Conceptual Cloud Computing Employing Identity Matrix and Knowledge Warehouse

Sarah Shafqat[1], Muhammad Naeem Ahmed Khan[1] and Qaisar Javaid[2]

[1]Shaheed Zulfiqar Ali Bhutto Institute of Science and Technology (SZABIST), Islamabad, Pakistan
[2]International Islamic University (IIU), Islamabad, Pakistan

sarah.shafqat@gmail.com, mnak2010@gmail.com, qaisar@iiu.edu.pk

Abstract

This paper presents a concept for devising a framework for Cloud Intellect (CI) to help organizations in curtailing data processing costs by outsourcing computations, maintaining confidentiality and integrity of the data sources, and computation processes. The prototype comprises a number of modules – Identity Matrix (IMx), Knowledge Warehouse among them. The architecture is designed in a way that it offers processes a virtual outlook of a network interface to support user level access to high-speed communication devices, like biometric devices, to support user level identification system. As the target audience are scrutinized and their corresponding interest groups are better understood, the framework further advances towards making a systematic data warehouse depicting meaningful information for public and private sector organizations and corporate sector etc. The right information would be directed towards right people at the right time; thus helping organizations to take right actions based on global network of trusted peers, as well as eliminating risks posed by identified malicious users. The main focus of this study is on software architecture.

Keywords: *Cloud Computing, Cloud Intellect, Architectural Framework, Knowledge Management*

1. Introduction

Cloud computing, envisioned as the next-generation architecture of IT Enterprise, is a new area of research that carries certain ambiguities pertaining to its potential use in the modern era where everyone is much concerned about privacy and security of digital data. Cloud computing utilizes dynamic, scalable and often visualized resources to provide computation and data management services over the Internet. The cloud users generally use HTTP as a medium to avail these services. In cloud computing, the knowledge, expertise or control over the technology infrastructure is ordinarily not needed by the user. The term “cloud computing” is used with relevance of diagrammatical representation of the cloud symbol which is often used to depict the Internet inflow charts. Thus, cloud is an abstraction symbol to represent the underlying complex infrastructure. In cloud computing, the provider builds solution (i.e., software, infrastructure or platform) over the Internet and these services are made available to the users on demand via subscription (Bhaskardeep, 2010).

Theoretically, there are three types of cloud. *Private clouds* are limited within an organization or a business having its own internal datacenter. Therefore, *private clouds* are not considered a part of cloud computing. The Internet services used by the general public worldwide is known as *public cloud. Hybrid cloud* is a mix of public cloud environment and isolated resources on private and controlled premises. In this type of cloud, certain aspects of IT infrastructure consisting of storage and computation devices etc. usually run on public cloud, while rest of classified IT infrastructure resides on secure and private premises. A few technologists believe in promoting the transfer of gigabytes of operational data on cloud to fully exploit business intelligence, but a large number of IT professional find the hybrid approach more practicable.

Cloud architecture works best with hybrid infrastructure as it offers a mix and match choice for keeping certain resources public and rest in the isolated environment, but it would incur a sunk cost and difficult to scale. In such a cloud architecture approach, applications and data are placed on the more appropriate platforms and are configured for processing. Hybrid computing substantiates the fact that not all the resources can be made public. Additionally, compliance, performance requirements and security concerns demand for a private platform. Cloud computing is categorized into three parts:- Software as a Service (SaaS) - Top Layer, Platform as a Service (PaaS) - Middle Layer, and Infrastructure as a Service (IaaS) - Base Layer.

SaaS refers to the top layer of cloud that directly interacts with customers. SaaS is a perfect solution for small and medium size businesses/organizations. Some of the SaaS examples include: SalesForce CRM (salesforce.com), Google Apps, monster.com (recruitment services) and mail.com.

PaaS comes in the middle of cloud stack and it mainly provides a platform to the developers and technical savvy individuals. Traditionally, a web developer would have to procure VS.net developer, SQL Server, Oracle licenses or third party reporting tools in order to develop and deploy software on a server within the fixed budget.

In such scenarios, using PaaS on cloud could be more appropriate choice as there would be no need to invest millions of dollars for making project environment ready. PaaS provider delivers the computing environment on the web and most of the time, web browser is sufficient to perform work. The simplicity of this platform enables small and medium-size companies or individual developers to launch their own SaaS application. PaaS is cost efficient solution in most of the software development scenarios. Examples of cloud OS are Google App Engine and Windows Azure. Cloud middleware consists of OrangeScape (orangescape.com) and Wolf PaaS (wolfframeworks.com).

