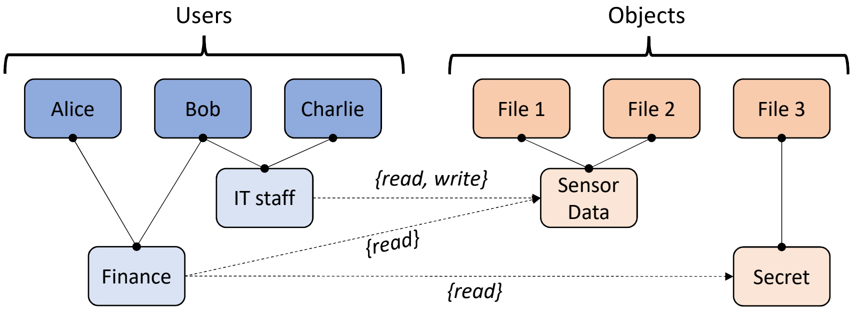


Figure 2.3: Example NGAC policy graph

to reduce single points of failure. To get an authorization decision the PEP sends an authorization query to the PDP, which has access to the stored policies via the Policy Retrieval Point (PRP), and attributes via the Policy Information Point (PIP). If, and only if, the user is authorized to access the requested data, the PEP will allow the request to go through and it will send the requested data back to the user.

A critical element needed to take advantage of the properties provided by ABAC is the access control policy language in which the policies are defined. Standards have been created to provide generic and standardized ways to express and enforce access control policies. One of the most common is eXtensible Access Control Markup Language (XACML), proposed in 2002 by the Organization for the Advancement of Structured Information Standards (OASIS). XACML is a XML-based language designed to provide mechanisms for expression, querying, and enforcement of access control policies [16]. In a similar way, the Security Assertion Markup Language (SAML) standard uses a XML-based language to provide users with authorization tokens to consume services [17].

Next Generation Access Control is another standard policy language gaining popularity in ABAC applications. Proposed by the National Institute of Standards and Technology (NIST) [18], it is designed as a relation-based flexible policy language. In NGAC policies are expressed as relations between user attributes and object attributes, where the connecting link contains the access rights associated with that relation, as illustrated in Figure 2.3 . The additional level of abstraction provided by the graph representation allows for a very flexible language proper for complex environments. Moreover, NGAC provides event processing features in the form of Obligations. Obligations can be seen as actions or responses that are triggered by a specified event. These responses contain a set of administrative operations, such as the creation of new relations, that are executed just after the event happens. The usage of obligations can prevent sensitive data leakage or enable history-based policies [19].

Access control in databases is usually managed at application level. Some database engines such as MySQL [20] or InfluxDB [21] includes some degree of authorization tools as part of their features. However, the level of granularity provided is often limited to individual tables or even databases. Moreover, particularly in time-series databases, it is challenging to setup fine-grained access control, mainly due to the large number of rows usually found in those types of databases.

2.3 Cloud Computing and Cloud Servers

As part of the set of emerging technologies that accompany Industry 4.0, cloud computing environments are offering convenient, scalable, agile, and reliable IT resources [22]. One of the main goals of cloud computing is to reduce the processing burden on the user side, delegating tasks to dedicated systems often located remotely over the network. Third party solutions such as Amazon AWS [23] or Microsoft Azure [24] offer high availability mechanisms for data management and storage. Taking advantage of this technology allows industries to enhance collaboration, availability, agility, and scalability while reducing up-front and management costs. Scalability in this context is of particular importance for industrial applications, it means that computational resources such as memory or processing power can dynamically increase as needed. Additionally, one can also distribute the resources horizontally for added reliability and availability [25]. However, the usage of third party cloud servers for data storage raises its own security concerns as the data owners lose full control of it while in transit or inside untrusted or trusted but curious servers. These security considerations are important deterrents for potential industry adopters. Hence, ensuring privacy, confidentiality, and data integrity of the data both in transit and at rest is critical [26, 27, 28].

2.4 Encryption in data sharing

To ensure the privacy of data in transit, mechanisms such as encryption are usually used. There are two basic categories of encryption algorithms: symmetric encryption, and asymmetric encryption. In symmetric encryption, the encryption and decryption key are the same, this key is known as the secret key. On the other hand, in asymmetric encryption, also known as public-key encryption, each party has a key pair of an encryption public key, and a decryption private key. The public key is, as the name implies, public and it is used by other users to encrypt messages that only the owner of the public key can decrypt, by using its private key.

