Available online at www.sciencedirect.com

ScienceDirect

Procedia Computer Science 68 (2015) 116 – 126

Procedia Computer Science

HOLACONF - Cloud Forward: From Distributed to Complete Computing

Secure Data Sharing and Processing in Heterogeneous Clouds

Bojan Suzic[a]*, Andreas Reiter[a], Florian Reimair[a], Daniele Venturi[b], Baldur Kubo[c]

[a] *Institute for Applied Information Processing and Communications, Inffeldgasse 16a, 8010 Graz, Austria*
[b] *Sapienza Università di Roma, Via Salaria 113 III Piano, 00198 Roma, Italy*
[c] *Cybernetica AS, Mäealuse 2/1, 12618 Tallinn, Estonia*

Abstract

The extensive cloud adoption among the European Public Sector Players empowered them to own and operate a range of cloud infrastructures. These deployments vary both in the size and capabilities, as well as in the range of employed technologies and processes. The public sector, however, lacks the necessary technology to enable effective, interoperable and secure integration of a multitude of its computing clouds and services. In this work we focus on the federation of private clouds and the approaches that enable secure data sharing and processing among the collaborating infrastructures and services of public entities. We investigate the aspects of access control, data and security policy languages, as well as cryptographic approaches that enable fine-grained security and data processing in semi-trusted environments. We identify the main challenges and frame the future work that serve as an enabler of interoperability among heterogeneous infrastructures and services. Our goal is to enable both security and legal conformance as well as to facilitate transparency, privacy and effectivity of private cloud federations for the public sector needs.

Peer-review under responsibility of Institute of Communication and Computer Systems.

Keywords: federated clouds; inter-cloud; authorization federation; security policy; access control; proxy re-encryption; attribute-based encryption

1. Introduction

The emergence of cloud services provided many benefits to organizations, including the possibility to consolidate infrastructures and allow for structured and efficient deployment of services. The public sector entities among the European Union adopted this wave by deploying new and transforming existing infrastructures to utilize the benefits of cloud technologies. However, the fact that public sector consists of many and dispersed institutions that operate cloud infrastructures prevents reaching the full level of efficiency in deployment and usage of these resources.

In this work we approach the challenge of federating private clouds among public sector institutions, considering

* Corresponding author. Tel.: +43 (316) 873 - 5553; *E-mail address:* bojan.suzic@iaik.tugraz.at

Peer-review under responsibility of Institute of Communication and Computer Systems.
doi:10.1016/j.procs.2015.09.228

the diverse characteristics of their infrastructures, heterogeneous environments and cross-organizational, cross-border (EU) context. The federation of private clouds should not only ensure the interoperability of various layers of infrastructures and collaborations. It should also comply with the range of security, legal and assurance requirements that deal with the data sharing and processing, satisfying particular requirements of public administrations. In the scope of this work therefore we focus on main enablers that will facilitate the secure and transparent federation of private clouds in the public sector. These include models and architectures for access control in distributed environments, data and security policy languages and cryptographic methods and services that support secure and efficient storage and processing of data in distributed and semi-trusted systems.

The structure of this paper is as follows. First, we elaborate on relevant initiatives and frameworks. Then we review access control approaches and architectures that enable data sharing and interoperability in connected clouds. The fourth chapter continues in the similar direction by dealing with security policy definitions, languages and enforcement, with the emphasis on fine-granularity, expressiveness and cross-system interoperability. The fifth chapter analyzes encryption methods that will serve as a basis for secure processing in inter-clouds. Finally, with each chapter we discuss the state-of-the-art and frame future work that will facilitate the goal of this project.

2. Background

During the previous years, there have been launched several initiatives supported by EU or its member states, with the aim to research and advance different aspects of cloud computing. The FP7 projects such as REMICS, NOVI, CONTRAIL, 4CaaSt and CHOReOS focused on the architectural dimensions of future clouds, aiming to provide better management facilities, improve the scalability and orchestration of the systems or support the transition from legacy systems. The SLA@SOI project provided the framework and tools to facilitate the implementation of service level agreements for service-oriented systems. Projects such as TClouds and PRACTICE aimed to build secure frameworks and infrastructures that deliver new levels of security, evaluated on specific use-cases.

The perspectives of security management have been investigated by other FP7 projects such as PRISM, DEMONS, CoMiFin and INTERSECTION, which focused on security monitoring in complex systems, including networking level, infrastructure protection and incident response management. PrimeLife and SWIFT aimed at privacy and trust related challenges, particularly in the scope of identity management and user-centric data protection. POSITIF, TAS3 and POSECCO approached the security from the management point of view, dealing with the decision support architectures, adaptive security services and dynamic security policy protection and management. The topics of interoperability and trust in heterogeneous systems are approached by the more recent project such as InterTrust and A4Cloud, which focus on dynamic secure policies and accountability in the cloud.

In this paper we present an ongoing work in the scope of SUNFISH, a project supported by EU Horizon 2020 R&I Framework Programme. Our work aims to investigate the challenge of data sharing and processing in the scope of private cloud federations among public administration entities, focusing on security and legal compliance. This initiative differs from the previous work in the sense that it reaches the problem considering diverse infrastructures and entities, focusing on security and legal requirements of the public sector not being previously addressed.

We aim to enable security policy management and processing both on the levels of organization, service and data, which will facilitate purposeful, accountable and traceable data sharing, transformation and processing among the institutions. Our approach will provide a framework and tools and perform their validation in a range of practical use cases that involve public authorities, inter-organizational and cross-border collaboration on different service and cloud layers.

