Uma aplicação de privacidade no gerenciamento de identidades em nuvem com *uApprove*

Daniel Ricardo dos Santos[1]*, **Carla Merkle Westphall**[1]

[1]Laboratório de Redes e Gerência - Departamento de Informática e Estatística
Universidade Federal de Santa Catarina (UFSC) - Florianópolis – SC – Brasil

{danielrs,carlamw}@inf.ufsc.br

Abstract. *Due to the continued growth in the use of cloud computing and the tendency to migrate services to this new paradigm, it becomes necessary to investigate security issues that might compromise its use. Identity Managament is an area in information security that is concerned with the management of users and their data, involving authentication, authorization and attribute release. One of the biggest issues when users' data are involved is privacy in the collection, manipulation, storage and destruction of these data. This paper presents a proposal for an identity management application with users' privacy protection implemented in a cloud computing environment.*

Resumo. *Com o crescimento da computação em nuvem e a tendência de migração de serviços para esse novo paradigma, torna-se necessário investigar questões de segurança que possam comprometer seu uso. O gerenciamento de identidades é um campo da segurança da informação que se preocupa com o gerenciamento de usuários e seus dados, envolvendo autenticação, autorização e liberação de atributos. Uma das questões mais preocupantes quando se envolvem dados de usuários é a privacidade na coleta, manipulação, armazenamento e destruição desses dados. Este trabalho apresenta uma proposta de aplicação de gerenciamento de identidades com proteção à privacidade dos usuários implementada em um ambiente de computação em nuvem.*

1. Introdução

Computação em nuvem é a entrega de recursos computacionais compartilhados, sejam de armazenamento, processamento ou mesmo de *software* para usuários através da Internet.

A segurança é importante para garantir o sucesso de ambientes de nuvem [Takabi et al. 2010] [Grobauer et al. 2011], com destaque para a proteção à privacidade, já que dados sensíveis passam a ficar sob a custódia de terceiros [Pearson 2009].

O gerenciamento de identidades cresce em importância conforme crescem os serviços que precisam utilizar autenticação e controle de acesso de usuários [Angin et al. 2010] [Bertino and Takahashi 2011]. Esse é o caso de muitos serviços que executam em ambientes de nuvem e precisam estabelecer a identidade de seus usuários ao mesmo tempo que devem proteger sua privacidade.

Este trabalho tem como objetivo identificar problemas de privacidade no gerenciamento de identidades em ambientes de computação em nuvem e mostrar uma solução

*Bolsista do CNPq - Brasil

para esses problemas através da implantação de uma estrutura de gerenciamento de identidades que garanta a privacidade dos usuários desses ambientes. O gerenciamento de identidade é realizado pelo *software* Shibboleth [Internet2 2011a] fazendo uso combinado do *plugin* de privacidade uApprove [SWITCH 2011]. Esta estrutura compõe um provedor de identidade executado em uma máquina virtual no ambiente de nuvem da Amazon [Amazon 2011].

O restante do artigo está organizado da seguinte forma: a seção 2 comenta os trabalhos científicos relacionados; na seção 3 são descritos os conceitos básicos sobre gerenciamento de identidades e computação em nuvem; a seção 4 aborda conceitos de privacidade e desafios que existem no ambiente de nuvem; na seção 5 são apresentadas a proposta e as ferramentas utilizadas; na seção 6 é descrito o desenvolvimento do trabalho e na seção 7 são feitas as considerações finais.

2. Trabalhos Relacionados

A privacidade está sendo pesquisada em diversos trabalhos da literatura [Orawiwattanakul et al. 2010], [Bertino and Takahashi 2011], [Goth 2011], [Tancock et al. 2010] e [Angin et al. 2010].

O artigo [Orawiwattanakul et al. 2010] descreve uma extensão do uApprove chamado uApprove.jp, que permite ao usuário individual do ambiente Shibboleth escolher quais serão os atributos liberados pelo provedor de identidade para o provedor de serviço.

O livro de [Bertino and Takahashi 2011] cita que a implantação de políticas de privacidade em sistemas de gerenciamento de identidades continua sendo um desafio.

Provedores de serviço e desenvolvedores das interfaces que atuam em favor dos usuários devem facilitar o entendimento. O formato destes cenários deve ser sucinto, conciso, breve e simples para que o usuário saiba o que está acontecendo [Goth 2011].

