Chapter 5

Federation, Presence, Identity, and Privacy in the Cloud

5.1 Chapter Overview

Building a seamless federated communications capability in a cloud environment, one that is capable of supporting people, devices, information feeds, documents, application interfaces, and other entities, is affected by the architecture that is implemented. The solution chosen must be able to find such entities, determine their purpose, and request presence data so that others can interact with them in real time. This process is known as discovery. Providing discovery information about the availability of various entities enables organizations to deploy real-time services and achieve significant revenue opportunities and productivity improvements.

The advent of on-demand cloud services is changing the landscape for identity management because most current identity management solutions are focused on the enterprise and/or create a very restrictive, controlled, and static environment. We are now moving into a new world, where cloud services are offered on demand and they continuously evolve to meet user needs. Previous models are being challenged by such innovations. For example, in terms of trust assumptions, privacy implications, and operational aspects of authentication and authorization, solutions that seemed to work before are now considered old, outdated, and clunky fixes to identity management. The fluid and omnipresent aspects of federation, presence, and identity in the cloud create new opportunities for meeting the challenges that businesses face in managing security and privacy in the cloud.

5.2 Federation in the Cloud

One challenge in creating and managing a globally decentralized cloud computing environment is maintaining consistent connectivity between untrusted components while remaining fault-tolerant. A key opportunity

for the emerging cloud industry will be in defining a federated cloud ecosystem by connecting multiple cloud computing providers using a common standard.

A notable research project being conducted by Microsoft, called the Geneva Framework, focuses on issues involved in cloud federation. Geneva has been described as a claims-based access platform and is said to help simplify access to applications and other systems. The concept allows for multiple providers to interact seamlessly with others, and it enables developers to incorporate various authentication models that will work with any corporate identity system, including Active Directory, LDAPv3-based directories, application-specific databases, and new user-centric identity models such as LiveID, OpenID, and InfoCard systems. It also supports Microsoft's CardSpace and Novell's Digital Me.

The remainder of this section focuses on federation in the cloud through use of the Internet Engineering Task Force (IETF) standard Extensible Messaging and Presence Protocol (XMPP) and interdomain federation using the Jabber Extensible Communications Platform (Jabber XCP),[1] because this protocol is currently used by a wide range of existing services offered by providers as diverse as Google Talk, Live Journal, Earthlink, Facebook, ooVoo, Meebo, Twitter, the U.S. Marines Corps, the Defense Information Systems Agency (DISA), the U.S. Joint Forces Command (USJFCOM), and the National Weather Service. We also look at federation with non-XMPP technologies such as the Session Initiation Protocol (SIP), which is the foundation of popular enterprise messaging systems such as IBM's Lotus Sametime and Microsoft's Live Communications Server (LCS) and Office Communications Server (OCS).

Jabber XCP is a highly scalable, extensible, available, and device-agnostic presence solution built on XMPP and supports multiple protocols such as Session Initiation Protocol for Instant Messaging and Presence Leveraging Extensions (SIMPLE) and Instant Messaging and Presence Service (IMPS). Jabber XCP is a highly programmable platform, which makes it ideal for adding presence and messaging to existing applications or services and for building next-generation, presence-based solutions.

Over the last few years there has been a controversy brewing in web services architectures. Cloud services are being talked up as a fundamental shift in web architecture that promises to move us from interconnected silos to a

1. Jabberwas acquired by Cisco Systems in November 2008.

collaborative network of services whose sum is greater than its parts. The problem is that the protocols powering current cloud services, SOAP (Simple Object Access Protocol) and a few other assorted HTTP-based protocols, are all one-way information exchanges. Therefore cloud services aren't real-time, won't scale, and often can't clear the firewall. Many believe that those barriers can be overcome by XMPP (also called Jabber) as the protocol that will fuel the Software-as-a-Service (SaaS) models of tomorrow. Google, Apple, AOL, IBM, Livejournal, and Jive have all incorporated this protocol into their cloud-based solutions in the last few years.

Since the beginning of the Internet era, if you wanted to synchronize services between two servers, the most common solution was to have the client "ping" the host at regular intervals, which is known as polling. Polling is how most of us check our email. We ping our email server every few minutes to see if we have new mail. It's also how nearly all web services application programming interfaces (APIs) work.

XMPP's profile has been steadily gaining since its inception as the protocol behind the open source instant messenger (IM) server jabberd in 1998. XMPP's advantages include:

- It is decentralized, meaning anyone may set up an XMPP server.
- It is based on open standards.
- It is mature—multiple implementations of clients and servers exist.
- Robust security is supported via Simple Authentication and Security Layer (SASL) and Transport Layer Security (TLS).
- It is flexible and designed to be extended.

XMPP is a good fit for cloud computing because it allows for easy two-way communication; it eliminates the need for polling; it has rich publish-subscribe (pub-sub) functionality built in; it is XML-based and easily extensible, perfect for both new IM features and custom cloud services; it is efficient and has been proven to scale to millions of concurrent users on a single service (such as Google's GTalk); and it also has a built-in worldwide federation model.