Cloud stack has IaaS as its base layer. It serves as the founding layer for the other two layers to execute. *Virtualization* is the buzzword to denote this stack and Amazon EC2 (Elastic Compute Cloud) is an example of it as an instance of virtual computer executes Amazon EC2 application. An instance of virtual computer is initialized in accordance with the user’s choice/requirement while selecting particular configuration of CPU, storage and memory to run user applications. The cloud infrastructure made available on choice by IaaS Provider includes routers, servers, hardware based load-balancing, storage and firewalls etc. (Singh, 2010).

The prominent risks associated to the cloud computing adoption include unauthenticated user level access to data, identity theft and fraud, cyber bullying, weak network structure that gives access to eavesdropping on confidential data, Latest technology facilitates service providers to unite their infrastructure to address a broader business space. It is also possible that a consumer maintains accounts with multiple service providers like e-bay, Gmail etc. The visibility and scope of attributes for every identity needs to be verified against a central trusted policy regulatory authority assumed by the systems. In such systems, extreme precaution is required to handle identities in order to avoid untoward incidents. Hence, identity management (IDM) holds an upper hand in the whole area of cloud security and can be considered as superset of all the issues encompassing cloud computing. An IDM in cloud has to manage control points, dynamic composite/decommissioned machines, virtual device or service identities etc. Today’s cloud requires dynamic governance of typical IDM issues like provisioning/de-provisioning, synchronization, entitlement, lifecycle management etc. (Bertino et al., 2009).

2. Literature Review

Cloud computing has become an area of extensive research in the recent past. There has been ongoing research to provide authorization service for group collaboration in a community. There are distributed communities of resource providers and consumers that employ complex and dynamic policies to govern the utilization of resource for a specific purpose. Pearlman et al. (2002) proposed a scalable method for enforcing these policies and built a data management application to allow resource providers to delegate some fine-grained access control authority to communities while maintaining control over the resources. Kamvar et al. (2003) developed algorithms for devising a solution for reputation management in peer-to-peer (P2P) networks and *EigenRep* (Kamvar et al., 2003) is one of the proposed solutions. As P2P networks have gained attention for the purpose of file sharing, yet they have opened the door for malicious anonymous attacks that found almost ideal environment for the spread of self-replicating inauthentic files. *EigenRep* offers a solution for a distributed and secure reputation mechanism for global computing based on *power iteration*. When selected peers use this reputation based system, the system identifies malicious peers and isolates them from the network. This reputation system results in significantly decreasing the forged files being stimulated on the network even if malicious peers found the way to sabotage the system. However, *EigenRep* is only effective for decreasing unauthorized entrance to the system but it does not keep logs for identified malicious peers to impede them sabotaging the system.

Terra (Garfinkel et al., 2003) is a virtual machine based platform for trusted computing that provides a flexible architecture allowing applications to run parallel on commodity hardware with wide range of security parameters. While running on general purpose computing platform with normal applications, its semantics is to provide a dedicated, separate and tamper-resistant platform to the applications simultaneously. To achieve this, Terra uses Trusted Virtual Machine Monitor (TVMM) that partitions the temper-resistant hardware platforms into several isolated virtual machines on a single general purpose platform serving as multiple boxes. On TVMM, each VM contain functionality of an *'open box'* or a *'closed box'* modes. '*Closed box'* mode acts as opaque special-purpose platform that protects the privacy and integrity of its contents. The TVMM with associated hardware acts as a trusted party allowing closed-box VMs to cryptographically identify the software running on the remote clients through attestation mechanism. But, for a successful attestation, there is a limitation imposed by any hardware or software resource to either shutdown or temporarily become unavailable on the client side. Moreover, when it attests which hardware and software is in use at the client side, a security concern arises too as the system becomes vulnerable to external access.

Trust is of significant importance for business communities on the web. The certainty of getting informed about the quality of the product or service is greatly emphasized by Ackerlof (Guha et al., 2004). Ackerlof also showed how vital it is for seller to be trustworthy for gaining edge in the competitive market. Being trusted in the online communities is an important factor and a considerable amount of research is being conducted on propagating trust through networks. Despite the aforesaid advantages, still there is no conformance of user identity as a fraudulent user can easily forge a trusted identity and can enter into a trusted network of friends.