Na computação em nuvem a privacidade deve seguir as leis e os contratos feitos entre as partes [Tancock et al. 2010] e também proteger dados de usuários em provedores de serviço [Angin et al. 2010], de acordo com as políticas de privacidade definidas.

3. Conceitos Básicos

3.1. Identidade e Gerenciamento de Identidades

Identidades digitais são coleções de dados que representam atributos ou características de uma entidade [Windley 2005]. Um serviço de gerenciamento de identidades pode ser definido como "o processo de criação, gerenciamento e utilização de identidades de usuários e a infraestrutura que suporta esse processo"[Lee et al. 2009].

Os seguintes papéis existem num sistema de gerenciamento de identidades [Bertino and Takahashi 2011]:

Usuário É a entidade que possui uma identidade e utiliza os serviços tanto do provedor de identidades quanto do provedor de serviços.

Provedor de Identidades (IdP - *Identity Provider*) Fornece os serviços de gerenciamento de identidades necessários para que o usuário use o provedor de serviços.

Provedor de Serviços (SP - *Service Provider*) Fornece os serviços que o usuário efetivamente deseja utilizar. O provedor de serviços delega a autenticação e autorização dos usuários que acessam seus serviços a um IdP.

3.2. Computação em Nuvem

O trabalho de [Marston et al. 2011] define computação em nuvem da seguinte forma: "É um modelo de serviço de tecnologia da informação onde os serviços computacionais (ambos hardware e software) são entregues sob demanda para os usuários através de uma rede na forma de auto-atendimento, independente de dispositivo e de localização. Os recursos necessários para fornecer os diferentes níveis de qualidade de serviço são compartilhados, dinamicamente escaláveis, alocados rapidamente, virtualizados e liberados com interação mínima com o provedor de serviço".

Três tipos diferentes de serviços são mencionados quando se considera computação em nuvem [Cloud Security Alliance 2010]: *Software as a Service* (SaaS), *Platform as a Service* (PaaS) e *Infrastructure as a Service* (IaaS). No modelo SaaS a empresa assina um serviço de uso do *software* que funciona como um aluguel, tanto do *software* como de toda a estrutura necessária para executá-lo. No modelo PaaS o vendedor do serviço oferece aos clientes uma plataforma de desenvolvimento de aplicativos, que o usuário utiliza tanto no desenvolvimento quanto na posterior disponibilização do serviço. No caso do IaaS o que o cliente procura é a própria infra-estrutra de computação: poder de processamento, capacidade de armazenamento e taxa de transmissão. Nesse tipo de serviço geralmente o cliente tem controle sobre a máquina através de acesso remoto.

4. Privacidade

A privacidade relaciona-se com a capacidade de um indivíduo proteger informações sobre si [Mather et al. 2009]. Uma política de privacidade é um documento que expressa a forma como uma entidade coleta, utiliza, administra e libera informações de seus usuários.

O *Fair Information Practice Principles (FIPS)* é um conjunto de regras para manipulação de informações com proteção à privacidade criado pela Comissão de Comércio Americana que regula o uso de informações privadas nos Estados Unidos e serve de base para regras de outros países [Federal Trade Comission 2011]. Os FIPs definem cinco princípios básicos: a **consciência** significa que o usuário deve ser avisado e entender como suas informações serão liberadas; a **escolha** significa que o usuário deve escolher como suas informações serão usadas; a **participação** permite ao usuário acessar e alterar suas informações; a **integridade** deve garantir que os dados dos usuários estejam corretos e o **cumprimento** garante que os princípios são respeitados.

No Brasil, a privacidade é uma garantia constitucional, mas não existe uma lei específica, como ocorre em outros países [CulturaDigital 2011].

A privacidade é um aspecto crítico da segurança em ambientes de nuvem [Mather et al. 2009], [Pearson 2009], [Bertino and Takahashi 2011], [Goth 2011], [Tancock et al. 2010] e [Angin et al. 2010].

