UNIVERSIDADE FEDERAL DE SANTA CATARINA

María Elena Villarreal

TOKEN DE PRIVACIDADE: UM MECANISMO DE ESPECIFICAÇÃO DE PREFERÊNCIAS DE PRIVACIDADE PARA SISTEMAS DE GERENCIAMENTO DE IDENTIDADE EM NUVEM

Florianópolis

2017

María Elena Villarreal

TOKEN DE PRIVACIDADE: UM MECANISMO DE ESPECIFICAÇÃO DE PREFERÊNCIAS DE PRIVACIDADE PARA SISTEMAS DE GERENCIAMENTO DE IDENTIDADE EM NUVEM

Dissertação submetida ao Programa de Pós-Graduação em Ciência da Computação para a obtenção do Grau de Mestre em Ciência da Computação. Orientadora: Prof[a]. Dr[a]. Carla Merkle Westphall

Florianópolis

2017

Ficha de identificação da obra elaborada pelo autor,
através do Programa de Geração Automática da Biblioteca Universitária da UFSC.

Villarreal, María Elena
Token de privacidade : um mecanismo de especificação de preferências de privacidade para sistemas de gerenciamento de identidade em nuvem / María Elena Villarreal ; orientadora, Carla Merkle Westphall, 2017.
97 p.

Dissertação (mestrado) - Universidade Federal de Santa Catarina, Centro Tecnológico, Programa de Pós Graduação em Ciência da Computação, Florianópolis, 2017.

Inclui referências.

1. Ciência da Computação. 2. Privacidade. 3. Gerenciamento de Identidade. 4. Computação em Nuvem. I. Westphall, Carla Merkle. II. Universidade Federal de Santa Catarina. Programa de Pós-Graduação em Ciência da Computação. III. Título.

AGRADECIMENTOS

À minha orientadora, Carla M. Westphall, por ter me acolhido no LRG e me ajudado em todas as etapas da pesquisa e, principalmente, por seu apoio e compreensão.

Ao professor Carlos B. Westphall e aos demais colegas do LRG por sempre me receberem muito bem no laboratório e pelos conhecimentos compartilhados.

Em especial ao meu pai Sergio, mestre em Ciência da Computação, não só pelo grande exemplo de pai e profissional de TI mas também pela enorme ajuda e orientação no desenvolvimento deste trabalho.

Ao meu namorado Cássio, designer gráfico excepcional, pela sua ajuda profissional no desenvolvimento das imagens e interfaces gráficas, bem como pelo seu imenso suporte e amor.

Ao meu sobrinho e afilhado Miguel, que me acompanhou e inspirou, embora dormindo em seu carrinho, em muitas manhãs e tardes de escrita desta dissertação.

À minha mãe Eliana e minha irmã Daniela pelo importantíssimo apoio, incentivo e carinho.

A Deus.

Les plumes ornent le paon; la science orne l'homme.

(Charles Dubois, 1857)

RESUMO

Com a crescente quantidade de dados pessoais armazenados e processados na nuvem, surgiram incentivos econômicos e sociais para coletar e agregar tais dados. Consequentemente, o uso secundário de dados, incluindo o compartilhamento com terceiros, tornou-se uma prática comum entre os provedores de serviço e pode levar a violações de privacidade e causar danos aos usuários, uma vez que envolve o uso de suas informações de forma não consensual e possivelmente indesejada. Apesar da existência de inúmeros trabalhos relativos à privacidade em ambientes de nuvem, os usuários ainda não possuem recursos para controlar como seus dados pessoais podem ser usados, por quem e para quais propósitos. Este trabalho apresenta um mecanismo para sistemas de gerenciamento de identidade que instrui os usuários sobre os possíveis usos secundários de seus dados pessoais, permite que eles definam suas preferências de privacidade e envia tais preferências ao provedor de serviço juntamente com seus dados de identificação em uma estrutura padronizada e legível por máquina, chamada token de privacidade. Esta abordagem baseia-se em uma classificação tridimensional dos possíveis usos secundários dos dados, quatro perfis de privacidade predefinidos e um personalizável, e um token seguro para a transmissão das preferências de privacidade. A aplicabilidade e a utilidade da proposta foram demonstradas mediante um estudo de caso e a viabilidade técnica e o correto funcionamento do mecanismo foram verificados através de um protótipo desenvolvido em Java para ser incorporado, em trabalhos futuros, a uma implementação do protocolo OpenID Connect. As principais contribuições deste trabalho são o modelo de especificação de preferências e o token de privacidade, que, ao permitirem que o usuário defina suas preferências e que estas sejam transmitidas ao SP para que alinhe suas ações, invertem o cenário atual onde o usuário é forçado a aceitar as políticas definidas pelos provedores de serviço.

Palavras-chave: Privacidade. Gerenciamento de Identidade. Computação em Nuvem.

ABSTRACT

With the increasing amount of personal data stored and processed in the cloud, economic and social incentives to collect and aggregate such data have emerged. Therefore, secondary use of data, including sharing with third parties, has become a common practice among service providers and may lead to privacy breaches and cause damage to users since it involves using information in a non-consensual and possibly unwanted manner. Despite the existence of numerous works regarding privacy in cloud environments, users still do not have means to control how their personal information can be used, by whom and for which purposes. This work presents a mechanism for identity management systems that instructs users about the possible secondary uses of their personal data, allows them to set their privacy preferences and sends these preferences to the service provider along with their identification data in a standardized, machine-readable structure, called privacy token. This approach is based on a three-dimensional classification of the possible secondary uses of data, four predefined privacy profiles and a customizable one, and a secure token for transmitting the privacy preferences. The applicability and the usefulness of the proposal were demonstrated through a case study, and the technical viability and the correct operation of the mechanism were verified through a prototype developed in Java in order to be incorporated, in future work, to an implementation of the OpenID Connect protocol. The main contributions of this work are the preference specification model and the privacy token, which invert the current scenario where users are forced to accept the policies defined by service providers by allowing the former to express their privacy preferences and requesting the latter to align their actions.

Keywords: Privacy. Identity Management. Cloud Computing.

LISTA DE FIGURAS

LISTA DE TABELAS

LISTA DE ABREVIATURAS E SIGLAS

PII	Personally Identifiable Information
SP	Service Provider
IdM	Identity Management
JWT	JSON Web Token
NIST	National Institute of Standards and Technology
IaaS	Infrastructure as a Service
PaaS	Platform as a Service
SaaS	Software as a Service
SSO	Single Sign-On
MIT	Massachussets Institute of Technology
CAFe	Comunidade Acadêmica Federada
RNP	Rede Nacional de Ensino e Pesquisa
SAML	Security Assertion Markup Language
FIM	Federated Identity Management
OIDC	OpenID Connect
JSON	JavaScript Object Notation
REST	Representational State Transfer
PET	Privacy-Enhancing Technology
PMRM	Privacy Management Reference Model and Methodology
P3P	Platform for Privacy Preferences
W3C	World Wide Web Consortium
XML	eXtensible Markup Language
EPAL	Enterprise Privacy Authorization Language
P2P	Purpose-to-Use
IP	Identificação Pessoal
CPP	Características Pessoais e Preferências
LO	Localização
AH	Atividades e Hábitos
RS	Relacionamentos
MS	Melhoria de Serviço
CI	Científico
CO	Comercial

PP	Outorgante das PIIs
TP	Terceira Parte
HMAC	Hash-based Message Authentication Code
DTO	Data Transfer Object
DAO	Data Access Object
BO	Business Object
HTML	HyperText Markup Language
CSS	Cascading Style Sheets
AES	Advanced Encryption Standard
CBC	Cipher Block Chaining
PKCS	Public-Key Cryptography Standards
IV	Initialization Vector

SUMÁRIO

1 INTRODUÇÃO

A Computação em Nuvem oferece infraestrutura, plataforma de desenvolvimento e aplicativos como serviço, sob demanda e cobrados conforme o uso. Por um lado, este paradigma fornece aos usuários maior flexibilidade, desempenho e escalabilidade sem a necessidade de manter e gerenciar sua própria infraestrutura de TI. Por outro, agrava o problema da aplicação e da verificação de segurança e faz com que os usuários percam, pelo menos parcialmente, o controle sobre seus dados e aplicações (ZHAO et al., 2016).

Com a crescente quantidade de dados pessoais armazenados e processados na nuvem, incluindo informações de identificação pessoal (*Personally Identifiable Information - PII*) dos usuários, surgiram incentivos econômicos e sociais para coletar e agregar tais dados. Consequentemente, o uso secundário de dados, incluindo o compartilhamento com terceiros, tornou-se uma prática comum entre os provedores de serviço (*Service Providers* - SPs) (IYILADE; VASSILEVA, 2014). No entanto, os usuários, em geral, não são cientes do uso secundário dos seus dados e da existência de terceiras partes. Isto ocorre porque interagem diretamente apenas com SPs, cujas políticas de privacidade são complexas e extensas, o que dificulta sua compreensão e leva os usuários, na maior parte das vezes, a ignorá-las (KOLTER, 2010).

De acordo com a taxonomia de privacidade definida por Solove (2006), o uso secundário consiste na utilização de dados para fins diferentes daqueles para os quais foram inicialmente coletados sem o consentimento do usuário, como o uso de dados pessoais recolhidos em redes sociais para oferecer publicidade personalizada. Esta prática pode, portanto, ferir a privacidade do usuário e causar danos, uma vez que envolve utilizar informações de forma não consensual e possivelmente indesejada.

Segundo Westin (1967), a privacidade é uma necessidade do ser humano e diz respeito ao direito dos indivíduos de determinarem por si mesmos quando, como e qual tipo de informação sobre eles pode ser revelada a outros. Desta forma, visto que se determinada ação viola ou não a privacidade de um usuário depende da percepção do mesmo e de sua vontade de compartilhar determinados tipos de dados, existe a necessidade de coletar e respeitar suas preferências (SOLOVE, 2006).

Um aspecto essencial para a implementação da segurança e da privacidade na nuvem é o Gerenciamento de Identidade (*Identity Management* - IdM), que engloba as tecnologias e os processos necessários

para a criação, o gerenciamento e o uso de identidades digitais e permite centralizar os dados de identificação do usuário nos provedores de identidade (*Identity Providers* - IdPs), que os enviam aos SPs para habilitar os processos de autenticação e controle de acesso (BENANTAR, 2006). Os sistemas de IdM, como OpenID Connect e Shibboleth, permitem a criação de federações, isto é, relacionamentos de confiança que tornam possível que usuários autenticados em um IdP acessem serviços fornecidos por vários SPs pertencentes a diferentes domínios administrativos. Por exemplo, quando os usuários se autenticam em serviços diferentes por meio de suas contas do Facebook, este age como um provedor de identidade.

Embora as políticas de privacidade das organizações afirmem que as informações podem ser usadas de forma secundária, a maioria dos usuários não as leem ou não as entendem. Deste modo, não podem tomar uma decisão informada sobre os usos secundários, uma vez que não estão cientes da variedade de usos potenciais (SOLOVE, 2006). Assim, faz-se necessário incorporar aos sistemas de IdM mecanismos que ofereçam uma maneira simples e eficaz para que usuários comuns possam expressar suas preferências de privacidade e controlar o uso de suas PIIs (ZHAO et al., 2016).

Esta dissertação apresenta um mecanismo para sistemas de gerenciamento de identidade, focado no OpenID Connect, que instrui os usuários sobre os possíveis usos secundários de seus dados pessoais e lhes permite definir suas preferências de privacidade. Estas preferências são convertidas em uma estrutura padronizada, legível por máquina, chamada token de privacidade, que é enviada ao SP juntamente com outros dados de autenticação para alinhar suas ações.

1.1 PROBLEMA

É comum os provedores de serviço usarem de forma secundária os dados pessoais dos usuários sem o consentimento ou a ciência dos mesmos. Embora existam diversas abordagens que visam permitir aos usuários definirem suas preferências de privacidade e às organizações expressarem suas práticas, segundo Zhao et al. (2016), elas são escassamente adotadas por não oferecerem métodos práticos e a maioria não considera a natureza descentralizada dos ambientes federados em nuvem. Consequentemente, os sistemas de IdM não oferecem mecanismos eficazes para coletar preferências de privacidade e enviá-las ao SP e os usuários ainda não possuem recursos para controlar como suas PIIs

podem ser usadas, por quem e para quais propósitos.

Werner e Westphall (2016) propõem um modelo de gerenciamento de identidade com privacidade para nuvem no qual IdPs e SPs interagem em ambientes dinâmicos e federados para gerenciar as identidades e garantir a privacidade de usuários. O modelo permite que os usuários escolham um escopo de disseminação de dados e criptografem tais dados, porém não define um mecanismo para determinar as preferências de privacidade dos usuários quanto ao uso e compartilhamento dos mesmos.

1.2 OBJETIVOS

O objetivo deste trabalho é viabilizar o controle do usuário sobre o uso secundário de seus dados de forma que sejam respeitadas suas preferências de privacidade pelos provedores de serviço. Para alcançar este objetivo, são definidos os seguintes objetivos específicos:

- Criar um modelo que represente de forma universal as preferências do usuário sem a necessidade de listar os dados extensivamente;
- Definir uma estrutura de dados legível por máquina que permita expressar as preferências do usuário e que possa ser enviada ao SP, e entendida e aplicada por ele;
- Desenvolver um mecanismo para coletar as preferências de privacidade do usuário e transmiti-las ao SP com base no modelo e na estrutura de dados definidos; e
- Ampliar o modelo de Werner e Westphall (2016) com o mecanismo desenvolvido.

1.3 JUSTIFICATIVA

O uso indevido ou não desejado dos dados pessoais de um usuário pode ferir sua privacidade. De acordo com Solove (2006), os danos causados pela violação de privacidade podem ser ainda maiores do que aqueles que poderiam ser infligidos por lesão corporal. Problemas de privacidade abrangem também a criação ou o aumento do risco de ocorrência de danos, isto é, de que um sujeito possa ser prejudicado no futuro. Atividades envolvendo PIIs, por exemplo, podem criar um risco maior de o usuário ser vítima de roubo de identidade ou fraude e

aumentam a probabilidade de que o indivíduo sofra danos morais, bem como danos monetários ou físicos.

Ainda segundo Solove (2006), o uso secundário potencial gera medo e incerteza sobre como as informações serão usadas no futuro, o que cria um sentimento de impotência e vulnerabilidade. O dano causado é moral e surge de negar às pessoas o controle sobre o uso futuro de seus dados, o que pode ter efeitos significativos em suas vidas.

1.4 MÉTODO

O método proposto para a obtenção do objetivo descrito na Seção 1.2 consiste nos seguintes passos:

1. Realizar um levantamento bibliográfico em livros e artigos científicos sobre Computação em Nuvem, Gerenciamento de Identidade, Privacidade, e Políticas e Preferências de Privacidade;

2. Analisar os trabalhos relacionados à proposta;

3. Definir os conceitos básicos necessários para a compreensão do assunto e a definição do mecanismo;

4. Com base em políticas de privacidade de provedores de serviço, no modelo conceitual proposto pela OASIS, o *Privacy Management Reference Model and Methodology* (PMRM) (OASIS, 2016) e nas normas ISO/IEC 29100 (ISO/IEC, 2011) e ISO/IEC 29101 (ISO/IEC, 2013), definir um modelo multidimensional para classificar os possíveis usos secundários de PIIs a fim de que seja possível representar as preferências de privacidade de forma abrangente;

5. Determinar perfis que representem diferentes tipos de usuários de acordo com seus níveis desejados de privacidade e suas preferências;

6. Definir uma estrutura de dados que permita representar tais preferências e transmiti-las ao SP;

7. Criar um mecanismo para anexar as preferências de privacidade às PIIs do usuário gerenciadas pelo IdP. Este mecanismo será implementado através de um *JSON Web Token* (JWT), semelhante aos tokens de ID e de acesso utilizados pelo OpenID Connect, e chamado de token de privacidade;

8. Demonstrar a aplicabilidade e a utilidade do mecanismo mediante a aplicação do modelo em um estudo de caso; e

9. Validar a viabilidade técnica e a correta operação do mecanismo através do desenvolvimento de um protótipo na linguagem de programação orientada a objetos Java de forma que, em trabalhos futuros, possa ser incorporado como uma extensão a uma implementação do OpenID Connect.

1.5 RESULTADOS ESPERADOS

Com o desenvolvimento deste trabalho, espera-se que o modelo seja implementado em sistemas de gerenciamento de identidade e utilizado em ambientes de identidade federada em nuvem para viabilizar a privacidade do usuário através do controle sobre suas PIIs e, assim, evitar possíveis danos morais, financeiros e físicos.

1.6 ORGANIZAÇÃO DO TRABALHO

Após esta introdução, no Capítulo 2, é apresentada a fundamentação teórica e o Capítulo 3 expõe os trabalhos relacionados a esta proposta. O Capítulo 4 descreve o mecanismo desenvolvido, que inclui o modelo que permite classificar os possíveis usos secundários das PIIs e definir as preferências de privacidade, bem como o token de privacidade. No Capítulo 5, é apresentada a validação do mecanismo através de um estudo de caso criado para demonstrar a aplicabilidade e a utilidade da proposta e um protótipo desenvolvido para verificar sua viabilidade técnica e correta operação. Finalmente, no Capítulo 6, são expostas as conclusões, as contribuições, as limitações e as propostas de trabalhos futuros.

2 FUNDAMENTAÇÃO TEÓRICA

Neste capítulo, são descritos conceitos fundamentais para o entendimento da proposta deste trabalho.

2.1 COMPUTAÇÃO EM NUVEM

A Computação em Nuvem, segundo Mell e Grance (2011), é um modelo que possibilita o acesso ubíquo, conveniente e sob demanda através de uma rede a um conjunto compartilhado de recursos computacionais configuráveis que podem ser rapidamente provisionados e liberados com esforço mínimo de gestão ou de interação com o provedor de serviço. Desta forma, abrange tanto as aplicações que são fornecidas como serviço através da Internet quanto o hardware e o software dos *datacenters* responsáveis pelo fornecimento de tais serviços (ARMBRUST et al., 2010).