P2P file sharing networks – being much polluted by malware, decoys and other fraudulent contents – ordinarily combat with a newly built object-based decentralized reputation system called '*Credence*' (Walsh & Sirer, 2005). Without having any trusted entities in the network, it separates trustworthy contents from contaminated data. Problem encountered in file sharing network is common where clients get interacted with unknown peers. The analysis of emergent behavior of users is currently being done in such systems with the view of *Credence* approach besides file sharing. Though *Credence* can distinguish trustworthy and contaminated contents, but it does not guarantee the intrusion point or source of this pollution.

Security problems in the industry are delineated in (Zimmer et al., 2009) and a set of trust concepts with respect to security architecture are defined. The required roots-of-trust in hardware and firmware including TCG-based hardware and UEFI-based firmware are discussed in the literature. The comprehension of architectural goals is mandatory to understand the technology platform that deploys richer interoperability and greater trust in the system. IBM System-X is one such technology to deliver security architecture. Features like observed behaviors of entities, disk encryption, malware resistance etc. are assigned greater consideration in this regards.

Online P2P communities like orkut, facebook etc. equally bring both prospects and threats (Xiong & Ling). The threats are countered by using community based reputations for estimating the trustworthiness of peers. PeerTrust, a framework to support reputation-based trust, includes an eloquent model for generating and quantifying the trusted peers with transaction-based feedback system forming its base; and is preferred for decentralized implementation over a structured P2P network. Three basic trust parameters and two adaptive factors are the leading features in computing trustworthiness of peers. A trusted channel made between peers, the context of transactions performed by a peer, the feedback sources for peers' credibility, transactions performed, context factor for the community and a general trust metric fuse these parameters.

Jacobs & Poll (2010) advocate that e-passport combines the concept of using three identification technologies biometrics, smartcards and radio frequency. However, e-passport using biometric identification is limited to be used as a service and cannot be referred as a medium for authenticating users on cloud.

3. Proposed Architectural Framework and Solution

Cloud computing being the hot topic nowadays has grabbed attention of researchers but still it has not been made systematic. Nevertheless, researchers are putting their efforts to streamline its processes and control the channel of information flow and the user traffic that makes it complex and congested.

When a user surfs the cloud and signs-up with different services, he/she ends up with multiple credentials and access permissions across different applications provided by different service providers. These fragmented user details present a challenge to the both users and service providers in the forms of synchronization of shared identities, security etc. Also, there is a chance for malicious users making multiple fraudulent identities to get unauthorized entrance to variety of services. Thus, it stimulates a strong desire for developing an intrinsic identity system that unambiguously identifies users and is trusted across the web and within the enterprises. Several service providers use different factors for authentication like account number, email ID, PayPal ID etc. Moreover, when transactions traverse multiple tiers of service hosted in the cloud, the semantics of the context of identity information has to be properly maintained, constrained and relaxed as per the specific needs of the users.

Until the identities are not associated to the published jargon, developing trustworthy meaning of the content remains a challenge in non-trusted relationships. Therefore, conceptualization of relating digital presence with physical personas in real life is intriguing for bringing trustworthiness in P2P networks in terms of end-to-end technical, information assurance and governance perspectives.

Architectural Framework

The problem areas related to architectural framework mainly cover the aspects of service outages, authenticity of information, auditability and intrusion detection. Detail description of these problems/issues is briefly discussed below.

a. Hardware or software resources are frequently shutdown or restarted for attestation purposes resulting in temporary service outages on the client side. The attestation process at client side may make the client vulnerable to threats as the client system is accessible at that particular time for attestation.

b. The authenticity of information is judged by intelligently reviewing and analyzing the opinions of many users in order to take appropriate decisions related to data security.

c. To date, there is no auditability mechanism in place for validating a user identity. A fraudulent user can still forge a trusted identity and get access to a trusted network of friends.

d. Although *Credence* can identify between trustworthy and polluted contents, but it does not trace back the intrusion point or source of pollution.

e. The identification system of *IBM System-X* despite bearing high potential is yet limited to firmware and can be merged with the cloud computing to bring about added trust.

f. Decentralization of *PeerTrust* is another limitation as, ideally, there should be a centralized system to authenticate activities of every user on the cloud. The fact that a single user ends up with multiple logins constricts the ease of use of cloud services due to difficulty faced by the user to remember and manage login details. Therefore, a single point of entry is needed to be introduced for cloud users.

g. Introduction of e-passport bearing biometric identification is limited to be used as a service as it is not referred as a medium for authenticating users on the cloud. Furthermore, every individual user on the cloud may not possess a passport which will limit his/her identification options.