3. Federating Access Control and Data Sharing in Connected Clouds

Access control, as a form of authorization, enforces confidentiality and integrity of data, or controls the usage of a service. An access control system consists of a collection of methods and components that determine and enforce the correct admission to activities by the users that are deemed legitimate for the system (Samarati & Di Vimercati).

In this chapter we identify relevant concepts and present a state-of-the art in the access control approaches that enable secure data sharing in interconnected systems. We particularly focus on models and architectures that support

the description of structures that reflect the complex organizational landscape and diverse requirements of public bodies. Finally, we analyze the challenges and propose the future work that intends to bridge the identified gaps.

3.1. Access control models

Traditional access control models that gained broad adoption include Mandatory Access Control (MAC), Discretionary-based Access Control (DAC) and Role-based Access Control (RBAC) (Sandhu & Samarati, 1994). While MAC and DAC each consider access capabilities from the perspectives of organization or the user which owns the data, RBAC builds on centralized MAC view, introducing the concept of *roles* derived from the notion of organizational duties. The standardized Core RBAC defines the concepts of *users*, *roles*, *objects*, *operations*, *permissions* and *sessions* (INCITS, ANSI, 2004).

OrBAC model by El Kalam et al. (2003) extends RBAC with an additional layer of abstraction, mapping the concepts of *subject*, *action* and *object* into the notions of *role*, *activity* and *view* in the context-specific setting. It also extends the permissions in RBAC with *prohibitions* and *obligations* as well as with the *context*.

The vast majority of systems today, even the ones present in distributed and cloud environments, conceptually depend either on these models, or on their derivations.

3.2. Enabling access control for distributed and risk-aware systems

The evolution of distributed and federated computing imposed the need for access control models that enable the definition and evaluation of access control from the perspective of connected systems in cross-domain context. The models that rely on user identities, such as DAC or MAC, are not completely suitable for decentralized and distributed systems. In the general scenario, the user or its identity has to be known at the time of policy definition and enforcement, which is not trivial to accomplish in multi-domain environments. Furthermore, the user's identity itself can encapsulate more information than it is actually needed or allowed to accomplish the transaction, or to conform to security and legislative requirements. Following that, Attribute-based Access control (ABAC) model shifted focus from user identities and applied an intensional approach that relies on the properties of principals. These properties are provided as attributes that can include beliefs about principals or serve as the basis for trusting these beliefs. Additionally, the attributes can be used to characterize the contextual conditions and requirements (Schneider, 2013; Jin, et al., 2012).

It is generally assumed that ABAC can provide benefits from other models, such as DAC, RBAC or MAC, while overcoming some of their limitations. With their ABAC-α model, Jin et al. (2012) provided contributions in that direction. UCON family of models introduced the concepts of *obligations*, *conditions*, *continuous enforcement* and *mutability of attributes* (Park & Sandhu, 2004).

The scopes of other contributions in the field can be summarized under the following categories: 1) *authorization process*, 2) *policy specification languages*, 3) *enforcement models* and 4) *model implementations* (Jin, 2014). The summary and recommendations over various related contributions are provided in the guide published by NIST (Hu, et al., 2014).

The integration of risk management practices in access control has been a subject of work of McGraw et al. Their Risk Adaptable Access control (RAdAC) incorporates the dimensions of *operational need* and *security risk* (McGraw, 2009; Britton & Brown, 2007). The operational need is determined at the time of the request and quantified using predefined conditions and the organizational policy set. The probabilistic measure of security risk in RAdAC is evaluated by taking into account the access environment. This includes the context and components that store and deliver the assets, their properties as well as the properties of the adversarial, such as its trustworthiness, role or physical location. Based on that, RAdAC derives information sharing decision from organizational policies, by considering normal and special circumstances and by confronting dimensions of need and a risk.

3.3. Authorization and trust models in multi-clouds

While traditional access control models found their application in cloud environments as well, their capabilities to answer the multifaceted requirements of complex and integrated systems are limited. One of notable contributions

towards the definition of access control architectures for federated cloud systems is proposed by Almutairi et al. (2011). Based on degrees of interoperation, privacy and verification complexity, they first characterized cloud collaborations as *ad-hoc*, *loosely coupled* and *federated*. The architecture proposed in their work relies on RBAC model and focuses on multi-layered horizontal and vertical integration of the clouds. The following integration types were identified: 1) *horizontal peer-to-peer interoperation* on the same cloud architectural level, 2) *vertical integration* of different layers within the same cloud, and 3) *cross-layered interoperation* between different clouds. The architecture of Almutairi et al. defines three main components to support these types: a virtual resource manager (VRM), a distributed access control module (ACM) and a service level agreement component (SLA).

Connected, these components handle the provision of a virtualized view of resources in the interoperation, storage, analysis and enforcement of SLA and access policies on the level of each domain. Considered separately, the VRM component is deployed at each layer of cloud architecture and is responsible for the provision of virtual resources and to ensure the compliance with SLA. The ACM component is deployed at each layer of cloud infrastructure and serves as a policy base, decision and enforcement point for its domain. The last component, SLA, provides a bridge between different domains instantiated across the cloud. It presents a virtualized view of cloud resources and performs role mapping between domains, specifying isolation constraints for resource sharing.