De acordo com [Mather et al. 2009], [Takabi et al. 2010], [Angin et al. 2010], [Marcon Jr. et al. 2010] existem alguns aspectos que podem ser levantados quando se pesquisa privacidade em ambientes de nuvem. O usuário deve ter o direito de saber quais informações suas estão mantidas na nuvem e poder solicitar a remoção dessas informações; deve também ter garantias de que seus dados são armazenados e transferidos de forma segura. Já os provedores de serviços de nuvem: precisam seguir leis, normas e regulamentos quando lidam com informações privadas; precisam saber onde e

como os dados privados são armazenados e de que forma podem ser transmitidos; devem manter políticas que tratem da retenção de dados na nuvem; devem garantir que não há cópias dos dados armazenados em outros locais após sua destruição; devem garantir que estão cumprindo os requisitos de privacidade; devem manter *logs* de acesso a dados; e, caso haja um caso de violação de privacidade ou vazamento de informações deve-se saber quem é o culpado e como controlar o caso.

5. Proposta e Ferramentas Utilizadas

A aplicação desenvolvida neste trabalho tem como objetivo implantar uma estrutura de gerenciamento de identidade que garanta a privacidade dos usuários autenticados em um ambiente de computação em nuvem para acessar provedores de serviço (Figura 1).

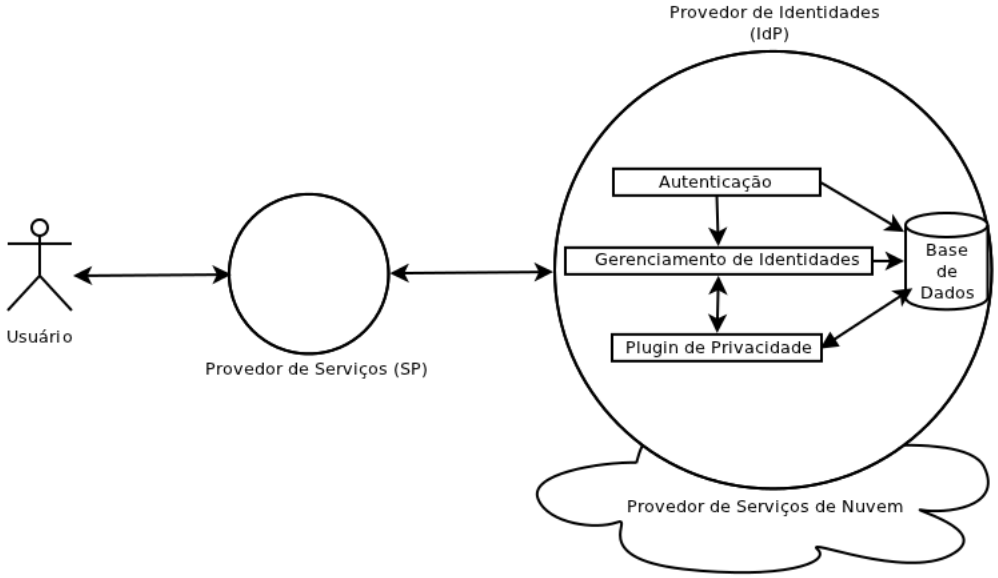

Figura 1. Diagrama geral da proposta

Neste cenário, iniciamente o usuário acessa o provedor de serviços. O provedor de serviços então redireciona o usuário para o seu respectivo provedor de identidades, que é informado pelo usuário e deve ter a confiança do provedor de serviços. O provedor de identidades está executando em um ambiente de nuvem, o que é transparente para o usuário. O provedor de identidades pede a autenticação do usuário e acessa seus atributos em sua base de dados. Quando o usuário está autenticado e antes de ser novamente redirecionado para o provedor de serviços, seus dados passam por um *plugin* de privacidade, momento no qual o usuário fica ciente e deve consentir com a liberação de seus atributos.

5.1. Amazon EC2

O EC2 foi o provedor de serviços de nuvem utilizado no trabalho. O EC2 provê uma Infraestrutura como um Serviço, em que é possível instanciar máquinas virtuais a partir de imagens de sistemas pré-definidas ou próprias. É possível configurar características da máquina como capacidade de processamento, memória e armazenamento.

No EC2 o usuário pode atribuir endereços IP estáticos às máquinas instanciadas e configurar a liberação de portas de acesso. A persistência dos dados é feita utilizando-se volumes *Elastic Block Storage* (EBS), que agem como discos rígidos das máquinas.

5.2. Shibboleth

Entre os diversos sistemas de gerenciamento de identidades disponíveis, optou-se pelo Shibboleth devido à sua popularidade em ambientes acadêmicos e boa documentação, além de ser um *software* de código aberto.