Transparency is a key to implementing accountability and setting up an international regulatory body for authenticating identities over the Internet through biometric devices – e.g., fingerprint sensors or smart cards – and confirming identity card and passport details in order to form a single knowledgebase to mesh multiple identities of a user over the cloud.

For building Cloud Intellect (CI) on the existing cloud computing architecture, the key artifact which needs to be transformed through reengineering of cloud infrastructure using reverse engineering techniques is the Architectural Framework. Keeping in view the key problem area related to the formation of Cloud Intellect (CI), we endeavor to present a model that amicably addresses the aforesaid problem. The salient features of the proposed architectural framework are:

a. There is a need to introduce a virtual layer of cloud that encompasses every machine on the network attested for its software and hardware. Even when a machine is temporarily down, its instance still appears on the cloud that is trusted for its security through a generalized script composed of *Credence* in order to combat identified pollution and untrustworthy content on the virtual network. The script also includes intrusion detection mechanism to identify the intrusion point of malicious contents and halt it from the source point.

b. For Cloud Intellect (CI), the authenticity of information is judged by analyzing and intelligently reviewing the opinions of many users to take various decisions. For making intelligent decisions, there should be a single instance of user that should operate anywhere on the cloud having different profiles linked up at a single point of occurrence. The users would have to physically attest to biometric device and provide their authorization credentials for further validation by the IRB. The proposed approach would preserve the auditability of user identity, reduce the use of forged identities to almost nil and help build trust in the P2P network.

c. To enable biometric attestation, the identification system of *IBM System-X* has to be molded for inclusion of every computing device that connects user to the cloud.

d. Identity Matrix (IMx) would give a single platform to users for access to the cloud and would help eliminate decentralization of *PeerTrust*. A user would have only one login that would keep all his/her profile linked to a single point. This way the user would not have to manage multiple accounts using several types of identification mechanisms like paypal, ebay etc.

e. Introduction of e-passport using biometric identification is limited to be used as a single service; therefore, it has not been referred as a medium for authenticating users on the cloud. However, e-passport can be linked with IMx to validate users having passports other than identity card details. Such an approach would also restrict minor’s access to adult websites. Ultimately, the proposed architecture would be *hybrid* in nature.

4. Proposed Design of Architectural Framework

High level DFD is made to better elaborate the functionality of the proposed Framework and is described at length in the subsequent section.

Context Diagram

As mentioned earlier, IMx is one of the modules of which Cloud Intellect is comprised of. Thus IMx is used to develop the main system. It is connected to knowledge warehouse and IRB data warehouse when triggered by any cloud user who interacts with it (Figure 1).