The challenge of authorization federation for multi-clouds is approached by Pustchi et al. (2015). They identified four potential types of trust relations between *trustee* and *trustor,* relevant for cross-cloud peer-to-peer access on IaaS level (Tang & Sandhu, 2013). These relations are characterized by dimensions of *responsibilities* and *capabilities* of involved parties, establishing cross-cloud, cross-domain and cross-project trust levels.

Trust models and architecture by Tang et al. (2013) have been implemented and analyzed in OpenStack environment, enabling cross-domain trust in the single cloud. Although it focuses primarily on IaaS resources and peer-to-peer federation restricted on RBAC model, the contribution of this work is valuable as it establishes formalized and practical approach to trust establishment and mapping of users and resources between clouds.

3.4. Interoperability in cross-domain and distributed environments

Due to domain-tied policies and party-specific configurations, the establishment of an effective federation of security policies and configurations among entities in distributed cloud environments presents a challenging task. One of promising directions relies on semantic technologies, enabling the cooperation of entities and the exchange of information on a higher abstraction level. A semantic backed cooperation of entities enables the exchange of information on the higher level of abstraction, allowing the parties to bridge different lexical descriptions, concepts or terminologies. Furthermore, the application of semantic technologies allows for the usage of reasoners that are able to derive the knowledge not explicitly specified in the policies. The formal validation and verification of models, as well as the detection and resolution of conflicts, are further possible with this approach.

Considering the need to facilitate mutual understanding of access control policies, Hu et al. (2009) proposed *Access Control Oriented Ontology System* (ACOOS), which serves as the basis for *Semantic Access Control Policy Language* (SACPL) for interconnected systems. ACOOS includes four ontologies for *subjects*, *objects*, *actions* and *attributes*. These are used to specify the entities and rules in access control policies.

The SACPL by Hu et al. (2009) extends the traditional access control 3-tuple with rules that specify categories for *conditions*, *confidentiality* and *priority*. Besides integrating the standard functionality with 3-tuple, additional categories included in the language enable the specification of requirements that have to be fulfilled in order to decide successful access request. The elements of SACPL are applied as annotations to XACML rules (Rissanen, 2012), which serve as a basis for interoperable security policy.

The direction suggested by Hu et al. (2009) has been continued by Bernabe et al. (2014). Their solution, however, focuses on multi-tenant environments and the broader family of role-based models, such as RBAC, hierarchical RBAC (hRBAC), conditional RBAC (cRBAC) and hierarchical objects (HO). Barnabe et al. use semantic technologies to describe both the infrastructure and the authorization model, including the respective rules. For this purpose they employ OWL 2 ontologies (W3C OWL WG, 2009), SWRL rules (Horrocks, et al., 2004) and Common Information Model (Bumpus, W. et al).

Takabi and Joshi (2012) approached the problem from the user's perspective. One of the limitations of existing policy management systems identified by them relates to the area of heterogeneity and interoperation of policies.

Namely, it is common that each CSP implements its own authorization mechanism, resulting in distributed and heterogeneous authorization policies across the systems. This diversity of mechanisms, policy models and their notations hinders the interoperation and cooperation among different service providers. The users on the other side do not possess the ability to ensure holistic and consolidated view on access policies in different cloud providers, nor are they enabled to alter and federate these policies easily.

In order to address these challenges, they proposed semantic-based policy framework that supports distributed and heterogeneous policies. This approach, similarly as the one proposed by Hu et al (2009), relies on OWL 2 ontologies and SWRL rules to enable interoperable rule and entity descriptions.

3.5. Discussion

In this chapter we reviewed the state of the art in access control models. We perceived the challenge of integrating access control in distributed architectures that cross organizational boundaries, as traditional models focused on closed and proprietary, in-house systems and context, no longer fit for increasingly cross-organizational data consumption and service interactions. The broad adoption of cloud solutions imposes the challenge of externalizing access control processes and their transparent integration in the wider context. The architecture of Almutairi et al. presents a promising direction for this purpose, however, the gaps should be bridged both in the terms of its practical implementation, as well as in the terms of interoperability of infrastructures. The system by Tang et al. promises as well, however, it lacks the support for other integration types.

We find the work presented in section 3.4 as a valuable basis that enable bridging the semantic gaps arising from integration of heterogeneous and different infrastructures. We intend to extend these approaches and apply them in the multilayered federation of semi-trusted and heterogeneous clouds, specially targeting eGovernment use cases and focused public administrations. The concept of access control in the scope of our work is tightly related with *policy definition languages* and *cryptographic mechanisms*. We believe that the integration of these mechanisms can provide additional value and enable secure and transparent federation on the levels of data, service and infrastructure integrations.

4. Policy Definition Languages in Distributed Cloud Environments

As the emergence of cloud computing redefined the data propagation and processing patterns, in and between applications, the requisite to specify security requirements in fine-grained and transparent manner arised as well. This also raised a challenge of automated understanding and trustable enforcement of these requirements across the boundaries of different systems and organizations. Those systems should conform to various levels of obligations, considering the dimensions of organizational, legal, user or domain-specific requirements. The policies that govern such processes operate on different layers, ranging from the high-level to low-level policies applied in the scope of particular subsystems. The enforcement of policy languages is tightly coupled with the language definition and its capabilities, executed at several points in the system, in a *centralized*, *a distributed* or in *a hybrid* way.

In this chapter we introduce and evaluate existing security policy languages, focusing on the data context and applicability in interconnected institutional facilities.