Those basic encryption building blocks can be combined in mechanisms such as OpenPGP security standard [29]. Pretty Good Privacy (PGP) uses symmetric-key and public-key encryption mechanisms to guarantee data privacy while sending and receiving data, and it is widely used in tasks such as texts, emails, or file sharing applications. The general process when preparing data to be sent is to encrypt the message with a random key, called session key, to create a ciphertext which is then packaged alongside the session key encrypted with the receiver's public key. This package is then sent to the final user which uses its own private key to get the session key, to finally get the underlying message [29].

Identity-based Encryption (IBE) emerges as a public-key scheme where instead of using public keys from users to encrypt messages, it uses some unique identifier for that user. This identifier can be strings such as its social security number, email address, or phone number. IBE allows data owners to encrypt data without explicitly knowing the receiving user's public key, simplifying tasks such as managing keys and guaranteeing their integrity and authenticity [30]. Attribute-based Encryption (ABE) is a public-key encryption mechanism targeted at groups rather than individual users firstly proposed by Sahai, et al. [31]. ABE follows the same design principle as IBE in that the data owner does not need to know the public keys of the targeted users. Rather, in ABE the data owner defines a set of attributes and use it to encrypt a message, only the users whose attributes overlap with those in the encryption's set can get the underlying message.

In ABE, the access policy under which the data is encrypted is called access structure. The access structure contains the set of authorized attributes, and can be represented as access trees. In an access tree, each node represents an attribute and non-leaf nodes represent threshold gates. If the attributes assigned to a user satisfy the access tree, that means that the user is allowed to decrypt the ciphertext. Where to embed the access structure and where to associate the attributes depends on the two main ABE schemes: Key-policy ABE (KP-ABE) and Ciphertext-policy ABE (CP-ABE). In KP-ABE the user's private keys contain the access structure while each ciphertext is associated with a set of attributes. In CP-ABE each user's private key is labeled with its associated attributes, while messages are encrypted with their related access structure [32].

2.5 ABAC-enabled ABE

In paper D the implications and possible approaches for a combination of ABAC with ABE are discussed. By the nature of how ABE access structures are expressed, the policy management and flexibility are negatively affected. ABAC-enabled ABE is an emerging idea targeted toward covering the weaknesses of each individual mechanism. Some of the main design goals for ABAC-enabled ABE are to provide flexibility to the expression of access policies over encrypted data while keeping a small storage overhead, and to keep the management related tasks as simple as possible while allowing for efficient access revocation.

Chapter 3

Contributions

3.1 Paper A

Title	Access control model for Time Series Databases using NGAC
Authors	Alex Chiquito, Ulf Bodin, Olov Schelén
Summary	In paper A a proof of concept architecture for enforcing fine-granular access control over time series data in industrial settings is proposed. For this architecture we chose an Attribute-based Access Control (ABAC) model using Next-Generation Access Control (NGAC) policy language. Four main requirements for industrial data sharing were defined to evaluate the effectiveness of our solution: 1) it should be capable of efficient data selection, 2) it should be possible to add and remove data sources in run-time, 3) it should be capable of managing and enforcing access control policies with different levels of granularity, and finally 4) it should be easy to properly maintain and potentially automate. The proposed architecture concept is based on query rewriting to achieve fine-granularity. Moreover, the idea of an on-boarding procedure is presented for sensors entering the system.
Motivation	This work was conceptualized as an extension to the Next-generation Database Access Control (NDAC) proposed by Ferraiolo et al[33] for its applications over industrial time-series data.
Published in	2020 International Conference on Emerging Technologies and Factory Automation (ETFA)
Contribution	The requirements and problem formulation were discussed with Ulf Bodin and Olov Schelén. I wrote a first draft, and the analysis was a collaborative effort of all the authors.

3.2 Paper B

Title Fine-grained Access Control for Time-series Databases using NGAC

Authors Alex Chiquito, Ulf Bodin, Olov Schelén

Summary In paper B, we formalize an architecture designed to express and enforce row-level policies while keeping a minimal performance impact, as an extension of the work done in paper A. To achieve row-level policies we propose two NGAC extension approaches based on storing filter strings as part of the policy definition. These filter strings are used in a later query rewriting step before retrieving data from the database. The first of those approaches take advantage of obligations and the event processing capabilities of NGAC. The second approach extends the graph assignments to allow the inclusion of filter strings as part of the authorized access rights list. Finally, this paper presents a quantitative and qualitative analysis of the two described approaches.

