UNIVERSIDADE FEDERAL DE SÃO CARLOS

CENTRO DE CIÊNCIAS EXATAS E TECNOLOGIA

PROGRAMA DE PÓS-GRADUAÇÃO EM

CIÊNCIA DA COMPUTAÇÃO

Sistema de Integração de Técnicas de Proteção de Privacidade que Permitem Personalização

ALUNO: Robson Eduardo De Grande

São Carlos-SP
Novembro/2006

Ficha catalográfica elaborada pelo DePT da Biblioteca Comunitária da UFSCar

G751si

Grande Robson Eduardo de.
Sistema de integração de técnicas de proteção de privacidade que permitem personalização / Robson Eduardo De Grande. -- São Carlos : UFSCar, 2007.
138 f.

Dissertação (Mestrado) -- Universidade Federal de São Carlos, 2006.

1.Redes de computação. 2.Privacidade e personalização. 3.Gerenciamento de serviços World Wide Web. 4. Navegação anônima. I. Título.

CDD: 004.6 (20ª)

Universidade Federal de São Carlos

Centro de Ciências Exatas e de Tecnologia

Programa de Pós-Graduação em Ciência da Computação

"Sistema de Integração de Técnicas de Proteção de Privacidade que Permitem Personalização"

ROBSON EDUARDO DE GRANDE

Dissertação de Mestrado apresentada ao Programa de Pós-Graduação em Ciência da Computação da Universidade Federal de São Carlos, como parte dos requisitos para a obtenção do título de Mestre em Ciência da Computação.

Membros da Banca:

Prof. Dr. Sérgio Donizetti Zorzo
(Orientador – DC/UFSCar)

Prof. Dr. Luis Carlos Trevelin
(DC/UFSCar)

Prof. Dr. Wagner Meira Junior
(UFMG)

São Carlos
Novembro/2006

“What we do in life echoes in eternity”
M. D. M.

Dedico este trabalho
a José, a Maria, a Wilson, a Priscila, a amigos e àqueles que acreditam que o esforço e o trabalho árduo, justo e sincero trazem bons resultados para a vida.

Agradecimentos

A Deus pela vida, pela saúde e por todas as oportunidades que me ofereceu, oferece e oferecerá.

A meus pais (José e Maria) e a meu irmão (Ricardo) que me apoiaram e confiaram em mim durante toda a minha vida.

À Priscila, minha noiva, pelo apoio, pelos conselhos, pelo carinho, pela confiança em minha capacidade e pela paciência.

Ao meu orientador Professor Dr. Sérgio Donizetti Zorzo, pela amizade, pelo apoio, pela confiança, pela orientação e pelos conselhos que possibilitaram a conclusão desse trabalho.

Aos professores do Departamento de Computação por seus conselhos, por sua ajuda e por suas anedotas.

Aos colegas e amigos do programa de Pós-graduação em Ciência da Computação pelas conversas, pelas ajudas, pelas opiniões, pelo companheirismo, por sua irreverência e pelas brincadeiras que suavizaram as frustrações e os momentos de desânimo do trabalho de pesquisa.

À Universidade Federal de São Carlos que me acolheu em minha graduação e em meu mestrado, ofereceu infra-estrutura e oportunidades para minha formação profissional, bem como, para o lazer e a descontração, e auxiliou na construção de diversas características de meu caráter.

A todos aqueles que contribuíram para a ocorrência de todos os fatos em minha vida, sem os quais eu não estaria aqui escrevendo esse texto e não seria da maneira que sou.

Resumo

A implementação da interface em aplicações de *e-commerce* deve considerar dois aspectos: as necessidades de privacidade do usuário e a utilização de *marketing* através da aplicação de personalização em serviços disponibilizados por *sites* na *Web*. Nesse contexto, existem vários mecanismos desenvolvidos tanto para realizar a coleta de dados quanto para trazer para o usuário certo nível de privacidade. Os *sites* usam desde formulários em páginas *Web* até técnicas como *clickstream*, *cookies* e *data mining* para obter informações, analisá-las e efetuar personalização.