4.1. Ponder

Ponder is a policy management framework that includes a policy specification language, a general deployment architecture and a policy administration toolkit (Damianou, 2002; Damianou, et al., 2001). Although aimed to support RBAC, the integration with other models such as MAC and DAC has been demonstrated as possible. Ponder differentiates policies in security and management category, enabling governance through both positive and negative policies. The first category relates to access control, including authorization, information filtering and delegation policies, while the management category specifies obligation and refrain policies. Following declarative and object-oriented approach, Ponder applies the concepts of reuse and extension. The policies can be structured in the form of *composite policies, roles* and *management structures*, allowing grasping the organizational complexity of large

enterprises. The scope of the policies is determined by *domains*, the structures used to manage collections of objects, and *meta-policies*, application specific constraints on a group of policies used to prevent policy conflicts.

Some drawbacks of Ponder are its lack of generality, category-specific syntax and difficult dynamic policy update (Phan, T. et al., 2008). Twidle et al. (2009) proposed a major system redesign as Ponder2, aiming to address disadvantages of Ponder's centralized model, which is not well suited for autonomic computing and large federations. For this purpose, Ponder2 relies on *self-managed cells*, an autonomous set consisting of hardware and software components that form an autonomous administrative domain capable of self-management. They also introduced a high-level object-oriented language PonderTalk, used to configure and control Ponder2 deployments.

4.2. eXtensible Access Control Markup Language

In contrast to Ponder, eXtensible Access Control Markup Language (XACML) represents a more generalized, broadly adopted and standardized approach that specifies policies in XML format (Rissanen, 2012). It defines the basic entities *Rule*, *Policy* and *PolicySet*. Rule entities implement the actual authorization logic and are the most atomic units of the policy. The policy entity in contrast combines multiple rules for evaluation by the XACML engine. It furthermore has an associated algorithm for combining multiple rules and optionally a set of obligations or advices, which can be used to execute particular actions once a policy evaluation succeeds. A policy set is positioned one level above and incorporates multiple policies, other policy sets, obligations and advices.

XACML not only defines the policy language with its structure, but also a data flow model and a methodical approach on how to evaluate and enforce certain policies. The core components of the data-flow model include Policy Enforcement Point (PEP) and Policy Decision Point (PDP) as well as Policy Administration Point (PAP) and Policy Information Point (PIP). The XACML approach has evolved to the de-facto standard in the field of data-security policy languages and therefore was adapted, modified and improved extensively.

With the intention to address the lack of formal semantics in XACML and answer strict requirements from eHealth domain, Masi et al. (Masi, et al., 2012) proposed the Formal Access Control Policy Language (FACPL). Being heavily inspired by XACML 3.0, FACPL in the first line offers a simple and clear syntax established on BNF context-free grammar, providing a more compact notation that is easier to read. The application of formal semantics enables policy expression based on a solid mathematical foundation, allowing the analysis and formal reasoning on policies (Margheri, A. et al., 2013). By supporting the export to XACML policies FACPL also maintains interoperability with XACML. The support for formal validation of policies is, however, an ongoing work.

The further enhancement of XACML that considers the support for usage control based model (UCON) (Park & Sandhu, 2004) has been proposed by Colombo et al. (2010). The UCON model enhances traditional access control with *mutable* attributes of subjects, objects and environment, enabling the continuous control of the policy enforcement while the access or consumption of resources is in progress. By integrating UCON facilities in the XACML approach, the U-XACML policy language by Colombo et al. enables the original model to support attribute updates and continuous policy evaluation. It also allows the definition and checking of conditions that govern ongoing attribute updates and obligations. The description and analysis of their proposal in event-based push and periodic invocation-based pull models have been provided in (Colombo, et al., 2010). The continuation of this work demonstrated the integration in OpenNebula and improved model that supports the management of concurrent user sessions (Lazouski, et al., 2012).

4.3. Data usage and consent control

Data processing in distributed systems represents a challenge from several views. First, the scalability in environments consisting of diverse organizational entities, cross-layered rules and policies might raise issues of efficiency and interoperability. As the second, the handling of data and PII (privacy identifiable information) should be performed on a privacy and legal conforming way, enabling users to provide informed consent and revocation.

The challenge of efficient and traceable data usage control and scalability in distributed systems has been approached by Kelbert and Pretschner (2014). In scenarios where the systems rely on centralized enforcement infrastructure, the necessity to contact a central point for each event might result in a high communication overhead, lowering the responsiveness, usability and scalability of systems. Kelbert and Pretschner addressed this issue by

providing a formal, distributed data usage control model that reduces the overhead in decentralized monitoring of multiple concurrently running systems. This infrastructure is supported by deploying local PDP/PIP instances that are partially able to make instant decisions. This model relies on formal methods to approximate all potentially relevant systems for evaluation of a particular policy, at each point in time. This way, the coordination is reduced to the set of identified systems relevant for the enforcement.

The user-controlled sharing and consumption of private data in complex interactions have been dealt in the scope of EnCoRe project (Ensuring Consent and Revocation), a research initiative from UK industry and academia sector. EnCoRe aimed to develop mechanisms that allow users to control the use, storing and sharing of their personal information held by others. Various levels of policy layers and legislative requirements were analyzed to deliver a comprehensive conceptual model for privacy policies (Pearson, 2009; Mont, et al., 2010). EnCoRe primarily resulted with technical architecture that extends XACML's PDP and PEP points with the ability to manage privacy-aware access control and obligation policies (Mont, et al., 2011). The user control of data sharing is enabled by *Consent and Revocation Privacy Assistant*, a client-side tool for definition and storage of privacy preferences.