Of course, XMPP is not the only pub-sub enabler getting a lot of interest from web application developers. An Amazon EC2-backed server can run Jetty and Cometd from Dojo. Unlike XMPP, Comet is based on HTTP,

and in conjunction with the Bayeux Protocol, uses JSON to exchange data. Given the current market penetration and extensive use of XMPP and XCP for federation in the cloud and that it is the dominant open protocol in that space, we will focus on its use in our discussion of federation.

The ability to exchange data used for presence, messages, voice, video, files, notifications, etc., with people, devices, and applications gain more power when they can be shared across organizations and with other service providers. Federation differs from peering, which requires a prior agreement between parties before a server-to-server (S2S) link can be established. In the past, peering was more common among traditional telecommunications providers (because of the high cost of transferring voice traffic). In the brave new Internet world, federation has become a *de facto* standard for most email systems because they are federated dynamically through Domain Name System (DNS) settings and server configurations.

5.2.1 Four Levels of Federation

Technically speaking, federation is the ability for two XMPP servers in different domains to exchange XML stanzas. According to the XEP-0238: XMPP Protocol Flows for Inter-Domain Federation, there are at least four basic types of federation[2]:

1. **Permissive federation**. Permissive federation occurs when a server accepts a connection from a peer network server without verifying its identity using DNS lookups or certificate checking. The lack of verification or authentication may lead to domain spoofing (the unauthorized use of a third-party domain name in an email message in order to pretend to be someone else), which opens the door to widespread spam and other abuses. With the release of the open source jabberd 1.2 server in October 2000, which included support for the Server Dialback protocol (fully supported in Jabber XCP), permissive federation met its demise on the XMPP network.
2. **Verified federation.** This type of federation occurs when a server accepts a connection from a peer after the identity of the peer has been verified. It uses information obtained via DNS and by

2. Peter Saint-Andre, "XEP-0238: XMPP Protocol Flows for Inter-Domain Federation," http://xmpp.org/extensions/xep-0238.html, retrieved 1 Mar 2009.

means of domain-specific keys exchanged beforehand. The connection is not encrypted, and the use of identity verification effectively prevents domain spoofing. To make this work, federation requires proper DNS setup, and that is still subject to DNS poisoning attacks. Verified federation has been the default service policy on the open XMPP since the release of the open-source jabberd 1.2 server.

3. **Encrypted federation.** In this mode, a server accepts a connection from a peer if and only if the peer supports Transport Layer Security (TLS) as defined for XMPP in Request for Comments (RFC) 3920. The peer must present a digital certificate. The certificate may be self-signed, but this prevents using mutual authentication. If this is the case, both parties proceed to weakly verify identity using Server Dialback. XEP-0220 defines the Server Dialback protocol,[3] which is used between XMPP servers to provide identity verification. Server Dialback uses the DNS as the basis for verifying identity; the basic approach is that when a receiving server receives a server-to-server connection request from an originating server, it does not accept the request until it has verified a key with an authoritative server for the domain asserted by the originating server. Although Server Dialback does not provide strong authentication or trusted federation, and although it is subject to DNS poisoning attacks, it has effectively prevented most instances of address spoofing on the XMPP network since its release in 2000.[4] This results in an encrypted connection with weak identity verification.

4. **Trusted federation.** Here, a server accepts a connection from a peer only under the stipulation that the peer supports TLS and the peer can present a digital certificate issued by a root certification authority (CA) that is trusted by the authenticating server. The list of trusted root CAs may be determined by one or more factors, such as the operating system, XMPP server software, or local service policy. In trusted federation, the use of digital certificates results not only in a channel encryption but also in strong authentication. The use of trusted domain certificates effectively prevents DNS poisoning attacks but makes federation

3. http://xmpp.org/extensions/xep-0220.html, retrieved 28 Feb 2009.
4. http://xmpp.org/extensions/xep-0220.html, retrieved 28 Feb 2009.

more difficult, since such certificates have traditionally not been easy to obtain.

5.2.2 How Encrypted Federation Differs from Trusted Federation

Verified federation serves as a foundation for encrypted federation, which builds on it concepts by requiring use of TLS for channel encryption. The Secure Sockets Layer (SSL) technology, originally developed for secure communications over HTTP, has evolved into TLS. XMPP uses a TLS profile that enables two entities to upgrade a connection from unencrypted to encrypted. This is different from SSL in that it does not require that a separate port be used to establish secure communications. Since XMPP S2S communication uses two connections (bi-directionally connected), encrypted federation requires each entity to present a digital certificate to the reciprocating party.

Not all certificates are created equal, and trust is in the eye of the beholder. For example, I might not trust your digital certificates if your certificate is "self-signed" (i.e., issued by you rather than a recognized CA), or your certificate is issued by a CA but I don't know or trust the CA. In either case, if Joe's server connects to Ann's server, Ann's server will accept the untrusted certificate from Joe's server solely for the purpose of bootstrapping channel encryption, not for domain verification. This is due to the fact that Ann's server has no way of following the certificate chain back to a trusted root. Therefore both servers complete the TLS negotiation, but Ann's server then require's Joe's server to complete server Dialback.