Published in 2021 IEEE 19th International Conference on Industrial Informatics (INDIN)

Contribution I made the code artifact, performed the evaluations, and wrote the first draft of the paper. The requirements, problem formulation, findings, and final analysis were discussed with Ulf Bodin and Olov Schelén.

3.3 Paper C

Title Application-scoped Access Control for the Construction Industry

Authors Ulf Bodin, André Christoffersson, Alex Chiquito, Johan Rodahl, Kåre Synnes

Summary This paper explores the use of delegated access using OAuth 2.0 and Attribute-based Access Control (ABAC) for the collaborative sharing of equipment at construction sites in a Service-Oriented Architecture environment. This paper proposes an IoT Application-scoped Access Control as a Service (IAACaaS) architecture model with capabilities of considering contextual attributes such as location, or urgency when authorizing booking construction lifts. This feature is supported by implementing and taking advantage of Next-generation Access Control obligations. Moreover, this work presents a guide to implement the proposed functionalities as reusable micro-services.

Published in 2021 International Conference on Emerging Technologies and Factory Automation (ETFA)

Contribution I contributed writing the guide for implementation, as well as providing expertise in ABAC and NGAC to conceptualize and describe the architecture. Moreover, I contributed in the problem description discussions.

3.4 Paper D

Title Attribute-Based Approaches for Secure Data Sharing in Industrial Contexts

Authors Alex Chiquito, Ulf Bodin, Olov Schelén

Summary Paper D defines key properties needed for secure data sharing in industrial settings, and investigates the use of attribute-based approaches to achieve them. Moreover, this work analyses Attribute-Based Access Control (ABAC) approaches applied in trusted domains to identify concepts that can be extended to work in untrusted cloud domains using encryption mechanisms. In paper D, a survey of Attribute-based Encryption (ABE) schemes is presented and a comparison against the previously described properties is performed to identify research gaps. Given the concepts identified from ABAC and the research gaps discussed in the ABE survey work, an ABAC-enabled ABE architecture concept is presented alongside an analysis of the similar related work. Finally, the knowledge gaps from the proposed architecture concept are discussed and paths for future work are presented.

Contribution I performed the literature review, analysis, and wrote a first draft of the paper. The problem description, key properties, and findings were discussed in collaboration with Ulf Bodin and Olov Schelén

Chapter 4

Conclusions

"I knew exactly what to do.
But in a much more real sense, I had no idea what to do."

-Michael Scott

The work in this thesis shows that Attributes-based approaches such as ABAC and ABE are able to provide mechanisms to define and enforce flexible, fine-granular, and dynamic policies while being easy to maintain. ABAC concepts such as contextual and environmental attributes provide important functionalities for potential enhancements in industrial use cases. Furthermore, we explore initial approaches for automation of access control policies. So, the essence of this work is to explore attribute-based mechanisms and approaches for secure data sharing inside industry environments.

Additional requirements based on industry use cases are discussed, leading to the definition of formal requirements for secure data sharing. The appended papers present architectural approaches using ABAC with the NGAC policy language to enforce fine-granular policies are presented and evaluated, covering a selected set of requirements. Moreover, a literature study of data sharing approaches in this context allows for the identification of knowledge gaps, and is complemented with an architecture concept proposed to cover the entirety of the identified requirements. The conceptual architecture presents additional knowledge gaps needed to be covered as part of future work.

4.1 Research Questions

In this section, the research questions are revised and answered based on the work presented in this thesis and the appended papers.

Q1 *How can data protection be provided in industrial settings and what are the essential properties in achieving such protection?*

This work considers two important aspects when providing data protection, the first being able to enforce access control policies to data access requests, and the second being able to guarantee the privacy of the data while at rest and in transit.

Given initial input from industry use-cases, paper A and B present initial concept approaches designed to enforce access control policies to data requests. In these initial approaches, the fine-granular and dynamic properties of access control are explored and NGAC is proposed as an appropriate policy language for this task. Moreover, given the nature of industrial environments, the importance of ease of management is considered early in this work. Furthermore, the use-case presented in Paper C further strengthens the motivation for dynamic access control and endorses the usage of NGAC for this type of application.