O *National Institute of Standards and Technology* (NIST) define cinco características essenciais que devem ser contempladas por uma nuvem (MELL; GRANCE, 2011). São elas:

- Autoatendimento sob demanda: uma nuvem permite que seus usuários utilizem seus serviços de forma automática e sob demanda, sem a necessidade de interação com operadores humanos;

- Amplo acesso à rede: os recursos são fornecidos através da rede e acessados por mecanismos padronizados, que permitem o uso por plataformas heterogêneas;

- Agrupamento de recursos: os recursos computacionais (físicos e virtuais) do provedor de serviços são agrupados de forma a atender dinamicamente as demandas dos consumidores, que, na maior parte das vezes, não têm conhecimento da localização exata de tais recursos, mas podem ser capazes de determiná-la em um nível de abstração mais alto (como país, estado ou *datacenter*);

- Rápida elasticidade: os recursos fornecidos através de uma nuvem podem ser provisionados e liberados rapidamente para ajustarem-se à variação da demanda. Como resultado, frequentemente, o usuário tem a impressão de que os recursos oferecidos são ilimitados; e

- Serviço medido: sistemas em nuvem controlam e otimizam o uso dos recursos automaticamente através de medições e podem monitorar e reportar a utilização de recursos de forma transparente aos consumidores e provedores.

2.1.1 Modelos de Serviços

Os serviços oferecidos em um ambiente computacional em nuvem são divididos em três modelos ou níveis de abstração, que podem ser vistos como uma arquitetura de camadas na qual os serviços de uma camada mais baixa compõem serviços de uma camada mais alta. Os modelos de serviços são descritos a seguir, de acordo com Voorsluys e Buyya (2011) e Mell e Grance (2011):

- Infraestrutura como Serviço (*Infrastructure as a Service* - IaaS): é considerada a camada inferior dos sistemas computacionais em nuvem e oferece recursos virtualizados de servidores de computação, armazenamento e comunicação sob demanda. Como exemplo, pode-se citar o EC2 da Amazon Web Services, que permite aos usuários utilizar servidores virtuais e configurar e escalonar a capacidade do ambiente computacional (AMAZON, 2017);

- Plataforma como Serviço (*Platform as a Service* - PaaS): esta camada, com nível de abstração intermediário, oferece um ambiente de desenvolvimento e hospedagem que permite aos usuários criar aplicações sem preocupar-se com requisitos de infraestrutura, como quantidade de processadores ou memória necessária. O Windows Azure, da Microsoft, é exemplo deste tipo de serviço e possibilita criar, implantar e gerenciar aplicativos através de uma rede global de *datacenters* (MICROSOFT, 2017);

- Software como Serviço (*Software as a Service* - SaaS): é a camada superior da arquitetura, na qual os serviços oferecidos consistem em aplicações que podem ser acessadas pelos usuários finais através da Web, sem a necessidade de instalá-las localmente e com as mesmas funcionalidades de programas tradicionais para *desktop*. O G Suite da Google constitui um exemplo deste modelo, uma vez que oferece como serviço aplicativos de e-mail, agenda, criação e compartilhamento de documentos, entre outros (GOOGLE, 2017).

2.1.2 Modelos de Implantação

Segundo Mell e Grance (2011), com base nas formas de implantação e acesso, e a despeito do modelo de serviço oferecido, uma nuvem pode ser classificada como:

- Pública: modelo de nuvem disponibilizado por organizações empresariais, acadêmicas ou governamentais para o público em geral;

- Privada: a infraestrutura da nuvem pode ser utilizada apenas por membros de uma determinada organização, embora possa ser gerenciada por esta ou por terceiros;

- Comunitária: é compartilhada por um grupo de consumidores de organizações distintas, mas com objetivos ou políticas em comum e pode ser gerenciada e operada por uma ou mais organizações pertencentes à comunidade ou por terceiros; e

- Híbrida: é composta por dois ou mais tipos diferentes de nuvens (pública, privada ou comunitária), que, apesar de unidas por tecnologias proprietárias ou padronizadas, permanecem como entidades únicas.

2.1.3 Nuvem Federada

A Agência de Segurança de Rede e Informação (*European Network and Information Security Agency* - ENISA) define, ainda, como um modelo de implantação a Nuvem Federada, que pode ser construída através da agregação de duas ou mais nuvens (CATTEDDU; HOGBEN, 2009) e é o âmbito de aplicação da proposta deste trabalho.

Segundo Kertesz (2014), o conceito de Federação de Nuvem diz respeito a um conjunto de provedores de nuvem interconectados através de padrões abertos com o objetivo de fornecer um ambiente de computação universal e descentralizado no qual as decisões são guiadas por acordos e restrições entre os participantes da federação.

As Federações de Nuvem proveem um ambiente distribuído e heterogêneo, composto por várias infraestruturas de nuvem, através da agregação de diferentes capacidades de provedores pertencentes tanto à área comercial quanto acadêmica (KERTESZ, 2014).

2.2 GERENCIAMENTO DE IDENTIDADE (IDM)

O Gerenciamento de Identidade engloba as tecnologias e os processos necessários para a criação, o gerenciamento e o uso de identidades digitais. Assim, o IdM é responsável pelo estabelecimento da identidade de um usuário ou sistema (autenticação), pelo gerenciamento do acesso a serviços por esse usuário (controle de acesso) e pela manutenção de perfis de identidade do usuário (ALPÁR; HOEPMAN; SILJEE, 2011).

Alguns conceitos básicos do IdM são descritos a seguir, de acordo com Benantar (2006), Bertino e Takahashi (2011) e ISO/IEC (2011):

- Entidade: termo genérico que designa um agente ativo que pode executar algum tipo de computação.
- Usuário: entidade externa a um sistema que pode ser uma pessoa ou um agente programável.
- Informação de Identificação Pessoal (PII): qualquer informação que pode ser usada para identificar o usuário ao qual a PII é relacionada, ou pode ser direta ou indiretamente ligada a ele. Assim, dependendo do escopo, podem ser consideradas PII informações como data de nascimento, localização GPS, endereço IP e interesses pessoais inferidos pelo rastreamento do uso de websites.
- Outorgante das PIIs: pessoa física a quem a PII é relacionada;
- Recurso: entidade, como uma aplicação ou um arquivo, à qual o acesso é requerido por um usuário.
- Provedor de Identidade (IdP): parte que provê identidades aos usuários e é, geralmente, responsável pelo processo de autenticação.
- Provedor de Serviço (SP): parte que provê serviços ou acesso a recursos aos usuários e, para isso, requer a submissão de credenciais válidas.

2.2.1 Identidade

De acordo com Benantar (2006), identidades são representações computacionais de entidades ativas em um sistema, como pessoas, dispositivos de rede ou agentes de programação. Assim, uma identidade

aponta para diversos atributos e direitos de uma entidade que, coletivamente, podem ser referidos como perfil. Ainda segundo o autor, devido ao advento das redes e dos sistemas de computação distribuída, uma identidade pode atravessar os limites de um único sistema computacional.

Para Bertino e Takahashi (2011), uma identidade digital engloba todas as propriedades e características disponíveis que podem ser utilizadas para identificar um usuário dentro de um contexto. Já a ISO/IEC 29100 define identidade como o conjunto de atributos que torna possível identificar o outorgante das PIIs (ISO/IEC, 2011).

Neste trabalho, considera-se como identidade uma representação digital de um usuário composta por um identificador único utilizado para referir-se a tal identidade e pelas suas PIIs.

2.2.2 Login Único (SSO)

O Login Único (*Single Sign-On* - SSO), de acordo com Benantar (2006), permite ao usuário estabelecer sua identidade uma única vez e acessar outras aplicações e serviços conectados pela mesma rede de forma contínua já que não exige novas autenticações a cada acesso. Isto é possível devido ao estabelecimento de um contexto de segurança.

Tecnologias de autenticação como o Kerberos, um protocolo parte do *Project Athena* do *Massachussets Institute of Technology* (MIT), permitem a implementação de SSO ao funcionarem como uma terceira parte responsável por autenticar o usuário e fornecer sua identidade a todos os sistemas que este deseja acessar (MIT, 2017).

Os sistemas tradicionais de SSO funcionam apenas quando o provedor de identidade e o provedor de serviço pertencem à mesma organização. Quando o SSO é viabilizado independentemente do fato de os dois componentes encontrarem-se na mesma instituição, esta implementação é chamada de *Federated Single Sign-On*, isto é, Login Único Federado (SHIBBOLETH, 2017b). Uma federação é formada quando um grupo de IdPs e SPs concorda em trabalhar em conjunto e é explicada a seguir.

2.2.3 Federação

Uma federação tem como base a confiança entre organizações e se manifesta através dos mecanismos usados para permitir que uma or-

ganização participante forneça serviços diretamente a outras entidades registradas em diferentes organizações pertencentes à federação (BENANTAR, 2006). Assim, um usuário autenticado em um determinado IdP pode acessar serviços oferecidos por SPs localizados em diferentes domínios administrativos em virtude da relação de confiança previamente estabelecida entre eles.

De forma geral, em um ambiente de identidade federada, o IdP registra os usuários, estabelece suas credenciais, os autentica e confirma o status da sua autenticação para o SP. Este, por sua vez, consome as asserções de identidade fornecidas pelo IdP e usa a informação do status de autenticação para autorizar o usuário a acessar serviços e aplicações. A confiança entre membros de uma federação de identidade, portanto, é fundamental para sua operação e é estabelecida através de um conjunto de acordos e regras associadas conhecido como seu *framework* de confiança (TEMOSHOK; ABRUZZI, 2016).

Em diversos âmbitos, é comum que determinados provedores de serviço desejem trabalhar com mais de um provedor de identidade, como, por exemplo, serviços comerciais com múltiplos clientes ou recursos usados por pesquisadores de diferentes instituições (SHIBBOLETH, 2017b). As federações de identidade, deste modo, permitem que diversas organizações colaborem entre si para oferecer serviços compartilhados de forma segura e confiável. Alguns exemplos são:

- CAFe (Brasil): A Comunidade Acadêmica Federada (CAFe), criada em 2007 pela Rede Nacional de Ensino e Pesquisa (RNP), reúne instituições de pesquisa e ensino brasileiras e adota tecnologias como *OASIS Security Assertion Markup Language* (SAML) e Shibboleth para a criação de ambientes federados (RNP, 2017).

- MATE (Argentina): O projeto, iniciado pela InnovaRed em 2014, é uma iniciativa de federação para as instituições argentinas de educação e pesquisa que utiliza Shibboleth como base (INNOVARED, 2014).

- SWITCHaai (Suíça): Criada em 2004 e coordenada pela *Swiss Education and Research Network* (SWITCH), esta federação é construída com Shibboleth e engloba todas as universidades suíças para fornecer às instituições participantes acesso a diversos sistemas de *e-learning* (SWITCH, 2017).

2.2.4 Modelos de Gerenciamento de Identidade

Com base nos conceitos apresentados nesta seção, podem-se definir dois modelos de gerenciamento de identidade: o IdM em domínio administrativo único e o IdM em ambientes federados com múltiplos domínios. No IdM em domínio único, ilustrado na Figura 1, o usuário é autenticado apenas uma vez (SSO) e essa identidade é usada para acessar recursos fornecidos por diferentes provedores de serviço pertencentes ao mesmo domínio administrativo (domínio A) que seu provedor de identidade (BENANTAR, 2006).

O Gerenciamento de Identidade Federado (*Federated Identity Management* - FIM), por sua vez, possibilita o uso de identidades através de diferentes organizações, segundo Landau e Moore (2012). O FIM é mostrado na Figura 2, em que os domínios administrativos A e B cooperam entre si através de uma federação. Desta forma, o IdP do domínio A, por exemplo, pode fornecer identidades para permitir que o Usuário 1 acesse recursos oferecidos por provedores de serviço do domínio B. As interações realizadas pelo Usuário 1 são representadas por linhas sólidas, enquanto as executadas pelo Usuário 2 são representadas por linhas tracejadas.

2.2.5 Sistemas de Gerenciamento de Identidade

De acordo com Alpár, Hoepman e Siljee (2011), os sistemas de gerenciamento de identidade típicos envolvem três partes: usuários, provedores de identidade e provedores de serviço. O usuário visita um SP, que, por sua vez, depende do IdP para lhe fornecer informações autênticas sobre o usuário. Nestes sistemas, de forma geral, o fluxo de autenticação e troca de atributos de um usuário que deseja acessar um serviço ocorre como apresentado na Figura 3 e descrito a seguir:

1. O usuário faz a requisição de um serviço do SP;
2. O SP pede para que o usuário se autentique em um IdP. O IdP realiza o processo de autenticação;
3. Se a autenticação é finalizada com sucesso, o IdP dá ao usuário um token ou outra asserção de identidade que é encaminhada ao provedor de serviço. O SP, então, verifica a asserção e, se esta é válida, aceita o usuário como autenticado;
4. Se for necessário obter mais informações sobre o usuário, o SP

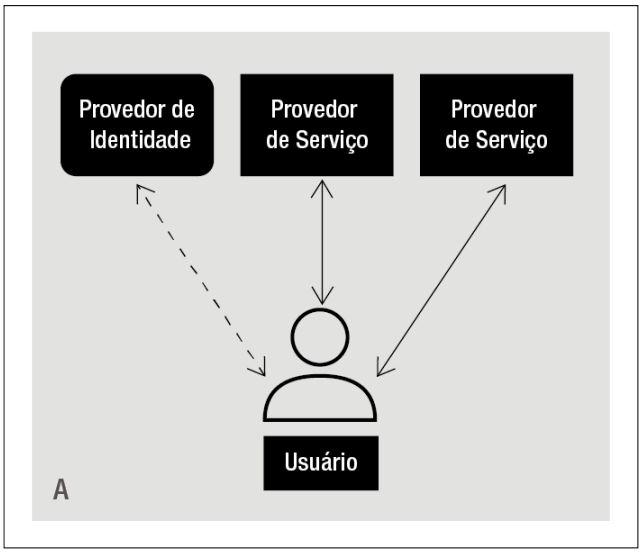

Figura 1 – Gerenciamento de identidade em um único domínio administrativo.

> pode contactar o IdP diretamente, usando a asserção como ponteiro para o perfil do usuário armazenado pelo IdP. Em alguns casos, esta última interação pode ser mediada pelo usuário.

Este processo genérico é utilizado em sistemas de gerenciamento de identidade para ambientes federados como o Shibboleth e o OpenID Connect, que são apresentados a seguir.

2.2.5.1 Shibboleth

O Shibboleth é uma ferramenta de código aberto criada e mantida pela organização Internet2 para fornecer um *framework* de SSO federado e de troca de atributos. Ele implementa padrões de identidade federada, como o SAML, e permite aos websites tomarem decisões de autorização informadas para cada acesso a recursos protegidos (SHIBBOLETH, 2017b).

Ao requisitar um acesso, o usuário é direcionado ao *Identity Provider Discovery*, um módulo de descoberta de provedores de identidade que auxilia os SPs que trabalham com múltiplos IdPs a identificarem

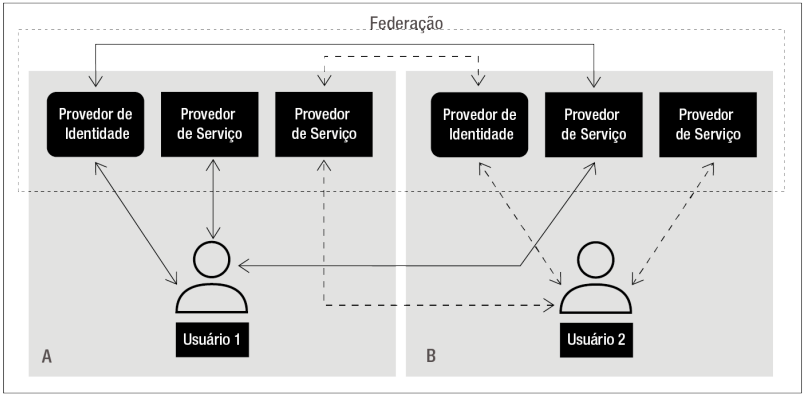

Figura 2 – Gerenciamento de identidade em ambiente federado com diferentes domínios administrativos.

para onde enviar a solicitação de autenticação. Este módulo oferece um seletor de provedores de identidade que deve ser integrado ao site do SP e no qual o usuário deve escolher aquele em que possui cadastro e que deseja usar para autenticar-se (SHIBBOLETH, 2017b).

O processo de SSO é realizado via HTTP e um documento com metadados que descreve vários aspectos técnicos de um IdP ou de um SP é utilizado para informar sobre as URLs que devem ser utilizadas durante a comunicação entre o provedor de identidade e o de serviços. Os metadados geralmente contêm informações como um ID de entidade, isto é, um identificador exclusivo; um nome e uma descrição legíveis por humanos; uma lista de URLs para as quais as mensagens devem ser enviadas e informações sobre quando usar cada uma; e informações de criptografia utilizadas ao criar e verificar mensagens.

De acordo com Weingartner (2014), o Shibboleth é amplamente adotado no meio acadêmico, porém sua documentação é incompleta, o que dificulta sua configuração e o entendimento de seu funcionamento.

2.2.5.2 OpenID Connect (OIDC)

O OpenID Connect (OIDC), criado e mantido pela fundação OpenID, é um protocolo de autenticação que utiliza o padrão JSON *(JavaScript Object Notation)* como formato de dados para o fluxo de mensagens (OPENID, 2015).

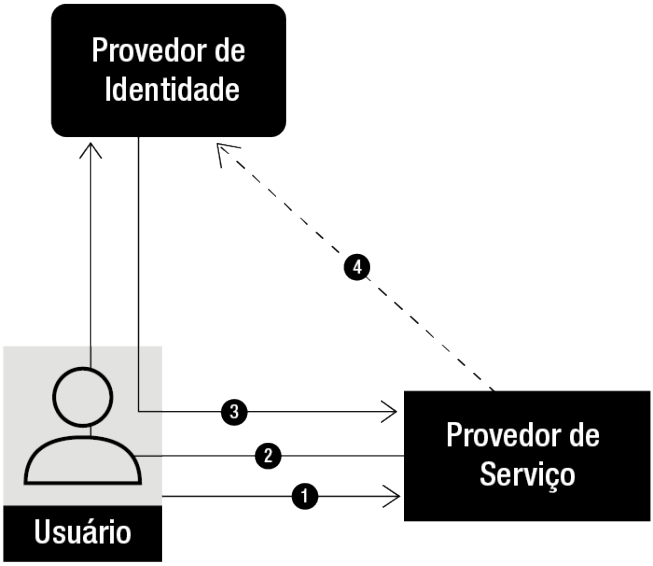

Figura 3 – Fluxo de autenticação e troca de atributos de um usuário em um sistema de IdM.

