Authentication in the Cloud: A Risk-based Approach

M.T. Dlamini\textsuperscript{1,2}, H.S. Venter\textsuperscript{1}, J.H.P. Eloff\textsuperscript{1,2} and Y. Mitha\textsuperscript{1}

Department of Computer Science
University of Pretoria\textsuperscript{1}, Pretoria 0002
Tel: +27 12 420 3035

and SAP Research Internet Applications & Services, Africa\textsuperscript{2}
Tel: +27 12 999 9100, Fax: +27 12 999 9131
email: [mdlamini, eloff, hventer]@cs.up.ac.za\textsuperscript{1}; [moses.dlamini, jan.eloff]@sap.com\textsuperscript{2}

Abstract—Most companies are moving their data and applications to the cloud in order to exploit the numerous benefits that this computing paradigm presents. Yet, there is still insufficient research on how user authentication is to be handled on cloud computing environments. Cloud computing challenges the way people think about authentication and how to manage user identity across multiple domains. Hence, this paper outlines the requirements for user authentication and handling identity on the cloud. It goes further to discuss real world scenarios that illustrate the multi-faceted nature of handling authentication within a cloud environment. The main contribution of this paper is our proposed cloud-based authentication architecture. Our architecture makes a proposal on how to provide flexible, robust and scalable authentication by taking a risk-based approach to user authentication on cloud environments.

Index Terms—Authentication, Cloud Computing, Federation, Identity, Identity Theft, Single Sign-On

I. INTRODUCTION

Cloud computing is a disruptive technology breakthrough that has greatly changed the way business is conducted in the 21st century. Cloud computing has created a new business environment that presents today’s organizations with a new generation of mobile and always-on users. This generation of users require access to corporate data and applications from wherever they sit (on-premise or hosted externally), at all times, from any location, using any device. Cloud computing presents a platform that does exactly what this type of users require.

However, cloud computing comes with new security challenges that must be addressed before organizations move on to exploit the opportunities that lies therein. For example, traditional authentication approaches such as Kerberos, Single Sign-On, Active Directory and Virtual Private Networks which have been extensively used in closed environments fall short and cannot scale in the new open cloud environment. Traditional authentication approaches need some tweaking if they are to work in the new environment, especially when we consider that in this new environment, users must be authenticated not only within their organizations boundaries but also beyond; i.e. across multiple domains where some of their organizations’ data and applications might be hosted.

The above challenge coupled with the ever increasing problem of identity theft is a major cause for concern, more so for organizations that are moving or planning a move to the cloud. Furthermore, cloud computing is provided via a hostile and vulnerable Internet backbone which cannot be trusted. In addition, mobile computing has also increased the number of devices that people use to access critical corporate data and application on-premise or hosted off-premise.

Advancements in social media, cloud and mobile computing are playing a major part in the increase of ID theft cases. For example, ID thieves are now using social media profiles to circumvent security questions like; “what is your mother’s maiden name?” Financial institutions require their customers to provide the answers to such questions as an extra layer for authentication. Social media avails this information to ID thieves for free. Hence, the 13% increase in the number of identity theft cases between 2010 and 2011 with an estimated impact of more than R418 Billion annually [1].

Legal and compliance regulatory authorities are beginning to pay attention to the growing threat of identity theft [2]. Legal mandates like the South African Protection of Personal Information Act of 2009, SOX, HIPAA and PCIDSS have now made it a requirement for organizations to implement an effective identity management system and strengthen their authentication mechanisms. This is meant to establish a minimum set of requirements for lawful processing of PII in order to combat and counter the threats to identity.

In summary the following factors have heightened the need for organizations to re-think and re-work their user authentication strategies.

- Increasing legal and compliance demands,
- Ever growing threat of identity theft and continual loss of PII,
- Plethora of vulnerable end point devices,
- Complexity of managing identity in a multi-domain cloud environment.

This paper outlines the requirements for effective user authentication on the cloud. Furthermore, this paper provides real world scenarios that illustrate the multi-faceted nature of handling user authentication within a cloud environment, and it goes further to propose and illustrate how strong authentication could be achieved within this environment.

The rest of the paper is structured as follows: Section II discusses relevant related work. Section III outlines the requirements for handling identity and access on the cloud.
environment gleaned from literature. Section IV describes some scenarios that could work in cloud computing environments. Section V briefly discusses federated identity architecture. Section VI concludes the paper and provides future research direction.

