Cloud-based identity and identity meta-data

Secure and control own data in globalization era

Grzegorz Spyra, Prof William J Buchanan, Peter Cruickshank, Dr Elias Ekonomou

Abstract – This paper proposes a new identity, and its underlying meta-data, model. The approach enables secure spanning of identity meta-data across many boundaries such as health-care, financial and educational institutions, including all others that store and process sensitive personal data. It introduces the new concepts of **Compound Personal Record** (CPR) and **Compound Identifiable Data** (CID) ontology, which aim to move toward *own your own data* model. The CID model ensures: authenticity of identity meta-data; high availability via unified Cloud-hosted XML data structure; and privacy through encryption, obfuscation and anonymity applied to Ontology-based XML distributed content. Additionally CID via XML ontologies is enabled for identity federation. The paper also proposes that access over sensitive data is strictly governed through an access control model with granular policy enforcement on the service side. This includes the involvement of relevant access control model entities which are enabled to authorize an ad-hoc *break-glass* data access which should give high accountability for data access attempts.

Keywords – identity; ID-based cryptography; health-care; obfuscation; anonymization; encryption; privacy; Compound Identifiable Data; access control; digital signing;

Introduction

Current technology and business trends are moving organizations, institutions and enterprises into the cloud, although many are aware of the risks, exemplified by the recent events which show that even highly protected personal data can be seized for processing, without the data owner's consent and knowledge (Kroes, 2013).

The risks increase: the latest leaks (January 2014) by Edward Snowden show that the USA's National Secure Agency (NSA) is involved in an on-going project called Penetrating Hard Targets (PHT), which aims to develop a super quantum computer enabling US government to decrypt information (Rich & Gellman, 2014). As this technology will inevitably become widespread over time, the conclusion has to be that any Cloud-based model or system that will process or store personal data will require not only multiple safeguards to protect the confidentiality, but also should deliver high accountability enforced by governments where personal data is protected by law.

Cloud-based services can often provide data security which delivers data protection sufficient to secure the data from outsider threats, but they cannot protect it from rogue Cloud Service Provider (CSP) employees (Mowbray, Pearson, & Shen, 2010). Currently, personal data stored by schools, hospitals and other organizations often does not meet baseline safeguards required to hold such data and is often not ready for Cloud-computing era (Hölbl, 2011).

In this paper, we define personal data as a piece of information that identifies an individual directly or indirectly (OECD, 2013a).

In health care, medical organizations store and process sensitive personal information, and also need access to sensitive data (X. Chen, 2004) to save their patients life at any time, without technological and jurisdictional constraints (OECD, 2013b). Unfortunately, access to such data is mostly restricted to one institution or very often a single building.

Even when personal data is stored and processed within secure and well-defined boundaries, problems can arise because the there is no oversight by the data owner (ie the patient): there are strong indications that Personal Health Record (PHR) owners would also like to have full access to their own information (eg Buchanan et al., 2013), and also to be able to control the rights of access to the records. PHRs and even more so the Electronic Health Record (EHR) which aggregate them require a platform that will allow secure data exchange (Zhang & Liu, 2010) preserving privacy across Cloud-based systems (Li, Yu, Zheng, Ren, & Lou, 2012).

Similar problems can be found in educational institutions where pupil (Buchanan, Lewis, Fan, & Uthmani, 2012) or student information cannot be shared due to legal and technological limitations, despite the data owner's expectations. In order to give people full ownership of their data and to enable institutions; organizations and enterprises around the world to securely share sensitive information security professional, as well as governments, we must deliver tools, platforms and knowledge base applicable to modern environments (Zhou, Varadharajan, & Hitchens, 2014).

In this paper, we will focus on the possibilities for an up-to-date identity meta-data model that could be designed to be ready for globally established secure data processing, cloud hosting and extensive authentication. With this model, owners would have control over their own data (or be able to delegate it to identified affiliated people) starting from their birth[1].

[1] We recognise that this raises the significant question of who can exercise 'ownership' of data for individuals who cannot exercise it for themselves (babies, the extremely ill, dementia patients). Space constrains mean that this issue of affiliated users (Neubauer & Heurix, 2011) cannot be addressed here.

This work builds on a number of projects including within health and social care, including with the UK Technology Strategy Board- (TSB) funded project with Chelsea and Westminster Hospital in London which focused on creating an e-Health Cloud within a hospital environment (Fan, 2011 and Fan, 2012). This used a novel method of defining the ownership of the data, and providing a rights infrastructure for the citizen (or patient) to define the rights of access to their data. This work has since been extended within a number of projects including the TSB Trusted Service project, which has focused on integrating both digital and human trust, to provide a fully integrated and holistic care infrastructure, and which integrates primary and secondary health care with assisted living (Ekonomou, 2011 and L, 2012).

IDENTITY META-DATA PARADIGM

Currently used digital identity models provide several levels of assurance (Buchanan et al., 2013). These levels depend on Identity Provider (IP) requirements for the authenticity of personal identification. Identities can be seen as the unique information sufficient to perform operations on objects. This information can be static in some identity models but can also vary dependent on claims by specific Service Providers (SPs) (Chappell, 2011).

Data held by the IP can most often be classified as either PII (Personally Identifiable Information) (Mccallister & Scarfone, 2010) or Personal Data. PII includes home addresses, social security numbers or maiden name – that is, enough to allow a unique identification of an individual. Depending on the authentication architecture supported by the IP, the PII meta-data is exchanged as claimed Verified Attributes (Buchanan et al., 2013) between IP and SP. Unfortunately, this framework has a structural weakness resulting from the lack of a unified, secure identity (assuming one is desirable), legally governed certification and baseline standards for IPs (Data Protection Working Party, 2012) and an easily accessible authoritative knowledge base for people registering with IP.

Identity verification methods require the highest level of assurance (Buchanan et al., 2013) with effective safeguards against unauthorised PII data divulgence. The new concept of Compound Identifiable Data (CID) supports the personal responsibility of the object owner over their digital identity and its authenticity. As illustrated in Figure 1 below, information that is exchanged while an SP is authenticating identity consists of an obfuscated unique identity identifier (Mowbray et al., 2010) and policies required for further authorization. Unlike claims-based identity authentication (Chappell, 2011) only conditions can be checked, - the identity meta-data cannot be passed to the SP. This new PII data flow asks to introduce clear security baselines and certification for SPs and Authentication, Authorization, and Accounting (AAA) mechanisms (Pearson & Wainwright, 2013). This model is still vulnerable to more sophisticated attacks such as data inference (Salamatian, Zhang, Calmon, & Bhamidipati, 2013).