O OIDC adiciona uma camada de identidade ao *OAuth 2.0*, um *framework* para desenvolvimento de protocolos de autenticação e autorização. Esta camada de identidade fornece informações sobre quem é o usuário que foi autenticado; onde, quando e como ele foi autenticado; quais atributos ele pode fornecer; e por que ele os está fornecendo (OPENID, 2015). Para tanto, o OIDC incorpora as concessões de autorização e os tokens de acesso definidos no *OAuth* e define um novo tipo de token, o token de ID, que contém informações para a autenticação do usuário. Ele adiciona, ainda, uma entidade chamada *UserInfo*, que fornece informações sobre o usuário a entidades que apresentam um token de ID válido (PEREZ-MENDEZ et al., 2014).

Para a implementação do token, o OpenID Connect utiliza estruturas de dados *JSON Web Token (JWT)*, um padrão aberto que permite representar atributos de forma compacta e segura para transferência entre duas partes. Estes atributos são codificados como um objeto JSON que é, então, usado como carga para gerar uma estrutura *JSON Web Signature (JWS)* ou como texto em claro de uma estrutura *JSON Web Encryption (JWE)*, permitindo que os atributos sejam, respectivamente, assinados digitalmente ou protegidos com um Código de

Autenticação de Mensagem (*Message Authentication Code* - MAC), e/ou criptografados (JONES; SAKIMURA, 2015).

O token de ID do OpenID Connect é composto por três partes, ilustradas na Figura 4: um cabeçalho, que contém informações sobre o tipo da estrutura e o algoritmo de segurança; um conjunto de atributos, que contém dados sobre a entidade representada pelo token, sobre o emissor e o receptor do mesmo, entre outros; e uma assinatura. Todos os dados do token são codificados em Base64 através de serialização compacta e o resultado final é uma sequência de caracteres, como mostrado na Figura 5, que pode ser transmitida através de URL e na qual cada seção é separada por um ponto final (CONNECT2ID, 2017b).

```
{
  "typ": "JWT",
  "alg": "HS256"
}
{
  "sub"           : "alice",
  "iss"           : "https://openid.c2id.com",
  "aud"           : "client-12345",
  "nonce"         : "n-0S6_WzA2Mj",
  "auth_time"     : 1488405982,
  "acr"           : "c2id.loa.hisec",
  "iat"           : 1488405983,
  "exp"           : 1488406983,
}
{
  D7SDQBpVCSRSqVUMP9PAungM0gh7JKjKgXYhUlKMr3Y
}
```

Figura 4 – Estrutura do token de ID do OpenID Connect adaptada de Connect2id (2017b).

O OIDC possui diversas implementações com documentação detalhada e de fácil compreensão e é utilizado por organizações como Google e MIT. Além disso, tem seus componentes especificados pela fundação OpenID e apresenta grande interoperabilidade por seguir a arquitetura *Representational State Transfer* (REST) e utilizar o padrão JSON (WEINGARTNER, 2014).

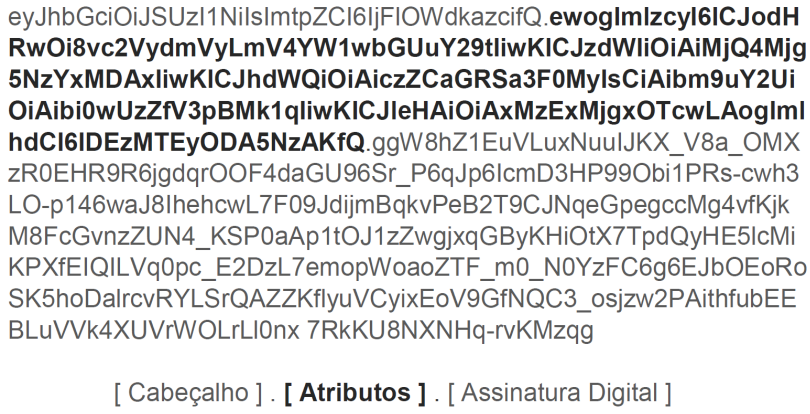

Figura 5 – Exemplo de um token de ID do OpenID Connect codificado em Base64 adaptado de Connect2id (2017b).

2.3 PRIVACIDADE

Westin (1967), em seu livro Privacidade e Liberdade, define privacidade como o direito de indivíduos, grupos ou instituições de determinarem por si mesmos quando, como e qual tipo de informação sobre eles pode ser revelada a outros. Segundo o autor, um aspecto essencial da liberdade do ser humano envolve o controle sobre suas informações pessoais. Assim, sua definição de privacidade destaca a habilidade das pessoas de decidirem sobre a quantidade, os destinatários e as condições para a divulgação de seus dados pessoais.

Bertino e Takahashi (2011) definem privacidade como o direito de um sujeito de controlar suas identidades em transações e a consideram parte importante do IdM. Para Xiao e Xiao (2013), a privacidade é uma questão crítica para a Computação em Nuvem devido ao fato de que os dados e a lógica de negócio dos usuários são armazenados em servidores que pertencem aos provedores de nuvem e são mantidos por eles e, assim, existem riscos de que informações pessoais dos usuários sejam indevidamente usados ou divulgados.

Segundo a taxonomia de privacidade definida por Solove (2006), o uso secundário de dados consiste no uso dos mesmos para propósitos diferentes daqueles para os quais foram inicialmente coletados sem o consentimento do sujeito. De acordo com Iyilade e Vassileva (2014),

os dados coletados por sistemas e aplicações podem revelar comportamentos, atividades e contexto dos usuários e o uso secundário destas informações pode representar riscos significativos à sua privacidade. Por exemplo, os dados pessoais podem ser usados para fins potencialmente prejudiciais, como para monitoramento ou para traçar o perfil do usuário para compartilhá-lo com potenciais empregadores ou empresas de seguros ou de empréstimos.

Neste trabalho, considera-se como privacidade o direito de um usuário de que seus dados de identificação pessoal não sejam usados ou divulgados para propósitos diferentes dos originais sem seu consentimento.

Uma forma de alcançar a privacidade em sistemas computacionais é através da implementação de Tecnologias de Reforço à Privacidade (*Privacy Enhancing Technologies* - PETs). Segundo a ISO/IEC (2011), as PETs são controles de privacidade que consistem em medidas, produtos ou serviços de Tecnologia da Informação que protegem a privacidade dos usuários através da eliminação ou redução de PIIs ou, ainda, mediante a prevenção de processamento de PIIs desnecessário ou indesejado.

2.3.1 Normas e Modelos de Referência

A principal preocupação de privacidade em sistemas de informação e comunicação diz respeito à proteção de PIIs. Desta forma, existem normas e modelos de referência internacionais que visam estabelecer padrões e metodologias para incentivar e guiar a implementação de mecanismos que protejam a privacidade dos usuários. Neste trabalho, foram usados como base para o desenvolvimento do mecanismo proposto as normas e os modelos descritos a seguir.

O padrão internacional ISO/IEC 29100:2011 tem como objetivo fornecer um *framework* de alto nível para a proteção de PIIs em sistemas de tecnologia da informação e comunicação. Para isso, define uma terminologia comum de privacidade, bem como os atores e seus papéis no processamento de PIIs, descreve requisitos de proteção de privacidade e fornece referências para princípios de privacidade conhecidos para tecnologia da informação. A norma define como determinar se um atributo do usuário que pode ser considerado PII e fornece exemplos de PIIs (ISO/IEC, 2011).

A ISO/IEC 29101:2013 baseia-se na ISO/IEC 29100 e descreve uma arquitetura de referência com boas práticas para a implementação

de requisitos de privacidade em sistemas que armazenam e processam PIIs. Para tanto, a arquitetura fornece uma abordagem consistente de alto nível para a implementação de controles de privacidade no processamento de PIIs; oferece orientação para o planejamento, o projeto e a construção de sistemas que protegem a privacidade do usuário através do controle do processamento, do acesso e da transferência de PIIs; e demonstra como as PETs podem ser usadas como controles de privacidade (ISO/IEC, 2013).

Já o *Privacy Management Reference Model and Methodology* (PMRM) é um modelo conceitual para auxiliar na implementação de funcionalidades de gerenciamento de privacidade e mecanismos capazes de executar controles de privacidade alinhados com políticas de privacidade. Este modelo de referência tem como objetivo melhorar a gestão, o cumprimento e a prestação de contas de privacidade em sistemas de informação complexos através de um modelo conceitual de gerenciamento de privacidade, que inclui definições de termos; uma metodologia; e um conjunto de Serviços e Funções necessários para suportar controles de privacidade (OASIS, 2016).

2.3.2 Políticas de Privacidade

Aplicações e sistemas computacionais coletam dados de forma passiva ou compartilhados voluntariamente pelos usuários. Para proteger estes dados e estar em conformidade com diferentes legislações, as organizações expressam suas práticas internas de segurança e privacidade na forma de declarações contidas em políticas de privacidade (KUMARAGURU et al., 2007). Segundo Caramujo e Silva (2015), estas políticas descrevem como os dados do usuário são gerenciados e divulgados em um sistema.

Neste trabalho, define-se como política de privacidade o documento que expressa as práticas das organizações em relação à coleta, ao uso e ao compartilhamento de dados do usuário.

Para representar as políticas de privacidade de maneira precisa e em um formato que possa ser entendido por máquinas, existem diversas linguagens, que, segundo Iyilade e Vassileva (2014), permitem tanto usuários e organizações quanto provedores de serviços expressarem seus controles e permissões de privacidade. Estas linguagens são consideradas trabalhos correlatos e as mais relevantes para esta proposta são descritas no capítulo seguinte.

De acordo com Kumaraguru et al. (2007), as políticas de privaci-

dade podem ser usadas para ganhar a confiança do usuário e aumentar os negócios, bem como para ajudar os clientes a tomarem decisões informadas. No entanto, pesquisas mostram que as políticas de privacidade, embora geralmente sejam as únicas fontes de informação sobre como os dados são usados pelos provedores, são vistas como complexas e incompreensíveis, e lidas apenas por uma pequena fração de usuários (KOLTER, 2010).

2.3.3 Preferências de Privacidade

De acordo com o padrão internacional ISO/IEC 29100, apresentado na Seção 2.3.1, preferências de privacidade são escolhas específicas feitas por um usuário sobre como suas PIIs devem ser processadas para um determinado propósito. Assim, podem-se citar como exemplos a preferência por anonimidade, a capacidade de restringir quem pode acessar uma PII específica ou a capacidade de restringir para qual propósito uma PII pode ser processada (ISO/IEC, 2011).

Neste trabalho, definem-se preferências de privacidade como as permissões de um usuário em relação ao uso secundário de suas PIIs, isto é, elas determinam para quais propósitos e para o benefício de quem um tipo de PII pode ser usado.

A ISO/IEC 29100 determina que, na medida do possível, deve ser dada ao outorgante das PIIs a possibilidade de escolher suas preferências quanto ao processamento de seus dados. O padrão sugere ainda que a capacidade de expressar preferências de privacidade seja implementada através de uma interface gráfica para ajudar o outorgante das PIIs a fazer uma escolha, apresentando um conjunto de opções predefinidas para as preferências de privacidade e usando linguagem facilmente compreensível. A implementação desta interface pode ser baseada em elementos como caixas de seleção ou menus suspensos (ISO/IEC, 2011).

3 TRABALHOS RELACIONADOS

Neste capítulo, são apresentados e analisados trabalhos nos quais é baseada a proposta desta dissertação e outras abordagens que têm como finalidade fornecer privacidade ao usuário em ambientes computacionais.

3.1 CLASSIFICAÇÃO DE USUÁRIOS POR SUAS PREFERÊNCIAS DE PRIVACIDADE

Chanchary e Chiasson (2015) realizaram uma pesquisa para entender como os usuários percebem o rastreamento online para publicidade comportamental. Os participantes, provenientes de diversos países, foram recrutados através de um serviço de *crowdsourcing* online. Assim, eles demonstraram que os usuários têm preferências claras em relação a quais classes de informação eles gostariam de divulgar online e que alguns estariam mais propensos a compartilhar dados se tivessem controle prévio de ferramentas de proteção de rastreamento. Os autores também identificaram três grupos de usuários de acordo com a forma como suas atitudes em relação à privacidade influenciavam sua vontade de compartilhar informação. Estes grupos são usados como base para os perfis de privacidade do mecanismo proposto nesta dissertação e são apresentados a seguir:

- Fundamentalistas (30.4%): consideram a privacidade como um aspecto muito importante e têm um sentimento muito forte em relação a ela;

- Pragmáticos (45.9%): consideram privacidade como um aspecto muito importante mas também gostam dos benefícios de abdicar um pouco de privacidade quando acreditam que sua informação não será usada incorretamente; e

- Despreocupados (23.6%): não consideram a privacidade como um aspecto importante ou não se preocupam sobre como as pessoas e organizações usam suas informações.

3.2 LINGUAGENS DE POLÍTICA DE PRIVACIDADE

Como mencionado no capítulo anterior, existem diversas linguagens que têm como finalidade representar as políticas de privacidade de maneira precisa e em um formato legível por máquinas. As três mais utilizadas e mais relevantes para esta proposta são apresentadas nos próximos parágrafos.

O P3P (*Platform for Privacy Preferences*) é um protocolo proposto pela W3C (2006) para informar aos usuários sobre as práticas de coleta e uso de dados de websites. Uma política P3P consiste em um conjunto de declarações no formato XML (*eXtensible Markup Language*) aplicadas a recursos específicos, como páginas, imagens ou *cookies*. Quando um usuário acessa um site que possui suas políticas definidas em P3P, cada vez que este deseja coletar algum dado, as preferências daquele são comparadas à política correspondente. Se esta é aceitável para ele, a transação continua automaticamente; se não, o usuário é notificado e pode decidir entre permitir (*opt-in*) ou rejeitar (*opt-out*) a transação. Esta linguagem fornece uma base para a coleta de preferências de privacidade, mas requer que todos os usuários e SPs definam suas políticas de privacidade nesta linguagem e não atende às necessidades dos ambientes de nuvem federada.

A EPAL (*Enterprise Privacy Authorization Language*) é uma linguagem formal criada pela IBM (2003) para suprir a necessidade da indústria de expressar as políticas de privacidade internas das organizações, que definem as práticas de uso de dados de acordo com direitos de autorização negativos e positivos. Uma política EPAL define uma lista de hierarquias de categorias de dados, categorias de usuários e propósitos, bem como conjuntos de ações, obrigações e condições. Estes elementos são usados para formular regras de autorização de privacidade que permitem ou rejeitam ações sobre categorias de dados por categorias de usuários para determinados propósitos sob dadas condições enquanto exigindo certas obrigações. Porém, por ser específica para políticas internas de empresas, não considera as preferências do usuário e não é adequada para privacidade em ambientes de identidade federada.

Estas linguagens de política de privacidade tradicionais, no entanto, preocupam-se apenas com alertar e fornecer controle ao usuário durante a coleta inicial de dados e, portanto, não oferecem meios para definir preferências que dizem respeito ao uso secundário dos dados. Para preencher esta lacuna, Iyilade e Vassileva (2014) propuseram a P2U (*Purpose-to-Use*), inspirada na P3P, mas que permite a especi-

ficação de políticas de privacidade editáveis pelo usuário e negociáveis. Estas políticas definem o propósito de uso, o tipo, o período de retenção e o preço dos dados compartilhados. Esta linguagem, embora permita a criação de políticas editáveis e negociáveis, é complexa para os usuários uma vez que assume que estes possuem políticas de privacidade e são capazes de defini-las em P2U. Ela também requer que os SPs expressem suas políticas em tal linguagem.

3.3 PERFIS UML DE PRIVACIDADE

Enquanto as linguagens de política de privacidade são abordagens de baixo nível que podem ser lidas e entendidas por máquinas, os perfis UML são especificações de alto nível que têm como objetivo fornecer um melhor entendimento dos requisitos de privacidade no desenvolvimento de aplicações e sistemas.

Caramujo e Silva (2015) estendem e validam um modelo conceitual suportado por um perfil UML que pode ser aplicado para representar políticas de privacidade para auxiliar no desenvolvimento de aplicações e sistemas integrados a redes sociais e melhorar o cumprimento de tais políticas. O modelo é composto pelos elementos: *PrivacyPolicy*, que representa o documento que os usuários devem aceitar para usar os serviços fornecidos pelas empresas; *Statement*, que descreve as regras ou ações especificadas na política de privacidade; *Recipient*, que representa uma ou mais entidades para as quais a política dá acesso às informações do usuário; *PrivateData*, que representa o tipo de informação do usuário que é coletada e gerenciada pelo SP; *Service*, que define quais serviços são oferecidos pelo SP; e *Enforcement*, que são os mecanismos disponíveis para fazer cumprir os *statements* descritos na política de privacidade.

Basso et al. (2015) definem um perfil UML para auxiliar no desenvolvimento de aplicações e serviços que precisam ser consistentes com as declarações de suas políticas de privacidade. Os autores identificam elementos de privacidade, como políticas e declarações, através das quais os usuários podem definir suas preferências quanto à coleta, ao uso, à retenção e à liberação de seus dados, e organizam suas relações em um modelo conceitual. Este modelo é, então, mapeado para um perfil UML definido por estereótipos, atributos e restrições que permite modelar declarações de políticas de privacidade reais.

Embora estes perfis ajudem os desenvolvedores a verificarem conflitos e inconsistências em relação às políticas de privacidade envolvidas

durante a implementação de sistemas e aplicações integradas, não oferecem meios práticos para os usuários definirem suas preferências de privacidade e transmiti-las aos SPs para que estes alinhem suas ações de coleta e uso de PIIs.