II. RELATED WORK

This section provides a review of existing authentication solutions for open cloud and other distributed systems. Authentication is defined as the process of checking to see if user credentials such as username and password are valid [5]. This validity check is a prerequisite for all access to resources. Research on this subject has been limited [11].

Cloud computing is similar to other traditional open distributed systems such as grid computing, service oriented architectures, peer-to-peer and client-server architectures [3]. Hence, it makes sense to discuss authentication on the cloud in relation to other traditional distributed systems.

In a grid computing environment, the grid security infrastructure (GSI) employs a local credential provider which supplies the client’s credentials to any service provider. The service provider requests client credentials from a trusted credential provider every time it has to authenticate a client [3]. This implies that mutual trust must be established between the client and the service provider beforehand [4].

The approach taken in grids is much better that the idea of a centralized Certificate Authority (CA) that has been extensively used in SSL (Secure Socket Layer) which suffers from a single point of failure i.e. the CA. This problem could be addressed through the elasticity of the cloud by initiating multiple instances of the CA as the need arises. This approach also suffers from the fact that mutual trust must be established beforehand and it limits its applicability to multiple domains similarly to cloud computing.

In terms of authentication in the cloud [5] argues that organizations must address challenges such as credential management, strong user authentication, and delegated authentication across all the types of cloud delivery models (public, private and hybrid) and service models (Saas, PaaS and IaaS – Software as a Service, Platform as a Service and Infrastructure as a Service respectively). Each of the delivery platforms and service models will have different requirements.

For example, within a public cloud existing user credentials for OpenID or Yahoo ID can be used to gain access to other public clouds [5, 9]. However, the same cannot be done on an externally hosted private cloud that holds critical business applications which require constant auditing. Hence, authentication on the cloud cannot be done the same way as other distributed systems. It requires a different mindset altogether.

Several researchers agree that cloud computing requires a new approach towards authenticating users [6, 10, 11]. The work of [10] assumes that strong user authentication in this environment could be achieved by just concatenating passwords in multiple levels i.e. organization, team and user level. This approach of just stacking passwords could work well within an on-premise cloud setting or in a public cloud that requires basic authentication measures. Passwords in their literal sense provide the first line of defense and nothing more. They have been broken a number of times through brute force and impersonation and, multiple layers of passwords cannot cope with the level of authentication required in hosted private cloud.

Unlike the work of Dinesha & Agrawal [10], and Chow et al. [6] extend the traditional authentication paradigm to include behavioral patterns or habits (which they call implicit authentication) beyond the traditional user credentials”. They leverage on mobile devices for their rich behavioral data to enhance cloud-oriented authentication techniques. They argue that past behavior put together with recent behavior can help compute a probability score which can be used to determine if a client device is in the legitimate hands of its rightful owner to make an authentication decision.

The major shortcoming of this work is related to the fact that cybercriminals could forge or manipulate the collected context information. This view is supported by the work of [7] which also uses behavioral patterns to authenticate users. Even though this work collects the history of behavioral patterns using IP addresses as opposed to the mobile device location and behavioral data, they acknowledge that users will be authenticated as long as the system does not pick up discrepancies on their behavioral data. The work of [6] does not explain some scenarios such as when the context does not match. This might be a loophole that could potentially lock out even legitimate users just because their context has changed or does not correlate with their past behavior.

Even though not convincing, this work does allude to the fact that different cloud deployments and service models have different risk profiles which require different authentication levels that are in line with the sensitivity of the applications being accessed.

Discrepancies in user context are more likely with the new mobile workforce that demands always-on access to corporate data and apps. These employees work flexible hours and from anywhere, which means that they could log on to their corporate portals at any given time of the day, using whatever devices, and from anywhere in the world. This approach takes another dimension if we consider the BYOD (Bring Your Own Device) trend where organizations do not provide any device but encourage their employees to bring their own device to do their work and are compensated for it. The complexity of working with user context for authentication decisions will start to increase dramatically.

Choudhury et al. [8] propose a strong authentication framework for cloud computing. On the client side, they use a smartcard and hashes the user credentials with a one-way hash functions to request access. The cloud service provider uses an Out Of Band (OOB) channel to deliver one time key which is sent as an SMS via a mobile network server back to the client for access to the resource. According to the authors this system is not susceptible to impersonating, replay, DDoS and man-in-the-middle attacks. This work [8] takes a good approach towards providing strong authentication for the cloud. However, the introduction of a mobile network server provides a single point of failure. If the mobile network server goes down the whole system comes to a stand-still. The scalability and reliability properties of the cloud can rectify this by replicating the
mobile network server. Another solution would be to consider a secondary software-based one time key generator. Software-based tokens ease the burden of carrying an extra gadget.