Finally, in the CID model, an access object can act as a subject, depending on the activity context as shown in Fig. 2. In other words, personal identity (user) may act as an object which is also defined under the CID sub-ontology. Furthermore, the SP can act as an identity (user) and a subject in the data access attempt. Both SP and a personal identity require identity and identity meta-data to be legally registered to perform specific activities in the shared Cloud space. With generic IAM framework, all activities of the subject over an access object are logged for further legal audits. The technology can benefit from a single, secure model where each entity of access control operation is equally accountable, as an identity instance inherits generic schema whether it is a real person or an automated robot.

Fig. 1 Linked identity with identity meta-data using obfuscated references across several security boundaries and contexts

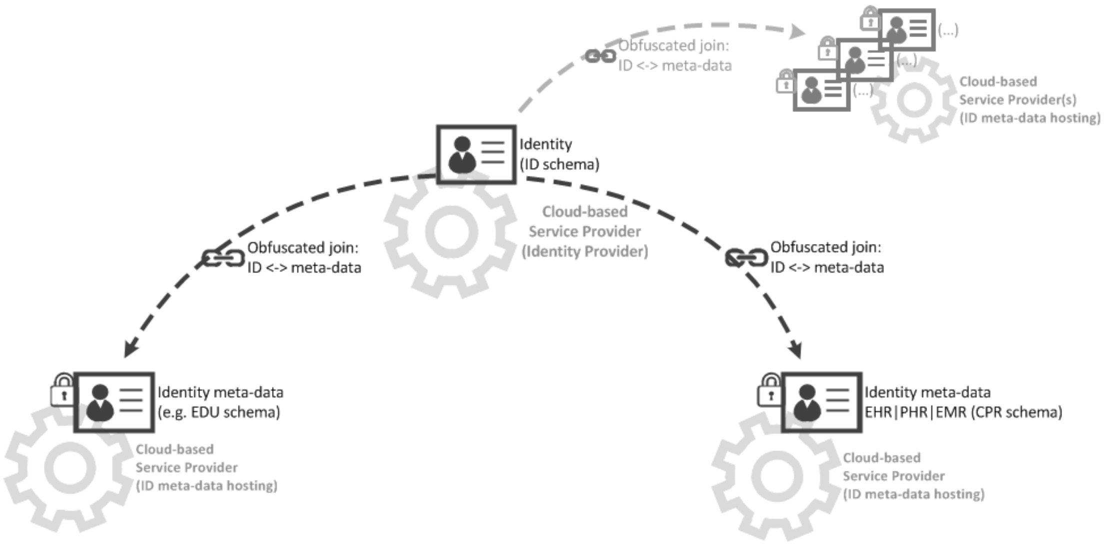
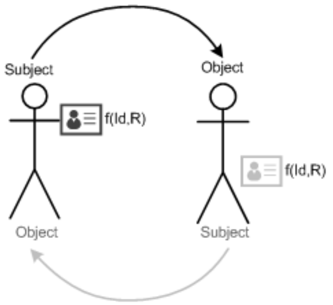


Fig. 2 Two different subjects access contexts with identity meta-data; f – function matching identity Id with its rights R in given context



CID Distribution

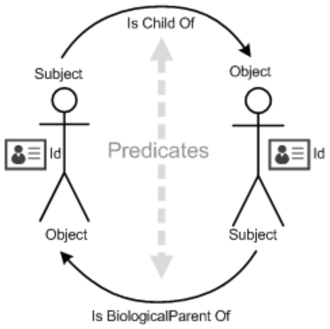
To understand the CID model's flexibility, its cross-platform integration abilities and its capacity for identity and access management (IAM) and federated identity management (FIM), we need to define models for CID ontology enablement and content distribution. An XML identity meta-data schema requires a technique which allows different parties to share an XML schema for a particular type of content that is attached to a CID. As CID content will be spanned across different systems, it has to share the core identity element so as maintain its unique reference to a single person, while also maintaining several different schemas for contextual interpretation. In this way, CID shares a core identity element that subjects can use for self-identification in the process of accessing objects. The core identity ontology should not only identify but also represent identity access and operations entitlements in the Cloud for various services enabled for different access control models.

One such XML-based ontology is Web Ontology Language (OWL) which was designed to define the semantics of the relationships between entities. OWL defines what is semantically correct in XML, and both deliver data framework for the Web as well as for Cloud-based systems. OWL has been successfully used for access control systems implementation (Finin, Joshi, Niu, Sandhu, & Winsborough, 2008) as well as with encrypted distributed XML content (Rahaman, Roudier, Miseldine, & Schaad, 2009). An alternative approach was introduced to deliver structures: the Resource Description Framework (RDF) together with RDF Schema (RDFS) that defines classes for RDF. However, although it seems to be easier to adapt, its limitations mean RDF cannot provide the functionality sufficient to build reliable Cloud-based CID framework.

Single ontology is mostly defined by class, sub-class, properties and relationships between these within OWL ontology (Rahaman et al., 2009). It is also possible to define OWL class relationships, where different ontologies can share a common parent.

Using OWL it is possible to define an ontology for our CID, which would allow identity to refer to other identities with simple predicate definitions (see Fig. 3). CID, as a compound XML model defined under several ontologies, can be spanned across several contexts (see Fig. 1). In other words, different parts of CID can be stored and processed by different organizations. XML parts can be defined under various ontologies with different OWL-defined, XML schemas. A model where different part of identity meta-data are distributed to different service providers enables it to make use of a range of existing XML schemas, for instance Microsoft HealthVault XML . Only such an approach can guarantee that data access to personal information can be distributed and efficiently maintained by different Cloud service providers. In the health-care sector, the CID model aligns with the EHR concept where a range of different (and possibly competing) health-care repositories can hold patient information (Zhang & Liu, 2010). Secure access to distributed data is possible thanks to one identity and identity meta-data framework. Securely linked identity-metadata is an assurance of data integrity and authenticity, what is required from new Cloud-based systems.

Fig. 3 Two subjects in refer to each other in triplet (Subject – Predicate – Object) represent linked identities



Distributed XML parts from within CID need to be linked in order to refer only to one identity. This secure XML linking defines that part of the identity meta-data that belongs only to one identity. Such linking is highly vulnerable to several types of attack including impersonation attack and man-in-the-middle attack. In response, the identity-based encryption (IBE) model has been successfully utilized to create secure dynamic reference for hierarchical data structures similar to the CID structure and is discussed further below.

In summary, individual parts of CID share a common ontology designed to support secure links. A mandatory obfuscated link is maintained from the main identity XML to sub parts of the identity meta-data and back from identity meta-data to the main identity. Requests from separate parts of identity meta-data can be hosted in a Service Oriented Architecture SOA implementation as shown in Fig. 4. Web Service(s) exposed as part of dedicated Cloud-based services can process distributed requests using encryption, obfuscation and anonymization. Next, each Cloud-based service can effectively support such distributed XML model with effective XML clusters (Costa & Ortale, 2012), where single XML document can be partitioned into several clusters.