3.4 MECANISMOS DE PRIVACIDADE DE SISTEMAS DE IDM

Nesta seção, são apresentados os mecanismos de privacidade disponíveis nos sistemas de gerenciamento de identidade federada apresentados no Capítulo 2.

3.4.1 User Consent do Shibboleth

O Shibboleth, até sua segunda versão, possuía uma extensão para o IdP desenvolvida pela SWITCH e chamada uApprove, que adicionava um fluxo para obter o consentimento do usuário quanto à liberação de seus dados. A partir da terceira versão, com as mudanças de projeto do Shibboleth, o mecanismo de consentimento passou a ser fornecido como parte padrão do software e tornou o uApprove obsoleto (SHIBBOLETH, 2017a).

Este mecanismo requer que os usuários aceitem a liberação de atributos para provedores de serviços durante autenticações que incluem dados de atributo na resposta (SHIBBOLETH, 2017a). Assim, os usuários são solicitados a dar consentimento para a liberação de atributos:

- No primeiro acesso aos recursos de um SP;
- Na liberação de um atributo para o qual consentimento não foi dado antes;
- Quando um atributo para o qual já foi dado consentimento não é mais liberado; e
- Quando o valor de um atributo para o qual já foi dado consentimento muda.

É possível habilitar o consentimento por atributo para permitir que o usuário selecione os atributos que ele deseja liberar e definir listas de atributos para os quais o usuário deve ser sempre solicitado a consentir (*whitelist*), de atributos para os quais o usuário não deve ser

solicitado a consentir (*blacklist*) e de atributos correspondentes a uma expressão regular para os quais o usuário deve ser solicitado a consentir (*Regex*).

Quanto à duração do consentimento para liberação de dados, os usuários podem escolher entre três opções: ser solicitado a cada login, ser solicitado se os atributos fornecidos para determinado serviço mudaram desde quando o consentimento foi dado (opção padrão) e nunca ser solicitado (opcional). Com a última opção, chamada de consentimento global, todos os atributos são liberados para qualquer provedor de serviço.

O *User Consent*, no entanto, possui algumas limitações. Para muitos serviços, por exemplo, a lista de atributos para os quais o usuário deve dar consentimento pode ser muito extensa, o que aumenta a complexidade e, muitas vezes, leva os usuários a liberarem todos os dados e optarem por não serem mais solicitados nos próximos acessos. Além disso, a permissão solicitada é apenas para a liberação dos dados que serão enviados explicitamente ao provedor de serviço pelo IdP para viabilizar o serviço e, portanto, não inclui informações que podem ser induzidas pelo SP e não conscientiza os usuários sobre os possíveis usos secundários dos seus dados.

3.4.2 End-User Consent/Authorization do OpenID Connect

O OpenID Connect possui um mecanismo de privacidade integrado que permite aos usuários consentirem ou negarem a liberação de determinados tipos de dados ao provedor de serviço (OPENID, 2014).

O Servidor de Autorização do OIDC, após a autenticação do usuário, precisa obter autorização antes de liberar informações ao SP. Este utiliza escopos para especificar quais privilégios de acesso são solicitados para um determinado recurso e o usuário os usa para determinar quais conjuntos específicos de atributos estarão disponíveis para o provedor de serviço. Uma aplicação pode solicitar as permissões específicas que necessita através do parâmetro *scope*.

O OpenID Connect define os seguintes escopos de dados:

- *openid*: este escopo é obrigatório e informa ao Servidor de Autorização que o SP está fazendo uma requisição OpenID Connect;

- *profile*: este escopo solicita acesso aos atributos padrão do perfil do usuário, como nome, sobrenome, nome de usuário, foto, gênero e data de nascimento;

- *email*: este escopo solicita acesso aos atributos referentes ao email do usuário;
- *address*: este escopo solicita acesso ao atributo de endereço;
- *phone*: este escopo solicita acesso aos atributos referentes ao telefone do usuário; e
- *offline_access*: este escopo solicita que seja concedido acesso ao *UserInfo* do usuário mesmo quando este não está logado.

Assim como o mecanismo presente no Shibboleth, o *End-User Consent/Authorization* do OpenID Connect solicita apenas autorização do usuário para que o IdP forneça os dados solicitados pelo SP para prestar o serviço e, portanto, não inclui informações que podem ser induzidas pelo provedor de serviço e não considera os possíveis usos secundários dos mesmos.

3.5 ARQUITETURA DE PRIVACIDADE CENTRADA NO USUÁRIO

Kolter (2010) propõe uma arquitetura de privacidade centrada no usuário e independente do provedor de serviço. Esta arquitetura é composta por uma comunidade colaborativa de privacidade e três componentes de gerenciamento de privacidade do lado do usuário, que são explicados a seguir.

A Comunidade de Privacidade permite que os usuários troquem informações sobre privacidade, avaliações e experiências sobre provedores de serviço, como quantidade de dados pessoais necessários para preencher um formulário e terceiras partes com quem o SP compartilha informações. O Gerador de Preferências de Privacidade oferece uma ferramenta para que os usuários definam suas preferências de privacidade. Já o Agente de Privacidade mostra ao usuário informações relevantes da Comunidade de Privacidade, a avaliação da política de privacidade e a reputação dada pela comunidade ao website visitado e compara as preferências geradas pelo Gerador de Preferências de Privacidade com políticas de privacidade legíveis por máquina. O componente Log de Divulgação de Dados, por sua vez, registra transferências de dados pessoais e fornece uma visão geral de fluxos de dados pessoais passados (KOLTER, 2010).

Esta arquitetura, no entanto, é complexa e não é completamente independente do SP uma vez que demanda que os provedores representem suas políticas em P3P. O Agente de Privacidade exige que o usuário

instale uma extensão em seu navegador; a Comunidade de Privacidade requer que os usuários a mantenham e forneçam informações confiáveis e explicações sobre as políticas de privacidade dos provedores; e a ferramenta de geração de preferências de privacidade exige que os usuários definam preferências específicas para doze tipos de serviços oferecidos por SPs e lista extensivamente os dados divididos em nove tipos. Se um tipo de serviço não é configurado, a ferramenta entende que o usuário não deseja interagir ou disponibilizar qualquer dado pessoal com SPs que oferecem esse tipo de serviço.

3.6 MODELO PARA IDM COM PRIVACIDADE EM NUVEM

Werner e Westphall (2016) propõem um modelo de gerenciamento de identidade com privacidade para nuvem no qual IdPs e SPs interagem em ambientes dinâmicos e federados para gerenciar as identidades e garantir a privacidade de usuários pertencentes a diferentes domínios administrativos. Os autores propõem configurações de privacidade predefinidas e personalizáveis que ajudam o usuário a declarar seu nível de privacidade desejado ao permitir que ele escolha um escopo de disseminação de dados (acesso anônimo, com pseudônimo, com atributos parciais ou com atributos totais) e ao alertar sobre o grau de confiança do SP.

O modelo de interação, mostrado na Figura 6, propõe o registro no IdP dos atributos e credenciais do usuário, que podem ser criptografados (Passo 1), bem como das políticas de privacidade para regular o uso e a disseminação de suas PIIs (Passo 2). Tanto os dados quanto as políticas registradas são colocados em um pacote chamado *sticky policies*, que é enviado ao SP em conjunto com um modelo de disseminação de dados e obrigações que devem ser cumpridas pelo provedor de serviço. A ideia das *sticky policies* é que as PIIs sejam sempre disseminadas com as políticas que regem o seu uso e sua disseminação de modo que as preferências de privacidade do usuário sejam cumpridas por qualquer SP. Os autores, no entanto, não definem um mecanismo para recolher essas preferências, anexá-las aos seus dados e enviá-las ao SP para que as cumpra.

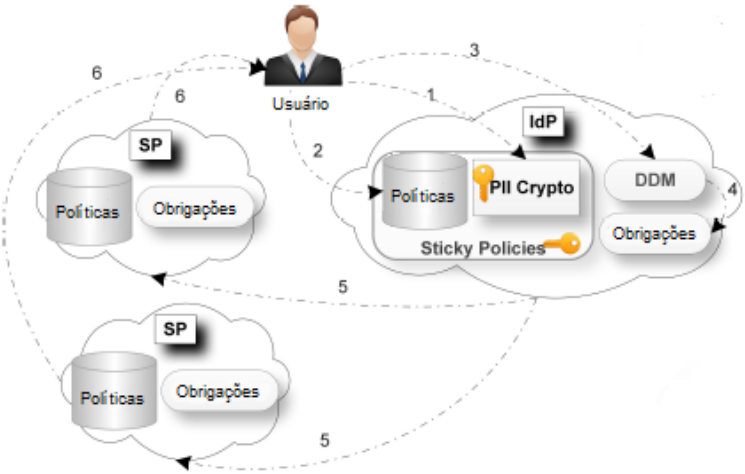

Figura 6 – Modelo de interação entre usuário, IdP e SP adaptado de Werner e Westphall (2016).

3.7 COMPARAÇÃO DOS TRABALHOS RELACIONADOS

A Tabela 1 apresenta a comparação de alguns dos trabalhos descritos e analisados neste capítulo com base nas características desejáveis de um mecanismo para especificar as preferências de privacidade do usuário em ambientes de nuvem federada. Foram incluídos nesta tabela apenas os trabalhos que são comparáveis ao mecanismo proposto ou que poderiam ser usados como base para sua criação. As características desejáveis são:

- Centrado no usuário: permite ao usuário definir suas preferências de privacidade. Os provedores de serviço devem adequar-se a elas e não o usuário às políticas dos SPs;
- Conscientiza sobre uso secundário: alerta o usuário sobre a possibilidade de usos secundários de seus dados;
- Controla uso secundário: possibilita ao usuário definir os usos secundários permitidos de seus dados pessoais;
- Adequado para nuvem federada: oferece recursos para ser implementado em ambientes de nuvem federada;

- Independente da representação de políticas: não impõe ao SP o uso de tecnologias específicas para representar e implementar suas políticas de privacidade;
- Fácil adoção: é simples e de fácil implantação, tanto por parte do usuário quanto do IdP e do SP. Não requer ferramentas e conhecimentos específicos do usuário e é implantado com tecnologias que o IdP e o SP já utilizam; e
- Usabilidade: permite ao usuário identificar sua necessidade de privacidade e configurar suas preferências de forma simples, intuitiva e eficiente.

Com base na tabela, pode-se dizer que as propostas existentes não oferecem métodos práticos para que os usuários definam suas preferências e que estas sejam enviadas ao SP e, na maior parte dos casos, obrigam o provedor de serviço a adotar tecnologias específicas para representar suas políticas de privacidade. Assim, a maioria não é centrada nos usuários, não os conscientiza sobre o uso secundário de seus dados nem lhes permite controlá-los, não são de fácil adoção e não são adequados para ambientes de nuvem federada.

Tabela 1 – Comparação dos trabalhos relacionados.

TRABALHOS	CARACTERÍSTICAS						
	Centrado no Usuário	Independente de Representação de Política	Adequado para Nuvem Federada	Conscientiza sobre Uso Secundário	Controla Uso Secundário	Fácil Adoção	Usabilidade
P3P (W3C, 2006)	✓						
EPAL (IBM, 2003)							
P2U (Iyilade e Vassileva, 2014)				✓	✓		
Shibboleth (2017a)		✓	✓			✓	
OpenID (2014)		✓	✓			✓	✓
Kolter (2010)	✓			✓			
Werner e Westphall (2016)	✓	✓	✓				
Token de Privacidade	✓	✓	✓	✓	✓	✓	✓

4 MECANISMO DE ESPECIFICAÇÃO DE PREFERÊNCIAS DE PRIVACIDADE

Este trabalho visa fornecer ao usuário controle sobre o uso secundário de suas PIIs e, consequentemente, protegê-lo contra a utilização indevida dos seus dados, através de um modelo para classificar e representar as preferências de privacidade e um mecanismo para implementá-lo (VILLARREAL et al., 2017).

O modelo consiste em uma classificação abrangente e tridimensional dos possíveis usos das PIIs que dá origem a um conjunto de quarenta e cinco preferências, e em quatro perfis de privacidade predefinidos baseados nestas preferências.

O mecanismo, por sua vez, possibilita aos usuários selecionarem um perfil de privacidade predefinido ou criarem um perfil personalizado escolhendo aceitar ou rejeitar cada preferência de privacidade, ou seja, autorizando ou não o uso de certo tipo de dados em diferentes condições. Este perfil é, então, transformado em um token de privacidade, um JWT seguro semelhante aos tokens de ID e de acesso já usados pelo protocolo OpenID Connect, e enviado ao provedor de serviço.

4.1 MODELO DE DEFINIÇÃO DAS PREFERÊNCIAS DE PRIVACIDADE

Devido à grande quantidade de ações e métodos possíveis para coletar e compartilhar dados, não é viável listá-los exaustivamente. Este trabalho propõe um modelo genérico que, com um conjunto restrito de regras, representa de forma universal os possíveis usos secundários dos dados. Este modelo, por um lado, é útil para que os usuários estabeleçam suas preferências de privacidade e, por outro, funciona como uma referência para os SPs avaliarem se suas aplicações de uso de dados atendem a estas preferências. O modelo é composto por uma classificação dos possíveis usos das PIIs do usuário e por perfis de privacidade predefinidos, que são descritos a seguir.

4.1.1 Classificação dos Possíveis Usos das PIIs

Os possíveis usos das PIIs do usuário foram classificados em um modelo tridimensional. As dimensões propostas são Tipo de Dados,

Propósito e Beneficiário e definem uma estrutura na qual cada posição representa uma regra que expressa a preferência de privacidade de um usuário que deve ser respeitada pelo SP. Desta forma, cada regra compreende três partes: o tipo de dados a que se refere a regra, para quais propósitos e para o benefício de quem eles podem ser usados. Por exemplo, um usuário pode definir que seus dados de localização podem ser usados com a finalidade de melhorar os serviços em benefício do outorgante das PIIs (benefício próprio) e, em outra regra, definir que o mesmo tipo de informação para o mesmo propósito não pode ser usado para o benefício de uma terceira parte.

Desta classificação, surgem quarenta e cinco preferências de privacidade, que representam de forma abrangente e universal os usos secundários possíveis e que podem ser coletadas de forma detalhada ou através de quatro perfis predefinidos, que são apresentados na Seção 4.1.2. As subseções seguintes descrevem as dimensões definidas, juntamente com seus atributos e suas respectivas abreviaturas.

4.1.1.1 Tipo de Dados

Esta dimensão representa a categoria das PIIs à qual se refere uma determinada preferência. Os atributos e abreviaturas desta dimensão são:

- Identificação Pessoal (IP): abrange qualquer tipo de informação que represente o outorgante das PIIs, como nome, identificadores nacionais, email, número de celular e foto.

- Características Pessoais e Preferências (CPP): são atributos e opções pessoais do outorgante das PIIs, como sexo, idade, peso, nível de educação, classe socioeconômica, crenças religiosas ou filosóficas e orientação sexual;

- Localização (LO): refere-se a qualquer informação sobre onde o usuário está ou aonde foi, bem como suas trajetórias, com qualquer grau de precisão e obtido por qualquer meio, como GPS, redes Wi-Fi ou sistemas de telecomunicações;

- Atividades e Hábitos (AH): são quaisquer atividades realizadas pelo usuário e hábitos inferidos através de rastreamento, como sites visitados, compras realizadas e perfil comportamental; e

- Relacionamentos (RS): pessoas com quem o outorgante das PIIs

está em um momento específico ou interage através de meios como redes sociais, emails e mensageiros instantâneos.

4.1.1.2 Propósito

O propósito representa o fim para o qual as PIIs podem ser utilizadas. Os valores desta dimensão são:

- Melhoria de Serviço (MS): refere-se ao uso dos dados para a implementação de melhorias nos serviços oferecidos, como personalização de funcionalidades, maior usabilidade e aumento da segurança;
- Científico (CI): diz respeito à concessão de dados para a realização de pesquisas acadêmicas e científicas; e
- Comercial (CO): representa o uso de PIIs do usuário para desenvolver ou oferecer novos produtos e serviços com a finalidade de obter benefícios comerciais.

4.1.1.3 Beneficiário

O beneficiário diz espeito à parte envolvida que se beneficia com o uso da PII. Os atributos definidos são:

- Outorgante das PIIs (PP): corresponde ao usuário que acessa o serviço e a quem se relacionam as PIIs;
- Provedor de Serviço (SP): refere-se à parte responsável por oferecer os serviços acessados pelo usuário e por processar seus dados; e
- Terceira Parte (TP): representa uma parte interessada nos dados diferente e independente do outorgante das PIIs e do provedor de serviço.

4.1.2 Perfis de Privacidade

Por meio da classificação apresentada na seção anterior, as preferências de privacidade dos usuários podem ser coletadas de forma detalhada ou através de quatro perfis predefinidos. Estes perfis foram

definidos com base no trabalho de Chanchary e Chiasson (2015), apresentado no Capítulo 3, que classifica os usuários em três grupos de acordo com suas preocupações em relação à sua privacidade. Tendo em conta que o grupo *Pragmáticos* apresenta a maior porcentagem de usuários e para oferecer opções mais representativas, este foi subdividido em *Conscientes* e *Pragmáticos*.

Os quatro perfis predefinidos são apresentados nas subseções seguintes e os valores das preferências de cada um podem ser visualizados na Tabela 2. Nesta tabela, os perfis de privacidade são indicados pela sua letra inicial e cada linha corresponde a uma preferência quanto à liberação de um tipo de dado, para um determinado propósito e em benefício de uma parte específica. Assim, as preferências assinaladas com um "✓" representam o consentimento do usuário para o uso dos dados.

4.1.2.1 Fundamentalista (F)

Este perfil destina-se a usuários que têm preocupações muito altas com a sua privacidade e não desejam que seus dados sejam usados para qualquer propósito diferente daquele para o qual foram coletados. Algumas funcionalidades, no entanto, podem não funcionar ou não funcionar corretamente quando este perfil é escolhido. Além disso, serão perdidas quaisquer oportunidades de melhorias de serviço e de ofertas personalizadas de produtos e serviços.