O Shibboleth é formado por duas partes principais: o IdP e o SP, que se encontram separados, mas se comunicam para prover o acesso seguro aos serviços. O fluxo de funcionamento do Shibboleth é representado na Figura 2.

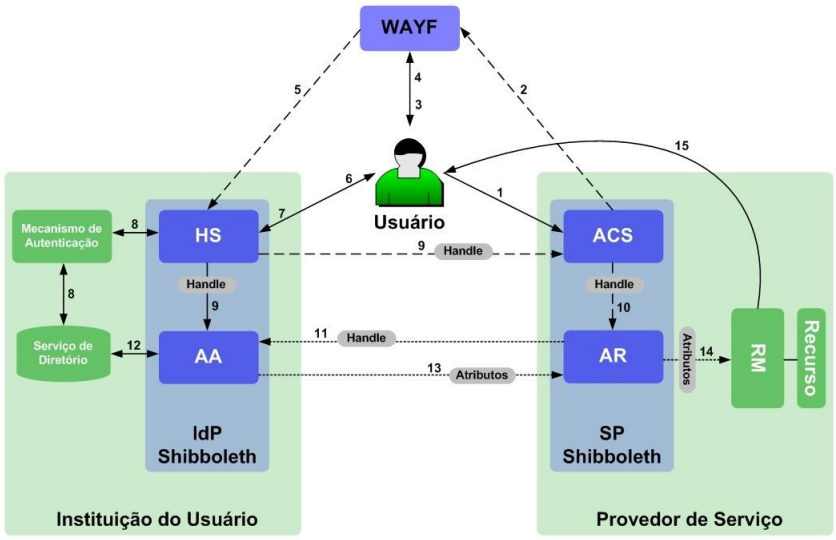

Figura 2. Funcionamento do Shibboleth. [de Cordova 2006]

No Passo 1 o usuário navega para o provedor de serviços para acessar um recurso protegido. Nos Passos 2 e 3 o Shibboleth redireciona o usuário para a página *Where are you from?* (WAYF), onde ele deve informar qual o seu provedor de identidades. No Passo 4 o usuário informa seu IdP e no Passo 5 ele é redirecionado para o *site*, que é o componente *Handle Service* (HS) do seu IdP. Nos Passos 6 e 7 o usuário informa seus dados e no Passo 8 o componente HS verifica a validade dos seus dados. O HS cria um *handle* para identificar o usuário e registra-o no *Attribute Authority* (AA). No Passo 9 esse *handle* confirma a autenticação do usuário. O *handle* é verificado pelo *Assertion Consumer Service* (ACS) e transferido para o *Attribute Requester* (AR) e no Passo 10 é criada uma sessão. No Passo 11 o AR utiliza o *handle* para requisitar os atributos do usuário ao IdP. No passo 12 o IdP verifica se pode liberar os atributos e no Passo 13 o AA responde com os valores dos atributos. No Passo 14 o SP recebe os atributos e os passa para o *Resource Manager* (RM), que no Passo 15 carrega o recurso [de Cordova 2006].

5.3. uApprove

Nos FIPS (descritos na seção 4) os princípios mais importantes da privacidade são a consciência dos usuários de que seus dados são coletados e armazenados e a possibilidade de escolha do usuário quanto a liberação desses dados. Uma ferramenta que implementa esses dois princípios é o uApprove [SWITCH 2011], um *plugin* de privacidade para o Shibboleth que encontra-se na versão 2.2.1.

O uApprove é dividido em três componentes principais: o ***IdP plugin*** é um filtro do Shibboleth, que testa se a ferramenta deve obter o consentimento do usuário para a

liberação de seus atributos; o ***Viewer*** apresenta ao usuário uma página *web* com os termos de uso que o usuário deve aceitar quando utiliza o provedor de identidades; e o ***Reset approvals*** permite que o usuário reinicie as liberações que já foram concedidas.

A Figura 3 mostra o fluxo de execução do *IdP plugin* para decidir se o *Viewer* deve ser invocado.