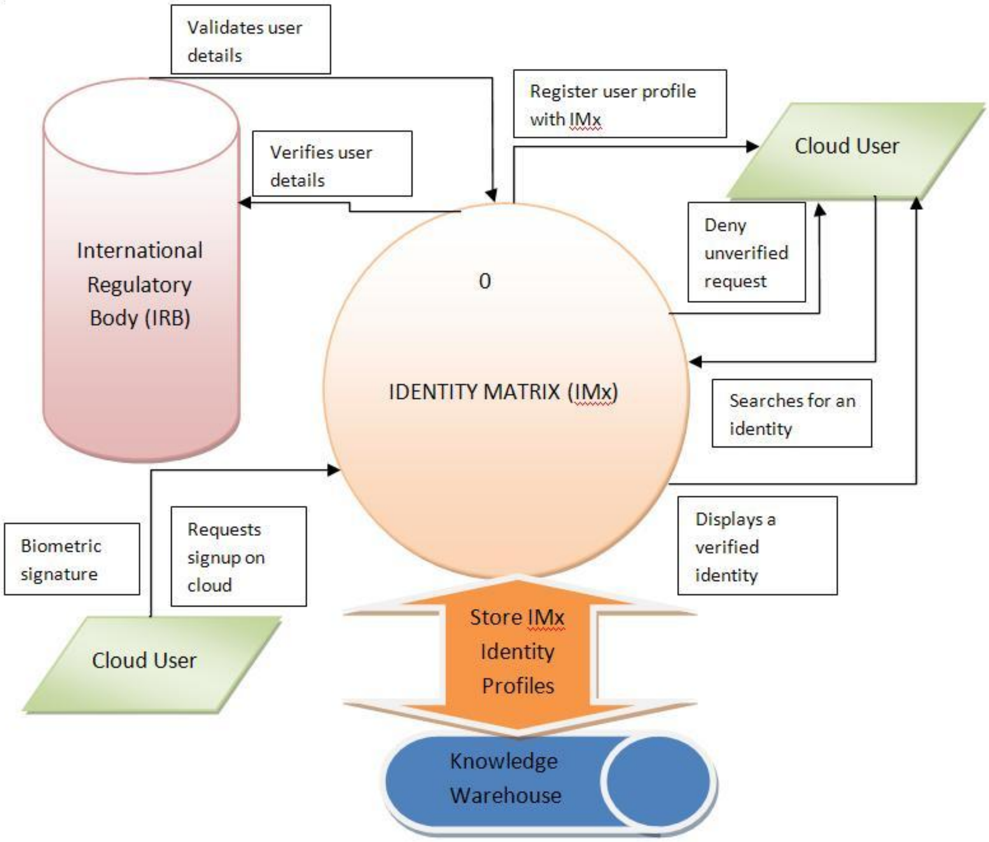


Figure 1. Context Diagram for Cloud

Proposed Architecture

Four cloud platforms – i.e., any cloud application, IMx, IRB and Knowledge warehouse – form a hybrid cloud multitude with different hosting service providers on separate virtual platforms. IRB symbolizes an isolated data warehouse where data obtained from identity card and passport issuers in different countries is stored. Cloud application refers to where user registers himself/herself giving some specific identity information like identity card and passport number and thumb impression etc. These set of values are verified by IRB. If the identities are validated then the user will be registered both on the cloud application and IMx automatically and the user's profile would be recorded in secure isolated knowledge warehouse where it could be retrieved as required. On IMx, any user's activities on the cloud can be viewed and kept transparent by maintaining the needed confidentiality. This is illustrated in Figure 2.