Fugkeaw et al. [11] propose the integration of strong multi-factor authentication with the SAML (Security Assertion Markup Language) authentication protocol to support Single Sign On federation in multiple distributed domains [12]. The SAML protocol works with the concept of an Identity Provider (IdP) and Service Provider (SP) to federate client authentication in multiple domains. The IdP is a source site that authenticates and transmits client profiles within SAML assertions to the Service Provider (SP) which holds applications or cloud services that the client requested [5]. Figure 1 below illustrates the SAML validation process.

![Figure 1. SAML Assertion Validation Process](image)

The client starts by sending an access request to a resource hosted in the SP. The SP redirect to the assertion back to the client, which then seamlessly relays it to the IdP for authentication. The IdP authenticates the client and sends the response back to the client using the assertion, which autonomously relays it back to the SP. The SP verifies the assertion and access to the requested resource is finally granted.

The SAML authentication protocol is widely used for both commercial and open source federation products. Other standards-based protocols that enable identities to be federated include OAuth [14] and OpenID [9].

In terms of authentication on cloud systems, SAML [12], OAuth [14] and OpenID [13] can be applied in different cloud deployment and service models. For example [5] argues that the OpenID SSO protocol is more user centric and therefore it would be applicable in a public cloud environment as opposed to the SAML protocol which would be applicable to a private cloud environment that requires strong authentication for external cloud services.

In summary, most of the covered literature agrees that authentication on the cloud requires new insights that would make it robust and scalable to fit the new environment. The covered literature reflects on several ways on how this could be achieved, starting from concatenated password, to two-factor strong authentication with OOB token, and federated identity using the SAML protocol. However, most of this work fails to show the different risk profiles that could unfold in the cloud given its different deployment and service models. The work of [5] made some ground on this aspect but there is still room for improvement. This is where this paper will make attempts to close the gaps to advance the current state-of-the-art. It will go further to outline the requirements for effective authentication on the cloud. Building on the requirement and scenarios, this paper further shows how identity could be made robust and flexible to fit the cloud environment. The next section will briefly outline some of the scenarios for authenticating users in cloud environments.

III. SCENARIOS FOR CLOUD-BASED AUTHENTICATION

This section considers a number of scenarios for a user requesting a resource that lies either on a private or public cloud which are either hosted on-premise or externally. The scenarios illustrate how organizations could share resources within a cloud environment in a controlled and secure manner. The scope of this exercise covers only the private and public cloud environments, and it only applies to the SaaS service model.

A. User requests subsequent access to a resource within an internal private cloud using a registered device.

Considering all the variables, this user seems to be doing his daily task which requires that he access this particular resource. If the previous access requests were granted, the system should evaluate the risk to be low, and let the user have access without necessarily requiring a username and password.

![Figure 2. Risk Score A](image)

B. User requests access to a resource within an internal private cloud using a registered device for the first time.

Considering that the resource lies on premises and a registered user is using a known device. The risk evaluates to low, authentication with a username and password would suffice. In considering the context of the request (that this is a first request to the resources) the risk must be adjusted to medium low. Hence, we escalate the authentication approach by requesting the user to supply an OOB one time password.

![Figure 3. Risk Score B](image)

C. User requests access to a resources hosted in an external private cloud of a partner from within his company premises using a registered device.

Under these circumstances the external private cloud as a SP will create and send an assertion to the IdP, which could be a central Active Directory or a local one that resides within the company where this user is registered. The intricacies associated with placing the user directory is still to be discussed in coming sections when we propose our solution. The assertion should have all the fields that the SP requires to authenticate the user. The SP should do due diligence on its side to locate the context of the user (that is a priori behavior) and identify the conditions under which the IdP will the authenticate the user given the sensitivity of the requested resource and the level of security required.
The IdP should then authenticate the user accordingly and send back the assertion. The SP should then validate the assertion and if it meets the requirements, grant the user access to the resource. If validation fails, the SP should prompt the user for a one time password, after which it should grant assess.