Considering the emergence of cloud services technologies, the second aspect of data protection is considered for Paper D. Paper D explores the available mechanisms to add transit and rest privacy to data sharing schemes, and further discusses the set of key properties needed for industrial data sharing. Attribute-based Encryption (ABE) is proposed to add cryptographic properties to data sharing processes. These properties allow the protection of data privacy in not-fully-trusted environments, such as commercial cloud servers, contributing to the trustworthy communication aspect discussed as part of the IDS requirements. An analysis of the state of the art is presented considering our proposed set of key properties and the research gaps are discussed. The most important challenge with ABE schemes is the large setup work needed to implement them in a complex heterogeneous environment. Moreover, the lack of expressive and dynamic ABE policy languages, combined with the administrative burden associated with ABE policy management motivates the research of additional mechanisms. Finally, this work explores the implications of combining ABAC schemes with ABE properties to cover both aspects. A survey of the approaches for this combination is presented and a concept architecture is proposed.

Q2 *How can fine granular access control policy creation and maintenance be automated and how can those policies be efficiently enforced for secure IoT data sharing?*

Attribute-based access control is proposed as an effective way to guarantee the authorized flow of data in industrial contexts. The findings of this work show that ABAC provides flexible mechanisms to express and enforce access policies in industrial use cases. In paper A and B initial approaches using NGAC policy language are presented. Moreover, an extension for NGAC is proposed designed to express and enforce fine-granular policies for time-series data. This extension is proposed in Paper A and two approaches for implementation are presented and evaluated in Paper B. The results suggest that the enforcement of fine-grained policies using NGAC is possible while not increasing the number of policies or adding important computational overhead. By exploiting NGAC abstract policy definition and event processing capabilities, the proposed and evaluated architectures set a ground for access control of users' data request.

Paper A outlines an onboarding procedure for devices entering an IIoT environment. The goal of this procedure is to automate the creation of the new device's policies by automatically creating the needed object and user attributes based on its metadata. This procedure is particularly useful for the management of Service-Oriented Architectures (SOA) settings such as an Arrowhead local cloud.

4.2 Reflections

The work presented in this thesis focuses on the NIST Next Generation Access Control (NGAC) and Attribute Based Encryption (ABE). This choice of focus was made based on the assumption that NGAC would provide superior abstractions compared to the eXtensible Access Control Markup Language (XACML), and thus allow for easier policy management. This decision was later backed up with related literature, as NGAC presents important advantages for our set of requirements. The contributions in Paper A were partly motivated because standard NGAC lacked some important features that other policy languages had, such as value-constrained policies. At that time, the evaluation of available alternatives suggested that investing research time into NGAC to add the missing features was the better path, as the policy graph abstraction and flexibility features were valuable. However, experiments applying a different industry standard policy language such as XACML or SAML have not been done to justify such evaluation.

The analysis in this thesis confirms that encryption properties are of key importance in providing privacy for outsourced data. Paper D explores attribute-based encryption (ABE) as an alternative to traditional access control. The paper shows that ABE lacks the flexibility and ease of management required in an industrial setting. Nevertheless, ABE brings properties considered important such as allowing data owners to encrypt data once using attributes, even before the target users get their keys. However, related literature's' approaches to solve the challenges have mainly focused on the mathematical components of encryption schemes, specifically on creating more efficient and expressive access structures. The mathematical formulation of encryption schemes has been out of scope for this work, which instead has focused on architectural and usability aspects.

4.3 Ethical Considerations

The focus of this work is to create easy to implement and maintain mechanisms to securely share data in industrial settings. As stated in the introduction to this work, data is a valuable resource for individuals and organizations. Access to the right data could lead to important economical gains or competitive advantages, but it may also lead to unfair business practices. Moreover, personal data is protected must be protected according to EU's General Data Protection Regulation (GDPR), and it is in any case of interest for individuals to give consent on how their data is used. For researchers in this area, is of utmost importance to consider any ethical considerations that may arise, and to guarantee the strict confidentiality, secrecy, and privacy of data used for research. There is also a responsibility towards data owners to take every measure possible to combat security breaches. No measures against security breaches should be dropped for any reason such as efficiency or cost-reduction.