4.1.2.2 Consciente (C)

Este perfil representa usuários que se preocupam com sua privacidade mas ainda assim querem habilitar a maior parte das funcionalidades, possibilitar melhorias no serviço e receber algumas oportunidades de ofertas personalizadas de produtos e serviços sem que haja compartilhamento de dados com terceiros.

4.1.2.3 Pragmático (P)

Este perfil destina-se a usuários que ainda desejam alguma privacidade, mas querem habilitar todas as funcionalidades e melhorias do serviço do provedor e permitem um compartilhamento restrito de dados com terceiros para possibilitar diversas ofertas personalizadas de

Tabela 2 – Configuração das preferências de cada perfil de privacidade predefinido quanto ao uso secundário das PIIs.

PREFERÊNCIAS			PERFIS			
Tipo de Dados	Propósito	Beneficiário	F	C	P	D
Identificação Pessoal (IP)	Melhoria de Serviço	Principal da PII		✓	✓	✓
		SP			✓	✓
		Terceira Parte				✓
	Científico	Principal da PII		✓	✓	✓
		SP		✓	✓	✓
		Terceira Parte		✓	✓	✓
	Comercial	Principal da PII		✓	✓	✓
		SP			✓	✓
		Terceira Parte				✓
Características Pessoais e Preferências (CPP)	Melhoria de Serviço	Principal da PII		✓	✓	✓
		SP			✓	✓
		Terceira Parte				✓
	Científico	Principal da PII		✓	✓	✓
		SP		✓	✓	✓
		Terceira Parte		✓	✓	✓
	Comercial	Principal da PII			✓	✓
		SP				✓
		Terceira Parte				✓
Localização (LO)	Melhoria de Serviço	Principal da PII			✓	✓
		SP				✓
		Terceira Parte				✓
	Científico	Principal da PII			✓	✓
		SP			✓	✓
		Terceira Parte			✓	✓
	Comercial	Principal da PII			✓	✓
		SP			✓	✓
		Terceira Parte				✓
Atividades e Hábitos (AH)	Melhoria de Serviço	Principal da PII		✓	✓	✓
		SP		✓	✓	✓
		Terceira Parte			✓	✓
	Científico	Principal da PII		✓	✓	✓
		SP		✓	✓	✓
		Terceira Parte		✓	✓	✓
	Comercial	Principal da PII		✓	✓	✓
		SP			✓	✓
		Terceira Parte			✓	✓
Relacionamentos (RS)	Melhoria de Serviço	Principal da PII		✓	✓	✓
		SP		✓	✓	✓
		Terceira Parte			✓	✓
	Científico	Principal da PII		✓	✓	✓
		SP		✓	✓	✓
		Terceira Parte		✓	✓	✓
	Comercial	Principal da PII			✓	✓
		SP			✓	✓
		Terceira Parte				✓

produtos e serviços.

4.1.2.4 Despreocupado (D)

Este perfil representa usuários que não se preocupam com sua privacidade ou com como suas PIIs são usadas e, portanto, permite que quaisquer dados sejam utilizados para qualquer propósito e em benefício de qualquer parte de acordo apenas com a política de privacidade de cada SP. Todas as funcionalidades, melhorias de serviço e ofertas personalizadas de produtos e serviços são habilitadas com este perfil.

4.2 TOKEN DE PRIVACIDADE

O mecanismo proposto consiste em definir as preferências do usuário conforme o modelo apresentado na seção anterior e, juntamente com informações adicionais do usuário, convertê-las em um objeto JSON, que é usado para criar um JWT assinado ou protegido com *Hash-based Message Authentication Code* (HMAC), e criptografado, chamado token de privacidade. Este token é gerado pelo IdP e transmitido ao SP, que deve solicitar sua validação para verificar sua integridade e descriptografá-lo para usar suas informações a fim de orientar o comportamento de suas aplicações de uso de dados.

4.2.1 Estrutura do token

O token de privacidade é um JWT e, assim, sua estrutura, ilustrada na Figura 7, compreende três seções:

- Cabeçalho: parte que declara que a estrutura de dados é um JWT e define o algoritmo de segurança escolhido e implementado pelo IdP;
- Conjunto de atributos: seção que contém os atributos do token, incluindo dados herdados do token de ID e as preferências de privacidade do usuário; e
- Assinatura do token: assinatura digital ou HMAC gerado com o algoritmo definido no cabeçalho para garantir a integridade do token.

```
{
  "typ": "JWT",
  "alg": "HS256"
}
{
  "sub"           : "alice",
  "iss"           : "https://openid.c2id.com",
  "aud"           : "client-12345",
  "iat"           : 1488405983,

  "IP_MS_PP"      : true,
  "IP_MS_SP"      : false,
  "IP_MS_TP"      : true,
  "IP_CI_PP"      : true,
  "IP_CI_SP"      : true,
  "IP_CI_TP"      : false,
      ...
}
{
  D7SDQBpVCSRSqVUMP9PAungM0gh7JKjKgXYhUlKMr3Y
}
```


Figura 7 – Estrutura do token de privacidade.

A seção do conjunto de atributos do token de privacidade é constituída por duas partes. A primeira define os seguintes atributos herdados do token de ID do OpenID Connect:

- *sub*: é o identificador do sujeito, isto é, uma sequência de caracteres que identifica de forma exclusiva o outorgante das PIIs;
- *iss*: representa a autoridade que emite o token, isto é, o IdP;
- *aud*: identifica a audiência para a qual o token deve ser enviada, isto é, o SP; e
- *iat*: declara a data e a hora em que o token foi emitido em formato Unix *timestamp*.

A segunda parte do conjunto de atributos define as preferências de privacidade do usuário. Assim, cada atributo corresponde a uma

posição da estrutura apresentada na Subseção 4.1.1, ou seja, a uma preferência de privacidade, e tem um valor booleano. A estrutura de uma preferência é a seguinte: a primeira abreviatura representa o tipo de dados, a segunda refere-se ao propósito e a última representa o beneficiário. Por exemplo, se o valor do atributo *IP_MS_PP* for verdadeiro, os dados de identificação pessoal (IP) do usuário podem ser usadas com propósito de melhoria de serviço (MS) em benefício próprio, isto é, do outorgante das PIIs (PP). De forma semelhante, se o valor do atributo *LO_CO_SP* for falso, significa que os dados de localização (LO) não podem ser usados para fins comerciais (CO) em benefício do provedor de serviço (SP).

4.2.2 Segurança e Transmissão do Token

Para garantir sua integridade, o token de privacidade deve ser protegido através de um algoritmo de HMAC ou de assinatura digital e, depois, criptografado a fim de proteger seu conteúdo e manter sua confidencialidade, bem como dificultar sua adulteração. A criptografia pode ser simétrica ou assimétrica de acordo com a escolha e implementação de cada provedor de identidade. O uso de assinatura digital ou de HMAC também depende da escolha do IdP, o qual é responsável pelo compartilhamento da chave secreta no segundo caso.

O token é protegido através do método *Sign-then-Encrypt* para evitar ataques em que a assinatura é removida deixando apenas uma mensagem criptografada, fornecer privacidade ao signatário e garantir que a assinatura seja sempre aceita, uma vez que assinaturas sobre texto criptografado não são consideradas válidas em algumas jurisdições.

Depois de assinado e criptografado, o token de privacidade é enviado ao SP através do *browser* do usuário. Para fazer esta transmissão de forma eficiente e sem comprometer o desempenho do sistema, o token é codificado em uma *string* Base64, que pode ser incorporada a uma URL. Depois de receber o token, o SP deve enviá-lo novamente ao IdP e solicitar sua validação. Este, após verificar a integridade do token de privacidade, envia uma confirmação de validade ao provedor de serviço.

Sempre que o token de ID do OIDC é transmitido, o token de privacidade deve acompanhá-lo, por exemplo, quando o token de ID expirou e um novo é solicitado ao IdP ou ao passar a identidade a terceiras partes. Isto é necessário para garantir que as PIIs dos usuários sejam sempre acompanhadas pelas preferências de privacidade corres-

pondentes.

4.2.3 Fluxo de IdM Modificado

Com a adição do token de privacidade, o fluxo modificado do OpenID Connect proposto por Werner e Westphall (2016) seria estendido, como mostrado na Figura 8, para abranger as seguintes etapas:

1. O usuário solicita acesso ao recurso no SP;
2. O gerenciador de segurança no SP pede ao IdP escolhido pelo usuário para autenticá-lo;
3. O IdP pede ao usuário suas credenciais;
4. O usuário fornece suas credenciais;
5. O IdP valida as credenciais do usuário e retorna os tokens de ID e de privacidade ao usuário, que os passa ao SP;
6. O SP envia os tokens de ID e de privacidade ao IdP para a prova de validação;
7. O IdP verifica os tokens e confirma sua validade ao SP;
8. O SP solicita atributos adicionais do usuário ao IdP;
9. O IdP mostra os escopos de dados suportados pelo SP para o usuário escolher;
10. O usuário escolhe um dos escopos e o informa ao IdP;
11. O IdP fornece os dados ao SP de acordo com o escopo selecionado;
12. O SP autoriza o usuário a acessar o recurso desejado.

Os passos 5, 6 e 7 do fluxo de Werner e Westphall (2016) foram modificados para incorporar a geração, a transmissão e a validação do token de privacidade e o perfil que é usado para gerá-lo é escolhido ou personalizado pelo usuário durante o processo de cadastro no IdP. A fim de oferecer mais flexibilidade, os usuários podem mudar sua escolha a qualquer momento, solicitando-a ao IdP.

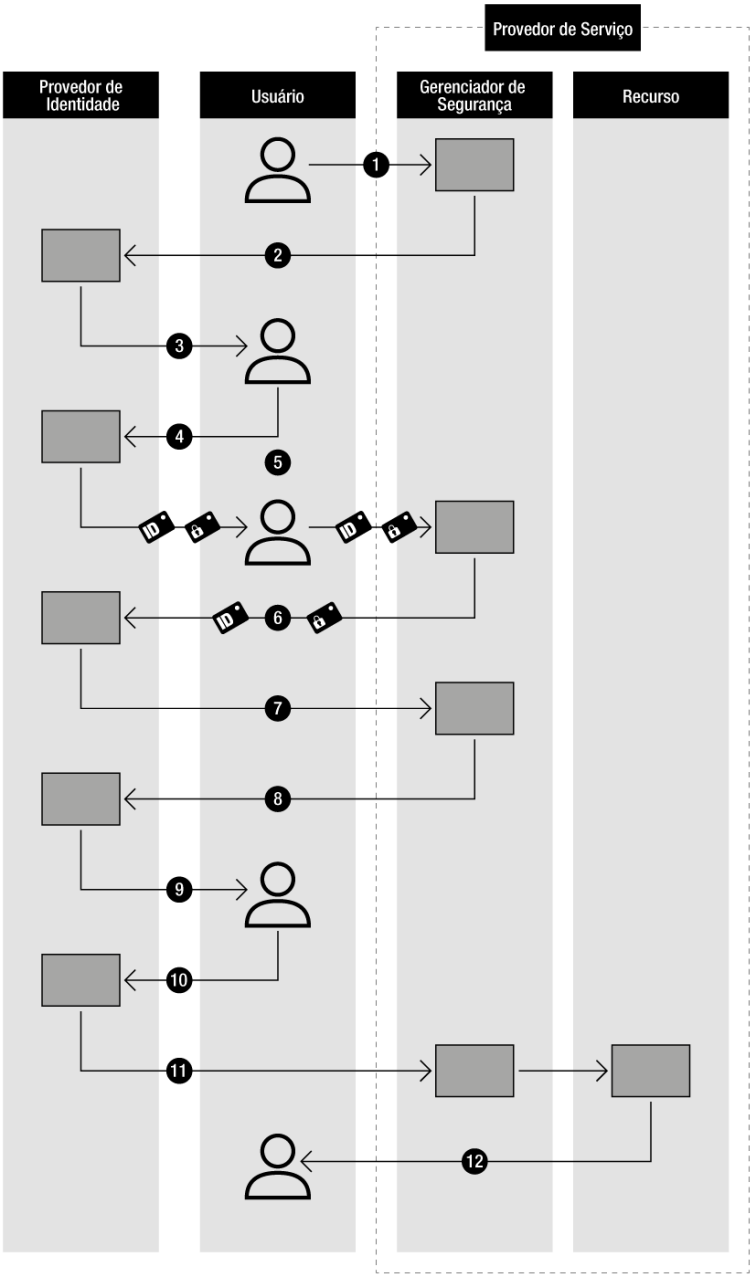

Figura 8 – Extensão do fluxo de IdM proposto por Werner e Westphall (2016) com a adição do token de privacidade.

5 VALIDAÇÃO DO MECANISMO

Neste capítulo, o mecanismo proposto é validado mediante sua aplicação em um estudo de caso e o desenvolvimento de um protótipo que implementa o mecanismo proposto.

5.1 ESTUDO DE CASO

Para demonstrar a aplicabilidade, a utilidade e o potencial do modelo apresentado na Seção 4.1, o mesmo é aplicado em um caso hipotético de um serviço online de inscrição em eventos. Este exemplo permite ver como, apesar da simplicidade do serviço e da pequena quantidade de dados fornecidos diretamente pelo usuário, existe um grande potencial de usos secundários dos dados, e como a aplicação do modelo pode limitar os abusos pelo provedor de serviço e a invasão à privacidade do usuário.

5.1.1 Descrição do Cenário

O caso analisado consiste em um SP que oferece um serviço online de inscrição em eventos. Para tanto, o organizador cadastra seu evento e o provedor de serviço fornece aos usuários uma página online com informações sobre o mesmo e meios para realização e pagamento das inscrições. Para poder utilizar este serviço, o usuário deve estar cadastrado em um IdP pertencente à mesma federação que o provedor do serviço e deve autenticar-se no SP através deste IdP. Além disso, pode ser necessário fornecer dados adicionais ao provedor de serviço, que mantém um registro das inscrições realizadas pelos usuários para facilitar a inscrição em futuros eventos.

O serviço prestado é gratuito tanto para o usuário quanto para o organizador do evento e para obter lucro, o SP publica anúncios em seu website, oferece serviços de transporte e hospedagem aos participantes e também comercializa os dados a terceiras partes interessadas.

Neste estudo de caso, utiliza-se como exemplo uma corrida de rua e são solicitados pelo SP os seguintes dados ao organizador do evento para cadastrá-la:

- Nome do organizador;

- CNPJ ou CPF;
- Nome do responsável;
- Nome oficial do evento;
- Tipo de corrida;
- Categorias;
- Local da corrida;
- Data e hora de início do evento; e
- Custo da inscrição.

Para permitir ao usuário usar o sistema, o provedor de serviço necessita os seguintes dados, que podem ser disponibilizados diretamente pelo IdP ou solicitados explicitamente ao usuário:

- Nome completo;
- Data de nascimento;
- Sexo;
- CPF; e
- Email.

Para realizar sua inscrição na corrida, o usuário é solicitado a fornecer os seguintes dados adicionais ao SP:

- Número de celular;
- Altura;
- Peso;
- Distância a ser corrida; e
- Forma e dados de pagamento.

Além dos dados fornecidos pelo usuário, o SP pode coletar dados de contexto no momento da inscrição, como:

- Local;
- Horário em que a inscrição foi feita;

- Dispositivo utilizado;
- Tipo de conexão; e
- Grupo de pessoas que fizeram inscrição do mesmo dispositivo e no mesmo momento.

Com base nos dados fornecidos e coletados na inscrição, no histórico de inscrições do usuário e nas inscrições realizadas no evento é possível, ainda, inferir outras informações, como:

- Horário em que o usuário costuma fazer as inscrições;
- Dispositivos e tipos de tecnologia mais utilizados para fazer as inscrições;
- Grupo de pessoas que pagaram com o mesmo cartão de crédito;
- Grupo de pessoas que participarão juntas do evento (inscrições realizadas em um mesmo contexto);
- Integrantes do grupo familiar;
- Se o usuário participa das corridas com o grupo familiar;
- Frequência de participação do usuário em corridas;
- Categorias preferidas pelo usuário;
- Nível de condicionamento físico do usuário;
- Hábito do usuário de participar de atividades ao ar livre;
- Local em que se encontrará o inscrito na data do evento;
- Pessoas que estarão juntas no local no dia do evento;
- Forma de pagamento habitual do usuário;
- Bandeira do cartão de crédito do usuário;
- Se o usuário possui mais de um cartão de crédito;

Finalmente, os dados recolhidos podem ser agregados a dados de outros eventos em que o usuário se inscreveu e de outras fontes para inferir novas informações.

5.1.2 Possíveis Usos Secundários dos Dados

A partir dos dados coletados e das informações inferidas, surge uma grande quantidade de usos secundários possíveis, alguns dos quais, definidos com base em políticas de privacidade reais, são apresentados a seguir classificados de acordo com o tipo dos dados.

5.1.2.1 Identificação Pessoal

Para os dados do tipo Identificação Pessoal, foram definidos os seguintes possíveis usos secundários:

- Adicionar nome e email a uma lista de notificações sobre o evento em que o usuário se inscreveu;
- Adicionar nome e email às listas de *mailing* do SP para divulgar novos eventos;
- Adicionar nome e email a uma lista de emails válidos a ser vendida a terceiras partes; e
- Armazenar CPF, nome e telefone celular do participante no banco de dados do SP para facilitar a inscrição em próximos eventos.

5.1.2.2 Características Pessoais e Preferências

Os possíveis usos secundários elencados para os dados do tipo Características Pessoais e Preferências são:

- Adicionar CPF e idade a uma lista a ser cedida a um projeto de pesquisa de uma universidade para elaborar estatísticas de participação em corridas por idade;
- Adicionar nome e email de pessoas que possuem mais de um cartão de crédito a uma lista a ser vendida para publicidade de um projeto imobiliário;
- Adicionar email e idade a uma lista de pessoas a ser vendida a uma empresa que comercializa suplementos alimentares;
- Adicionar email, altura e peso a uma lista de homens que medem mais de 1,90m e mulheres que medem mais de 1,80m a ser oferecida a lojas especializadas em tamanhos grandes; e

- Armazenar CPF, sexo e idade do participante no banco de dados do SP para facilitar a inscrição em próximos eventos.