**Figure 4. Risk Score C**

**D. User registered in the IdP make an attempt to access a resource hosted in an external private cloud from a external public cloud.**

The user initiates a request to the SP from an external public cloud. When the SP gets the request, it determines that it comes from a legitimate user of a trusted partner who is not in his/her usual place of work. The risk in this case is high and requires strong authentication levels.

**Figure 5. Risk Score D**

**E. User under the same conditions as scenario D uses an unregistered device from a new location.**

If the same user as depicted in scenario D decides to use a new device that has not been registered under his/her profile, and from a new location; the risk will increase to be very high. The level of authentication should be escalated accordingly. In such a scenario, the system must give the user limited read only access for a short time period only to publicly shared applications and data.

**Figure 6. Risk Score E**

This section discussed some of the different scenarios that could occur in cloud-based collaborative networks. These are not meant to be an exhaustive list of all the possible scenarios, but merely meant to illustrate the point. They reflect that authentication in this environment need to scale up and down depending on the user context and sensitivity of the requested resources.

Building from these scenarios, the next section lists some of the requirements that must be considered in implementing effective authentication in a cloud-based collaborative environment.

**IV. REQUIREMENTS FOR AUTHENTICATION ON THE CLOUD**

With the numerous benefits that cloud computing brings come additional risks that require new approaches to authentication. In this new environment, not only does authentication have to be done in a seamless manner and in accordance to the sensitivity of the data or application resources, but it also has to cater for all the different ways by which users could access resources using any device, from any location to get an always-on connectivity experience. Moreover, authentication in this new cloud environment must cater for access to resources not only for internal users but also to external users from other domains.

Below we briefly discuss the requirements for authentication on the cloud as gleaned from literature and the possible scenarios in the previous sections.

**A. Flexibility**

Authentication on the cloud must be flexible enough to cater for a wide variety of scenarios and use appropriate mechanisms to grant users access to shared resources in a controlled and secure manner. The authentication systems must be flexible to cater for the different cloud deployments (private, public and hybrid) and service (SaaS, PaaS and IaaS) models. These systems must be flexible to ensure that they cater for a diverse user base with different roles coming from different domains with varying security levels.

**B. Robustness and Scalability**

Authentication on the cloud must be made strong and robust enough to prevent unauthorized access to resources. Authentication must be made to scale according to the demand, sensitivity and user context and provide SSO (Single Sign-On). For example, access to public resources should not require intensive authentication mechanisms (username and password can suffice) and external access must be protected with the necessary authentication mechanisms (geo-location, device identity along with username and password would suffice). In terms of SSO, users must authenticate once and be able to access other resources based on the initial authentication. This is not to say that the SP must not require further authentication attributes. For example, a user who is authenticated within his/her internal private cloud with just a username and password should not be allowed to use only these credentials to access sensitive applications from an external private cloud SP.

**C. Context-aware**

Authentication in the cloud must be able to factor in the context of the user who is requesting access to make the final decision to either grant or deny access to the shared resources. Such context might include access devices, geo-location or positioning and behavioral history, or successful and failed authentication attempts. The system must use such context to determine the level of authentication required for a specific user to access resources. The system must adapt to changing user context and act accordingly to grant and deny access as need arises.

**D. Based on Widely Adopted Standards**

Authentication in the cloud must be based on open and standardized protocols.

**E. Logging and Auditing**

All access should be logged for reporting purposes and for digital forensics investigations. Logging should be done in such a way that logs cannot be tempered within. The logs
must be stored in a digital forensic ready manner i.e. they must be correctly indexed, time-stamped, encrypted and stored in a separate virtual instance for auditing and in preparation for an investigation when the need arises.

V. FEDERATED IDENTITY ARCHITECTURES

This section briefly discusses federated identity architectures which have been widely used in different distributed systems to share identity information from multiple domains i.e. centralized and distributed systems. This paper will approach these two aspects in terms of cloud computing.

A. Centralized Identity Architecture

Within the cloud environment, a centralized architecture can represent two scenarios as discussed below i.e. a centralized identity provisioning or a centralized service provisioning.

1) Centralized identity provisioning

This approach has a global IdP where all user profiles are stored, managed, maintained and updated. All authentication requests are sent to this central IdP which then help validate users and then pass assertions to the relevant SP holding the requested resource. This type of an architecture is now being referred to as Identity as a Service (IDaaS) [15].