ENCRYPTION, OBFUSCATION AND ANONYMIZATION

A major problem with the Cloud is when it comes to data protection, as it seems to be local and personal when, in fact, unencrypted data once stored in the Cloud might be the subject of direct or indirect processing by third parties. Lack of encryption mechanisms used to protect the information and the global character of the Cloud causes data leaks without data owner control (Mowbray et al., 2010). Even encrypted data still brings several risks of data leakage because encryption techniques which ensure sufficient security require high processing power. This processing power is needed to effectively encrypt and decrypt data 'live' in memory, revoke encryption keys to actually achieve truly personal data in the Cloud (Pagano & Pagano, 2011).

The need for Identity Based Encryption

Encryption delivers the best security possible for identity meta-data. CID data before it is hosted by any Cloud-based service should be encrypted by default. The encryption of CID is required for every single XML node (Rahaman et al., 2009), as they from a sensitive part of hierarchical identity meta-data. It should be also applied respectively to the XML ontology definition to ensure access control granularity (Zhou et al., 2014).

Encrypted data require secure key repositories able to perform revocation when necessary. One approach to effective key lifecycle management is Identity-based Encryption (IDE) with self-expiring keys (Boneh & Franklin, 2003). In Cloud-based implementations, IDE works efficiently by introducing ephemeral cryptographic keys. In IBE with public-key encryption the public key can be derived from unique identity identifier, therefore, this approach reduces the need for certificate authorities and public key certificates (Yao, Fazio, Dodis, & Lysyanskaya, 2004). Different types of IBE implementations provide benefits for different system models. Forward-secure hierarchical ID-based encryption (fs-HIBE) has successfully been used for several identity and access management IAM implementations; it allows secure dynamic joins between identities, making use of time constraints and dynamic key revocation (Yao et al., 2004).

Multiple hierarchical ID-based encryption scheme (MHIBE) is another concept derived from generalization of fs-HIBE. MHIBE is not only highly suitable for federated identity management systems such as this but, because of the ability of encryption with multiple ID-tuples, it can be efficiently used with the role-based access control (RBAC) systems implementations (Yao et al., 2004) discussed later.

The main technological problem with effective data protection is the efficiency of encryption algorithms when encrypted information requires indexing (Mowbray et al., 2010). Several methods can be used in parallel for encryption to effectively index unencrypted XML data (Luk et al., 2002), although implementation would require further techniques to protect the PII data in the CPR which actually consists of PHR and EHR data.

How then can the CPR (following its owner's consent) or other data from identity meta-data be effectively found and used for processing from among billions of other records that are stored in the Cloud? Answers include XML obfuscation (Mowbray et al., 2010) and anonymization (Ye, Wu, Hu, & Hu, 2013), techniques which provide high performing searching and indexing algorithms, ensuring the accessibility required.

Obfuscation

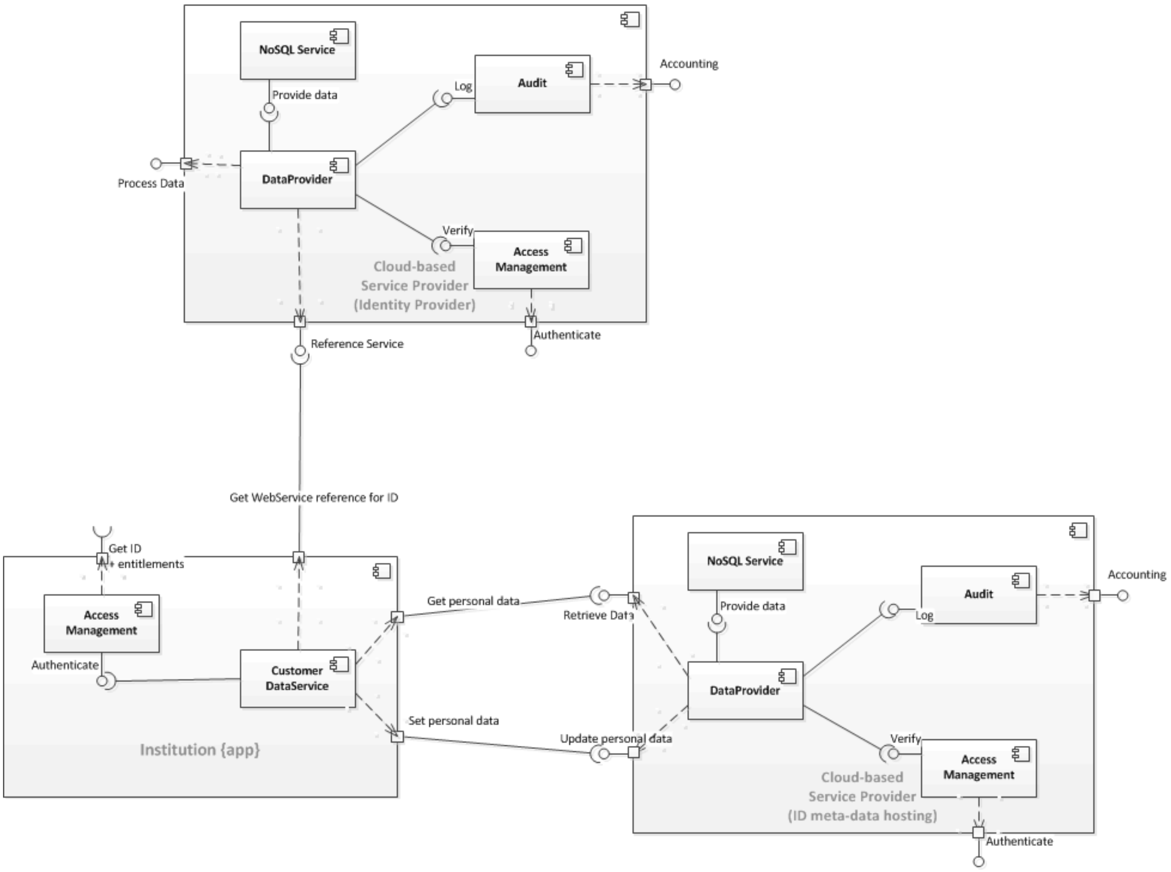
Obfuscation methods aim to hide data so it cannot be directly processed. Obfuscated data allows the object owner to reveal only the necessary information required to execute an operation on that information without exposing PII information. For example, a health Cloud-based services provider could introduce a technique where patient's identity is not a subject of exchange between parties; instead, only unique pseudonyms are exchanged between parties to securely satisfy claims (Mowbray et al., 2010).

