Kevin Hamlen, Peng Liu, Murat Kantarcioglu, Bhavani Thuraisingham, and Ting Yu. "Identity Management for Cloud Computing: Developments and Directions." In *Proceedings of the 7th Annual Cyber Security and Information Intelligence Research Workshop (CSIIRW)*, October 2011.

IDENTITY MANAGEMENT FOR CLOUD COMPUTING: DEVELOPMENTS AND DIRECTIONS

Kevin Hamlen[1], Peng Liu[2], Murat Kantarcioglu[1], Bhavani Thuraisingham[1], Ting Yu[3]

1. **The University of Texas at Dallas**
2. **Pennsylvania State University**
3. **North Carolina State University**

ABSTRACT

Cloud computing technologies have been rapidly adopted by organizations to lower costs and to enable flexible and efficient access to critical data. As these new cloud technologies emerge, cyber security challenges associated with these technologies have increased at a rapid pace. One of the critical areas that needs attention for secure cloud computing is identity management where the multiple identities of cloud users operating possibly in a federated environment have to be managed and maintained. In this paper, we first explore identity management technologies and secure cloud computing technologies. We will then discuss some of the security balances for cloud computing with respect to identity management.

1. INTRODUCTION

There is a critical need to securely store, manage, share and analyze massive amounts of complex (e.g., semi-structured and unstructured) data to determine patterns and trends in order to improve the quality of healthcare, better safeguard the nation and explore alternative energy. The emerging cloud computing model attempts to handle massive amounts of data. Google has now introduced the MapReduce framework for processing large amounts of data on commodity hardware. Apache's Hadoop distributed file system (HDFS) is emerging as a superior software component for cloud computing combined with integrated parts such as MapReduce. However, state-of-the-art cloud computing systems are not sufficient due to the fact that (i) they do not provide adequate security mechanisms to protect sensitive data and (ii) they do not have the capability to process massive amounts of semantic web and geospatial data.

To address the limitations of current cloud computing platforms, researchers have utilized state-of-the-art hardware, software and data components based on Hadoop and MapReduce technologies and are developing a secure cloud computing framework. For example, modern hardware parts (e.g., secure coprocessors) to improve the performance due to incorporating additional security functionalities, integrated open source software parts, as well as custom developed software parts to support secure cloud query operations on complex data, provide fine-grained access control and reference monitor support, as well as provide strong authentication mechanisms.

Some recent work examines an XACML-based access control model with SAML for security assertion for a cloud computing framework. However, due to the fact that numerous identities for millions of users may have to be managed in a cloud environment, we need to reexamine the entire concept of identity management for the cloud. Identity management t or digital identity management is closely intertwined with web services. Users as well as web services have to be authenticated before accessing resources. Single Sign-on is the popular solution where one time sign-on gives a user of a service access to the various resources. Furthermore, SAML currently provides authentication facilities for web services. However, with regulatory requirements for e-business, and with the emergence of the cloud computing paradigm, one needs a stronger mechanism for authentication and this mechanism has come to be known as identity management [1].

Federated identity "describes the technologies, standards and use cases which serve to enable the portability of identity information across otherwise autonomous security domains." [2]. The goal is to ensure that users of one domain take advantage of all the technologies offered by another domain in a seamless manner. Note that federation is about organizations working together to carry out a task (such as B2B operations) or solving a particular problem. While the idea has been around for many years, it is only recently with the emerging standards of four web services that we can now have secure federations. In such federations, access to the resources by users has to be managed without burdening the user. With appropriate federate identity management, users should be able to share data across domains, support single sign-on as well as enable cross-domain user attribute management

This paper will provide an overview of the various developments with identity management as well as secure cloud computing and then examine identity management for cloud computing. In Section 2 we will discuss identity management technologies. In Section 3 we will discuss security for cloud computing. Issues on identity management for cloud computing will be discussed in Section 4. Standards efforts are discussed in Section 5. The paper is concluded in Section 6.

2. IDENTITY MANAGEMENT

Two concepts that are at the foundations of Digital Identity Management are (i) Single sign-on and (ii) federated identity management. As stated in [2], Single sign-on (SSO) is a property where a user logs in once and gains access to all systems possibly in a federation. This way the user has to log in once and has access to the resources in the federation or coalition or organization, without being prompted to log in again at each of them. Two types of SSO mechanisms are Kerberos-based and smart card-based. With Kerberos mechanism, Kerberos ticket granting ticket TGT is used to grant credentials. In the smart card based sign-on, the user uses the smart card for sign-on. Enterprise Single Sign-on (E-SSO), provides the support for minimizing the number of passwords and user-IDs when accessing multiple applications. As stated earlier "federated identity, or the 'federation' of identity, describes the technologies, standards and use-cases which serve to enable the portability of identity information across otherwise autonomous security domains [2]. The use cases include typical use-cases including cross-domain, web-based single sign-on.