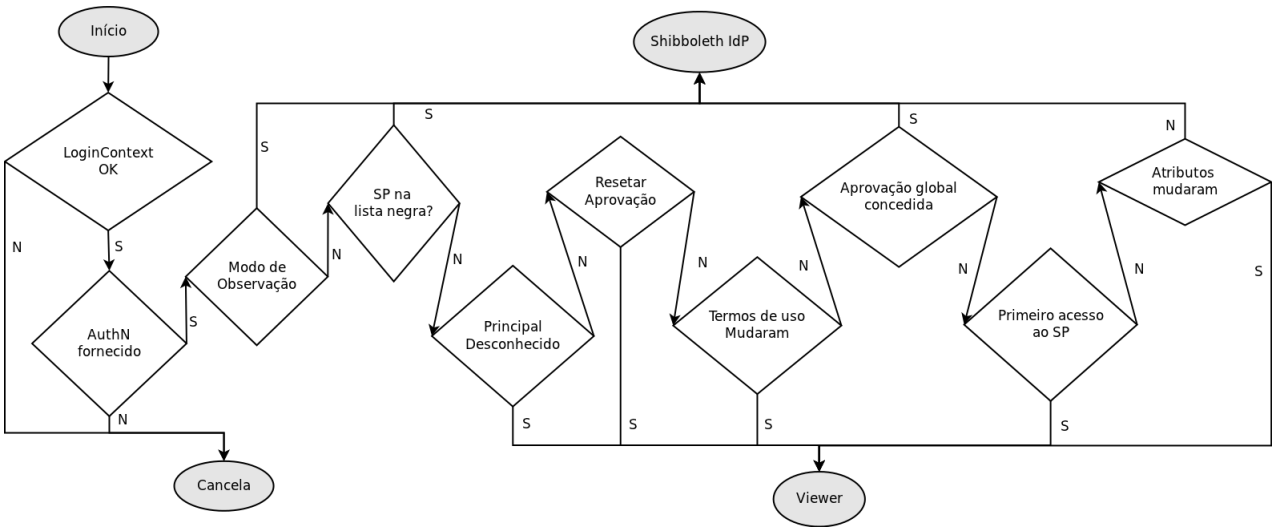

Figura 3. Fluxograma de execução do uApprove. Adaptado de: [SWITCH 2011]

Primeiramente o *plugin* verifica se o *LoginContext*, que é um objeto Java criado quando uma autenticação é requisitada, está correto. Caso o *LoginContext* esteja correto é verificado se o *Shibboleth Authentication Request* (AuthN), uma mensagem enviada pelo SP para o IdP para iniciar uma sessão, foi fornecido. Se alguma dessas verificações for negativa a execução é cancelada e o processo de autenticação terminado.

Caso as duas primeiras verificações sejam positivas o *plugin* verifica se está executando em modo de observação, onde só registra os atributos que serão liberados, sem interagir com o usuário. Caso esteja nesse modo o fluxo segue para o Shibboleth IdP. Em caso negativo o *plugin* continua seu fluxo, verificando se o SP se encontra na lista negra, uma lista de SPs nos quais o uApprove deve assumir automaticamente o consentimento do usuário.

Se o SP se encontrar na lista o fluxo segue para o Shibboleth IdP, senão o *plugin* verifica se o *Principal*, o identificador único de um usuário, é conhecido (já usou o *plugin*). Se o *Principal* for desconhecido (nunca utilizou o *plugin*), o *Viewer* será invocado.

Se o *Principal* for conhecido e tiver reiniciado seus consentimentos, o *Viewer* será invocado, senão o *plugin* continua seu fluxo. Na sequência, caso os termos de uso tenham sido alterados desde o último acesso o fluxo segue para o *Viewer*, senão o *plugin* continua.

Depois é verificado se o usuário concedeu aprovação global para a liberação de seus atributos. Em caso afirmativo o fluxo segue para o Shibboleth IdP, em caso negativo segue para a próxima verificação. Então verifica-se se o usuário está acessando o SP pela primeira vez, em caso afirmativo o *Viewer* é invocado, em caso negativo é feita a última verificação, que se refere aos atributos sendo requisitados pelo SP. Se eles tiverem sido alterados o *Viewer* é invocado, senão o fluxo segue para o IdP.

Em todos os casos em que o fluxo for para o Shibboleth IdP a execução do *plugin* é ignorada pelo usuário. Em todos os casos em que o *Viewer* for invocado, o usuário deve interagir e fornecer seu consentimento.