In the trusted federation scenario, Dialback can be avoided if, after using TLS for channel encryption, the server verifying identity proceeds to use the SASL protocol for authentication based on the credentials presented in the certificates. In this case, the servers dispense with server Dialback, because SASL (in particular the EXTERNAL mechanism) provides strong authentication.

5.2.3 Federated Services and Applications

S2S federation is a good start toward building a real-time communications cloud. Clouds typically consist of all the users, devices, services, and applications connected to the network. In order to fully leverage the capabilities of this cloud structure, a participant needs the ability to find other entities of interest. Such entities might be end users, multiuser chat rooms, real-time

content feeds, user directories, data relays, messaging gateways, etc. Finding these entities is a process called discovery.

XMPP uses service discovery (as defined in XEP-0030) to find the aforementioned entities. The discovery protocol enables any network participant to query another entity regarding its identity, capabilities, and associated entities. When a participant connects to the network, it queries the authoritative server for its particular domain about the entities associated with that authoritative server.

In response to a service discovery query, the authoritative server informs the inquirer about services hosted there and may also detail services that are available but hosted elsewhere. XMPP includes a method for maintaining personal lists of other entities, known as roster technology, which enables end users to keep track of various types of entities. Usually, these lists are comprised of other entities the users are interested in or interact with regularly. Most XMPP deployments include custom directories so that internal users of those services can easily find what they are looking for.

5.2.4 Protecting and Controlling Federated Communication

Some organizations are wary of federation because they fear that real-time communication networks will introduce the same types of problems that are endemic to email networks, such as spam and viruses. While these concerns are not unfounded, they tend to be exaggerated for several reasons:

- Designers of technologies like XMPP learned from past problems with email systems and incorporated these lessons to prevent address spoofing, unlimited binary attachments, inline scripts, and other attack tactics in XMPP.
- The use of point-to-point federation will avoid problem that occur with multihop federation. This includes injection attacks, data loss, and unencrypted intermediate links.
- Using certificates issued by trusted root CAs ensures encrypted connections and strong authentication, both of which are currently feasible with an email network.
- Employing intelligent servers that have the ability to blacklist (explicitly block) and whitelist (explicitly permit) foreign services, either at the host level or the IP address level, is a significant mitigating factor.

5.2.5 The Future of Federation

The implementation of federated communications is a precursor to building a seamless cloud that can interact with people, devices, information feeds, documents, application interfaces, and other entities. The power of a federated, presence-enabled communications infrastructure is that it enables software developers and service providers to build and deploy such applications without asking permission from a large, centralized communications operator. The process of server-to-server federation for the purpose of interdomain communication has played a large role in the success of XMPP, which relies on a small set of simple but powerful mechanisms for domain checking and security to generate verified, encrypted, and trusted connections between any two deployed servers. These mechanisms have provided a stable, secure foundation for growth of the XMPP network and similar real-time technologies.

5.3 Presence in the Cloud

Understanding the power of presence is crucial to unlocking the real potential of the Internet. Presence data enables organizations to deploy innovative real-time services and achieve significant revenue opportunities and productivity improvements. At the most fundamental level, understanding presence is simple**:** It provides true-or-false answers to queries about the network availability of a person, device, or application. Presence is a core component of an entity's *real-time* identity. Presence serves as a catalyst for communication. Its purpose is to signal availability for interaction over a network. It is being used to determine availability for phones, conference rooms, applications, web-based services, routers, firewalls, servers, appliances, buildings, devices, and other applications. The management of presence is being extended to capture even more information about availability, *or even the attributes associated with such availability*, such as a person's current activity, mood, location (e.g., GPS coordinates), or preferred communication method (phone, email, IM, etc.). While these presence extensions are innovative and important, they serve mainly to supplement the basic information about an entity's network connectivity, which remains the core purpose of presence.

Presence is an enabling technology for peer-to-peer interaction. It first emerged as an aspect of communication systems, especially IM systems such as ICQ, which allowed users to see the availability of their friends. The huge role that IM has had in establishing presence is evident with the protocols

available today, such as Instant Messaging and Presence Service (IMPS), Session Initiation Protocol (SIP) for Instant Messaging and Presence Leveraging Extensions (SIMPLE), the Extensible Messaging and Presence Protocol (XMPP), first developed in the Jabber open source community and subsequently ratified as an Internet standard by the IETF.

Implementation of presence follows the software design pattern known as publish-and-subscribe (pub-sub). This means that a user or application publishes information about its network availability to a centralized location and that information is broadcast to all entities that are authorized to receive it. The authorization usually takes the form of a subscription. In IM implementations, contacts or buddies are the authorized entities. The popularity of these services among millions of people validated the value of the concept of presence.

For enterprise solutions, the limits of consumer-based IM services quickly became clear when enterprises tried to integrate presence into business-critical systems and services. Because business organizations require a great deal more control and flexibility over the technologies they deploy, they needed a presence solution that could provide separation between the presence service and the communication mechanisms (e.g., IM or VoIP) that presence enables. Any solution had to be scalable, extensible, and support a distributed architecture with its own presence domain. It should not overload the network and should support strong security management, system authentication, and granular subscription authorization. Also, any device or application should be able to publish and subscribe to presence information. Enterprise solutions should have the ability to federate numerous cross-protocol presence sources and integrate presence information from multiple sources. Any solution should be able to access presence data via multiple methods. The ability to integrate presence information with existing organizational infrastructure such as active directory is very important. Being able to publish content and allow other people and/or applications to subscribe to that information ensures that updates and changes are done in real time based on the presence/availability of those people/applications.