Obfuscation uses basic cryptographic techniques to hide rather than encrypt data. These methods use keys and functions to derive obfuscated information that corresponds to identity meta-data, but which do not disclose actual information (Mowbray et al., 2010). To decide which part of identity meta-data should be obfuscated, policy-based obfuscation can be used, where different policies enforces obfuscation of specific fields before these are made available for Cloud-based processing. The privacy manager implementation proposed by Mowbray et al (2010) for personal data obfuscation can almost transparently integrate with existing applications, thus this is a reasonable safeguard that ensures data security due-care principals.

Anonymization

Another approach is anonymization. Often, for example, in research and monitoring aggregate data, it is acceptable to process anonymized data. k-anonymization techniques are widely studied as part of artificial intelligence research. They apply to dataset processing where sophisticated attack techniques like data linking (data inference) can be used to uniquely identify individual from among other records that are not directly exposed for processing (Kisilevich, Rokach, Elovici, & Shapira, 2010). Quasi-identifiers (QIs) can be derived using k-anonymity from the table of k number of records, where derived k-anonymous table ensures anonymity of the QI among the other k-1 records (Ye et al., 2013). k-anonymization can be effectively used to securely deliver statistical data, therefore, all personal data processing, which requires generalized information rather than identity specific data should be delivered via anonymization.

Fig. 4 Calling identity meta-data by obfuscated reference under SOA (possible use case)



INTEGRITY & AUTHENTICITY

Identity meta-data must provide the most accurate information possible. It should ensure not only data quality, but also data integrity and authenticity. Data quality can be maintained with well-designed XML ontologies applied at different CID contexts. Data integrity gives assurance that data has not been amended since the last valid data change was committed. Authenticity ensures that the subject identified as the last data processor actually initiated the data transaction. Because changes made over CID are not accountable at the CID level and require dedicated functionality responsible for accounting, the identity meta-data needs to deliver basic integrity and authenticity assurance. This assurance can be guaranteed with a digital signature applied to the part of information that requires data integrity. As the digital signature can be derived not only from the information, but can be bound with a unique identifier, it can be used for information which requires data authenticity (Bartel, Boyer, Fox, LaMacchia, & Simon, 2008).

For a CID model, we need to ensure that a malicious or ignorant subject did not amend the information, that information was changed in the current identity context and an entitled subject processed that information.

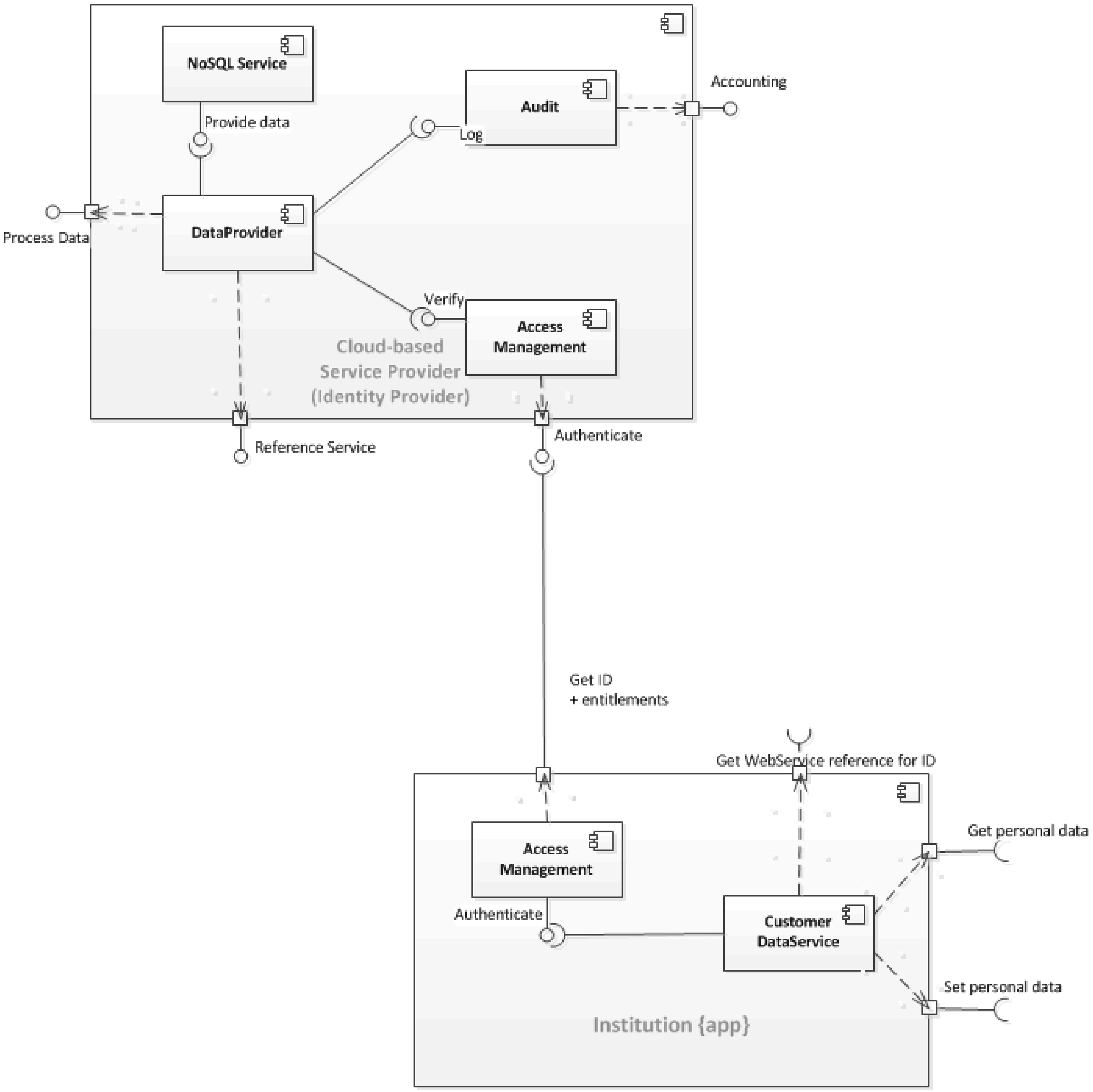
The access control models we described here use signing for nonrepudiation and integrity enforcement; however, identity meta-data requires same enforcement at the level of actual data. For instance using our emergency access example, where a medical professional needs to access a patient’s data to check their medical history, if an unauthorized subject (including the data owner) amended medical history, it may have critical consequences leading to patient’s death.

Digital signing cryptography requires secure keys to derive a signature. Public-key infrastructure (PKI) and ID-based signing (IBS) are two different approaches we can use to deliver keys (Kiltz, Neven, CWI Amsterdam, IBM Zürich Research laboratory, & Katholieke Universitet Leuven, 2009). While PKI involves trusted certification authorities (CAs) to certify public keys and bind them with digital identity, in IBS public key consists of an identity unique identifier, therefore, it simplifies implementation model by eliminating CA entity from key management lifecycle.