5.1.2.3 Localização

Paras os dados de Localização, foram definidos os seguintes possíveis usos secundários:

- Se a inscrição for feita de um dispositivo móvel, fornecer número de celular e localização a uma empresa especializada em oferecer serviços de publicidade por localização;

- Utilizar localização do usuário no momento da inscrição para oferecer ingressos para outros eventos próximos ao local;

- Quando a distância entre o local da inscrição e o local do evento for maior que um valor determinado, oferecer serviços de transporte e hospedagem;

- Utilizar localização do evento para oferecer ingressos para outros eventos próximos ao local; e

- Adicionar nome, email e data do evento a uma lista a ser vendida a restaurantes próximos ao local da corrida.

5.1.2.4 Atividades e Hábitos

Os possíveis usos secundários determinados para os dados do tipo Atividades e Hábitos são:

- Utilizar dados sobre a forma habitual de pagamento para autopreenchimento do formulário de inscrição;

- Adicionar email e tipos de corrida preferidas a listas a serem vendidas a empresas que comercializam artigos esportivos; e

- Se o usuário viaja com frequência para participar de eventos, adicionar email e frequência de participação em eventos a uma lista a ser vendida a empresas de segurança domiciliar.

5.1.2.5 Relacionamentos

Paras os dados da categoria Relacionamentos, foram elencados os seguintes possíveis usos secundários:

- Adicionar email e grupo familiar a uma lista a ser fornecida a um projeto de pesquisa do Ministério de Saúde sobre pessoas que participam de eventos de corrida em família; e

- Oferecer locação de um veículo coletivo a pessoas que realizaram a inscrição em um mesmo contexto.

5.1.3 Aplicação do Modelo

Nesta subseção, é apresentado como a aplicação do modelo e a escolha do perfil por parte do usuário modificam e restringem o comportamento do SP com relação ao uso de seus dados. Cada caso descreve como a aplicação de uso de dados do provedor de serviço age conforme o perfil de privacidade predefinido escolhido pelo usuário.

5.1.3.1 Caso 0: Sem Aplicação do Modelo

Caso o modelo não seja utilizado, o SP poderá fazer todos os usos secundários mencionados na subseção anterior sem a ciência ou o consentimento do usuário. Estes usos são condicionados apenas pelas políticas de privacidade do provedor de serviço, quando existentes.

5.1.3.2 Caso 1: Perfil Despreocupado

Neste caso, o perfil escolhido pelo usuário é o Despreocupado e, portanto, todas as permissões estão habilitadas. Desta forma, o provedor de serviço pode realizar todos os usos secundários elencados anteriormente. A diferença com a não aplicação do modelo, no entanto, consiste no fato de que o usuário, ao ser solicitado a selecionar um perfil, é conscientizado sobre os possíveis usos secundários e, ao escolher o perfil, consente a utilização de seus dados, o que garante que não haverá violação à sua privacidade.

5.1.3.3 Caso 2: Perfil Fundamentalista

Neste caso, o perfil selecionado pelo usuário é o Fundamentalista e, portanto, nenhum dos usos secundários mencionados é permitido. Ainda assim, a prestação do serviço seria possível, uma vez que não há objeção em utilizar os dados para seus propósitos primários (inscrição para a corrida, neste exemplo). No entanto, se o benefício econômico do SP baseia-se apenas no uso secundário dos dados, pode não ser interessante para este fornecer o serviço nestas condições. Assim, para viabilizar o serviço, o SP poderia solicitar permissão específica para utilizar os dados ou cobrar uma taxa do usuário ou do organizador do evento, por exemplo.

5.1.3.4 Caso 3: Perfil Consciente

Neste caso, o perfil escolhido pelo usuário é o Consciente e os usos secundários permitidos e os não permitidos são apresentados na Tabela 3.

5.1.3.5 Caso 4: Perfil Pragmático

Com perfil Pragmático, os usos secundários permitidos e os não permitidos são apresentados na Tabela 4.

5.1.4 Resultados

Com a aplicação deste estudo de caso, demonstrou-se a aplicabilidade e a utilidade da proposta mediante a aplicação do modelo em diferentes situações. Através da apresentação de possíveis usos secundários dos dados no caso apresentado, verificou-se que a escolha de diferentes perfis de privacidade resultam em diferentes comportamentos do provedor de serviço e que estas diferenças são significativas.

Verificou-se que a simples aplicação do modelo com todas as permissões de uso habilitadas, ou seja, com o perfil Despreocupado, significa uma melhoria importante em comparação com a não aplicação do modelo, já que conscientiza o usuário em relação ao uso secundário de suas PIIs e garante que não haja violação de privacidade, já que o uso é autorizado.

Tabela 3 – Possíveis usos secundários permitidos e não permitidos com o perfil de privacidade Consciente.

PERFIL CONSCIENTE	
USOS SECUNDÁRIOS PERMITIDOS	**PREFERÊNCIA**
Adicionar nome e email a uma lista de notificações sobre o evento em que o usuário se inscreveu.	IP_MS_PP
Armazenar CPF, nome e telefone celular do participante no banco de dados do SP para facilitar a inscrição em próximos eventos.	IP_MS_PP
Adicionar CPF e idade a uma lista a ser cedida a um projeto de pesquisa de uma universidade para elaborar estatísticas de participação em corridas por idade.	CPP_CI_TP
Armazenar CPF, sexo e idade do participante no banco de dados do SP para facilitar a inscrição em próximos eventos.	CPP_MS_PP
Utilizar dados sobre a forma habitual de pagamento para autopreenchimento do formulário de inscrição.	AH_MS_PP
Adicionar email e grupo familiar a uma lista a ser fornecida a um projeto de pesquisa do Ministério de Saúde sobre pessoas que participam de eventos de corrida em família.	RS_CI_TP
USOS SECUNDÁRIOS NÃO PERMITIDOS	**PREFERÊNCIA**
Adicionar nome e email às listas de *mailing* do SP para divulgar novos eventos.	IP_CO_SP
Adicionar nome e email a uma lista de emails válidos a ser vendida a terceiras partes.	IP_CO_TP
Adicionar nome e email de pessoas que possuem mais de um cartão de crédito a uma lista a ser vendida para publicidade de um projeto imobiliário.	CPP_CO_TP
Adicionar email e idade a uma lista de pessoas a ser vendida a uma empresa que comercializa suplementos alimentares.	CPP_CO_TP
Adicionar email, altura e peso a uma lista de homens que medem mais de 1,90m e mulheres que medem mais de 1,80m a ser oferecida a lojas especializadas em tamanhos grandes.	CPP_CO_TP
Se a inscrição for feita de um dispositivo móvel, fornecer número de celular e localização a uma empresa especializada em oferecer serviços de publicidade por localização.	LO_CO_TP
Utilizar localização do usuário no momento da inscrição para oferecer ingressos para outros eventos próximos ao local.	LO_CO_SP
Quando a distância entre o local da inscrição e o local do evento for maior que um valor determinado, oferecer serviços de transporte e hospedagem.	LO_CO_SP
Utilizar localização do evento para oferecer ingressos para outros eventos próximos ao local.	LO_CO_SP
Adicionar nome, email e data do evento a uma lista a ser vendida a restaurantes próximos ao local da corrida.	LO_CO_TP
Adicionar email e tipos de corrida preferidas a listas a serem vendidas a empresas que comercializam artigos esportivos.	AH_CO_TP
Se o usuário viaja com frequência para participar de eventos, adicionar email e frequência de participação em eventos a uma lista a ser vendida a empresas de segurança domiciliar.	AH_CO_TP
Oferecer locação de um veículo coletivo a pessoas que realizaram a inscrição em um mesmo contexto.	RS_CO_SP

Tabela 4 – Possíveis usos secundários permitidos e não permitidos com o perfil de privacidade Pragmático.

PERFIL PRAGMÁTICO	
USOS SECUNDÁRIOS PERMITIDOS	PREFERÊNCIA
Adicionar nome e email a uma lista de notificações sobre o evento em que o usuário se inscreveu.	IP_MS_PP
Adicionar nome e email às listas de *mailing* do SP para divulgar novos eventos.	IP_CO_SP
Armazenar CPF, nome e telefone celular do participante no banco de dados do SP para facilitar a inscrição em próximos eventos.	IP_MS_PP
Adicionar CPF e idade a uma lista a ser cedida a um projeto de pesquisa de uma universidade para elaborar estatísticas de participação em corridas por idade.	CPP_CI_TP
Armazenar CPF, sexo e idade do participante no banco de dados do SP para facilitar a inscrição em próximos eventos.	CPP_MS_PP
Utilizar localização do usuário no momento da inscrição para oferecer ingressos para outros eventos próximos ao local.	LO_CO_SP
Quando a distância entre o local da inscrição e o local do evento for maior que um valor determinado, oferecer serviços de transporte e hospedagem.	LO_CO_SP
Utilizar localização do evento para oferecer ingressos para outros eventos próximos ao local.	LO_CO_SP
Utilizar dados sobre a forma habitual de pagamento para autopreenchimento do formulário de inscrição.	AH_MS_PP
Adicionar email e tipos de corrida preferidas a listas a serem vendidas a empresas que comercializam artigos esportivos.	AH_CO_TP
Se o usuário viaja com frequência para participar de eventos, adicionar email e frequência de participação em eventos a uma lista a ser vendida a empresas de segurança domiciliar.	AH_CO_TP
Adicionar email e grupo familiar a uma lista a ser fornecida a um projeto de pesquisa do Ministério de Saúde sobre pessoas que participam de eventos de corrida em família.	RS_CI_TP
Oferecer locação de um veículo coletivo a pessoas que realizaram a inscrição em um mesmo contexto.	RS_CO_SP
USOS SECUNDÁRIOS NÃO PERMITIDOS	PREFERÊNCIA
Adicionar nome e email a uma lista de emails válidos a ser vendida a terceiras partes.	IP_CO_TP
Adicionar nome e email de pessoas que possuem mais de um cartão de crédito a uma lista a ser vendida para publicidade de um projeto imobiliário.	CPP_CO_TP
Adicionar email e idade a uma lista de pessoas a ser vendida a uma empresa que comercializa suplementos alimentares.	CPP_CO_TP
Adicionar email, altura e peso a uma lista de homens que medem mais de 1,90m e mulheres que medem mais de 1,80m a ser oferecida a lojas especializadas em tamanhos grandes.	CPP_CO_TP
Se a inscrição for feita de um dispositivo móvel, fornecer número de celular e localização a uma empresa especializada em oferecer serviços de publicidade por localização.	LO_CO_TP
Adicionar nome, email e data do evento a uma lista a ser vendida a restaurantes próximos ao local da corrida.	LO_CO_TP

É possível visualizar, também, que a aplicação do perfil Fundamentalista, mesmo impedindo totalmente o uso secundário dos dados, não inviabiliza tecnicamente o fornecimento de serviços. No entanto, pode criar a necessidade de adequar o modelo comercial do SP.

A aplicação dos perfis Consciente e Pragmático dão origem a comportamentos diferentes do SP, uma vez que o Consciente restringe o uso secundário de seus dados por terceiras partes, mesmo que algumas oportunidades de oferta de serviços sejam perdidas; e o Pragmático, por outro lado, permite um maior uso de seus dados por terceiros a fim de possibilitar o acesso a uma maior quantidade de oportunidades e serviços personalizados. Desta forma, foi possível visualizar como perfis mais exigentes, ao mesmo tempo que aumentam a privacidade do usuário, o fazem perder oportunidades, o que justifica que alguns usuários tenham preferência pelo perfil Despreocupado.

Finalmente, observou-se que o mecanismo pode servir como base para criar meios de negociação entre o usuário e o SP, através do qual o SP possa oferecer remuneração ou algum outro benefício ao usuário pelo uso de seus dados.

5.2 PROTÓTIPO

A fim de verificar a viabilidade técnica e o correto funcionamento do mecanismo proposto, bem como servir de base para uma futura extensão de uma implementação do protocolo OpenID Connect, foi desenvolvido um protótipo. Este é uma aplicação Web implementada na linguagem de programação Java que permite visualizar através de interfaces gráficas o processo de geração e criptografia do token de privacidade, bem como a comunicação entre o IdP e o SP.

O protótipo executa os processos que devem ser realizados pelo IdP para cadastrar usuários, coletar suas preferências de privacidade e armazená-las, e, quando solicitado pelo SP, gerar, proteger com HMAC, criptografar e validar o token de privacidade.

A aplicação representa também um SP que oferece um serviço online de inscrições para eventos e suas aplicações de coleta e uso de dados, com o objetivo de mostrar os efeitos das diferentes preferências de privacidade do usuário no comportamento do provedor de serviço com relação ao uso secundário de suas PIIs. Esta funcionalidade foi incluída para implementar o estudo de caso apresentado na seção anterior.

5.2.1 Especificação do Protótipo

Para atender aos objetivos do protótipo, foram definidos os requisitos funcionais e não-funcionais que são apresentados nas Tabelas 5 e 6, respectivamente.

Tabela 5 – Requisitos funcionais do protótipo.

REQUISITOS FUNCIONAIS
RF01: Cadastrar usuários no provedor de identidade.
RF02: Coletar e armazenar as preferências de privacidade dos usuários.
RF03: Editar o cadastro dos usuários.
RF04: Gerar, proteger com HMAC e criptografar o token de privacidade quando solicitado.
RF05: Validar o token quando solicitado pelo provedor de serviço.
RF06: Mostrar passo a passo os eventos do processo de geração e validação do token e os resultados.
RF07: Mostrar a estrutura do token de privacidade gerado.
RF08: Representar a autenticação do usuário no provedor de serviço através do provedor de identidade.
RF09: Representar a inscrição de usuários em uma corrida no provedor de serviço.
RF10: Coletar dados do usuário e informações de contexto no momento da inscrição.
RF11: Representar o comportamento das aplicações de uso de dados do provedor de serviço.
RF12: Mostrar as variações de comportamento destas aplicações conforme os diferentes perfis de privacidade.

Tabela 6 – Requisitos não funcionais do protótipo.

REQUISITOS NÃO FUNCIONAIS
RNF01: Usar as mesmas tecnologias e procedimentos que o OpenID Connect.
RNF02: Interface gráfica para definição de perfis que atenda os objetivos do modelo e as recomendações da ISO/IEC 29100 (2011).

Para modelar o protótipo e as classes necessárias, foi criado um Diagrama de Classe, cuja forma simplificada é apresentada na Figura 9. Para facilitar a interpretação, são mostradas apenas as principais classes e seus atributos e métodos mais relevantes. As classes *Issuer* e *ClientID* pertencem à implementação Connect2id do OpenID Connect, cuja biblioteca Java foi utilizada para o desenvolvimento do protótipo.

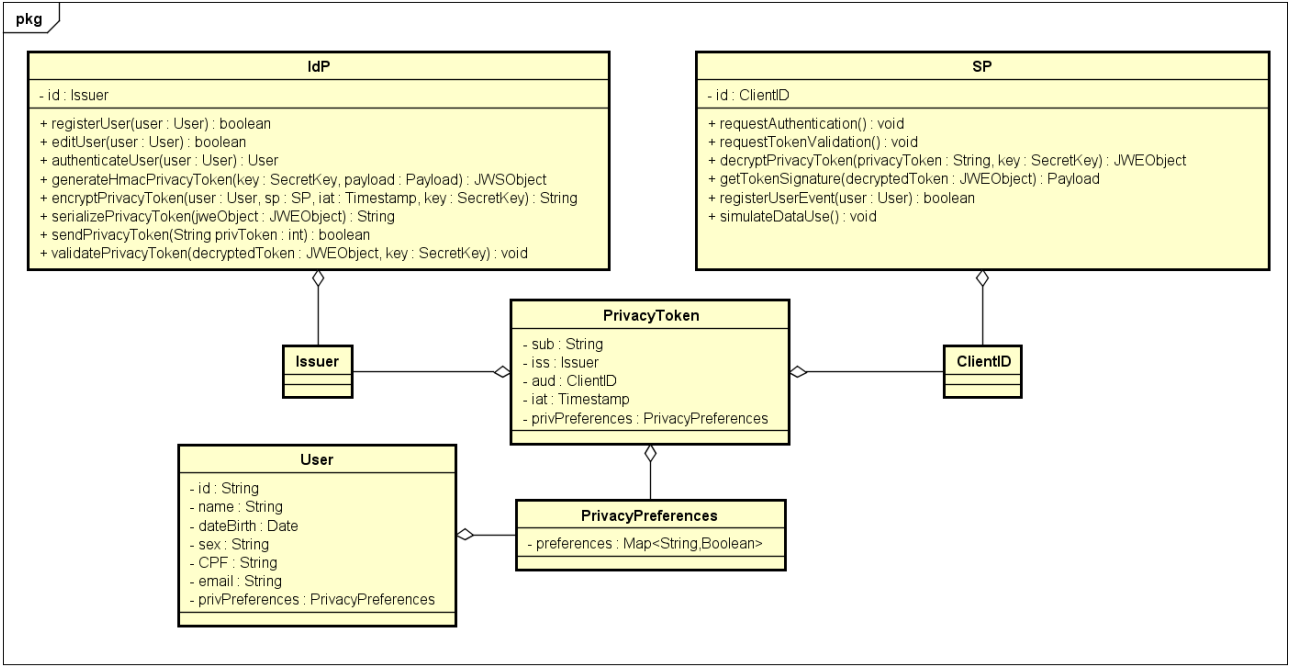

Figura 9 – Diagrama de Classes UML simplificado do protótipo.

5.2.2 Implementação

O protótipo foi desenvolvido como uma aplicação Web na linguagem de programação orientada a objetos Java, no ambiente de programação Netbeans 8.0.2 com Servidor Web Apache Tomcat 8 e Banco de Dados MySQL. Para organizar as classes implementadas e separar as diferentes camadas do projeto, foram utilizados os padrões de projeto *Data Transfer Object* (DTO), *Data Access Object* (DAO) e Business Object (BO). As páginas Web foram definidas em HTML (*HyperText Markup Language*) e estilizadas com CSS (*Cascading Style Sheets*), e as interfaces foram desenvolvidas com o programa Adobe Illustrator CC.