Moreover, fairness and justice considerations may arise in view of the level of information access any particular organization or individual may have over their competitors. Data sharing mechanisms within data driven ecosystems should not endorse or facilitate any type of monopolistic or unethical business practices. These can be such as unfairly benefiting the biggest or wealthiest participant, or limiting the level of information access of new competitors to a point where it is impossible for them to generate value [34].

In addition, it is important to provide accountability and traceability measures to link users' actions with their consequences. Nevertheless, further ethical considerations arise from surveying users and potentially vulnerating users' privacy.

4.4 Future and Ongoing Work

Based on the findings and knowledge gaps of this work, there are some tasks ongoing and a range of possible future work for the research.

4.4.1 Ongoing Work

1. **Implementation of a policy server system in a SoA architecture**: As a contribution to the Arrowhead Tools project, the implementation of an NGAC-based policy server system is being developed. This system is planned to serve as a policy decision point for local cloud wide policies. Moreover, an application library for policy enforcement is being built to interact with the policy server, and a graphical Policy Administration Point (PAP) front-end was created to simplify the creation and management of policies. The goal of these systems and libraries is to provide a centralized mechanism for creation, management, and enforcement of data sharing access control policies inside an Eclipse Arrowhead local cloud.

2. **Contractual automation of access control policies using Ricardian contracts**: The application of the access control policy automation concept discussed in Paper A is being explored in the area of contractual negotiations. Given a successful negotiation between a data provider and a data consumer, the creation of access control policies for the just acquired datasets is automated. For this application, the NGAC policy language was chosen thanks to its flexibility. These policies are created following the data embedded within the digital Ricardian contracts[1] generated by the negotiation system just after the deal is signed by the involved

[1] I. Grigg, The ricardian contract. URL (accessed 2022-05-08): https://doi.org/10.1109/WEC.2004.1319505

parties. A research paper is being written, and further work applying this concept to a SoA framework such as Eclipse Arrowhead is planned.

4.4.2 Future Work

1. **Generation of ABE access structures from ABAC policies**: As discussed in Paper D, a mechanism to get appropriate access structures and attribute lists for encryption and key generation process from ABAC policies is still an open research gap. This process is key for granting ABAC's expressiveness and flexibility properties to ABE. The generation of such cryptographic parameters is planned to be achieved by exploiting the processing capabilities of the PDP in combination with the expressiveness of the graph-based NGAC policy language. This design would allow data owners to enforce cryptographic access control to their data while keeping the policy management benefits of ABAC systems.

2. **ABE model evaluation for ABAC-enabled ABE architecture**: The ABAC-enabled ABE conceptual architecture described in Paper D requires quantitative and qualitative evaluations to formalize an optimal configuration for industrial data sharing. Whether to use Ciphertext-policy ABE (CP-ABE) or Key-policy ABE (KP-ABE) model for the specific use case is still an open question. Important performance implications may arise from the model of choice, in addition to qualitative trade-offs for the data owner and users both during run-time and in the initial implementation.

References

[1] H. Lasi, P. Fettke, H.-G. Kemper, T. Feld, and M. Hoffmann, "Industry 4.0," *Business & information systems engineering*, vol. 6, no. 4, pp. 239–242, 2014.

[2] International Data Space (IDS), "International data space - reference architecture model," 2019.

[3] "What is considered personal data under the eu gdpr?," Feb 2019.

[4] S. N. I. A. (SNIA), "What is data protection?."

[5] Arrowhead, "Eclipse arrowhead™," 2022.

[6] projects.eclipse.org, "Eclipse arrowhead," Feb 2022.

[7] C. Science and C. Telecommunications Board, "Academic careers for experimental computer scientists and engineers," *Communications of the ACM*, vol. 37, no. 4, pp. 87–90, 1994.

[8] M. W. Maier, "Architecting principles for systems-of-systems," *Systems Engineering: The Journal of the International Council on Systems Engineering*, vol. 1, no. 4, pp. 267–284, 1998.

[9] M. Rosen, B. Lublinsky, K. T. Smith, and M. J. Balcer, *Applied SOA: service-oriented architecture and design strategies.* John Wiley & Sons, 2012.

[10] C. Paniagua, *Autonomous Runtime System of Systems Interoperability.* PhD thesis, Luleå University of Technology, 2020.

[11] P. C. van Oorschot, *Computer Security and the Internet.* Springer, 2020.