To keep the CID model as homogeneous as possible and therefore the final framework simple, we will focus on IBS as a preferred digital signing technique. IBS and ID-based encryption share the same concept for secure key management. IBS, unlike IBE, can use certificates issued by an involved trusted authority (TA) based on identity identifier and assigned public key. IBS certificate do not require a CA, as it is simply a digital signature derived from public key and identity unique identifier (Kiltz et al., 2009).

As an alternative to certificates, IBS can utilize hierarchical ID-based encryption (HIBE), the encryption, which was discussed above as a preferred CID encryption method. Hierarchical IBS (HIBS) schemas become very useful when combined with HIBE (Gentry & Silverberg, 2002) as HIBE schema derived from content encryption can be transformed into HIBS schema. The digital signing and verification processes are, therefore, simplified.

Fig. 5 CID authentication in SOA



ACCESS CONTROL

Role Based Access Control

Most mature modern access control models that are ready to securely protect the asset (such as PII), from unauthorized access, hardly span outside a simple boundary (Spyra, 2012). The best-known model, Role-Based Access Control (RBAC) can be easily adapted for highly secure end-to-end identity provisioning and revocation within specific security contexts. A role is assigned with an identity for a set of transactions, for example, as the ability for a doctor to access patient data and take further actions according to new circumstances and patient history (Ferraiolo & Kuhn, 1992) then update the patient record. This access control model controls the subject access over the object, based on roles assigned to the subject in the organization, which defines a security boundary (Zhou et al., 2014).

Roles can span across several systems and well-integrated infrastructures can ensure role change enforcement on the end system. Furthermore, well-defined roles and consistent RBAC system implementation are safeguards against several security threats such as collusion, creeping privileges and excessive privileges. Enforcement of separation of duties is a countermeasure for collusion attack (Sandhu, Ferraiolo, & Kuhn, 2012), while the principle of least privilege overcomes problems of creeping privileges and excessive privileges (Stewart, Tittel, & Chapple, 2011).

Although RBAC ensures high security within an organization, it does not introduce a name space that can be implemented across organizations, for example, in an open Cloud space. However, several research studies have shown that when combined with an OWL to deliver Ontology-based XML content (Rahaman et al., 2009), RBAC can be adapted for global Cloud-based IAM solutions.

The CID model requires the RBAC concept with its several variations to enforce control and secure identity with its meta-data in the Cloud. To protect policy that is applied on the object in the Cloud and accessed as a part of RBAC transaction the part that exposes the policy can be encrypted (Zhou et al., 2014). The new concept of Cryptographic RBAC for Cloud addresses several security threats that have roots in early RBAC architectures, where this access control model had closed security boundaries such as enterprises, organizations and institutions. Role-based Encryption is a model that allows data encryption before it is handed-over to the Cloud provider, thus ensuring that only data owners and identities that hold the required access role can decrypt the information.

CID requires also clearly defined ontology to reach Cloud maturity for RBAC. There are several approaches emerging that introduce standardized RBAC in different sectors, one, the Enhanced RBAC, is focused on clinical education, biomedical research, and patient care (Le, Doll, Barbosu, Luque, & Wang, 2012). This work clearly highlights the fact that there is a need to define strict ontologies where Enhanced RBAC could be applied. These ontologies would constrain and help to define CPR ontology to enable it for access control in the Cloud.

RBAC implementations can have disadvantages. One of them is the existence of a single point of failure in case of attack that aims to compromise the central access management system. XML-based data where RBAC is used to control access can still be protected using distributed access control system (López, Maña, & Yagüe, 2002). Distributed access control systems can be scaled and adapted for Cloud-based implementations. This problem was also addressed with Cryptographic RBAC (C-RBAC) (Zhou et al., 2014) and, unlike the distributed access control approach; this model was designed for Cloud-based IAM systems implementation. C-RBAC uses policies that are enforced via Cloud services and which can be controlled in a decentralized manner by the data owner.

Finally, to deliver the fully homogeneous model, there are ARBAC97 and SARBAC models, which aim to provide control over RBAC systems including granular role hierarchy amendments, new policy definitions and all other administrative operations which are fully controlled via a dedicated roles set (Zhou et al., 2014).

This new identity meta-data model requires a highly reliable and efficient encryption and access control model. The most vulnerable and sensitive part of the CID, the compound personal record (CPR), needs safeguards that will make data securely available not only to data owner but also to authorized data processors.

In summary, the RBAC model can become a very powerful access control model suitable for Cloud-based services but only when supported with relevant implementation guidelines and baselines, which enforce security policies and legal regulations.

Attribute Based Access Control

The Attribute-based Access Control (ABAC) provides another approach to govern access, by giving the data owner full control over their data. In the ABAC model, roles are bound into role attributes and are attached to a data element through attributes based encryption (ABE) (Yang & Jia, 2014). The ABAC model can coexist with RBAC and easily enables RBAC beyond a single security boundary (Spyra, 2012).

ABE allows the data owner to encrypt the personal data under specific attributes. Same attributes are attached to subjects who will process the data (Li et al., 2012). CID model and especially CPR meta-data have to use access control system with encryption applied to access control properties that are attached to data.

The ABE model has been proposed as the most suitable technology for Cloud-based global data access (Yang & Jia, 2014), although ABAC among other access control models described here have specific features in combination can satisfy CID model. There seems to be an increasing interest in ABE as demand on electronic health-care systems has grown in the last few years (Li et al., 2012).

Attribute-based infrastructures have been proposed as ready for handling PII information for instance a special implementation of ABE called cipher text policy ABE (CP-ABE) with message broadcasting enables an ABAC system to perform ad hoc direct revocation (Hur & Noh, 2011). As with RBAC, the main problem with CP-ABE is single trust authority (TA) that can be used to decrypt data. Key escrow enables a single TA to decrypt all the information and a compromised TA provides the potential attacker access to all the protected data. A way to overcome this problem is multiple-authority ABE (MA-ABE) where each TA releases only a partial secret key that is used to encrypt information. On the other hand key revocation under this approach creates a bottleneck where each TA needs to be involved in keys lifecycle (Li et al., 2012).

While CP-ABE allows data owners to decide on attribute structure defining permissions before encrypting data sent into the Cloud, the other approach key-policy ABE (KP-ABE) uses policies to define permissions and the data owner assigns attributes to define encrypted data (Hur & Noh, 2011). Service managing policies for KP-ABE, automatically generates access structure for the data then combines access policies into keys (Li et al., 2012).

All of the ABE techniques described here struggle with weak revocation thus there are on-going research projects to create an effective and efficient keys revocation algorithm for ABAC systems (Yang & Jia, 2014).