Os mecanismos de proteção de privacidade existentes são limitados em fornecer garantias de privacidade na navegação do usuário, e alguns não permitem a coleta de dados dos *sites Web*. O sistema MASKS gerencia anonimato para o usuário e permite a coleta implícita. Ele não fornece segurança na obtenção explícita e não permite sessões de navegação pela divisão dos *cookies* em diferentes grupos de interesse. O Projeto 3P introduz um mecanismo para manter os usuários cientes sobre políticas de privacidades enquanto navegam por *sites*. Ele insere pouca confiabilidade para a coleta implícita. Nesse trabalho, um sistema é desenvolvido para combinar uma extensão do sistema MASKS, uma extensão do P3P e uma arquitetura de certificação de privacidade. Esse sistema combinado concilia as qualidades de cada mecanismo e supre suas limitações. O sistema MASKS estendido inclui sessões no *proxy* de mascaramento para possibilitar a criação de sessões de navegação. O P3P estendido adiciona informações sobre os benefícios do usuário ao divulgar seus dados para serviços *Web*. Os certificados de privacidade garantem o cumprimento das políticas declaradas pelos *sites*. Testes com usuários são realizados para comparar e avaliar as vantagens de uso desse novo sistema. Os resultados mostraram que o sistema combinado fornece maior confiança na navegação do usuário. Assim, a implementação desse sistema prova ser útil em prover privacidade durante a navegação na *Web*, sem impedir a coleta de dados.

Palavras-chave: Privacidade, Personalização e Web.

Abstract

The interface implementation in e-commerce applications must consider two aspects: the privacy necessities of the user and the marketing utilization by the personalization application in services provided by *sites* in the Web. In this context, there are several mechanisms developed as to accomplish the data gathering as to bring to the user certain privacy level. *Sites* use since formularies in Web pages until techniques as clickstream, cookies and data mining to obtain information, to analyze them and to accomplish personalization.

Current mechanisms for privacy protection are limited in their ability to provide privacy guarantees to data gathered on the Web, and some of them don't allow the data gathering by Web *sites*. The MASKS system manages the anonymity of its users and allows the implicit collecting of data. It neither provides security in the explicit gathering and nor allows the creation of browsing sessions via grouping of cookies into groups of different interests. Project 3P introduces a mechanism to keep users aware of privacy policies while they navigate Web *sites*. It adds low reliance to the implicit gathering. Here, we introduce a new system that combines an extension of MASKS system, an extension of P3P and an architecture of privacy certification. This combined system gathers the best qualities of each mechanism, and removes their limitations. The extended MASKS system includes sessions in the masking proxy to enable the creation of browsing sessions. The extended P3P adds information about the user benefits in disclosing his/her data to Web services. The privacy certificates guarantees the accomplishment of the policies stated by *sites*. Tests with users were carried out in order to compare and to evaluate the advantages of using this newly combined system. The results showed that the combined system provides users' navigation with higher reliability. Thus, the implementation of this system proves useful for providing privacy during Web browsing, without impeding data gathering.

Keywords: Privacy, Personalization, Web.

Lista de Figuras

Lista de Tabelas

Sumário

1 Introdução

A privacidade e a personalização têm grande importância para os empreendimentos *on-line*: elas são requisitos fundamentais para a atratividade de um *site*. No acesso a serviços da *Web*, a segurança das informações do usuário tem conseqüências em sua confiança. Um *site* que oferece mecanismos confiáveis de acesso a seus serviços, fornece maiores garantias de segurança de informação para os visitantes. Da mesma forma, a modelagem de conteúdo de acordo com as características dos usuários provê comodidade e facilidade na navegação pelas páginas de um *site*. Esses benefícios somente são fornecidos através da caracterização direcionada dos serviços de um *site*. A identificação dos interesses dos usuários permite criar essa caracterização. Assim, apesar do crescimento de adeptos à Internet, a privacidade do usuário e a personalização de serviços devem ser consideradas na implementação *on-line* de empreendimentos.

A Internet em seu estado atual é vastamente explorada por usuários e organizações. Sua adoção se apresenta de forma exponencial ao longo do tempo. Ela evidencia uma taxa de crescimento de 2,3% ao mês em 2003 [1] e um aumento de 20 milhões de novos usuários entre os meses de março e de junho do ano de 2006 [2]. Em conseqüência desse crescimento, a *Web* comporta usuários de diferentes perfis. Ela abriga interesses variados identificados no fornecimento de produtos e de serviços por organizações e na busca de informações e de serviços por usuários.