One important concept in identity management is the Identity Metasystems. Identity Metasystem is an "interoperable architecture for digital identity that enables people to have and employ a collection of digital identities based on multiple underlying technologies, implementations, and providers." Essentially with this approach, users can continue to maintain their identities and choose the identity system that will work for them so that the system will manage their identities when migrating to different technologies. The roles of the Identity Metasystem are identity provider, relying parties and subjects. Identity providers issue digital identities. Relying parties are the ones who require identities such as various services. Subjects include the end users and organizations. Information card is an implementation of the Identity Metasystem. Information cards are personal digital identities that people can use online which support single sign-on as users can sign in at one place and have access to the a variety of resources on the web.

The various web sites are now implementing federated identity management with Open ID. OpenID is an open, decentralized user identification standard, allowing users to log onto many services with the same digital identity. OpenID is essentially a URL and the user is authenticated by their OpenID provider. Many corporations such as Symantec and Microsoft support OpenID. For example, Microsoft provides interoperability between OpenID and its Windows CardSpace. OpenID extends the entities of the Identity Metasystem and consists of the following: End-user: The person who wants to assert his or her identity to a site. Identifier: The URL chosen by the end-user as their OpenID identifier. Identity provider or OpenID provider: This entity provides the service of registering OpenID URLs and provides OpenID authentication. Relying party: The site that wants to verify the end-user's identifier. (this is essentially the service provider). Server or server-agent: The server that verifies the end-user's identifier. User-agent: Users access the identity provider or a relying party through the user agent (e.g., the browser). The use of OpenID is as follows. A user visits a relying party's (e.g. service provider) web site to request a service. This relying party has an OpenID form which is the login for the user. User would then give his identity which is provided by an Identity prior to the logic process. From this information the relying party will discover the identity provider web site.

Another key system in identity management is Shibboleth. Shibboleth is a distributed web resource access control system that allows federations to cooperate together to share web-based resources It defines a protocol for carrying authentication information and user attributes from a home to a resource site. The resource site can then use the attributes to make access control decisions about the user. This web-based middleware layer uses SAML. Access control is carried out in stages. In stage one, the resource site redirects the user to their home site, and obtains a handle for the user that is authenticated by the home site. In stage two, the resource site returns the handle to the attribute authority of the home site and it returns a set of attributes of the user, upon which to make an access control decision.

There are some issues with single sign-on with Shibboleth. How does the resource site know the home site of the user? How does it trust the handle returned? The answer is, it is handled by the system trust model. Authentication procedure is as follows.

When the resource site asks for home site from the user, he selects it from the list of trusted sites which are already authenticated by Certificates. Handles are validated by the SAML signature along with the message. User selects the home site from the list. Home site authenticates the user if he is already registered. After home server authentication, it returns a message with SAML sign to the Target Resource site. Resource site (if sign matches) then provides a pseudonym (handle) for the user and sends an assertion message to home page to find out if the necessary attributes are available with the user. To ensure privacy, the system provides a different pseudonym for the user's identity each time. It needs the release attribute policy from the user attributes each time to provide control over the authority attributes in the target site. Agreement attribute release policy is between the user and the administrator.

Finally, an organization called the Liberty Alliance was formed to promote standards for identity management. Two major efforts released by this consort are the Liberty Identity Federation (also called identity federation) and the Liberty identity web services (also called identity web services). Liberty Identity Federation enables the web users (e.g. e-commerce users) to authenticate and sign-on a domain and from there have access to multiple services. This is the basis of SAML 2.0. The identity web services standard is an open framework for deploying and managing identity-based Web services. These web services applications include Geo-location, Contact Book, Calendar, Mobile Messaging and Liberty People Service. With these services, one can manage bookmarks, blogs, and photo sharing and related social services on the web in a privacy-preserving manner. Privacy and policy management are key aspects of the work of Liberty Alliance.

3. SECURE CLOUD COMPUTING

A layered framework for assured cloud computing consisting of the secure virtual machine layer, secure cloud storage layer, secure cloud data layer, and the secure virtual network monitor layer. Cross-cutting services are provided by the policy layer, the cloud monitoring layer, the reliability layer and the risk analysis layer. We discuss the developments in some of the layers [3].