One of the outcomes of EnCoRe project are sticky policies as well, the machine-readable specifications that are integrated with the encrypted data and govern its further usage across the systems (Pearson, S. et al, 2011). The authors denote the possibility to implement a range of operations using techniques such as *search-enabled* or *indexable encryption*, enabling the activities such as searching and indexing over encrypted data.

4.4. Discussion

Elaborating on different concepts, we notice the lack of languages that are generalized enough to support a range of deployment types and dimensions of security, which is necessary for establishment of cloud based collaborations that encompass various architectural types, heterogeneous systems and workflows present in public sector.

From a language perspective, the real-world integration of these systems demands the capability to bridge the policy requirements defined on the levels of particular organization, on inter-organizational level and on the level of legal requirements that deal with data processing and cross-border collaborations. This applies to various dimensions of security as well. The capability of security policy should go beyond the pure specification of access control rules and encompass advanced expressivity that enables the definition and execution of actions that should be taken in particular data sharing and processing steps, inside and outside of organizational or jurisdictional boundaries. Such actions may include data transformation, masking or hiding, as well as the execution of obligations that would enable traceability, assurance and legal conformance.

We further notice that data-security policy languages are tailored to permit-deny access control decisions. In the context of distributed and cloud computing a more fine-grained decision is required in the form of e.g. "*Permission to read the last four digits only on systems that do not belong to the trusted agency*" or "*Permission to check the existence of a particular value for a purpose of salary transfer*". Based on previous findings, we identify XACML as potential basis for further work that would provide architectural, process and semantic enhancements with the purpose to bridge the gaps and enable policy federations across infrastructures and services in public sector.

5. Cryptography Mechanisms for Cloud-based Data Sharing

The collaboration in multi-layer federated environments requires the storage, sharing and processing of data that exhibits various levels of sensitivity, raising the issues of conformance with security and legislative requirements or obligations. Depending on levels of trustfulness between cooperative environments, the appropriate trade-off should be applied considering the dimensions of data security, its processibility by the target system, and overall efficiency.

In this chapter we present the approaches that enable data sharing and processing in outsourced environments that support different levels of trust and access granularity. We consider both perspectives that enable the data sharing using advanced encryption schemas, as well as using novel architectures that deliver cryptographic services on various platforms and protocols.

5.1. Outsourced storing and processing of data

The application of standard encryption schemes, such as AES, can significantly alter the data format, causing disruptions both in storing and processing of data. Indeed, when storage devices and applications are designed to operate on unencrypted data, they may not be able to operate on encrypted data. The family of *Format-Preserving Encryption* (FPE) schemes aims to address this issue by bridging privacy and confidentiality with data processing requirements of applications. These schemes encrypt messages into cipher texts with the same format, enabling secure and legal compliant storage and processing of data in legacy and distributed systems.

The initial contribution of Black and Rogaway (2002) considered the enciphering on arbitrary domains with finite message space and providing the first detailed analysis of this challenge. The Rank-then-Encipher (RtE) method suggested by Bellare et al. (2009) reduces the task of designing an FPE for format $\mathcal{F}$ to the task of designing an FPE for an integral domain. In particular, the RtE framework allows one to apply the same encryption logic to all formats, eliminating the need to design format-specific encryption schemes (Goldberg & Sipser, 1985). However, as the scheme relies on the efficiency of ranking and unranking, the challenge of designing efficient methods for general formats still remains. Several works suggested FPEs for specific formats considering more practical message-domains such as social-security numbers (Hoover, 2015), credit-card numbers, and dates (Liu, et al., 2010).

To address the drawbacks of previous approaches, Luchaup et al. (2014) developed libFTE – a unifying format preserving and format-transforming encryption scheme. In format-transforming encryption, all ciphertexts are guaranteed to have format $\mathcal{F}_0$, which may differ from the message format. libFTE also employs the RtE method, where regular expressions represent formats, that are ranked using either a corresponding deterministic finite automaton (DFA) or non-deterministic finite automaton (NFA). To allow the use of more efficient regex-to-NFA transformation, the authors relax the ranking method, such that it can also be based on an NFA.

Tokenization schemes are also used to protect data privacy. Tokenization is the process of the substitution of a sensitive data with a non-sensitive surrogate value, also referred as token. There is no direct "relation" between the sensitive data and the tokens, while in encryption there is some kind of dependence between the cipher text and the sensitive data (e.g., mathematical relation in RSA). Hence, the only way to derive sensitive data from the token is to have a mapping table, which maps the token to the original sensitive data. As a result, when given a token it is infeasible to get the original data without the mapping table.

The scenarios that involve outsourcing of data storage and processing might, however, require more advanced processing capabilities to be carried out remotely. Foresti identified challenges in this domain and classified them into following categories: 1) *efficient querying of outsourced data*, 2) *prevention of data modifications*, 3) *overhead in query execution*, 4) *enforcement of the access control* and 5) *different data protection needs of cooperating servers* (Foresti, 2010). *Order Preserving Encryption Schemes* (OPES) deal with this challenge, enabling the processing of encrypted data in outsourced and untrusted environments.