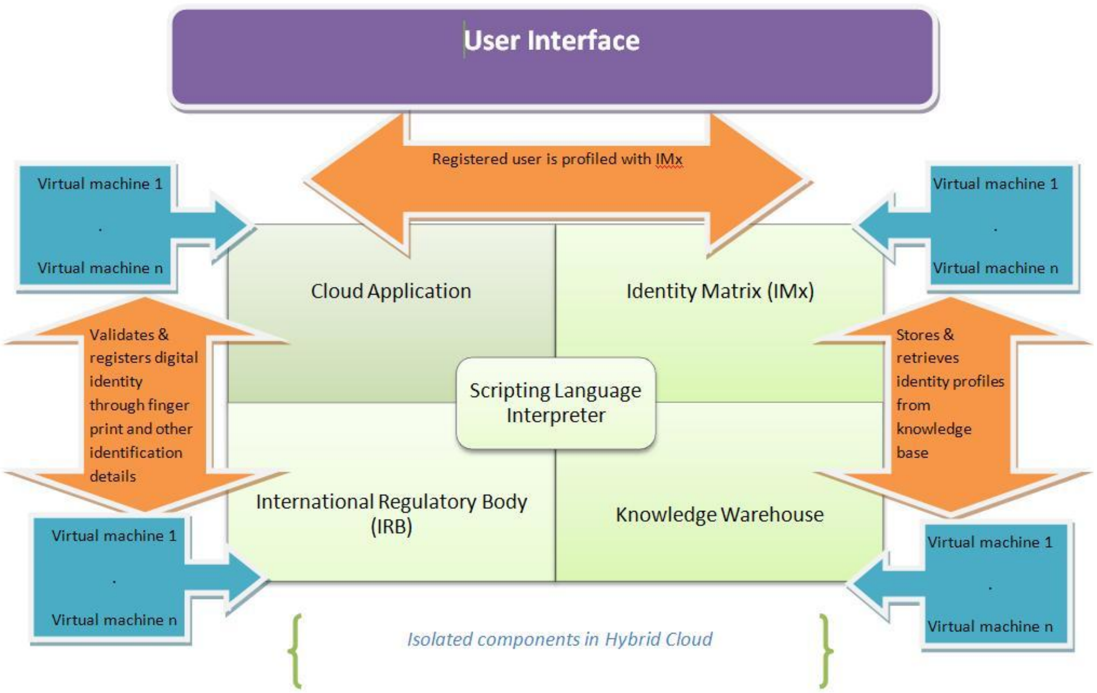

User Interface
Registered user is profiled with IMx
Virtual machine 1
.
Virtual machine n
Virtual machine 1
.
Virtual machine n
Cloud Application
Identity Matrix (IMx)
Validates & registers digital identity through finger print and other identification details
Scripting Language Interpreter
Stores & retrieves identity profiles from knowledge base
International Regulatory Body (IRB)
Knowledge Warehouse
Virtual machine 1
.
Virtual machine n
Virtual machine 1
.
Virtual machine n
Isolated components in Hybrid Cloud

Figure 2. Hybrid Framework for Cloud

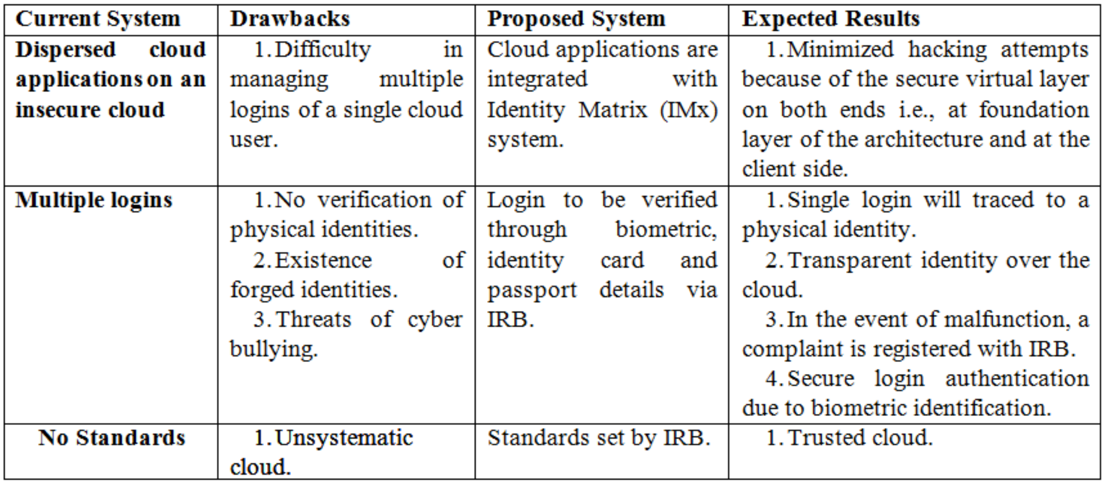
5. Validation and Discussion

Validation of the prototype model is done by comparing it with the existing cloud architectures. A comparative summary is provided in Table I. The validation is primarily based on three main aspects of the current systems, namely, dispersed cloud application on an insecure cloud, multiple logins and absence of standards. The foremost drawback with the dispersed cloud application on an insecure cloud pertains to addressing the issue of managing multiple logins for a single cloud user. In our proposed framework, this problem is addressed through integrating cloud applications with IMx system to support single sign-on. The proposed solution minimizes the chances of hacking by introducing a secure virtual layer at foundation layer and at the client side.

The second aspect of multiple logins in the existing systems suffers the drawbacks of lack of verification of physical identities, the existence of forged identities and possible threats of cyber bullying. In our proposed framework, we suggest to use biometric credentials or acceptable proofs of identification like identity card or passport details that are duly verified by IRB to resolve this issue. The key benefits obtained through this approach could be multifaceted. Firstly, a single login can be easily traced back to physical identity. Secondly, in the event of malfunctioning, a complaint can be registered with IRB to resolve the issue. In addition the proposed approach will support transparent identities over the cloud due to the mechanism of secure login authentication through biometric identification.