6. Desenvolvimento Prático

Usando o Amazon EC2 foi instanciada uma máquina virtual executando Windows Server 2008 e atribuído à máquina o IP estático 50.19.108.64, com DNS público ec2-50-19-108-64.compute-1.amazonaws.com. Para persistência dos dados utilizou-se um volume EBS de 30GB (Figura 4).

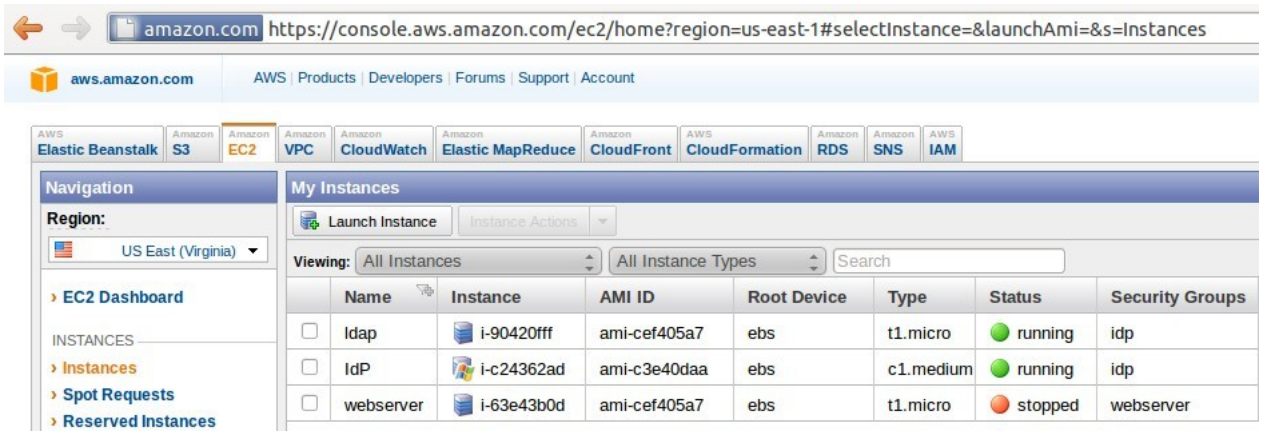

Figura 4. Máquinas virtuais instanciadas no EC2

As portas liberadas no *firewall* foram: 3306 para acesso ao MySQL, 3389 para acesso remoto, 8009 para uso do Shibboleth e 8080 para uso do Tomcat.

Na máquina instanciada e em execução foi instalado o servidor *web* Apache 2.2. O servidor aceita conexões não-SSL (na porta 80) e conexões SSL (nas portas 443 e 8443).

Depois foi instalado o servidor de aplicações Apache Tomcat 6.0.22, no qual devem ser executadas as aplicações de autenticação, gerenciamento de identidades e o *plugin* de privacidade. Foi então configurado um *proxy* no Apache para repassar os pedidos dessas aplicações para o Tomcat.

Foi instalado o mecanismo de autenticação JASIG CAS Server [JASIG 2011], versão 3.3.2, que autentica usuários através de *login* e senha e então repassa os usuários autenticados para o Shibboleth. O CAS foi configurado para procurar os usuários em um diretório LDAP, instalado em outra máquina virtual, executando Ubuntu Server 10.10.

Na instalação do provedor de identidades Shibboleth, a federação escolhida para ser utilizada foi a TestShib [Internet2 2011b]. Para utilizá-la foi necessário cadastrar o IdP, informando o endereço DNS e o certificado gerado, configurando também o Shibboleth para utilizar os metadados da federação.

Na configuração da liberação de atributos do usuário foi usado o esquema brEduPerson, uma extensão do eduPerson para federações brasileiras.

Seguiu-se para a instalação do uApprove. O *plugin* precisa armazenar informações sobre o consentimento dos usuários e a liberação de seus atributos e para isso foi utilizado o MySQL, versão 5.5, instalado na mesma máquina do Shibboleth.

Foi então gerado um arquivo que contém um exemplo de Termos de Uso e, com as configurações prontas, criado um filtro para ativar o uso do IdP plugin com o Shibboleth.

Com a instalação concluída, uma visão detalhada da aplicação pode ser resumida na Figura 5, que representa a visão detalhada da parte do IdP da Figura 1.