Sticky-policies

The ABAC model gives more control over owned data via attributes that define access domain; sticky-policies approach go further by giving the data owner granular control over each piece of personal information. PII data seeks for access control model which can enforce legal regulations regarding data protection (OECD, 2013a) and which can be easily adapted for Cloud-based implementations. CID requires technology which will easily integrate with encrypted XML content, while CPR seeks for highly reliable access control model to protect PII and grant access to personal data only when access entitlement is followed by relevant legal consent.

Sticky-policies make use of trust authorities (TA), which validate compliance with policies in order to lease decryption keys. Policies likewise cover data owner consent, give subject rights to process data. Model where TA has to be contacted by the service provider (SP) to access PII data delivers high accountability. Each personal data access attempt is a subject of auditing (Pearson, Bramhall, & HP Laboratories, 2003) and can be tracked in case of data leakage incident.

The data owner can then feel that they own the data released into the Cloud because of not only the policies associated with data following data owner approval, but also for the TA, which specifies where the policy can be interpreted, and is pre-selected by the data owner (Pearson, Mont, & Kounga, 2011). Information about the TA is attached to the policy and is passed to SP. XML schema that can store sticky-policy definition can be easily integrated into CID. The content of the policy definitions can be encrypted using ID-based encryption (IBE) (Pearson et al., 2003). Both policies and data encrypted with IBE add security on top of the sticky-policies model, however the bottleneck of this method is that encryption applied this way makes data *heavy-weighted* (Pearson et al., 2011).

The CID model requires the IBE to securely bound identity meta-data together across different security contexts. In both cases, Cloud-based implementation would need introducing additional entity, which takes care of IBE related operations. Having in mind a larger picture, the system model for identity meta-data can be simplified and the Cloud services hosting CID can be generalized.

Purpose-based Access Control

All the above access models control access based on entitlements granted and detailed access policies. Another model, a Purpose-based access control allows long-term maintenance of access granted at some point of time (Sun & Wang, 2012) and enforces need-to-know and need-to-have principles. In more traditional access control model from the moment when access is granted to subject via either role or direct assignment this access relationship from subject to object is preserved over time unless relevant auditing procedures enforce access control review and revoke creeping privileges (Stewart et al., 2011). This purpose justifies the subject to store, process or access an object (Sun & Wang, 2012). It can be defined under intended purpose and access purpose categories. Therefore, the access decision is made based on the correlation between the intended purpose and the access purpose. Intended purpose fall into three components: Allowable Intended Purpose (AIP), Conditional Intended Purpose (CIP) and Prohibited Intended Purpose (PIP) (M. Chen, Yang, & Hwang, 2013). Where AIP defines unrestricted data access, CIP conditional data access and PIP denies any access for given purpose. Combined with access purpose, which can consist of single RBAC assignment, the data access is enhanced by very granular control (M. Chen et al., 2013).

As RBAC model was successful in delivering effective access control functionality and became widely adopted in many enterprises, it is reasonable to consider integration of RBAC with policy-based access control model (M. Chen et al., 2013).

The concept of access purpose is not an integral part of any of the previously described access control models. Although it does not mean that related security procedures cannot define circumstances where the subject becomes entitled to process data under the defined access control model. The purpose-based access control model shows that there is a need for legal baselines and guidelines for Cloud-based IAM implementations. In a global model, access to PII should only be allowed through a single legal framework to prevent data redundancy, therefore avoiding inconsistent access management system where data can be processed because of conflicting definitions of legally justified access purposes.

Break-Glass - emergency access

Complete identity and access management (IAM) system consists not only of technologies but also of relevant security policies and procedures built to support access control and provide reliable accountability of a subject's activities over an object. In most generic scenarios, a subject is entitled to process data when it is granted rights at some point of time. Rights are granted based on subject roles assignment, or based on direct permissions applied to the object. In a secured environment, before PII data can be processed, the subject requires a consent (OECD, 2013a).

Now let us analyse a person's experience of an accident abroad, and where a medical professional needs to access the patient's personal record - which is CPR data - and due to injury, the patient cannot approve doctor's access to it. This scenario requires a dedicated and strictly controlled **Break-Glass** process allowing access to personal data to be subject to post-processing approval (Li et al., 2012). Such access attempt should trigger communication channels that inform the relevant authorities (e.g. supervisor of the person performing a Break-Glass access, local health-care practice where patient is registered). Next, in most cases, access needs to be justified by the person performing the emergency access, and then afterwards by the relevant authorities. Break-Glass action thus requires legal enforcement to account each occurrence of the emergency access.

Whichever access control model is be used with the new identity model, an identity service provider SP should obtain legal approval to perform several types of 'break-glass' authentication like Claim-based identity (Chappell, 2011) where Verified Attributes, that are classified as personal identifiable information (PII), are claimed. PII information cannot be passed to SP without legal consent. Most of the modern SPs do not respect data protection principals (Hölbl, 2011). Often a client provides PII data to SP without understanding the implications. With the CID model person would not need to provide any data online to the SP. When individual approves access to own PII the SP would be allowed to only claim this information, which has a legal justification and authoritative approval.

Related Work

This paper provides an overview of secure models, which it is argued combine to give a solid foundation for a new reliable Cloud-based identity meta-data model. Many identity and PII data related models recognize security issues that come with Cloud computing (Mowbray et al., 2010; Pearson & Wainwright, 2013) propose methods to enforce accountability over data retrieved from the shared space. Others (Li et al., 2012) also focused on data protection in the health-care context propose to find the most suitable encryption and access control model but also propose a mature framework for Cloud-based implementations.

Several works (Jain & Farkas, 2013; Le et al., 2012) aim to deliver a unified XML model, which can be adapted under several security contexts like health-care, education institutions, enterprises and other. Here is worth mentioning Microsoft, which initiated its HealthVault development project which were defined in detail several XML schemas ready to adapt in medical institutions. The World Wide Web Consortium (W3C) leads in addressing security aspects of XML. The W3C has created the W3C XML Signature Working Group focused on digital signatures and W3C XML Encryption Working Group specialized in encrypted content.

Publicly exposed data requires several safeguards, and, in this field, works (M. Chen et al., 2013; Sun & Wang, 2012) related to purpose-based access control models play an important role as they aim to fill an existing gap not addressed with any previously existing access control model. Currently the most important non-technically related work is done by OECD (OECD, 2013b) and aims to deliver legal frameworks that ensure data protection and address privacy concerns related to Cloud-based computing era.

Conclusions and Future Work

This paper shows that there are techniques and technologies ready to be used in modern Cloud-based systems to deliver a global secure distributed service to everyone, independent of their location but ideally within a common legal framework, their assigned health-care system or the internal business model of organizations. The model presented thus delivers secure functionality that enables global access to identity-centric sensitive data.