[12] P. A. Loscocco, S. D. Smalley, P. A. Muckelbauer, R. C. Taylor, S. J. Turner, and J. F. Farrell, "The inevitability of failure: The flawed assumption of security in modern computing environments," in *In Proceedings of the 21st national information systems security conference*, Citeseer, 1998.

[13] C. S. Jordan, *Guide to Understanding Discretionary Access Control in Trusted Systems.* DIANE Publishing, 1987.

[14] A. Elliott and S. Knight, "Role explosion: Acknowledging the problem.," in *Software Engineering research and practice*, pp. 349–355, Citeseer, 2010.

[15] D. Servos and S. L. Osborn, "Current research and open problems in attribute-based access control," *ACM Computing Surveys (CSUR)*, vol. 49, no. 4, pp. 1–45, 2017.

[16] S. Godik and T. Moses, "Oasis extensible access control markup language (xacml)," *OASIS Committee Secification cs-xacml-specification-1.0*, 2002.

[17] J. Hughes and E. Maler, "Security assertion markup language (saml) v2. 0 technical overview," *OASIS SSTC Working Draft sstc-saml-tech-overview-2.0-draft-08*, vol. 13, 2005.

[18] D. F. Ferraiolo, L. Feldman, and G. A. Witte, "Exploring the next generation of access control methodologies," tech. rep., 2016.

[19] D. Ferraiolo, R. Chandramouli, D. Kuhn, and V. Hu, "Extensible access control markup language (xacml) and next generation access control (ngac)," pp. 13–24, 03 2016.

[20] A. MySQL, "Mysql 8.0 reference manual," 2018.

[21] Influxdata, "Influx 1.8 documentation," 2021.

[22] T. Dillon, C. Wu, and E. Chang, "Cloud computing: Issues and challenges," in *2010 24th IEEE International Conference on Advanced Information Networking and Applications*, pp. 27–33, 2010.

[23] A. E. C. C. U. Guide, "Amazon web services," *Inc. Dec*, 2012.

[24] M. Copeland, J. Soh, A. Puca, M. Manning, and D. Gollob, "Microsoft azure," *New York, NY, USA:: Apress*, 2015.

[25] D. Dempsey, F. Kelliher, *et al.*, "Industry trends in cloud computing,"

[26] C. Alliance, "Security guidance for critical areas of focus in cloud computing v3. 0," *Cloud Security Alliance*, vol. 15, pp. 1–176, 2011.

[27] L. Wei, H. Zhu, Z. Cao, X. Dong, W. Jia, Y. Chen, and A. V. Vasilakos, "Security and privacy for storage and computation in cloud computing," *Information Sciences*, vol. 258, pp. 371–386, 2014.

[28] S. Kamara and K. Lauter, "Cryptographic cloud storage," in *International Conference on Financial Cryptography and Data Security*, pp. 136–149, Springer, 2010.

[29] J. Callas, L. Donnerhacke, H. Finney, D. Shaw, and R. Thayer, "Openpgp message format," tech. rep., 2007.

[30] D. Huang, Q. Dong, and Y. Zhu, *Attribute-Based Encryption and Access Control.* CRC Press, 2020.

[31] A. Sahai and B. Waters, “Fuzzy identity-based encryption,” in *Advances in Cryptology – EUROCRYPT 2005* (R. Cramer, ed.), (Berlin, Heidelberg), pp. 457–473, Springer Berlin Heidelberg, 2005.

[32] J. Bethencourt, A. Sahai, and B. Waters, “Ciphertext-policy attribute-based encryption,” in *2007 IEEE symposium on security and privacy (SP’07)*, pp. 321–334, IEEE, 2007.

[33] D. Ferraiolo, S. Gavrila, G. Katwala, and J. Roberts, “Imposing fine-grain next generation access control over database queries,” pp. 9–15, 03 2017.

[34] M. Christen, B. Gordijn, and M. Loi, *The ethics of cybersecurity.* Springer Nature, 2020.

Department of Department of Computer Science, Electrical and Space Engineering
Division of EISLAB

ISSN 1402-1757
ISBN 978-91-8048-092-5 (print)
ISBN 978-91-8048-093-2 (pdf)

Luleå University of Technology 2022

FSC
www.fsc.org
MIX
Papper från
ansvarsfulla källor
FSC® C010493

SVANENMÄRKET
Trycksak
3041 3042

Print: Lenanders Grafiska, 428768

LULEÅ
UNIVERSITY
OF TECHNOLOGY