5.3.1 Presence Protocols

Proprietary, consumer-oriented messaging services do not enable enterprises or institutions to leverage the power of presence. A smarter approach is to use one of the standard presence protocols, SIMPLE or XMPP. is an instant

messaging and presence protocol suite based on SIP and managed by the Internet Engineering Task Force (IETF). XMPP is the IETF's formalization of the core XML messaging and presence protocols originally developed by the open source Jabber community in 1999. These protocols have been in wide use on the Internet for over five years. Both of these protocols will be explained in greater detail in Chapter 7.

The modern, reliable method to determine another entity's capabilities is called *service discovery,* wherein applications and devices exchange information about their capabilities directly, without human involvement. Even though no framework for service discovery has been produced by a standards development organization such as the IETF, a capabilities extension for SIP/SIMPLE and a robust, stable service discovery extension for XMPP does exist.

The SIMPLE Working Group is developing the technology to embed capabilities information within broadcasted presence information. A capability already exists in a widely-deployed XMPP extension. Together, service discovery and capabilities broadcasts enable users and applications to gain knowledge about the capabilities of other entities on the network, providing a real-time mechanism for additional use of presence-enabled systems.

5.3.2 Leveraging Presence

The real challenge today is to figure out how to leverage the power of presence within an organization or service offering. This requires having the ability to publish presence information from a wide range of data sources, the ability to receive or embed presence information in just about any platform or application, and having a robust presence engine to tie ubiquitous publishers and subscribers together.

It is safe to assume that any network-capable entity can establish presence. The requirements for functioning as a presence publisher are fairly minimal. As a result, SIP software stacks are available for a wide range of programming languages and it is relatively easy to add native presence publishing capabilities to most applications and devices. Enabling devices and applications to publish presence information is only half of the solution, however; delivering the right presence information to the right subscribers at the right time is just as important.

5.3.3 Presence Enabled

What does it mean to be "presence-enabled"? The basic concept is to show availability of an entity in an appropriate venue. Some modern applications aggregate presence information about all of a person's various connections. For communication devices such as phones and applications such as IM, presence information is often built into the device itself. For less communication-centric applications, such as a document or web page, presence may be gathered by means of a web services API or channeled through a presence daemon. Providing presence data through as many avenues as possible is in large measure the responsibility of a presence engine, as described below.

The presence engine acts as a broker for presence publishers and subscribers. A presence broker provides aggregation of information from many sources, abstraction of that information into open and flexible formats, and distribution of that information to a wide variety of interested parties. In the realm of presence, the qualities of aggregation, abstraction, and distribution imply that the ideal presence broker is trustworthy, open, and intelligent. As presence becomes more prevalent in Internet communications, presence engines need to provide strong authentication, channel encryption, explicit authorization and access control policies, high reliability, and the consistent application of aggregation rules. Being able to operate using multiple protocols such as IMPS, SIMPLE, and XMPP is a basic requirement in order to distribute presence information as widely as possible. Aggregating information from a wide variety of sources requires presence rules that enable subscribers to get the right information at the right time.

5.3.4 The Future of Presence

It will remain to be seen if XMPP is the future of cloud services, but for now it is the dominant protocol for presence in the space. Fixing the polling and scaling problems with XMPP (which we will discuss in Chapter 8, has been challenging but has been accomplished by providers such as Tivo, and the built-in presence functionality offers further fascinating possibilities. Presence includes basic availability information, but it is extensible and can also include abilities such as geo-location. Imagine cloud services taking different actions based on *where* the client initiated a connection.

5.3.5 The Interrelation of Identity, Presence, and Location in the Cloud

Digital identity refers to the traits, attributes, and preferences on which one may receive personalized services. Identity traits might include government-issued IDs, corporate user accounts, and biometric information. Two user attributes which may be associated with identity are presence and location. Over the last few years, there has been an aggressive move toward the convergence of identity, location, and presence. This is important because a standard framework tying identity to presence and location creates the ability to develop standards-based services for identity management that incorporate presence and location. Identity, presence, and location are three characteristics that lie at the core of some of the most critical emerging technologies in the market today: real-time communications (including VoIP, IM, and mobile communications), cloud computing, collaboration, and identity-based security.

Presence is most often associated with real-time communications systems such as IM and describes the state of a user's interaction with a system, such as which computer they are accessing, whether they are idle or working, and perhaps also which task they are currently performing (reading a document, composing email etc.). Location refers to the user's physical location and typically includes latitude, longitude, and (sometimes) altitude. Authentication and authorization mechanisms generally focus on determining the "who" of identity, location defines the "where," and presence defines the "what"—all critical components of the identity-based emerging technologies listed above, including cloud computing.