Focusing on efficient querying of database systems, Agrawal et al. (2004) introduced an OPES for numeric data which allows any comparison operator to be applied directly on encrypted data without yielding false positives. The approach is tailored to SQL databases and therefore supports operators such as MAX, MIN and COUNT, and operations like GROUP BY and ORDER BY. Their solution can easily be integrated with existing database systems without applying modifications. The schema as proposed by Agrawal et al. produces an output that adheres to a user defined distribution, regardless of the distribution of the input values. Hence, an attacker cannot conclude on particular entries based on the distribution of the values. However, as shown by Boldyreva et al. (2009), this system cannot achieve indistinguishability against chosen-plaintext attack in a practical case.

A practical approach that enables the execution of SQL queries over encrypted data is provided by Popa et al. (2011). It implements a collection of SQL-aware encryption schemas, including the OPES introduced by Boldyreva et al. (2009) which is the first provable and secure scheme. Designed as a proxy system intermediating between an application and database management systems, CryptDB tries to address two threats. The first one is a curious database administrator who intends to learn private data. The second threat is an adversary that gains a complete access over application and database server. This threat is only partially addressed by ensuring the confidentiality of the data of users not logged-in. The overhead in the practical application of CryptDB system has been evaluated by Popa et. al. (2011) demonstrating the promising results in the case of online bulletin board in the range of 14.5%.

5.2. Controlled data sharing

While the data security in distributed and semi-trusted environments can be enforced by applying encryption, the more demanding scenarios that involve further data propagation and processing in cooperative inter-domain perspective require techniques to control the usage of encrypted data beyond the initial domain. One of the approaches in this direction has been proposed by Blaze et al. (1998). Their scheme based on ElGamal utilizes a semi-trusted *proxy* that converts cipher texts encrypted for one party to cipher texts encrypted for another party without revealing the plaintext. The drawbacks of this schema are its bi-directionality and lack of quorum control of the involved parties. This issue has been addressed by Jakobsson (1999) with an asymmetric proxy re-encryption schema under quorum control, as long as there is no dishonest quorum.

Ateniese et al. contributed further by providing a summarization of most useful properties of proxy re-encryption schemas. Their first attempt utilizes a two stage decryption scheme as proposed by Dodis and Ivan (2003). The key is divided into two shares and the decryption process is performed on the proxy and on the particular party. This approach requires the decrypting party to use custom secret keys and generally does not conform to the list of properties. Without going into the very details of the algorithms, Ateniese et al. (2006) achieved to develop a proxy re-encryption system adhering to the above properties and only putting limited trust in the proxy.

While proxy re-encryption schemes enabled controlled sharing of encrypted data, the more advanced scenarios of fine-grained access control over distributed untrusted parties have been further enabled by using *Attribute-Based Encryption* (ABE) (Sahai & Waters, 2005). The simplest setting for ABE involves an authority generating a set of public parameters including a binary relation (x, y). A sender can use the public parameters to encrypt a message m under some string x, obtaining a ciphertext c_x. A user might receive a secret key sk_y from the authority (for some string y). In case $(x, y) = 1$, decrypting c_x under sk_y yields the original plaintext m; otherwise, in case $(x, y) = 0$, the decryption fails. By restricting x to represent a user's identity (instead of a set of attributes), the process can be transformed to *identity-based encryption* (Boneh & Franklin, 2001) as a special case. Similarly, it should be noted that ABE is a special case of *functional encryption* (Boneh, et al., 2011).

Depending on the role of the parameters x and y, two main types of ABE can be identified. In *Ciphertext-Policy ABE* (CP-ABE) the parameter x is used to specify an access policy and y to denote a set of attributes. In this case attribute sets are assigned to private keys, allowing the senders to specify an access policy that receivers' attribute sets should comply with (Bethencourt & Sahai, 2007; Waters, 2011). *Key-Policy ABE* (KP-ABE) on the contrary specifies an access policy with y and a set of attributes with x. This way the ciphertexts can be annotated with attribute sets and private keys associated with access structures that specify which ciphertexts the user will be entitled to decrypt (Sahai & Waters, 2005; Goyal, V. et al., 2006). Additionally, by conjunctively combining KP-ABE and CP-ABE, so-called *Dual-Policy ABE* schemes can be obtained (Attrapadung & Imai, 2009) .

Standard ABE assumes a single trusted authority that manages all attributes; each user wishing to obtain a secret key for a set of attributes must prove its identity to the trusted authority. Chase (2007) defined a multi-authority ABE scheme that supports many authorities operating simultaneously, each handing out secret keys for a different set of attributes. The limitations of this work have been further addressed in (Lewko & Waters, 2011).

Further contributions enabling secure data sharing that combine ABE and proxy re-encryption have been proposed by Katai et al. (2013) and Lie et al. (2014), which enable time-based constraints.

The controlled data sharing has been addressed from another perspective in the work of Reimair et al. (2015). Their work on Cryptographic Service Interoperability Layer (CrySIL) proposes an architecture that supports distributed usage of cryptography services for environments consisting of diverse solutions and complex workflows characterizing long-evolving organizations. CrySIL facilitates interoperability by integrating various forms of authentication, deploying distributed crypto execution and emulation environment (Reimair & Feichtner, 2015) and leveraging the range of platforms, such as smartphones, to serve as cloud crypto service providers (Reimair, et al., 2015).