A aplicação implementa classes que representam o IdP, o SP, o usuário, as preferências de privacidade do usuário e o token de privacidade. O objeto *User* é definido pelos dados pessoais do usuário coletados através de um formulário de cadastro no IdP e os atributos do objeto *PrivacyPreferences* recebem os valores correspondentes ao perfil de privacidade selecionado ou personalizado pelo usuário. Já os valores dos atributos do objeto *PrivacyToken* são definidos pelos IDs do IdP, do SP e do usuário, pelas preferências de privacidade do usuário e por um *timestamp* do momento em que foi gerado.

A classe *IdP* possui métodos para gerar, proteger com HMAC, criptografar, serializar e transmitir um objeto *PrivacyToken*, que são chamados quando o usuário utiliza o seu cadastro no IdP para se autenticar no provedor de serviço. A classe *SP*, por sua vez, possui métodos para receber o token de privacidade, solicitar sua validação ao IdP, descriptografá-lo e usá-lo para definir quais usos secundários são permitidos.

Quando o usuário deseja utilizar o serviço, o SP solicita o login ao IdP, que o autentica e cria um objeto *PrivacyToken*. Este objeto é codificado em um objeto JSON, conforme o trecho de código apresentado na Figura 10, com o auxílio da biblioteca Google GSON, que possibilita converter objetos Java em suas representações JSON, bem como converter uma string JSON em um objeto Java equivalente (GOOGLE, 2017).

Para ser transmitido com segurança e eficiência, a representação JSON do token de privacidade é usada como carga para criar um JWS com a biblioteca Nimbus JOSE + JWT, que permite a criação e a verificação de JWTs (CONNECT2ID, 2017a). Este JWS é protegido com HMAC usando o algoritmo SHA-256 e uma chave secreta. O trecho de código que gera o HMAC é mostrado na Figura 11.

Depois de gerado o HMAC, o JWS é usado como carga para criar

```
116     // Cria um GSON builder
117     Gson gson = new GsonBuilder().setPrettyPrinting().create();
118
119     // Cria um objeto PrivacyToken
120     PrivacyToken privToken = new PrivacyToken(sp.getId(),
121                                               this.getIss(),
122                                               user.getId(),
123                                               user.getPrivPreferences(),
124                                               iat);
125
126     // Transforma o objeto PrivacyToken em objeto JSON
127     String tokenJsonString = gson.toJson(privToken);
```

Figura 10 – Trecho de código da classe IdP responsável por transformar o objeto *PrivacyToken* em um objeto JSON.

```
375     // Cria carga com a string JSON do token de privacidade
376     Payload payload = new Payload(tokenJsonStr);
377
378     // Cria um objeto signer
379     JWSSigner signer = new MACSigner(key.getEncoded());
380
381     // Cria um objeto JWS protegido por HMAC com um objeto JSON como carga
382     JWSObject jwsObject = new JWSObject(new JWSHeader(JWSAlgorithm.HS256),
383                                         payload);
384
385     // Aplica o HMAC
386     jwsObject.sign(signer);
```

Figura 11 – Trecho de código da classe IdP responsável por proteger o token de privacidade com HMAC.

um JWE, que é criptografado com o algoritmo *Advanced Encryption Standard* (AES) no modo de operação *Cipher Block Chaining* (CBC), com PKCS #7 (*Public-Key Cryptography Standards*) e um IV (*Initialization Vector*) de 128 bits. O código responsável por este processo de criptografia é ilustrado na Figura 12.

Finalmente, o objeto é submetido a uma serialização compacta que o transforma em uma string Base64, para poder ser transmitida de forma fácil e eficiente ao SP através de URLs, por exemplo. O método responsável por esta codificação é mostrado na Figura 13. O processo completo de geração e preparação para transmissão do token de privacidade pode ser visualizado na Figura 14, que mostra os sucessivos estados do token e suas transições, bem como as tecnologias utilizadas.

```
137     // Cria um objeto JWE com um JWS assinado como carga
138     JWEObject jweObject = new JWEObject(
139         new JWEHeader.Builder(JWEAlgorithm.DIR,
140                 EncryptionMethod.A128CBC_HS256)
141             .contentType("JWT")
142             .build(),
143         new Payload(jwsObject));
144
145     // Executa a criptografia
146     jweObject.encrypt(new DirectEncrypter(key.getEncoded()));
```

Figura 12 – Trecho de código da classe IdP responsável por criptografar o token de privacidade.

```
446     public String serializePrivacyToken(JWEObject jweObject){
447         // Serializa o objeto JWE na forma compacta
448         String jweString = jweObject.serialize();
449
450         // Retorna o token de privacidade codificado em Base64
451         return jweString;
452     }
```

Figura 13 – Método da classe IdP responsável por codificar o token de privacidade em Base64.

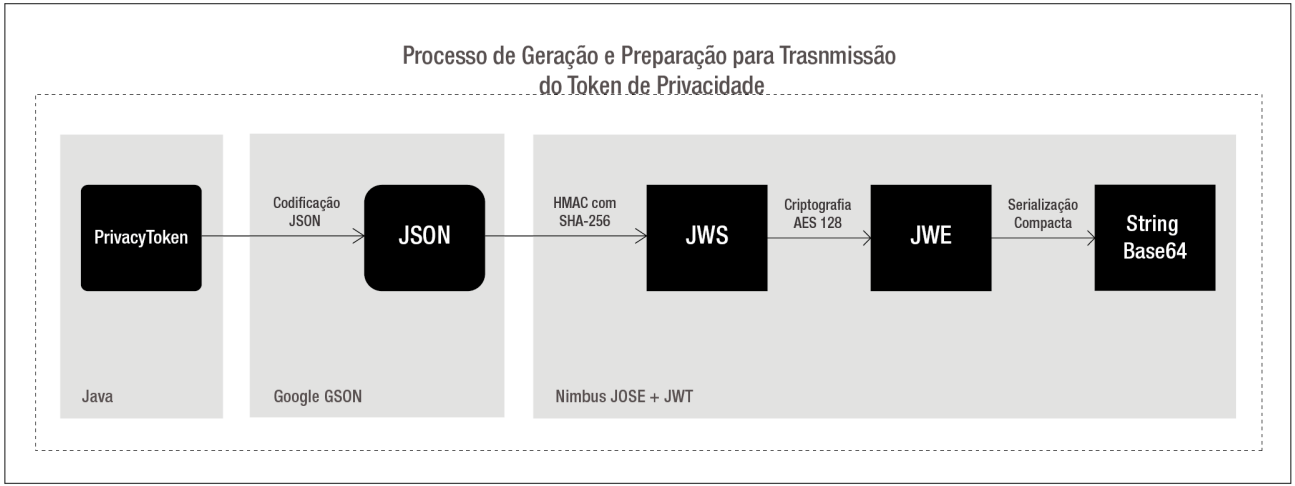

Figura 14 – Processo de geração do token de privacidade.

5.2.3 Interfaces e Uso do Protótipo

A tela inicial do protótipo, mostrada na Figura 15, é dividida em duas partes: uma correspondente ao IdP, com opções que permitem cadastrar, listar e editar os usuários; e a outra correspondente ao SP, com opções para fazer login e inscrever-se em uma corrida. A opção de login de usuário deflagra a sequência de geração, transmissão e validação do token de privacidade e mostra passo a passo o processo. Já a opção de inscrição é habilitada apenas quando o usuário está autenticado e permite que este se inscreva em uma corrida, o que gera um relatório sobre a coleta de dados e os usos secundários conforme o perfil armazenado no seu token de privacidade.

Figura 15 – Tela inicial do protótipo.

A página de cadastro do IdP solicita dados básicos do usuário e permite selecionar um perfil de privacidade predefinido ou criar um personalizado. Com a finalidade de verificar a utilidade do modelo para conscientizar o usuário a respeito do uso secundário de seus dados e permitir definir de forma simples suas preferências, as seções referentes à escolha do perfil e de personalização foram desenvolvidas com foco em boas práticas de design e usabilidade e com base na ISO/IEC 29100.

Desta forma, cada perfil é representado por um número, um nome, um ícone e uma descrição breve mas expressiva. Além disso, cores são usadas para ajudar a diferenciar os perfis e representar os

níveis de riscos para a privacidade em cada um deles, sendo vermelho para o perfil com os maiores riscos e verde para aquele com os menores riscos. Para fornecer ao usuário mais informações sobre os possíveis usos de suas PIIs e sobre o perfil escolhido, o botão *Ver detalhes* mostra o perfil completo, ou seja, todas as preferências de privacidade com as configurações correspondentes. A Figura 16 mostra a seção da tela de cadastro referente à escolha do perfil de privacidade.

Figura 16 – Tela do protótipo com os quatro perfis predefinidos e o personalizável.

A opção de perfil Personalizado leva à página mostrada na Figura 17, que apresenta um *checkbox* para cada preferência de privacidade e permite ao usuário assinalar os usos de seus dados que deseja autorizar. Para orientar e simplificar a escolha, o usuário pode ainda selecionar um dos quatro perfis como base para personalização. Esta mesma tela é mostrada na opção *Ver detalhes* dos perfis predefinidos, porém com as preferências já assinaladas e desabilitadas para edição.

Do lado do SP, a opção Login inicia o processo de autenticação do usuário e mostra, através da interface gráfica, todos os passos que são desencadeados e os resultados gerados. A Figura 18 mostra a tela que apresenta o passo de geração do HMAC do token de privacidade e a estrutura do mesmo.

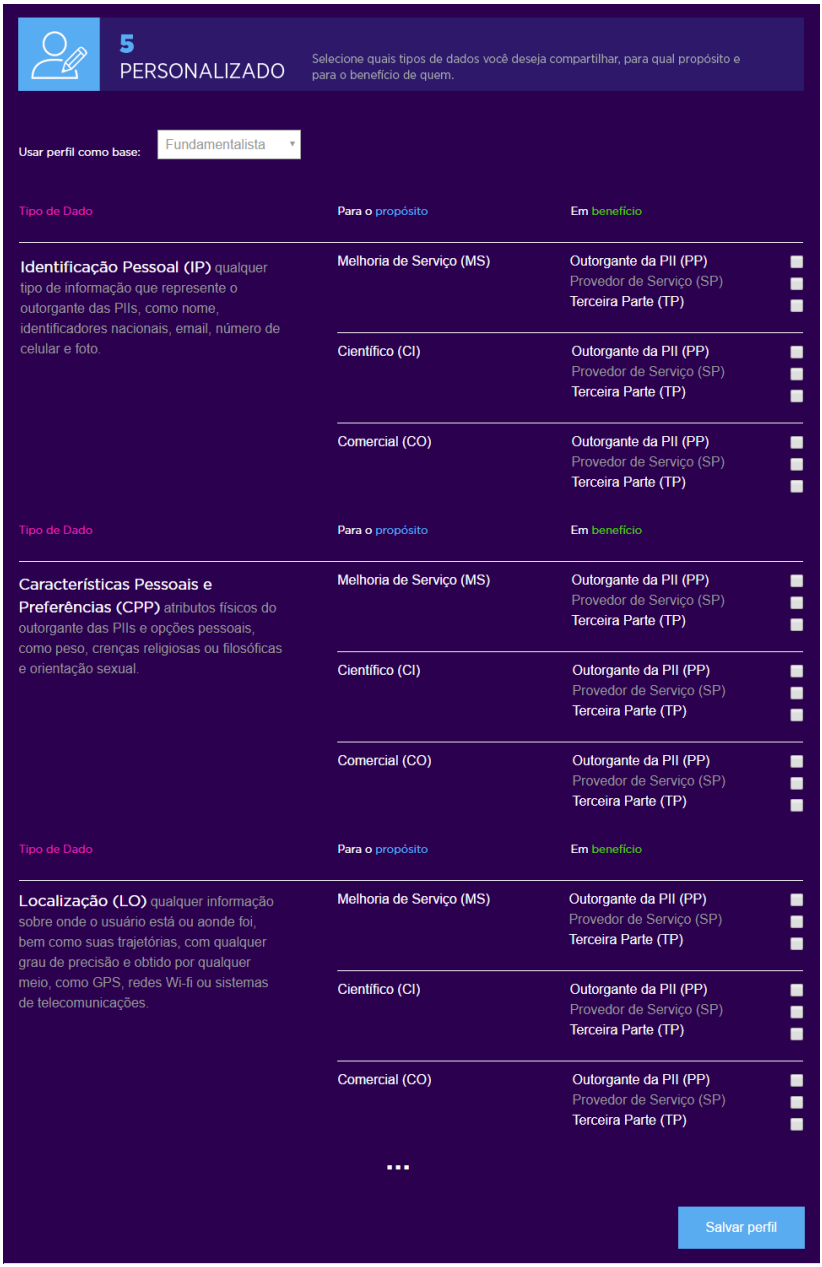

Figura 17 – Parte da tela do protótipo que permite personalizar as preferências de privacidade.

A segunda opção do lado do provedor de serviço abre uma página para a inscrição do usuário autenticado em uma corrida, na qual são solicitados alguns dados adicionais e também são coletadas informações de contexto. Terminada a inscrição, são mostrados os dados obtidos do IdP, os solicitados ao usuário e os coletados do contexto, conforme a Figura 19. Na sequência, é realizada uma análise dos possíveis usos secundários e o resultado é apresentado na forma de um relatório que lista os usos secundários dos dados permitidos e não permitidos para o perfil deste usuário. A tela que representa este relatório é mostrada na Figura 20.

5.2.4 Contribuições do Protótipo

Com o desenvolvimento do protótipo, foi demonstrada a viabilidade do uso das estruturas de dados e tecnologias escolhidas para a geração, a transmissão e o uso do token de privacidade. Sendo estas tecnologias também utilizadas pelo OpenID Connect, há a possibilidade de estendê-lo mediante o uso das mesmas.

As interfaces desenvolvidas permitem fazer testes para verificar a eficácia do modelo proposto para conscientizar o usuário sobre os possíveis usos secundários de seus dados, auxiliá-lo a definir suas preferências de privacidade e viabilizar e simplificar o processo de especificação de tais preferências.

O protótipo foi desenvolvido de forma a possibilitar o seu uso para fins didáticos, já que implementa um estudo de caso de maneira interativa e realista, informa sobre a existência de usos secundários e permite fazer experiências em que o usuário pode escolher diferentes perfis e visualizar os efeitos tanto na proteção da sua privacidade quanto no aproveitamento ou perda de oportunidades advindas da autorização do uso dos dados.

Este protótipo permite também realizar experimentos e testes de usabilidade para verificar os efeitos da aplicação do mecanismo nos usuários e determinar se o modelo realmente é eficaz para conscientizá-los sobre os usos secundários de suas PIIs, qual é a sua reação perante esta consciência e se os perfis e os resultados obtidos representam suas necessidades e desejos.

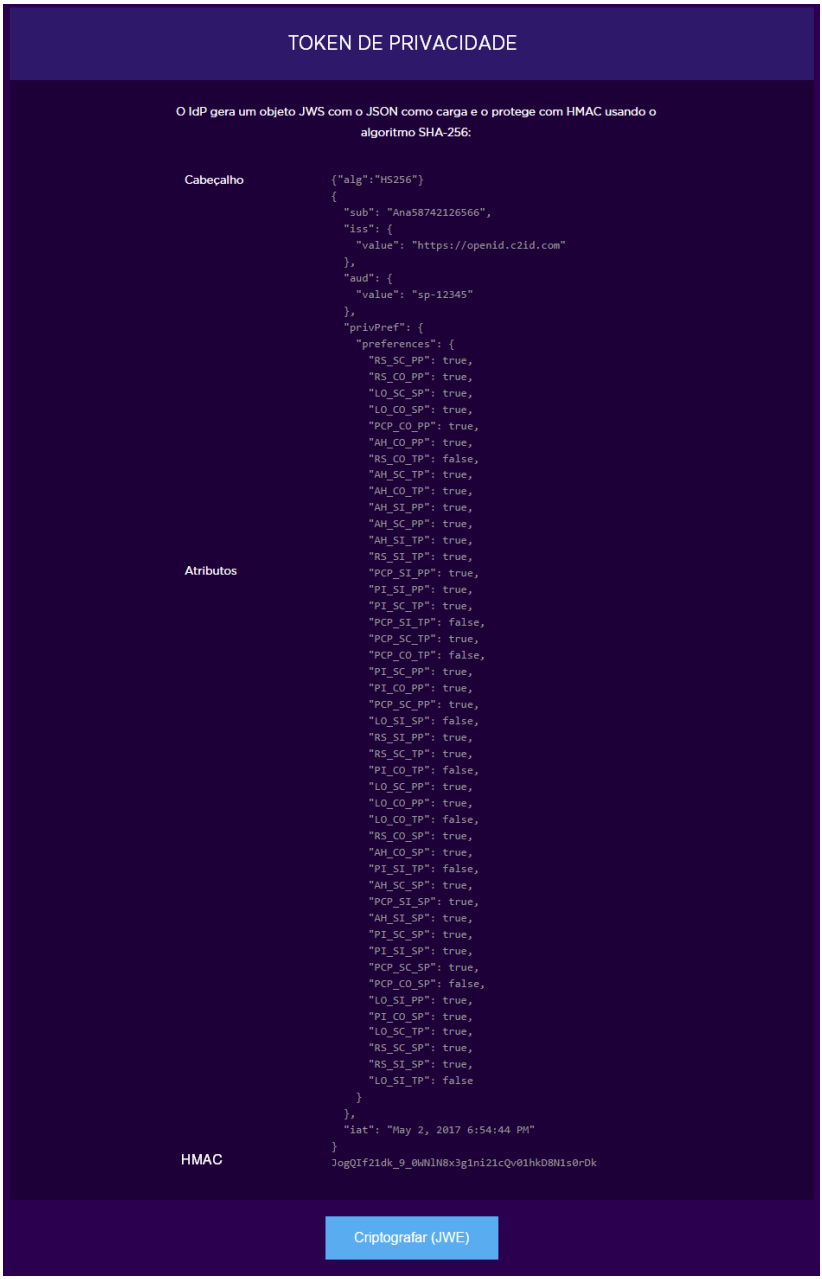

Figura 18 – Tela do protótipo que mostra o token de privacidade gerado e protegido com HMAC.