(i) Secure Hypervisors: Virtual machine (VM) technology is widely adopted as an enabler of cloud computing and provided through hypervisors. Ensuring the security of hypervisors is essential for assured cloud computing. Developments in secure hypervisors including secure VMware and Secure XEN platforms are being examined. Furthermore, researchers are also examining combinations of hardware/software approach is effective for system assurance. Solutions in virtual machines to defend against security threats, such as Key Logger, Buffer Overflow and Intrusions are also being explored.

(ii) Secure Cloud Storage Management: Security issues related to cloud storage systems include security for the Hadoop framework. Storage infrastructure which integrates resources from multiple providers to form a massive virtual storage system is being developed. When a storage node hosts the data from multiple domains, a VM will be created for each domain to isolate the information and corresponding data processing. Since data may be dynamically created and allocated to storage nodes, it is necessary to support secure VM management services such as pool management. The VM will be created dynamically to host data and support processing for each domain. The thread pool concept is leveraged to create VM pools. The VM pool will grow and shrink according to the demands and resource constraints. Implementations of the virtual global cloud storage infrastructure on top of Xen and VMware are also being carried out.

(iii) Secure Cloud Data Management: Various security issues related to cloud data management include cloud query optimization and query rewriting. For example, secure query processing algorithms for RDF (Resource Description Framework) data in clouds with an XACML-based (eXtensible Access Control Markup Language) policy manager utilizing the Hadoop/MapReduce Framework have been developed. In addition, algorithms for secure query processing based on the HIVE framework have also been developed. Some researchers are examining risk-aware access control query processing strategies for cloud computing as well as QoS for clouds.

Secure Cloud Network Management: The potential impact of network-based security threats on cloud computing systems and their hosted applications makes cloud computing systems prime targets for adversaries. Securing such systems requires a multi-level approach as potential security threats may come from various entities both internal as well as external to the system. In addition, a potential security attack on a hosted service application may also have a negative impact on other co-located services or applications.

Security Policy Management for Cloud Computing: Researchers are examining various types of policy management in cloud systems. In

addition results from the in-line reference monitor concept as applied to clouds is also being examined. For example, cloud frameworks often demand more sophisticated policy languages for fine-grained data confidentiality policies, accountability policies and identity management policies. To support such policies, customized OS's usually become necessary. Such OS's incur computational overhead, both in terms of resource consumption and process load-times. The need to customize the OS to support new policies introduces inflexibility to the policy language and could add to the trusted computing base of the system. To achieve more flexible, lighter-weight, yet high assurance protection for process-level cloud security, traditional hypervisor architectures are being extended with an extra level of security based on certified in-lined reference monitors (IRM's).

Cloud Monitoring: Data mining algorithms are being developed for malicious code detection and network traffic analysis for clouds. For example, some of these algorithms mine data streams and detect novel classes of malicious code. There are also tools being developed solely to monitor clouds. For example, for Infrastructure as a Service (IaaS) type of cloud computing applications, tools to monitor the utilization and load distribution in the underlying physical resources are being developed.

4. IDENTITY MANAGEMENT FOR THE CLOUD

In Section 2 we discussed identity management technologies while in Section 3 we discussed secure cloud computing technologies. It is increasingly being realized that effective identity management is critical for the secure operation of clouds. However, much of the recent work has focused on simple role-based access control models for secure clouds [3]. However, in a cloud environment, it will be difficult to define user roles across organizations. Furthermore, a user may have several identities not only over multiple systems, but also over multiple clouds. In this section we will explore some of the challenges.

In an article by Gopalakrishnan [4], the author argues that identity management in a cloud has to manage "control points in a dynamic composite decommissioned machines, virtual device or service identities." The services in a cloud may be dynamic in nature and therefore lifecycle management of the identities need to take into consecration aspects such as service provisioning and de--provisioning, Standards such as Service Provisioning Markup Language that is being used for web services provisioning have to take into consideration features of cloud such as real-time resource allocation. User-IDs in a cloud will be dynamic and therefore technologies such as OpenID have to be extended to function in cloud. One of the major challenges in adapting OpenID is establishing trust relationships in the cloud. Therefore, an appropriate trust model for the cloud is crucial.

As stated earlier, simple role-based access control (RBAC) is too limiting for a cloud environment. Attribute-based access control (ABAC), upon which standards such as XACML is based on, is widely adopted for service-oriented systems. However, in a recent article on identity management by Olden [5], the author argues that both RBAC and ABAC may not be suitable for the cloud. This is because in a cloud environment, "attributes and role memberships are decoupled from the operating systems and can be distributed across systems via a federation". With respect to authentication, technologies such as single sign on with Security Assertions Markup Language have received a lot of prominence in a service-oriented environment. However, the cost of building an identity infrastructure based on SAML has been extensive and its use in a cloud has to be examined.