Data protection is driven by technology, law and social convention. The focus of this paper has been technological solutions, but these make most sense in the context of strong data protection laws such as implemented by European countries which comply with data protection principals to that identifiable information about people is stored and processed with respect to people's fundamental rights to privacy (OECD, 2013b). A weakness in the current framework is that once stored in the Cloud, PII data or any information can be classified as divulged data (Pearson et al., 2011). Currently *loss of governance* and *data leakage* are the highest concerns related to Cloud services (Pearson & Wainwright, 2013).

The common problem with distributed XML data is that its implementation requires highly secure Cloud-based services for hosting. This includes encryption applied on data level to ensure that personal information will not be compromised, and become a subject of illegal processing in case of data leakage on the CSP side. The issue that requires further discussion is related to encryption aging when encrypted data is seized and stored until technology (such as quantum decryption) reaches a maturity level sufficient to break the encryption applied. This scenario may require well-designed dedicated Cloud-based services, which provide best security possible for CID identity meta-data but only in distinguished context. These could benefit from XML data clustering, which ensures high data availability (Costa & Ortale, 2012). Such services, if compromised, would guarantee that only fragments of personal data are exposed, but the PII remains un-compromised.

Accounting of CID it is another topic which requires further research. With this, digital signatures can validate information integrity, authenticity and non-repudiation only for the last committed change. An historical insight can be only be delivered with reliable accounting, which can either log each data fragment update, or take complete identity meta-data snap-shots. An important research area is thus the enforcement of proper accountability for CID.

Cloud-based system enabled to host global identity for Cryptographic RBAC require highly reliable revocation techniques where not only user, here the subject, is removed from the role, but also the role is either revoked or renewed in case of encryption keys or the role itself being compromised. There is thus a need for further research looking at all the security-related processes around existing access control models, and which are ready for CID management in the Cloud. Along with this further work is required to clearly defined Service Oriented Architecture (SOA) design listing baseline entities of authentication, authorization, accounting, keys management, access, identification, identity provisioning, identity revocation and data hosting and processing.

REFERENCES

Bartel, M., Boyer, J., Fox, B., LaMacchia, B., & Simon, E. (2008). XML Signature Syntax and Processing (Second Edition). Retrieved from http://www.w3.org/TR/xmldsig-core/

Boneh, D., & Franklin, M. (2003). Identity-Based Encryption from the Weil Pairing. *SIAM Journal on Computing*, *32*(3), 586–615.

Buchanan, P. W. J., Anderson, C., Smales, A., Varga, J., Uthmani, O., Fan, L., … Lawson, A. (2013). Who Would You Trust To Identify You In Accessing Your Health Record ? So who do we trust ?

Buchanan, P. W. J., Lewis, R., Fan, D. L., & Uthmani, O. (2012). Information Sharing Around Child Protection.

Chappell, D. (2011). Claims-based Identity for Windows; Technologies and Scenarios. DavidChappell & Associates.

Chen, M., Yang, C., & Hwang, M. (2013). Privacy Protection Data Access Control. *International Journal of Network Security*, *15*(6), 391–399.

Chen, X. (2004). *Identity Federation in Federated Trust Healthcare Network*. University of Virginia.

Costa, G., & Ortale, R. (2012). On Effective XML Clustering by Path Commonality: An Efficient and Scalable Algorithm. In *2012 IEEE 24th International Conference on Tools with Artificial Intelligence* (pp. 389–396). Ieee. doi:10.1109/ICTAI.2012.60

Data Protection Working Party. (2012). *Opinion 05/2012 on Cloud Computing* (pp. 1–27).

Ekonomou, E., etc al. (2011, November). An Integrated Cloud-based Healthcare Infrastructure. In Cloud Computing Technology and Science (CloudCom), 2011 IEEE Third International Conference on (pp. 532-536). IEEE.

Fan L, et al, DACAR Platform for eHealth Services Cloud, Cloud Computing (CLOUD), 2011 IEEE International Conference on, 219-226

Fan, L, et al. "SPoC: Protecting Patient Privacy for e-Health Services in the Cloud." eTELEMED 2012, The Fourth International Conference on eHealth, Telemedicine, and Social Medicine. 2012.

Ferraiolo, D. F., & Kuhn, D. R. (1992). Role-Based Access Controls. In *15th National Computer Security Conference (1992), Baltimore* (pp. 554–563).

Finin, T., Joshi, A., Niu, J., Sandhu, R., & Winsborough, W. (2008). ROWLBAC - Representing Role Based Access Control in OWL. In *ACM Symposium on Access Control Models and Technologies (SACMAT'08)*.

Gentry, C., & Silverberg, A. (2002). Hierarchical ID-Based Cryptography. In *International Conference on the Theory and Application of Cryptology and Information Security (ASIACRYPT 2002)* (pp. 548–566).

Hölbl, M. (2011). *Cloud Computing Security and Privacy Issues* (pp. 2–5).

Hur, J., & Noh, D. K. (2011). Attribute-Based Access Control with Efficient Revocation in Data Outsourcing Systems. *IEEE Transactions on Parallel and Distributed Systems*, *22*(7), 1214–1221. doi:10.1109/TPDS.2010.203

Jain, A., & Farkas, C. (2013). Ontology-Based Authorization Model for XML Data in Distributed Systems. In *Digital Rights Management*. IGI Global.

Kiltz, E., Neven, G., CWI Amsterdam, T. N., IBM Zürich Research laboratory, S., & Katholieke Universitet Leuven, B. (2009). Identity-Based Signatures. In *Cryptology and Information Security Series* (Vol. 2, pp. 31–44). IOS Press.

Kisilevich, S., Rokach, L., Elovici, Y., & Shapira, B. (2010). Efficient Multidimensional Suppression for K-Anonymity. *IEEE TRANSACTIONS ON KNOWLEDGE AND DATA ENGINEERING*, *22*(3), 334–347.

Kroes, N. (2013). How we're boosting trust in the cloud, post PRISM - European Commission. Retrieved October 11, 2013, from http://ec.europa.eu/commission_2010-2014/kroes/en/blog/trust-cloud-prism

Le, X. H., Doll, T., Barbosu, M., Luque, A., & Wang, D. (2012). An enhancement of the role-based access control model to facilitate information access management in context of team collaboration and workflow. *Journal of Biomedical Informatics*, *45*(6), 1084–1107. doi:10.1016/j.jbi.2012.06.001

Li, M., Yu, S., Zheng, Y., Ren, K., & Lou, W. (2012). Scalable and Secure Sharing of Personal Health Records in Cloud Computing using Attribute-based Encryption. *IEEE TRANSACTIONS ON PARALLEL AND DISTRIBUTED SYSTEMS*, *XX*(Xx), 1–14.