A confiança do usuário é dependente da privacidade oferecida. Ela é reconhecida como um item importante para levar ao sucesso de um marketing *on-line*. O medo do usuário do risco de divulgação não autorizada ou do mau uso de suas informações leva-o a deixar de acessar certos serviços ou a trocar por outros mais confiáveis. A coleta ou o uso das informações sem a ciência ou o consentimento do usuário da *Web* leva a uma perda de controle sobre suas informações. Desse modo, o aumento da percepção do usuário de controle sobre seus dados reflete no crescimento da adoção de serviços e produtos baseados na *Web*. Da mesma forma que a ocorrência de eventos em que a privacidade é invadida, o usuário cria um conceito de insegurança para os serviços *on-line*.

Para personalizar algum serviço, é preciso ter um conhecimento mínimo sobre quem o receberá. Dado que personalizar é tornar algo mais próximo das características de alguém, um

usuário somente receberá um serviço personalizado depois que alguma informação for obtida pelo *site*. Além disso, a identificação do usuário é necessária para possibilitar a realização de diversos serviços na *Web*. Dessa forma, a coleta de dados é essencial para que *sites* disponibilizem serviços na Internet.

Diversos mecanismos são utilizados para a coleta de dados na *Web*. Ela pode ocorrer tanto de maneira explícita quanto de maneira implícita. Para obter informações, são utilizados basicamente formulários, *cookies* [3], *Web bugs* e *clickstream* [4] [5]. Através de técnicas de *Data Mining*, o conhecimento sobre os usuários é extraído dos dados coletados, os quais são obtidos por mecanismos que observam as requisições de páginas de visitantes a sites.

Existem diversas técnicas de proteção com o propósito de evitar a invasão de privacidade na coleta de dados através do aumento do controle do usuário sobre suas informações. A criação de pseudônimos, o uso de contratos e selos de privacidade, o mascaramento de requisições, a divulgação contextualizada de práticas de privacidade e de benefícios de personalização e a introdução de anonimato para a navegação são exemplos dessas técnicas. Cada um desses procedimentos aborda de maneira diferente a privacidade do usuário.

Segundo a taxonomia de proteção de privacidade de Ishitani [6], não existe um mecanismo que abranja todas as camadas de proteção. Dessa forma, MASKS, P3P e selos de privacidade são unidos em um único sistema para englobar um maior número de camadas dessa classificação. As três ferramentas trazem privacidade e confiança da divulgação de informações, permitem a coleta de dados para personalização e não requerem mudanças nos protocolos de comunicação existentes.

O sistema MASKS fornece privacidade para a navegação e permite a obtenção implícita de informações. Esse servidor de mascaramento possibilita a identificação de perfis de grupos de interesse sem que o usuário seja identificado. Entretanto, em virtude de o *proxy* apresentar a propriedade de dividir os *cookies* em diferentes grupos de interesse, ele limita a determinação de perfis e impossibilita a manutenção de sessões.

O Projeto 3P divulga para os usuários informações sobre as práticas de privacidade de um *site*. Ele introduz um mecanismo de leitura e avaliação automáticas de contratos de privacidade. Para isso, um formato é especificado para a criação de políticas e para a construção de agentes de usuários. Apesar de esses contratos de privacidade informarem sobre as práticas dos *sites*, eles não podem trazer nenhuma garantia para o usuário, principalmente para a coleta implícita.

Com o intuito de aumentar o controle do usuário sobre suas informações, um novo sistema é desenvolvido para proteger a privacidade do usuário sem deixar de permitir a coleta de informação para a agregação de personalização a serviços da *Web*. A solução desenvolvida envolve o uso de várias abordagens presentes na literatura; ela procura oferecer no contexto de privacidade uma abrangência maior que a das ferramentas existentes. A abrangência mencionada contempla desde os aspectos legais, até níveis de segurança de informação presentes em serviços oferecidos pelas redes de computadores. As ferramentas incorporadas ao sistema provêm privacidade em seu escopo de abordagem e permitem o fornecimento de serviços personalizados.