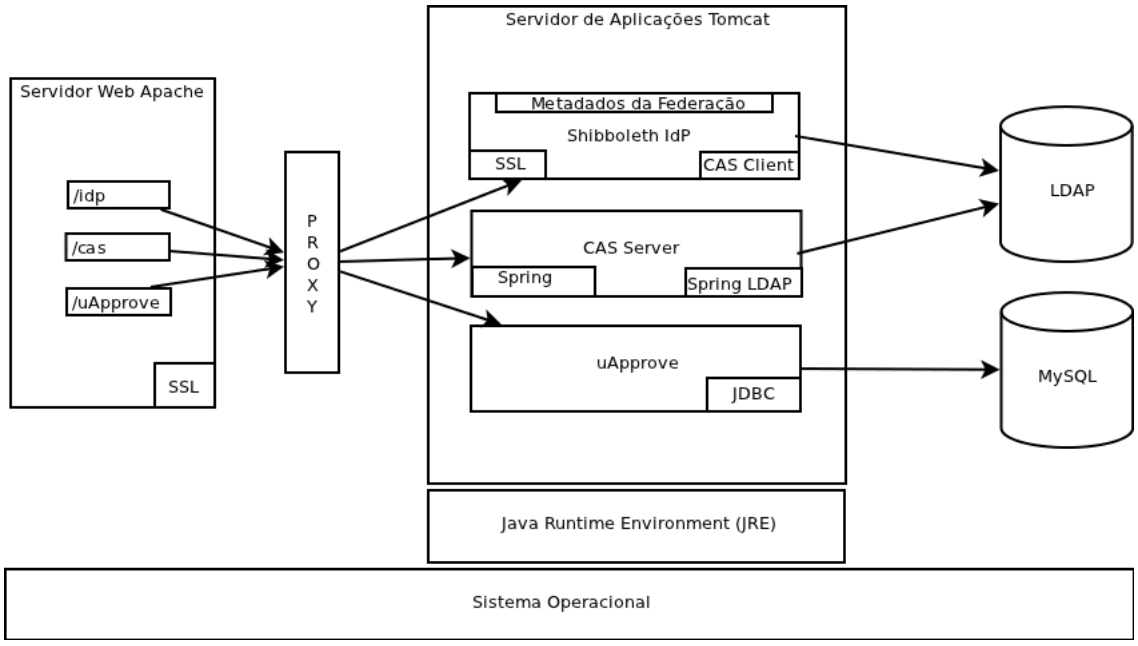

Figura 5. Visão detalhada da aplicação

Como ponto de acesso temos o servidor *web* Apache, que recebe as requisições HTTPS e as encaminha para o Tomcat, para que sejam recebidas pela aplicação correta. Dentro do Tomcat existem três aplicações sendo executadas: Shibboleth IdP, CAS Server e uApprove. O diretório LDAP se encontra na máquina com o Ubuntu Server 10.10. O restante dos componentes se encontram na máquina virtual com o Windows Server 2008.

Para realizar seu primeiro acesso ao SP o usuário acessa o provedor de serviços em `https://sp.testshib.org/` e informa o provedor de identidades `https://ec2-50-19-108-64.compute-1.amazonaws.com/idp/shibboleth` para ser então redirecionado para a página de autenticação, fornecida pelo CAS, onde faz sua autenticação por *login* e senha, que são buscados no LDAP.

Depois da autenticação o Shibboleth busca no diretório os atributos que devem ser liberados. Nesse momento o filtro do uApprove entra em ação e exibe uma página contendo os termos de uso do IdP. Caso o usuário aceite os termos de uso o *plugin* o redireciona para uma página que mostra os atributos que serão liberados (Figura 6).

O usuário autenticado é novamente requisitado a aceitar a liberação de seus atributos e, se concordar, é levado à página de acesso protegido do provedor de serviços.

Figura 6. Atributos que serão liberados

7. Conclusões e trabalhos futuros

Nesse trabalho foi possível tratar problemas específicos de privacidade no gerenciamento de identidades em ambientes de nuvem: a falta de consciência dos usuários quanto à liberação de seus atributos para provedores de serviço e a falta de preocupação dos provedores de identidades quanto à apresentação de seus termos de uso. Isso é importante, de acordo com [Goth 2011] [Bertino and Takahashi 2011] [Angin et al. 2010] e contribui para tratar os aspectos citados na seção 4.