5.3.6 Federated Identity Management

Network identity is a set of attributes which describes an individual in the digital space. Identity management is the business processes and technologies of managing the life cycle of an identity and its relationship to business applications and services. Federated identity management (IdM) refers to standards-based approaches for handling authentication, single sign-on (SSO, a property of access control for multiple related but independent software systems), role-based access control, and session management across diverse organizations, security domains, and application platforms. It is a system that allows individuals to use the same user name, password, or other personal identification to sign on to the networks of more than one entity in order to conduct transactions. Federation is enabled through the use of

open industry standards and/or openly published specifications, such that multiple parties can achieve interoperability for common use cases. Typical use cases involve things such as cross-domain, web-based single sign-on, cross-domain user account provisioning, cross-domain entitlement management, and cross-domain user attribute exchange.

Single sign-on enables a user to log in once and gain access to the resources of multiple software systems without being prompted to log in again. Because different applications and resources support different authentication mechanisms, single sign-on has to internally translate to and store different credentials compared to what is used for initial authentication. The most widely implemented federated IdM/SSO protocol standards are Liberty Alliance Identity Federation Framework (ID-FF), OASIS Security Assertion Markup Language (SAML), and WS-Federation.

Within a typical cross-carrier internetworking environment, federated IdM may be implemented in layers. For converged IP services, federated IdM may involve separate authentications at the application layer and the network layer. Increasingly, the application-layer authentications rely on any or all of the federated IdM standards mentioned above.

5.3.7 Cloud and SaaS Identity Management

As SaaS vendors and their customers sort through the security implications of the hybrid on-demand/on-premises model for cloud applications, they face a number of very interesting identity management challenges. The typical large enterprise IT shop has relatively mature production implementations for standard identity management functionalities such as user authentication, single sign-on, user management, provisioning/deprovisioning, and audit. Because these implementations were designed and deployed to support users accessing applications running inside the enterprise, they often do not transition well to a model that calls for users to access applications (such as Salesforce.com and GoogleApps) which are hosted outside the corporate firewall.

With the advent of cloud computing and the identity requirements that corporate IT departments are putting on SaaS providers, the line between on-demand applications and on-premises applications is blurring, and a hybrid model is emerging in which the goal is closer integration of SaaS applications and functionality within enterprise IT infrastructure. The result is that sometimes corporate IT may have deployed an effective common model for identity management within the enterprise, but that

common model breaks down when requirements call for integration with on-demand applications. This breakdown comes in the form of proliferating on-demand user name and password accounts for users, manual processes for provisioning and deprovisioning users to on-demand applications, limited audit visibility across on-demand applications, and constraints on data integration between external and internal applications.

With the success of single sign-on inside the enterprise, users are calling for interoperability outside the enterprise's security domain to outsourced services, including business process outsourcing (BPO) and SaaS providers, and trading partners, as well as within the enterprise to affiliates and subsidiaries.

As a result of business demands that employees be able to traverse the Internet with highly sensitive data, using secure connections that protect the user, the enterprise, and the service provider, Internet-based SSO has seen a substantial increase over the last few years. There are many options to consider for delivering a SSO that works over the Internet. Choosing the right technology is crucial to successfully implementing federated identity management and mitigating long deployment times. The typical options for SSO are either a proprietary SSO (web agents) or standards-based SSO (identity federation). The idea of SSO has been around for years; it was the reason why enterprise portal software was invented in the late 1990s, and why many companies built proprietary SSO solutions. However, proprietary solutions that had to be rolled out by IT departments proved to have serious time, cost, complexity, and security implications.

In June 2008, Salesforce.com disclosed that it was using Security Assertion Markup Language (SAML), an open identity federation standard from OASIS, to implement SSO. The key benefit of using SAML instead of a proprietary SSO is that with SAML the same solution a customer uses for SSO to Salesforce.com can be used with GoogleApps or any of the other hundreds of companies that now support the SAML standard. This eliminated the need for multiple one-offs for SSO. The fact that the leading on-demand application made the move to SAML is a signal that the SaaS/on-demand community is on the path to adopting common models for identity management and security. SAML is the dominant web services standard for federated identity management today. It defines a set of XML formats for representing identity and attribute information, as well as protocols for requests and responses for access control information.

The key principle behind SAML is an *assertion,* a statement made by a trusted party about another. For example, a federated identity management server produces assertions about the identity and rights of users. An individual application does not need to have direct access to the user repository or trust a user—it only needs to know and trust the assertions source. Assertions can be encoded in browser requests or included in web services transactions, enabling log-ins for both person-to-machine and machine-to-machine communications. This was another first, the ability to use the same standards protocol for both back-end transactions and web portal access control.

5.3.8 Federating Identity

Identity federation standards describe two operational roles in an Internet SSO transaction: the identity provider (IdP) and the service provider (SP). An IdP, for example, might be an enterprise that manages accounts for a large number of users who may need secure Internet access to the web-based applications or services of customers, suppliers, and business partners. An SP might be a SaaS or a business-process outsourcing (BPO) vendor wanting to simplify client access to its services. Identity federation allows both types of organizations to define a trust relationship whereby the SP provides access to users from the IdP. There are four common methods to achieve identity federation: Use proprietary solutions, use open source solutions, contract a vendor to do it, or implement a standards-based federated solution.