5.3. Discussion

In this chapter we reviewed various cryptographic techniques that enable secure data sharing and processing in cloud collaborations, considering the use-cases relevant for public authorities. Although there exist various encryption

schemas and system architectures that go beyond the scope of this survey, such as *homomorphic encryption*, we focus on practically attainable approaches that may exhibit applicability considering the data sets, performance and security requirements as well as the degree of integration in the workflows and infrastructures of public agencies. In this sense, we recognize the challenge for further work to apply and evaluate a holistic approach that enables the integration of surveyed techniques in standard workflow encompassing both the architectures and processes of federated infrastructures and services among various public administrations entities.

6. Conclusion

In this work we approached the issue of secure data sharing and processing in heterogeneous clouds. The focus of our work was the scenario that considers the federation of cloud services among various and dispersed infrastructures of public administrations. Our work further considered the integrations with public cloud services that mainly arise for the purpose of outsourcing for cost and performance effective data storage and processing.

In our work we revised the state of the art in access control models, architectures, policy languages and cryptographic approaches to secure data sharing and processing. We identified the gaps in each of these categories and defined the scope of our future work. We aim to provide an integrated, holistic approach to cloud integration that would both encompass advanced access control architectures, interoperable, semantic-rich, transparent and multi-dimensional security policies as well as novel cryptographic schemes for fine-grained and secure data sharing and processing.

This work will be further evaluated in the scope of real-world application that includes federations and cross-border scenarios of public administrations from several European countries.

Acknowledgements

This work has been sponsored by European Commission through SUNFISH project, executed under the framework of the Horizon 2020 Research and Innovation Programme.

References

1. Agrawal, R., Kiernan, J., Srikant, R. & Xu, Y., 2004. *Order preserving encryption for numeric data.* s.l., ACM.
2. Almutairi, A. et al., 2011. A distributed access control architecture for cloud computing. *IEEE software 2,* Band 2, pp. 36-44.
3. Ateniese, G., Fu, K., Green, M. & Hohenberger, S., 2006. Improved proxy re-encryption schemes with applications to secure distributed storage. *ACM Transactions on Information and System Security,* pp. 1-30.
4. Attrapadung, N. & Imai, H., 2009. Dual-policy attribute based encryption. *Applied Cryptography and Network Security.*
5. Bellare, M. et al., 2009. Format-preserving encryption. In: *Selected Areas in Cryptography.* s.l.:Springer Berlin Heidelberg.
6. Bernabe, J. et al., 2014. Semantic-aware multi-tenancy authorization system for cloud architectures. In: *Future Generation Computer Systems 32.* s.l.:s.n., pp. 154-167.
7. Bethencourt, J. & Sahai, A., 2007. Ciphertext-policy attribute-based encryption. In: *IEEE Symposium on S&P 07.* s.l.:IEEE.
8. Black, J. & Rogaway, P., 2002. *Ciphers with arbitrary finite domains.* s.l., Springer Berlin Heidelberg, pp. 114-130.
9. Blaze, M. et al., 1998. Divertible protocols and atomic proxy cryptography. *Advances in Cryptology—EUROCRYPT'98.*
10. Boldyreva, A., Chenette, N., Lee, Y. & O'neill, A., 2009. *Order-preserving symmetric encryption.* s.l., Springer Berlin Heidelberg.
11. Boneh, D. & Franklin, M., 2001. Identity-based encryption from the Weil pairing. *In Advances in Cryptology—CRYPTO 2001.*
12. Boneh, D., Sahai, A. & Waters, B., 2011. Functional encryption: Definitions and challenges. *Theory of Cryptograph.*
13. Britton, D. & Brown, I., 2007. *A security risk measurement for the RAdAC model.* s.l., Naval Postgraduate School Monterey.
14. Bumpus, W. et al, 2000. *Common information model: implementing the object model for enterprise management.* s.l.:J.W. & Sons.
15. Chase, M., 2007. Multi-authority attribute based encryption. *Theory of cryptography,* pp. 515-534.
16. Colombo, M., Lazouski, . A., Martinelli, F. & Mori, P., 2010. *A Proposal on Enhancing XACML with Continuous Usage Control Features.* s.l., Springer US.
17. Damianou, N., 2002. *A policy framework for management of distributed systems,* s.l.: Imperial College.
18. Damianou, N., Dulay, N., Lupu, E. & Sloman, M., 2001. *The ponder policy specification language.* s.l., Springer Berlin Heidelberg.
19. El Kalam, A. et al., 2003. *Organization based access control.* s.l., IEEE, pp. 120-131.
20. Foresti, S., 2010. *Preserving privacy in data outsourcing.* Springer Science & Business Media Hrsg. s.l.:s.n.
21. Goldberg, A. & Sipser, M., 1985. *Compression and ranking.* s.l., ACM, pp. 440-448.
22. Goyal, V. et al., 2006. Attribute-based encryption for fine-grained access control of encrypted data. *Proceedings of the 13th ACM conference on Computer and communications security.*