The third aspect observed in the existing systems pertains to the absence of standards due to which the cloud has become quite unsystematic. In our proposed architecture, we suggest to use IRB standards to transform the cloud into a trusted cloud.

Table 1. Comparative Summary



Current System	Drawbacks	Proposed System	Expected Results
Dispersed cloud applications on an insecure cloud	1. Difficulty in managing multiple logins of a single cloud user.	Cloud applications are integrated with Identity Matrix (IMx) system.	1. Minimized hacking attempts because of the secure virtual layer on both ends i.e., at foundation layer of the architecture and at the client side.
Multiple logins	1. No verification of physical identities. 2. Existence of forged identities. 3. Threats of cyber bullying.	Login to be verified through biometric, identity card and passport details via IRB.	1. Single login will traced to a physical identity. 2. Transparent identity over the cloud. 3. In the event of malfunction, a complaint is registered with IRB. 4. Secure login authentication due to biometric identification.
No Standards	1. Unsystematic cloud.	Standards set by IRB.	1. Trusted cloud.

6. Conclusion and Future Work

The development of evolving Cloud Intellect (CI) comprises several modules, mainly Knowledge Warehouse. As the prototype is understood and for the evolution of Cloud Intellect (CI), there is a need for storage, compatibility, scalability and performance. Therefore, the future work is likely to focus on developing all the artifacts as shown in the conceptual model that can be integrated e.g., IMx, knowledge management, storage, virtualization, scalability and data security modules. These modules are envisaged to be based on robust algorithms to provide efficient mechanisms as well as adhere to the proper governance standards implemented by IRB by exploiting neural computing functionalities.

Knowledge management system is likely to be the key expected outcome of this research in the future. As the target audiences are scrutinized and their interest groups are better understood, the system would be designed in such a way that it could further make a systematic data warehouse depicting meaningful information for various government and non-government organizations, particularly those dealing with national security, e-governance, businesses etc. The proposed model will help disseminating the right information to the right consumers at the right time, and this tactic would help organizations to take right actions based on informed decisions through global network of trusted peers. The proposed approach will also eliminate risks posed by formerly identified malicious users.

References

[1] Bhaskardeep, “Cloud Computing – SaaS”, http://www.c-sharpcorner.com/UploadFile/bhaskardeep/190/, **(2010)** January 18.

[2] B. N. Singh, “Infrastructure as a Service: Cloud Computing”**,** http://www.techno-pulse.com/2010/04/infrastructure-as-service-iaas-cloud.html, **(2010)** April 4.

[3] E. Bertino, F. Paci, R. Ferrini and N. Shang, “Privacy-preserving Digital Identity Management for Cloud Computing”, Purdue University, **(2009)**.

[4] M. Linden, “Provisioning and deprovisioning in an identity federation”, **(2008)**.

[5] A. Gopalakrishnan, “Cloud Computing Identity Management”, SETLabs Briefings, vol. 7, no. 7, **(2009)**.

[6] L. Pearlman, V. Welch, I. Foster, C. Kesselman and S. Tuecke, “A Community Authorization Service for Group Collaboration”, **(2002)**.

[7] S. D. Kamvar, M. T. Schlosser and H. Garcia-Molina, “EigenRep: Reputation Management in P2P Networks”, Stanford University, **(2003)**.

[8] T. Garfinkel, B. Pfaff, J. Chow, M. Rosenblum and D. Boneh, “Terra: A Virtual Machine-Based Platform for Trusted Computing, Stanford University, **(2003)**.

[9] R. Guha, R. Kumar, P. Raghavan and A. Tomkins, “Propagation of Trust and Distrust”, IBM Almaden Research Center, **(2004)**.

[10] K. Walsh and E. G. Sirer, “Fighting PeertoPeer SPAM and Decoys with Object Reputation”, Cornell University, **(2005)**.

[11] V. J. Zimmer, S. R. Dasari and S. P. Brogan, “Trusted Platforms UEFI, PI and TCG-based firmware”, Intel, IBM, **(2009)**.

[12] L. Xiong and L. Liu, “PeerTrust: Supporting Reputation-Based Trust for Peer-to-Peer Electronic Communities”, IEEE Computer Society, **(2009)**.

[13] B. Jacobs and E. Poll, “Biometrics and Smart Cards in Identity Management”, Radboud University Nijmegen, **(2010)**.