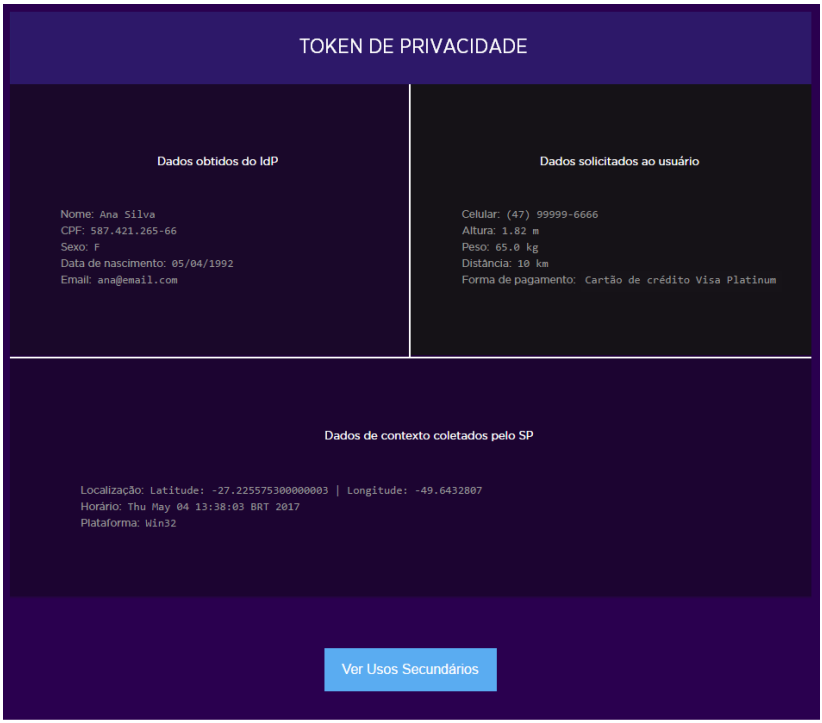

Figura 19 – Tela do protótipo que mostra os dados obtidos do IdP, solicitados ao usuário e coletados do contexto no momento da inscrição.

Figura 20 – Tela do protótipo que lista os usos secundários permitidos e não permitidos para o perfil selecionado pelo usuário.

6 CONCLUSÃO

Neste trabalho, foi apresentado um mecanismo prático que permite aos usuários controlarem como suas PIIs podem ser usadas em ambientes de identidade federada em nuvem. O mecanismo instrui o usuário sobre os possíveis usos secundários de suas PIIs pelos SPs, lhes permite definir um perfil de privacidade e envia suas preferências ao provedor de serviço.

Este mecanismo está baseado em um modelo que classifica de forma genérica e abrangente os possíveis usos secundários das PIIs em três dimensões, o que dá origem a um conjunto de quarenta e cinco preferências que permitem controlar tais usos. Estas preferências podem ser definidas individualmente ou através de quatro perfis predefinidos, são codificadas em uma estrutura de formato padronizado e legível por máquina denominado token de privacidade e enviadas ao SP juntamente com os dados de autenticação do usuário.

Os trabalhos existentes na área propõem abordagens de baixo nível, como as linguagens de políticas de privacidade, que podem ser executadas por máquinas (W3C, 2006; IBM, 2003; IYILADE; VASSILEVA, 2014); especificações conceituais e de alto nível, como os perfis UML, que fornecem um melhor entendimento sobre os requisitos de privacidade no desenvolvimento de sistemas e aplicações (BASSO et al., 2015; CARAMUJO; SILVA, 2015); ou arquiteturas e modelos completos que visam utilizar as abordagens anteriores para fornecer privacidade aos usuários (KOLTER, 2010; WERNER; WESTPHALL, 2016). Além disso, os sistemas de gerenciamento de identidade Shibboleth e OpenID Connect apresentam mecanismos de privacidade que limitam-se a restringir os dados que o usuário permite que o IdP envie ao SP (SHIBBOLETH, 2017a; OPENID, 2014).

As propostas mencionadas, entretanto, não oferecem métodos práticos para que os usuários definam suas preferências e para que estas sejam enviadas ao SP e, na maior parte dos casos, obrigam o provedor de serviço a adotar tecnologias específicas para representar suas políticas de privacidade. Além disso, a maioria não é adequada para ambientes de nuvem federada e não oferece recursos para que os usuários controlem o uso secundário de seus dados, obrigando-os a aceitarem as políticas dos SPs.

O mecanismo proposto neste trabalho, por sua vez, é centrado no usuário, o conscientiza sobre os usos secundários de seus dados e o auxilia no controle dos mesmos. Além disso, é de fácil adoção, tanto

por parte do usuário quanto do IdP e do SP, uma vez que não requer ferramentas e conhecimentos específicos do usuário e é implantado com as tecnologias que o IdP e o SP já utilizam. Assim, uma característica importante é que não exige que os provedores de serviço usem padrões específicos para expressar e implementar suas políticas de privacidade. Espera-se apenas que estes adaptem os seus sistemas de coleta e uso de dados para interpretar e cumprir as preferências expressas no token de privacidade, o qual os SPs já são capazes de ler e compreender já que possui o mesmo formato que os outros tokens utilizados pelo protocolo OpenID Connect.

A aplicabilidade e a utilidade da proposta foram demonstradas mediante a aplicação do modelo em um estudo de caso e a viabilidade técnica e a correta operação do mecanismo foram verificadas através de um protótipo desenvolvido em Java, que emprega as tecnologias para criação e transmissão do token de privacidade e implementa o estudo de caso. O protótipo tem como finalidade também servir como base para uma futura extensão de uma implementação do OpenID Connect.

As interfaces do protótipo foram desenvolvidas com foco em boas práticas de design e usabilidade e com base na ISO/IEC 29100 a fim de verificar a capacidade do modelo de conscientizar o usuário sobre a existência de usos secundários do seus dados e facilitar a configuração de suas preferências de privacidade. Desta forma, o protótipo permite realizar testes de usabilidade e experimentos para verificar os efeitos da aplicação do mecanismo nos usuários.

A proposta desta dissertação foi definida de forma a ampliar o modelo de Werner e Westphall (2016) e, portanto, pode ser ser incorporada a este como seu mecanismo de definição de preferências de privacidade quanto ao uso e compartilhamento dos dados pessoais do usuário. No entanto, por sua simplicidade e abrangência e por utilizar tecnologias e padrões abertos, o uso do modelo e do token de privacidade não está restrito a sistemas de gerenciamento de identidade federada e pode ser aplicado em qualquer ambiente em que seja necessário definir as preferências de privacidade do usuário.

6.1 CONTRIBUIÇÕES

Com a utilização do mecanismo proposto, os usuários podem expressar suas preferências de privacidade, que são transmitidas sempre com seus dados de autenticação e devem ser usadas pelo SP para alinhar suas ações com relação ao uso secundário dos dados. Desta forma, as

principais contribuições deste trabalho são o modelo de especificação de preferências e o token de privacidade, que invertem o cenário atual onde o usuário é forçado a aceitar as políticas definidas pelos SPs.

6.2 LIMITAÇÕES

O mecanismo proposto tem como única finalidade possibilitar ao usuário controlar o uso secundário de suas PIIs permitindo-lhes definir suas preferências de privacidade e enviando-as ao SP. Assim, não determina de que forma o provedor de serviço atenderá essas preferências e executará suas políticas de privacidade. Para isto, existem diversas abordagens que podem ser usadas e não há necessidade de que o SP modifique as já adotadas.

Além disso, o mecanismo não define quais dados o provedor de identidade pode passar para o provedor de serviço e de que forma. Outros mecanismos devem responsabilizar-se por essa defingão, como o proposto por Werner e Westphall (2016) e o já existente no OpenID Connect.

Embora o uso do token de privacidade possa criar uma necessidade de negociação entre o usuário e o provedor de serviço, o mecanismo proposto não inclui métodos para tal negociação, uma vez que esta é específica para um determinado serviço e deve ser realizada entre o SP e o usuário sem a necessidade de modificar o token de privacidade ou envolver o IdP.

6.3 TRABALHOS FUTUROS

Como trabalhos futuros, propõe-se estender uma implementação do OpenID Connect para suportar o mecanismo proposto, bem como analisar o impacto do tamanho do token na transmissão por URL e, caso necessário, implementar mecanismos de compactação. Propõe-se, também, realizar testes de usabilidade para verificar os efeitos do modelo nos usuários e para avaliar possíveis melhorias na classificação dos usos secundários das PIIs.

Finalmente, propõe-se aplicar o mecanismo em serviços reais para avaliar os impactos tanto nos usuários como nos provedores.

REFERÊNCIAS

ALPÁR, G.; HOEPMAN, J.; SILJEE, J. The Identity Crisis. Security, Privacy and Usability Issues in Identity Management. **Computing Research Repository**, abs/1101.0427, 2011. Disponível em: <http://arxiv.org/abs/1101.0427>. Acesso em: 16 mai. 2017.

AMAZON. **Amazon Elastic Compute Cloud (Amazon EC2)**. Seatle: [s.n.], 2017. Disponível em: <http://aws.amazon.com/pt/ec2/>. Acesso em: 16 mai. 2017.

ARMBRUST, M. et al. A View of Cloud Computing. **Communications of the ACM**, ACM, Nova York, EUA, v. 53, n. 4, p. 50–58, abr. 2010. ISSN 0001-0782.

BASSO, T. et al. Towards a Profile for Privacy-Aware Applications. In: **Computer and Information Technology; Ubiquitous Computing and Communications; Dependable, Autonomic and Secure Computing; Pervasive Intelligence and Computing (CIT/IUCC/DASC/PICOM), 2015 IEEE International Conference on**. [S.l.: s.n.], 2015. p. 371–378.

BENANTAR, M. **Access Control Systems: Security, Identity Management and Trust Models**. 1. ed. Nova York: Springer, 2006. ISBN 978-0-387-27716-5.

BERTINO, E.; TAKAHASHI, K. **Identity Management: Concepts, Technologies, and Systems**. 1. ed. Norwood: Artech House, 2011. ISBN 978-1-60807-039-8.

CARAMUJO, J.; SILVA, A. M. R. d. Analyzing Privacy Policies Based on a Privacy-Aware Profile: The Facebook and LinkedIn Case Studies. In: **2015 IEEE 17th Conference on Business Informatics**. [S.l.: s.n.], 2015. v. 1, p. 77–84. ISSN 2378-1963.

CATTEDDU, D.; HOGBEN, G. **Cloud Computing Risk Assessment: Benefits, Risks and Recommendations for Information Security**. nov. 2009. Disponível em: <http://enisa.europa.eu/publications/cloud-computing-risk-assessment>. Acesso em: 16 mai. 2017.

CHANCHARY, F.; CHIASSON, S. User Perceptions of Sharing, Advertising, and Tracking. In: **11th Symposium On Usable**

Privacy and Security (SOUPS), Ottawa. [S.l.]: USENIX Association, 2015. p. 53–67. ISBN: 978-1-931971-249.

CONNECT2ID. **Nimbus JOSE + JWT**. 2017. Disponível em: <http://www.connect2id.com/products/nimbus-jose-jwt/>. Acesso em: 16 mai. 2017.

CONNECT2ID. **OpenID Connect explained**. 2017. Disponível em: <http://connect2id.com/learn/openid-connect>. Acesso em: 16 mai. 2017.

GOOGLE. **G Suite by Google Cloud**. Mountain View: [s.n.], 2017. Disponível em: <http://gsuite.google.com>. Acesso em: 16 mai. 2017.

GOOGLE. **Google GSON**. 2017. Disponível em: <http://github.com/google/gson>. Acesso em: 16 mai. 2017.

IBM. **Enterprise Privacy Authorization Language (EPAL 1.2)**. 2003. Disponível em: <https://www.w3.org/Submission/2003/SUBM-EPAL-20031110/>. Acesso em: 16 mai. 2017.

INNOVARED. **Federacion MATE**. 2014. Disponível em: <http://www.federacionmate.gob.ar/?q=es/mate>. Acesso em: 16 mai. 2017.

ISO/IEC. **ISO/IEC 29100. International Standard - Information Technology - Security Techniques - Privacy Framework**. 2011. Disponível em: <http://www.iso.org/iso/iso_catalogue/catalogue_tc/catalogue_detail.htm?csnumber=45123>. Acesso em: 16 mai. 2017.

ISO/IEC. **ISO/IEC 29101. International Standard - Information Technology - Security Techniques - Privacy Framework**. 2013. Disponível em: <http://www.iso.org/iso/catalogue_detail.htm?csnumber=45124>. Acesso em: 16 mai. 2017.

IYILADE, J.; VASSILEVA, J. P2U: A Privacy Policy Specification Language for Secondary Data Sharing and Usage. In: **Proceedings of the 2014 IEEE Security and Privacy Workshops**. Washington, EUA: IEEE Computer Society, 2014. (SPW '14), p. 18–22. ISBN 978-1-4799-5103-1. Disponível em:

<http://dx.doi.org/10.1109/SPW.2014.12>. Acesso em: 16 mai. 2017.

JONES, J. B. M.; SAKIMURA, N. **JSON Web Token (JWT)**. maio 2015. Disponível em: <http://tools.ietf.org/html/rfc7519>. Acesso em: 16 mai. 2017.

KERTESZ, A. Characterizing Cloud Federation Approaches. In: _____. **Cloud Computing – Challenges, Limitations and R&D Solutions**. [S.l.]: Springer Series on Computer Communications and Networks, 2014.

KOLTER, J. P. **User-Centric Privacy: A Usable and Provider-Independent Privacy Infrastructure**. [S.l.]: BoD Books on Demand, 2010. ISBN 9783899369175.

KUMARAGURU, P. et al. A Survey of Privacy Policy Languages. In: **Proceedings of the 2007 Symposium on Usably Privacy and Security**. [S.l.: s.n.], 2007.

LANDAU, S.; MOORE, T. Economic Tussles in Federated Identity Management. **First Monday**, v. 17, n. 10, set. 2012. ISSN 13960466.

MELL, P.; GRANCE, T. **The NIST Definition of Cloud Computing**. September 2011. NIST, Gaithersburg, MD, United States.

MICROSOFT. **Windows Azure**. Redmond: [s.n.], 2017. Disponível em: <http://www.windowsazure.com/pt-br/>. Acesso em: 16 mai. 2017.

MIT. **Kerberos: The Network Authentication Protocol**. 2017. Disponível em: <https://web.mit.edu/kerberos/>. Acesso em: 16 mai. 2017.

OASIS. **Privacy Management Reference Model and Methodology (PMRM) Version 1.0**. 2016. Disponível em: <http://docs.oasis-open.org/pmrm/PMRM/v1.0/PMRM-v1.0.html>. Acesso em: 16 mai. 2017.

OPENID. **OpenID Connect Core 1.0 incorporating errata set 1**. 2014. Disponível em: <http://openid.net/specs/openid-connect-core-1_0.html>. Acesso em: 16 mai. 2017.

OPENID. **Welcome to OpenId Connect**. 2015. Disponível em: <http://www.openid.net/connect/>. Acesso em: 16 mai. 2017.

PEREZ-MENDEZ, A. et al. Identity Federations Beyond the Web: A Survey. **Communications Surveys Tutorials, IEEE**, v. 16, n. 4, p. 2125–2141, Fourthquarter 2014. ISSN 1553-877X.

RNP. **CAFe Comunidade Acadêmica Federada**. 2017. Disponível em: <http://portal.rnp.br/web/servicos/cafe>. Acesso em: 16 mai. 2017.

SHIBBOLETH. **ConsentConfiguration**. 2017. Disponível em: <https://wiki.shibboleth.net/confluence/display/IDP30/ConsentConfiguration>. Acesso em: 16 mai. 2017.

SHIBBOLETH. **What's Shibboleth?** 2017. Disponível em: <https://shibboleth.net/about/>. Acesso em: 16 mai. 2017.

SOLOVE, D. J. A Taxonomy of Privacy. **University of Pennsylvania Law Review**, JSTOR, p. 477–564, 2006.

SWITCH. **SWITCHaai**. 2017. Disponível em: <https://www.switch.ch/aai/>. Acesso em: 16 mai. 2017.

TEMOSHOK, D.; ABRUZZI, C. **Draft NISTIR 8149: Developing Trust Frameworks to Support Identity Federations**. 2016. NIST, Gaithersburg, EUA.

VILLARREAL, M. E. et al. Privacy Token: A Mechanism for User's Privacy Specification in Identity Management Systems for the Cloud. In: **ICN 2017: The Sixteenth International Conference on Networks, Veneza**. [S.l.]: IARIA XPS Press, 2017. p. 53–58. ISBN: 978-1-612085-463.

VOORSLUYS, J. B. W.; BUYYA, R. Introduction to Cloud Computing. In: _____. **Cloud Computing: Principle and Paradigms**. Hoboken, EUA: John Wiley & Sons, Inc., 2011.

W3C. **The Platform for Privacy Preferences 1.1 (P3P 1.1) Specification**. 2006. Disponível em: <https://www.w3.org/TR/P3P11/>. Acesso em: 16 mai. 2017.

WEINGARTNER, R. **Controle de Disseminação de Dados Sensíveis em Ambientes Federados**. Dissertação (Mestrado) — Universidade Federal de Santa Catarina, 2014.

WERNER, J.; WESTPHALL, C. M. A Model for Identity Management with Privacy in the Cloud. In: **2016 IEEE Symposium on Computers and Communication (ISCC)**. [S.l.]: IEEE, 2016. p. 463–468. ISBN: 978-1-5090-0679-3.

WESTIN, A. **Privacy and Freedom**. [S.l.]: The Bodley Head, 1967.

XIAO, Z.; XIAO, Y. Security and Privacy in Cloud Computing. **Communications Surveys Tutorials, IEEE**, v. 15, n. 2, p. 843–859, 2013. ISSN 1553-877X.

ZHAO, J. et al. Privacy Languages: Are We There Yet to Enable User Controls? In: **Proceedings of the 25th International Conference Companion on World Wide Web, Montreal, Quebec, Canada**. International World Wide Web Conferences Steering Committee, 2016. (WWW '16 Companion), p. 799–806. ISBN 978-1-4503-4144-8. Disponível em: <http://dx.doi.org/10.1145/2872518.2890590>. Acesso em: 16 mai. 2017.