López, J., Maña, A., & Yagüe, M. I. (2002). XML-based Distributed Access Control System. *E-Commerce and Web Technologies Lecture Notes in Computer Science*, *2455*(i), 203–213.

Luk, R. W. P., Leong, H. V., Dillon, T. S., Chan, A. T. S., Croft, W. B., & Allan, J. (2002). A survey in indexing and searching XML documents. *Journal of the American Society for Information Science and Technology*, *53(6)*(6), 415–437. doi:10.1002/asi.10056

Mccallister, E., & Scarfone, K. (2010). Guide to Protecting the Confidentiality of Personally Identifiable Information (PII) Recommendations of the National Institute of Standards and Technology. U.S. Department of Commerce. Retrieved from http://csrc.nist.gov/publications/PubsSPs.html

Mowbray, M., Pearson, S., & Shen, Y. (2010). Enhancing privacy in cloud computing via policy-based obfuscation. *The Journal of Supercomputing*, *61*(2), 267–291. doi:10.1007/s11227-010-0425-z

OECD. (2013a). OECD Guidelines on the Protection of Privacy and Transborder Flows of Personal Data. Organisation for Economic Co-operation and Development. Retrieved from http://www.oecd.org/internet/ieconomy/oecdguidelinesontheprotectionofprivacyandtransborderflowsofpersonaldata.htm

OECD. (2013b). The OECD Privacy Framework. Organisation for Economic Co-operation and Development of.

Lo, O., et al. (2012). Technical evaluation of an e-health platform. IADIS E-Health.

Pagano, F., & Pagano, D. (2011). Using in-memory encrypted databases on the cloud. *2011 1st International Workshop on Securing Services on the Cloud (IWSSC)*, 30–37. doi:10.1109/IWSSCloud.2011.6049022

Pearson, S., Bramhall, P., & HP Laboratories. (2003). Towards Accountable Management of Identity and Privacy : Sticky Policies and Enforceable Tracing Services Marco Casassa Mont. In *14th International Workshop on Database and Expert Systems Applications (DEXA'03)*. IEEE Computer Society.

Pearson, S., Mont, M. C., & Kounga, G. (2011). Enhancing Accountability in the Cloud via Sticky Policies. *Secure and Trust Computing, Data Management, and Applications Communications in Computer and Information Science*, *187*, 146–155.

Pearson, S., & Wainwright, N. (2013). An interdisciplinary approach to accountability for future internet service provision. *International Journal of Trust Management in Computing and Communications*, *1*(1), 52. doi:10.1504/IJTMCC.2013.052524

Rahaman, M. A., Roudier, Y., Miseldine, P., & Schaad, A. (2009). Ontology-Based Secure XML Content Distribution. *IFIP Advances in Information and Communication Technology*, *297*, 294–306.

Rich, S., & Gellman, B. (2014). NSA seeks to build quantum computer that could crack most types of encryption. *The Washington Post*. Retrieved from http://www.washingtonpost.com/world/national-security/nsa-seeks-to-build-quantum-computer-that-could-crack-most-types-of-encryption/2014/01/02/8fff297e-7195-11e3-8def-a33011492df2_story.html

Salamatian, S., Zhang, A., Calmon, P., & Bhamidipati, S. (2013). How to Hide the Elephant – or the Donkey – in the Room : Practical Privacy Against Statistical Inference for Large Data. In *1st IEEE Global Conference on Signal and Information Processing*. IEEE Signal Processing Society.

Sandhu, R. S., Ferraiolo, D., & Kuhn, R. (2012). The NIST Model for Role-Based Access Control: Towards A Unified Standard. In *5th ACM Workshop on Role Based Access Control* (pp. 47–63). Berlin.

Spyra, G. (2012). *Next Generation Authentication Infrastructures With Role Based Security For Cloud Computing*. Edinburgh Napier University.

Stewart, J. M., Tittel, E., & Chapple, M. (2011). Accountability and Access Control. In *CISSP®: Certified Information Systems Security Professional Study Guide, Fifth Edition* (Fourth., pp. 1–45).

Sun, L., & Wang, H. (2012). A purpose-based access control in native XML databases. *Concurrency and Computation: Practice and Experience*, *24*(10), 1154–1166. doi:10.1002/cpe

Yang, K., & Jia, X. (2014). ABAC: Attribute-Based Access Control. In *Security for Cloud Storage Systems* (pp. 39–58). New York, NY: Springer New York. doi:10.1007/978-1-4614-7873-7

Yao, D., Fazio, N., Dodis, Y., & Lysyanskaya, A. (2004). ID-based encryption for complex hierarchies with applications to forward security and broadcast encryption. In *Proceedings of the 11th ACM conference on Computer and communications security - CCS '04* (p. 354). New York, New York, USA: ACM Press. doi:10.1145/1030083.1030130

Ye, M., Wu, X., Hu, X., & Hu, D. (2013). Anonymizing classification data using rough set theory. *Knowledge-Based Systems*, *43*, 82–94. doi:10.1016/j.knosys.2013.01.007

Zhang, R., & Liu, L. (2010). Security Models and Requirements for Healthcare Application Clouds. *2010 IEEE 3rd International Conference on Cloud Computing*, 268–275. doi:10.1109/CLOUD.2010.62

Zhou, L., Varadharajan, V., & Hitchens, M. (2014). Cryptographic Role-Based Access Control for Secure Cloud Data Storage Systems. In S. Nepal & M. Pathan (Eds.), *Security, Privacy and Trust in Cloud Systems* (pp. 313–344). Berlin, Heidelberg: Springer Berlin Heidelberg. doi:10.1007/978-3-642-38586-5

Table 1. Selected access control models brief

Access Control	Advantages	Disadvantages
RBAC	▪ Separation of Duties ▪ Applied during identity provisioning/revocation	▪ Traditional RBAC implementations hardly scalable outside boundaries (e.g. enterprises) ▪ Missing globally established generic roles for Cloud ▪ No legal enforcement outside security boundary
ABAC	▪ Separation of Duties ▪ Applied during or after identity provisioning/revocation ▪ Easy to scale outside boundaries ▪ More granular access control than RBAC	▪ Missing globally established generic set of policies for Cloud ▪ No legal enforcement outside security boundary
Sticky-policies	▪ Enables the most granular access control model ▪ Gives owner full control over PII data access ▪ More granular access control than ABAC	▪ Policies may become hard to maintain over time ▪ Missing globally established generic set of policies for Cloud
Purpose-based Access Control with RBAC	▪ Introduces more specific RBAC framework for personal data processing ▪ Effectively enforces a need-to-know principle	▪ Missing globally established generic set of policies for Cloud ▪ No legal enforcement outside security boundary