Many attempt to write their own solution, only to find out there is a huge learning curve and a very high risk that the solution will be incompatible with the external applications and partners they want to connect to. Proprietary solutions rarely scale to connect with multiple partners. Open source libraries are often missing key abilities such as partner enablement and integration, rarely support the SAML 2.0 communication standard, and require significant continuous effort to adapt and maintain. If you choose to contract an identity management stack vendor, the federation component of the stack vendor's suite is usually the newest, least mature component, and its connection capabilities may be very limited in scope.

The most successful way to achieve identity federation is to choose a standalone federation vendor, whose sole focus is to provide secure Internet SSO through identity federation to numerous applications and partners. These vendors provide best-of-breed functionality, and they will work with

the identity management system you already have in place. Theses vendors should proactively go beyond the standards to address loopholes associated with underlying technologies such as XML digital signatures and provide centralizing management and monitoring of security credentials and identity traffic. Without a standards-based identity federation server, implementing SSO that works over the Internet can take 6 to 9 months. A properly configured standards-based identity federation server as provided by current SaaS cloud providers should facilitate an implementation in less than 30 to 45 days.

5.3.9 Claims-Based Solutions

Traditional means of authentication and authorization will eventually give way to an identity system where users will present claims that answer who they are or what they can do in order to access systems and content or complete transactions. Microsoft has developed a flexible claims architecture[5] based on standard protocols such as WS-Federation, WS-Trust, and the Security Assertion Markup Language (SAML), which should replace today's more rigid systems based on a single point of truth, typically a directory of user information. The claims model can grow out of the infrastructure users have today, including Public Key Infrastructure (PKI), directory services, and provisioning systems. This approach supports the shared industry vision of an identity metasystem that creates a single-user access model for any application or service and enables security-enhanced collaboration. Microsoft Geneva,mentioned at the beginning of the chapter, allows developers to use prebuilt identity logic and enables seamless interoperability between claims-based and non-claims-based systems.

5.3.10 Identity-as-a-Service (IaaS)

Identity-as-a-Service essentially leverages the SaaS model to solve the identity problem and provides for single sign-on for web applications, strong authentication, federation across boundaries, integration with internal identities and identity monitoring, compliance and management tools and services as appropriate. The more services you use in the cloud, the more you need IaaS, which should also includes elements of governance, risk management, and compliance (GRC) as part of the service. GRC is an increasingly recognized term that reflects a new way in which organizations can adopt an integrated approach to these three areas. However, this term

5. http://msdn.microsoft.com/en-us/security/aa570351.aspx.

is often positioned as a single business activity, when in fact it includes multiple overlapping and related activities, e.g., internal audit, compliance programs such as Sarbanes-Oxley, enterprise risk management, operational risk, and incident management.

IaaS is a prerequisite for most other aspects of cloud computing because you cannot become compliant if you cannot manage your identities and their access rights consistently in the cloud. That goes well beyond authentication. Approaches for consistent policy management across different cloud services will again require new standards, going beyond what federation standards such as SAML, authorization standards such as eXtensible Access Control Markup Language (XACML), and other standards such as the Identity Governance Framework (IGF) provide today. Some of the current IaaS vendors include Ping Identity, Symplified, TriCipher and Arcot Systems.

The biggest threat in cloud computing is manageability. The biggest threat to business by far is managing identities, authentication, authorization, and all of the regulatory auditing requirements. Within any cloud environment, an identity access strategy is a vital component and a prerequisite. GRC services are moving to the cloud as well, and these are the topic of the next section.

5.3.11 Compliance-as-a-Service (CaaS)[6]

Managed services providers historically have faced contractual difficulties with their customers in negotiating information assurance requirements, particularly regarding regulatory compliance verification. This problem becomes even more complex in a cloud computing environment, where physical resources can be geographically diverse, the regulatory landscape is vast and international in nature, and no single one-to-one relationship can determine the outcome of anything in the cloud.

Although this complexity may seem untenable at first glance, cloud computing potentially furnishes an exciting and cost-effective layer of opportunity in the creation of a "Compliance-as-a-Service" (CaaS) offering. CaaS could solve a number of problems that have been viewed as difficult or impossible, both by service providers and by their customers:

6. This section is based on email exchanges and input from Eddie Schwartz, CSO of Netwitness (www.netwitness.com), 12 Mar 2009.

- **Cost-effective multiregulation compliance verification:** A dominant percentage of all security and privacy regulations utilize a common base of security controls and best practices. These regulations, which have developed over many years, have been built on an identical, common body of knowledge augmented by a small percentage of nuance associated with industry-specific requirements. In a CaaS environment, next-generation network security monitoring technology could be deployed in the cloud to perform automated, rules-based data mining of cloud traffic flows. Compliance-oriented security services could be created to support verification of specific regulatory controls, from the network to the application layers, with commensurate alerting and reporting mechanisms.
- **Continuous audit:** A CaaS offering could provide continuous audit of security controls associated with the compliance domains within its scope. This approach would provide a higher level of information assurance than daily scans, quarterly spot audits, or statistical sampling methodologies. Additionally, the classic problem of third-party assurance and verification of a service provider's security would be resolved because of the transparency thatCaaS would provide into the service provider's security controls.
- **Threat intelligence:** Any CaaS offering would benefit from the aggregate threat intelligence and distributed security analytics associated with multiple cloud customers. This situational visibility would be invaluable in understanding and defending against current and emerging threats to the cloud computer environment.