23. Hoover, D. N., 2015. *Format-preserving encryption via rotating block encryption.* U.S., Patentnr. 8,948,376.
24. Horrocks, I. et al., 2004. *SWRL: A Semantic Web Rule Language,* s.l.: W3C.
25. Hu, L., Ying, S., Jia, X. & Zhao, K., 2009. Towards an approach of semantic access control for cloud computing. In: *Cloud Computing* . s.l.:Springer Berlin Heidelberg, pp. 145-156.
26. Hu, V. et al., 2014. *Guide to attribute based access control (ABAC) definition and considerations,* s.l.: NIST.
27. INCITS, ANSI, 2004. *Incits 359-2004. role-based access control.* s.l.:ANSI.
28. Ivan, A.-A. & Dodis, Y., 2003. Proxy Cryptography Revisited.
29. Jakobsson, M., 1999. On quorum controlled asymmetric proxy re-encryption. *Public Key Cryptography,* pp. 112-121.
30. Jin, X., 2014. *Attribute-based access control models and implementation in cloud infrastructure as a service,* s.l.: s.n.
31. Jin, X., Krishnan, R. & Sandhu, R., 2012. *A Unified Attribute-Based Access Control Model Covering DAC, MAC and RBAC.* s.l., s.n.
32. Kaitai, L. et al, 2013. *A ciphertext-policy attribute-based proxy re-encryption with chosen-ciphertext security.* s.l., IEEE.
33. Kelbert, F. & Pretschner, A., 2014. Decentralized Distributed Data Usage Control. In: *Cryptology and Network Security.* s.l.:Springer International Publishing, pp. 353-369.
34. Lazouski, A., Mancini, G., Martinelli, F. & Mori, P., 2012. *Usage control in cloud systems.* s.l., IEEE.
35. Lewko, A. & Waters, B., 2011. Decentralizing attribute-based encryption. *Advances in Cryptology–EUROCRYPT 2011,* pp. 568-588.
36. Liu, Q. et al, 2014. Time-based proxy re-encryption scheme for secure data sharing in a cloud environment. *Information Sciences 258*.
37. Liu, Z., Jia, C., Li, J. & Cheng, X., 2010. *Format-Preserving encryption for datetime.* s.l., IEEE, pp. 201-205.
38. Luchaup, D. et al., 2014. *LibFTE: a toolkit for constructing practical, format-abiding encryption schemes.* s.l., USENIX, p. 115.
39. Margheri, A. et al., 2013. On a formal and user-friendly linguistic approach to access control of electronic health data.
40. Masi, M., Pugliese, R. & Tiezzi, F., 2012. Formalisation and implementation of the XACML access control mechanism. *Engineering Secure Software and Systems*.
41. McGraw, R., 2009. *Risk-adaptable access control (radac).* s.l., NIST.
42. Mont, M. C. et al., 2010. *EnCoRe: towards a conceptual model for privacy policies.* s.l., s.n.
43. Mont, M. et al., 2011. *Technical architecture arising from the third case study,* s.l.: s.n.
44. Park, J. & Sandhu, R., 2004. The UCON ABC usage control model. *ACM Transactions on Information and System Security*.
45. Pearson, S. et al, 2011. *End-to-end policy-based encryption and management of data in the cloud.* s.l., IEEE.
46. Pearson, S., 2009. *Taking account of privacy when designing cloud computing services.* s.l., IEEE Computer Society.
47. Phan, T. et al., 2008. *A survey of policy-based management approaches for service oriented systems.* s.l., IEEE, pp. 392-401.
48. Popa, R. A. et al., 2011. *CryptDB: protecting confidentiality with encrypted query processing.* s.l., ACM, pp. 85-100.
49. Pustchi, N., Krishnan, R. & Sandhu, R., 2015. Authorization Federation in IaaS Multi Cloud. *Proceedings of the 3rd International Workshop on Security in Cloud Computing.*
50. Reimair, F. & Feichtner, J., 2015. *Attribute-based Encryption goes X.509.* s.l., s.n.
51. Reimair, F., Teufl, P., Kollmann, C. & Thaller, C., 2015. *MoCrySIL - Carry Your Cryptographic Keys in Your Pocket.* s.l., s.n.
52. Reimair, F., Teufl, P. & Zefferer, T., 2015. *WebCrySIL - Web Cryptographic Service Interoperability Layer.* s.l., s.n., pp. 35-44.
53. Rissanen, E., 2012. *eXtensible access control markup language (XACML) version 3.0 OASIS standard.* s.l.:OASIS.
54. Sahai, A. & Waters, B., 2005. Fuzzy identity-based encryption. *Advances in Cryptology–EUROCRYPT 2005,* pp. 457-473.
55. Samarati, P. & Di Vimercati, S., 2001. *Access control: Policies, models, and mechanisms.* s.l.:s.n.
56. Sandhu, R. & Samarati, P., 1994. Access control: principle and practice. *Communications Magazine,* pp. 40-48.
57. Schneider, F., 2013. Chapter 9: Credentials-based authorization. In: *Untitled Textbook on Cybersecurity.* s.l.:s.n.
58. Takabi, H. & Joshi, J., 2012. Semantic-based policy management for cloud computing environments. *Int Journal of Cloud Computing*.
59. Tang, B., Li, Q. & Sandhu, R., 2013. *A multi-tenant RBAC model for collaborative cloud services.* s.l., IEEE.
60. Tang, B. & Sandhu, R., 2013. *Cross-tenant trust models in cloud computing.* s.l., IEEE.
61. Twidle, K., Dulay, N., Lupu, E. & Sloman, M., 2009. *Ponder2: A policy system for autonomous pervasive environments.* s.l., IEEE.
62. W3C OWL WG, 2009. *OWL 2 Web Ontology Language Document Overview,* s.l.: s.n.
63. W3C, 2012. *OWL 2 web ontology language document overview,* s.l.: World Wide Web Consortium.
64. Waters, B., 2011. Ciphertext-policy attribute-based encryption: An expressive, efficient, and provably secure realization. *Public Key Cryptography–PKC 2011,* pp. 53-70.