Other features that an identity management system must provide are auditing and accountability. Auditing in a distributed environment comes with numerous challenges such as how much audit data to collect and techniques for analyzing the data. In a cloud environment, auditing becomes even more challenging, especially in a public cloud. One of the objectives of the cloud is to dynamically allocate the resources regardless of who has requested the resources and where resources may be. Therefore, in such an environment, capturing all of the activities of all of the users of the cloud becomes a challenge.

Semantic web technologies such as web ontology language are being examined to store and reason about the identities. For example, ontology alignment techniques are being explored to determine whether multiple identities of a user can be aligned. For example in [6], the authors proposed the SemID ontology for identity management. This ontology represents roles, policies, and access rules to control access to the resources. Such approaches are yet to be examined for the cloud.

Cloud services are often managed by providers from different domains. For a cloud user, which services she invokes and the sequence of invocation might be sensitive, which may not be suitable to be discovered by service providers, even when they

collude. Thus, identity management for cloud computing should support certain forms of privacy-preserving resource access, which, on the one hand, allows service providers to authenticat an entity and control its access according to policies, and meanwhile, provides isolation of service access traces to prevent linkage, if a user chooses to do so. This feature also needs to be balanced with auditing and accountability requirements of cloud computing to prevent misuse of cloud services.

In summary, what is needed for identity management for the cloud is a trust model that handles (i) various trust relationships, (ii) access control policies based on roles and attributes, (iii) real-time provisioning, (iv) authorization, and (v) auditing and accountability. Furthermore the identity architecture has to be integrated into the cloud architecture. Several technologies have to be examined to develop the trust model; these include service-oriented technologies, standards such as SAML and XACML, and identity management technologies such as OpenID. Finally, does one size fit all? That is, can we develop a trust model that will be applicable to all types of clouds such as private clouds, public clouds and hybrid clouds?

5. STANDARDS FOR IDENTITY MANAGEMENT IN CLOUDS

While W3C and OASIS have developed several standards for identity management, web services, XML, XACML, SAML and semantic web, standards for secure clouds are only in the beginning changes. However, OASIS has recently formed a technical committee (TC) on identity management, privacy and trust in cloud computing services. As stated in [7], "*The OASIS IDCloud TC works to address the serious security challenges posed by identity management in cloud computing. The TC identifies gaps in existing identity management standards and investigates the need for profiles to achieve interoperability within current standards. It performs risk and threat analyses on collected use cases and produces guidelines for mitigating vulnerabilities*."

The TC has stated that they will be using the various OASIS standards as their building blocks to developed identity management standards for the cloud. These building blocks include the following. Digital Signature Services, Extensible Resource Identifier (XRI), and the XRI Data Interchange. The use case categories examined by this TC include infrastructure trust establishment, Infrastructure identity management, federated identify management, authentication, authorization, account/attribute management, security tokens and audit and compliance.

6. SUMMARY AND DIRECTIONS

As we have stated in this paper, cloud computing technologies have been rapidly adopted by organizations to lower costs and to enable flexible and efficient access to critical data. As these new cloud technologies emerge, cyber security challenges associated with these technologies have increased at a rapid pace. Some efforts have been reported on the solutions to address the security problems for the cloud. However identity management for the cloud has received very little attention.

In this paper we have discussed identity management developments including the various concepts such as single sign-on, and opened as well as discussed security issues for cloud computing. We then discussed some of the challenges on providing identity management for the cloud. In particular we discussed identification and authentication, authorization and access control, auditing and accountability, and trust and privacy issues for the cloud. We also discussed the goals of the newly formed OASIS technical committee on identity management. As more progress is made by this technical committee and more research is carried out on this topic, we can expect promising solutions to the very difficult problem of identity management in the cloud.

REFERENCES

[1] http://www.opengroup.org/idm/

[2] Federated Identity, http://en.wikipedia.org/wiki/Federated_identity

[3] K. Hamlen et al, Security Issues for Cloud Computing, Journal of Information Security and Privacy, 2010.

[4] A. Gopalakrishnan, Cloud Computing Identity Management, SETLabs briefings, 2009

[5] E. Olden, Architecting a Cloud-Scale Identity Fabric, IEEE Computer, March 2011.

[6] M. Choudhury et al, SemID: Combining Semantics with Identity Management, SECUREWARE, 2009/

[7] http://www.oasis-open.org/committees/tc_home.php?wg_abbrev=id-cloud