A proposta de solução, com o uso dos softwares Shibboleth e uApprove, mostrou que é possível resolver os dois problemas de maneira eficiente e sem comprometer a usabilidade da aplicação. A proposta se mostrou viável e pôde ser implantada em uma nuvem pública, com a possibilidade de utilização em federações consolidadas. Por fim, este artigo também contribui para um melhor entendimento do funcionamento do uApprove.

A maior dificuldade para a realização do trabalho foi a falta de referências de implementações em ambientes de nuvem. Vários artigos apresentam modelos e propostas, mas praticamente não há exemplos de implementações reais. Automatização da verificação de compatibilidade entre políticas de privacidade de provedores e de usuários pode ser considerado um trabalho futuro.

Referências

Amazon (2011). Amazon elastic compute cloud. `http://aws.amazon.com/ec2/`.

Angin, P., Bhargava, B., Ranchal, R., Singh, N., Linderman, M., Ben Othmane, L., and Lilien, L. (2010). An entity-centric approach for privacy and identity management in cloud computing. In *IEEE SRDS, 2010*, pages 177 –183.

Bertino, E. and Takahashi, K. (2011). *Identity Management: Concepts, Technologies, and Systems*. Artech House.

Cloud Security Alliance (2010). Domain 12: Guidance for identity and access management v2.1.

CulturaDigital (2011). Os rumos da lei de proteção de dados. http://culturadigital.br/dadospessoais/os-rumos-da-lei-de-protecao-de-dados/.

de Cordova, A. S. (2006). Aplicação prática de um sistema de gerenciamento de identidades. TCC, Ciência da Computação, UNIVALI.

Federal Trade Comission (2011). Fair information practice principles. http://www.ftc.gov/reports/privacy3/fairinfo.shtm.

Goth, G. (2011). Privacy gets a new round of prominence. *Internet Computing, IEEE*, 15(1):13 –15.

Grobauer, B., Walloschek, T., and Stocker, E. (2011). Understanding cloud computing vulnerabilities. *IEEE Security and Privacy*, 9:50–57.

Internet2 (2011a). About shibboleth. http://shibboleth.internet2.edu/about.html.

Internet2 (2011b). Testshib two. https://www.testshib.org/testshib-two/index.jsp.

JASIG (2011). Jasig cas. http://www.jasig.org/cas.

Lee, H., Jeun, I., and Jung, H. (2009). Criteria for evaluating the privacy protection level of identity management services. *Emerging Security Information, Systems, and Technologies, The International Conference on*, 0:155–160.

Marcon Jr., A., Laureano, M., Santin, A., and Maziero, C. (2010). Aspectos de segurança e privacidade em ambientes de computação em nuvem. In *Livro-texto de minicursos do SBSeg 2010*, volume 1, pages 53 –102, Porto Alegre, RS. SBC.

Marston, S., Li, Z., Bandyopadhyay, S., Zhang, J., and Ghalsasi, A. (2011). Cloud computing – the business perspective. *Decision Support Systems*, 51(1):176 – 189.

Mather, T., Kumaraswamy, S., and Latif, S. (2009). *Cloud Security and Privacy: An Enterprise Perspective on Risks and Compliance*. O'Reilly Media, Inc.

Orawiwattanakul, T., Yamaji, K., Nakamura, M., Kataoka, T., and Sonehara, N. (2010). User-controlled privacy protection with attribute-filter mechanism for a federated sso environment using shibboleth. In *3PGCIC, 2010*, pages 243 –249.

Pearson, S. (2009). Taking account of privacy when designing cloud computing services. In *Proc. of the 2009 ICSE Workshop*, CLOUD '09, pages 44–52, Washington, DC, USA. IEEE Computer Society.

SWITCH (2011). uapprove. http://www.switch.ch/aai/support/tools/uApprove.html.

Takabi, H., Joshi, J. B., and Ahn, G.-J. (2010). Security and privacy challenges in cloud computing environments. *IEEE Security and Privacy*, 8:24–31.

Tancock, D., Pearson, S., and Charlesworth, A. (2010). A privacy impact assessment tool for cloud computing. In *IEEE CloudCom, 2010*, pages 667 –676.

Windley, P. (2005). *Digital Identity*. O'Reilly Media, Inc.