5.3.12 The Future of Identity in the Cloud

As more business applications are delivered as cloud-based services, more identities are being created for use in the cloud. The challenges of managing identity in the cloud are far-reaching and include ensuring that multiple identities are kept secure. There must be coordination of identity information among various cloud services and among enterprise identity data stores and other cloud services. A flexible, user-centric identity management system is needed. It needs to support all of the identity mechanisms and protocols that exist and those that are emerging. It should be capable of operating on various platforms, applications, and service-oriented architectural patterns. Users must be empowered to execute effective

controls over their personal information. In the future, they will have control over who has their personal data and how it is used, minimizing the risk of identity theft and fraud. Their identity and reputation will be transferable. If they establish a good reputation on one site, they will be able to use that fact on other sites as well.

5.4 Privacy and Its Relation to Cloud-Based Information Systems

Information privacy[7] or data privacy is the relationship between collection and dissemination of data, technology, the public expectation of privacy, and the legal issues surrounding them. The challenge in data privacy is to share data while protecting personally identifiable information. The fields of data security and information security design and utilize software, hardware, and human resources to address this issue. The ability to control what information one reveals about oneself over the Internet, and who can access that information, has become a growing concern. These concerns include whether email can be stored or read by third parties without consent, or whether third parties can track the web sites someone has visited. Another concern is whether web sites which are visited collect, store, and possibly share personally identifiable information about users. *Personally identifiable information* (PII), as used in information security, refers to information that can be used to uniquely identify, contact, or locate a single person or can be used with other sources to uniquely identify a single individual.[8]

Privacy is an important business issue focused on ensuring that personal data is protected from unauthorized and inappropriate collection, use, and disclosure, ultimately preventing the loss of customer trust and inappropriate fraudulent activity such as identity theft, email spamming, and phishing. According to the results of the Ponemon Institute and TRUSTe's 2008 Most Trusted Companies for Privacy Survey, privacy is a key market differentiator in today's cyberworld. "Consumer perceptions are not superficial, but are in fact the result of diligent and successful execution of thoughtful privacy strategies," said Dr. Larry Ponemon, chairman and founder of the Ponemon Institute. "Consumers want to do business with brands they believe they can trust."[9]

7. http://en.wikipedia.org/wiki/Information_privacy, retrieved 28 Feb 2009.
8. http://en.wikipedia.org/wiki/Personally_identifiable_information, retrieved 28 Feb 2009.
9. http://www.truste.org/about/press_release/12_15_08.php, retrieved 28 Feb 2009.

Adhering to privacy best practices is simply good business but is typically ensured by legal requirements. Many countries have enacted laws to protect individuals' right to have their privacy respected, such as Canada's Personal Information Protection and Electronic Documents Act (PIPEDA), the European Commission's directive on data privacy, the Swiss Federal Data Protection Act (DPA), and the Swiss Federal Data Protection Ordinance. In the United States, individuals' right to privacy is also protected by business-sector regulatory requirements such as the Health Insurance Portability and Accountability Act (HIPAA), The Gramm-Leach-Bliley Act (GLBA), and the FCC Customer Proprietary Network Information (CPNI) rules.

Customer information may be "user data" and/or "personal data." User data is information collected from a customer, including:

- Any data that is collected directly from a customer (e.g., entered by the customer via an application's user interface)
- Any data about a customer that is gathered indirectly (e.g., metadata in documents)
- Any data about a customer's usage behavior (e.g., logs or history)
- Any data relating to a customer's system (e.g., system configuration, IP address)

Personal data (sometimes also called personally identifiable information) is any piece of data which can potentially be used to uniquely identify, contact, or locate a single person or can be used with other sources to uniquely identify a single individual. Not all customer/user data collected by a company is personal data. Examples of personal data include:

- Contact information (name, email address, phone, postal address)
- Forms of identification (Social Security number, driver's license, passport, fingerprints)
- Demographic information (age, gender, ethnicity, religious affiliation, sexual orientation, criminal record)
- Occupational information (job title, company name, industry)
- Health care information (plans, providers, history, insurance, genetic information)

- Financial information (bank and credit/debit card account numbers, purchase history, credit records)
- Online activity (IP address, cookies, flash cookies, log-in credentials)

A subset of personal data is defined as sensitive and requires a greater level of controlled collection, use, disclosure, and protection. Sensitive data includes some forms of identification such as Social Security number, some demographic information, and information that can be used to gain access to financial accounts, such as credit or debit card numbers and account numbers in combination with any required security code, access code, or password. Finally, it is important to understand that user data may also be personal dasta.

5.4.1 Privacy Risks and the Cloud

Cloud computing has significant implications for the privacy of personal information as well as for the confidentiality of business and governmental information. Any information stored locally on a computer can be stored in a cloud, including email, word processing documents, spreadsheets, videos, health records, photographs, tax or other financial information, business plans, PowerPoint presentations, accounting information, advertising campaigns, sales numbers, appointment calendars, address books, and more. The entire contents of a user's storage device may be stored with a single cloud provider or with many cloud providers. Whenever an individual, a business, a government agency, or other entity shares information in the cloud, privacy or confidentiality questions may arise.