Market leaders in IDaaS include Intel, Okta, OneLogin, PasswordBank Technologies, and PingIdentity [15]. According to PasswordBank, IDaaS externalizes user authentication making it easy to control and monitor users activity [15]. This type of a system could be easily connected to an event and audit log repository. Gathering data on user activity from a central location can help provide harmonized and global audit trails of all the users on the system, facilitating fraud detection and digital forensic investigation.

However, this approach has its downsides. Similar to the other centralized systems, IDaaS systems suffers from a single point of failure. In addition, if corporate user identity is entrusted to an external cloud service provider, it is unclear as to who assumes liability for damages when unauthorized users access critical data, or the process by which organizations can prove their compliance to regulations in such instances.

2) Centralised service provisioning

In this approach all partners keep their user provisioning and de-provisioning internally and put their applications on a shared private or hybrid cloud to access at anytime. In such a setup, applications from multiple organizations share the same addressable space of a single cloud service provider – a concept called multi-tenancy in cloud computing terms.

This approach suffers from a single point of failure. A targeted DDoS to the SP can bring down the entire system. However, one can argue that the cloud provides the capability for replication which can help the system’s resilience. Moreover, there is a high probability of resource leakage when we consider that the resources reside in one addressable space and are only protected by virtualized infrastructure. Next we discuss the distributed architecture.

B. Distributed Identity and Service Provisioning

This approach represents a loosely coupled cloud computing environment where user profiles are provisioned and de-provisioned within internal private/public clouds and shared across domains using open standard protocols such as SAML, WS-Federation, SOAP, OAuth and OpenID. The source IdP is responsible for their user identity information, authentication of the users and federating user identity. Moreover, each IdP in this case also acts as a SP and control all access to their on-premise applications. This approach gives control back to the organizations that own the applications and host the user profiles. This approach provides the required flexibility to enable IdP/SP to effectively control access to their applications with the relevant authentication metrics.

The problem with this approach is the complexity of distributing identity and applications. It becomes very difficult to log and audit user activity on each SP/IdP if one does not have a global view of the entire system.

In summary, the above architectures come with both advantages and disadvantages which might make them applicable in some scenarios and not applicable in others.

VI. PROPOSED APPROACH

In light of the above related work, multi-faceted cloud scenarios, requirements, and both the advantages and disadvantages of the three architectures, this paper considers the distributed identity and service provisioning architecture. This architecture provides the flexibility, robustness and scalability that is required. This architecture can be best supported by the widely used standardized SAML 2.0 protocol. Below, we discuss our detailed cloud-based authentication architecture as shown in Fig. 7.

In this architecture a user within an internal private cloud uses one of his/her registered devices and use it to log on to the system. If the picked device is not registered the user must register it under his/her profile with its unique ID, which could be a MAC address or IMEI (International Mobile Equipment Identification). Before the user finalizes the device registration, he/she will be prompted to enter an OOB token after which the device will be registered and linked to the user profile by the user directory. This information will be relayed to the authentication engine to add on to the user context. Once the device is registered, it can now be used to log on to the system.

Basic login requires a username and password along with a device ID which is detected automatically by the system without the user having to worry about remembering it on top of their login credentials. This information is validated by the authentication engine and passed to the federation engine in an SAML assertion. All communications happen within a secured communication channel.

Once authenticated, the user is presented with all the links to the resources that he/she has access to. If he/she decides to click on a resource that is held internally, the resource server verifies that the user has the relevant role, permissions and access rights to the resource. This is the authorization part.

Authentication is decoupled from authorization, even though each one of these cannot work without the other. Authorization is outside the scope of this paper.

In case the user requests a resource that lies within an external private cloud, the system will contact its federation
Authentication

Authentication and logging will be discussed. Once that token is as requested, the authentication engine is hereby acknowledged. Opinions expressed and conclusions arrived at are solely those of the authors and should not necessarily be attributed to the University of Pretoria, the ICSA research group or SAP Research IA&S.

Future work will discuss the implementation of this framework and its application to mobile users. In Proceedings of the 2010 ACM workshop on Cloud computing security workshop (CCSW '10). ACM, New York, NY, USA, 1-6.

REFERENCES

Moses Dlamini received his BSc from the University of Swaziland. He received his Honours and MSc Computer Science from the University of Pretoria. Moses works at SAP Research Internet Applications & Services. His research interest lies in security of cloud computing, collaborative business networks, Internet of Things and Services.