A user's privacy and confidentiality risks vary significantly with the terms of service and privacy policy established by the cloud provider. For some types of information and some categories of cloud computing users, privacy and confidentiality rights, obligations, and status may change when a user discloses information to a cloud provider. Disclosure and remote storage may have adverse consequences for the legal status of or protections for personal or business information. The location of information in the cloud may have significant effects on the privacy and confidentiality protections of information and on the privacy obligations of those who process or store the information. Information in the cloud may have more than one legal location at the same time, with differing legal consequences. Laws could oblige a cloud provider to examine user records for evidence of criminal activity and other matters. Legal uncertainties make it difficult to assess the status of

information in the cloud as well as the privacy and confidentiality protections available to users.

5.4.2 Protecting Privacy Information

The Federal Trade Commission is educating consumers and businesses about the importance of personal information privacy, including the security of personal information. Under the FTC Act, the Commission guards against unfairness and deception by enforcing companies' privacy promises about how they collect, use, and secure consumers' personal information. The FTC publishes a guide that is a great educational tool for consumers and businesses alike, titled "Protecting Personal Information: A Guide for Business."[10] In general, the basics for protecting data privacy are as follows, whether in a virtualized environment, the cloud, or on a static machine:

- **Collection:** You should have a valid business purpose for developing applications and implementing systems that collect, use or transmit personal data.
- **Notice:** There should be a clear statement to the data owner of a company's/providers intended collection, use, retention, disclosure, transfer, and protection of personal data.
- **Choice and consent:** The data owner must provide clear and unambiguous consent to the collection, use, retention, disclosure, and protection of personal data.
- **Use:** Once it is collected, personal data must only be used (including transfers to third parties) in accordance with the valid business purpose and as stated in the Notice.
- **Security:** Appropriate security measures must be in place (e.g., encryption) to ensure the confidentiality, integrity, and authentication of personal data during transfer, storage, and use.
- **Access:** Personal data must be available to the owner for review and update. Access to personal data must be restricted to relevant and authorized personnel.
- **Retention:** A process must be in place to ensure that personal data is only retained for the period necessary to accomplish the intended business purpose or that which is required by law.

10. http://www.ftc.gov/bcp/edu/pubs/business/idtheft/bus69.pdf, retrieved 27 Feb 2009.

- **Disposal:** The personal data must be disposed of in a secure and appropriate manner (i.e., using encryption disk erasure or paper shredders).

Particular attention to the privacy of personal information should be taken in an a SaaS and managed services environment when (1) transferring personally identifiable information to and from a customer's system, (2) storing personal information on the customer's system, (3) transferring anonymous data from the customer's system, (4) installing software on a customer's system, (5) storing and processing user data at the company, and (6) deploying servers. There should be an emphasis on notice and consent, data security and integrity, and enterprise control for each of the events above as appropriate.[11]

5.4.3 The Future of Privacy in the Cloud

There has been a good deal of public discussion of the technical architecture of cloud computing and the business models that could support it; however, the debate about the legal and policy issues regarding privacy and confidentiality raised by cloud computing has not kept pace. A report titled "Privacy in the Clouds: Risks to Privacy and Confidentiality from Cloud Computing," prepared by Robert Gellman for the World Privacy Forum, provides the following observations on the future of policy and confidentiality in the cloud computing environment:

- Responses to the privacy and confidentiality risks of cloud computing include better policies and practices by cloud providers, more vigilance by users, and changes to laws.
- The cloud computing industry could establish standards that would help users to analyze the difference between cloud providers and to assess the risks that users face.
- Users should pay more attention to the consequences of using a cloud provider and, especially, to the provider's terms of service.
- For those risks not addressable solely through policies and practices, changes in laws may be needed.

11. Further details on privacy guidelines for developing software products and services can be found at http://www.microsoft.com/downloads/details.aspx?FamilyID=c48cf80f-6e87-48f5-83ec-a18d1ad2fc1f&displaylang=en.

- Users of cloud providers would benefit from greater transparency about the risks and consequences of cloud computing, from fairer and more standard terms, and from better legal protections. The cloud computing industry would also benefit.[12]

5.5 Chapter Summary

In this chapter, we covered the importance and relevance of federation, presence, identity, and privacy in cloud computing. We covered the latest challenges, solutions, and potential future for each area. Combined with the standards for cloud computing, the concepts of this chapter are the glue for the architectural elements that make the cloud a highly distributed, reliable, flexible, and cost-efficient functional medium in which to conduct business. The number-one concern and challenge concerning cloud computing and services is security It is a critical element of cloud computing and is associated with the other areas discussed in this chapter. In the next chapter, we will discuss the latest security vulnerabilities, challenges, and best practices for security in the cloud.

12. http://www.worldprivacyforum.org/pdf/WPF_Cloud_Privacy_Report.pdf, 23 Feb 2009, retrieved 28 Feb 2009.