Esse sistema é construído a partir da realização da análise sobre a situação da privacidade dos usuários de Internet. Através da combinação entre o servidor MASKS estendido, o Projeto 3P estendido e os certificados de privacidade, o sistema adiciona anonimato para coleta implícita e maior confiança no envio explícito de informações do usuário. Sessões de navegação são incluídas no servidor MASKS para estendê-lo. O P3P é estendido com a adição de informações sobre os benefícios do usuário na divulgação de dados.

O capítulo 2 conceitua personalização de forma abrangente, apresentando uma definição que reflete o contexto abordado neste trabalho. Apresenta ainda o estado-da-arte da personalização na *Web*, apontando as necessidades que levam a sua aplicabilidade, a forma como é empregada e as dificuldades encontradas na implementação efetiva.

O capítulo 3 apresenta as ferramentas utilizadas para a aplicação de personalização. Para coleta de informações e identificação do usuário, são mais comuns o uso de formulários, *Web bugs*, *cookies* e a análise de navegação, denominada *clickstream*. A armazenagem das informações coletadas ou manufaturadas pode ser realizada em bancos de dados ou nos computadores dos próprios usuários, como no caso dos *cookies*. Para a interpretação das informações coletadas, é utilizada de forma mais abrangente a técnica de mineração de dados [7] para atuar em um repositório com grande quantidade de informações.

O capítulo 4 conceitua privacidade de forma abrangente e apresenta uma definição que reflete o contexto abordado neste trabalho. Também apresenta a situação da privacidade dos usuários na *Web* por evidenciar os momentos em que ela pode ser perdida ou o prejudicada através da coleta e uso de informação. Ao final, é apresentado um enfoque para o problema da existência de privacidade sem o comprometimento de serviços personalizados, os quais têm grande importância no contexto virtual.

O capítulo 5 descreve uma taxonomia em camadas do contexto de proteção de privacidade [6]. Essa divisão é utilizada no detalhamento da proposta, que procura incorporar as características positivas de proteção de privacidade de cada conjunto de abordagem delimitado pela taxonomia. Assim, para conseguir atingir todo o contexto de proteção de privacidade, é necessário construir um sistema que englobe todas as camadas dessa classificação.

No capítulo 6, são apresentados os certificados e os selos de privacidade. Eles são utilizados para incrementar a visão de confiança dos usuários nos sites *Web* visitados. Os selos de privacidade são marcas de privacidade de confiança apresentadas nas principais páginas dos sites. Essas marcas dão garantias aos visitantes de que as práticas do site são vigiadas por entidades reguladoras de privacidade.

As Leis de Proteção de Privacidade são apresentadas no capítulo 7. Elas regulamentam as ações e os mecanismos para favorecer a proteção de privacidade. Entretanto, na *Web* não existe uma legislação única que a rege. Cada nação procura aplicar suas leis com relação à privacidade de seus cidadãos e também aos usuários da Internet. Essas aplicações podem parecer semelhantes, mas possuem diferentes características referentes à diversidade cultural e à soberania de cada país. Apesar dessas diferenças, existem também organizações que estão envolvidas com o controle e com a direção para facilitar a regulamentação de leis de privacidade na Internet.

O capítulo 8 exibe os mecanismos existentes e mais conhecidos para a proteção de privacidade do usuário. Alguns desses mecanismos utilizam o anonimato, que protege completamente a privacidade na navegação do usuário, mas prejudica a personalização e não é aplicável na divulgação explícita de informações pelo usuário. Outros permitem personalização, através da inserção de contratos de privacidades, criação de pseudônimos ou introdução de máscaras para a navegação do usuário.

No capítulo 9, é apresentado o trabalho de mestrado, cujo resultado foi o desenvolvimento de um sistema de integração de técnicas de privacidade que permitem personalização. Esse sistema engloba 5 camadas da taxonomia introduzida por Ishitani. A arquitetura do sistema é apresentada em módulos, e suas funcionalidades são descritas pelo comportamento de cada módulo e a relação entre eles.

No capítulo 10, é descrita a metodologia adotada para a execução dos testes com usuários e para realização da análise dos dados coletados. Esses testes comparam o uso do sistema proposto e o uso do Projeto 3P, bem como avaliam a percepção de confiança que o

usuário possui ao utilizar o mecanismo desenvolvido. Eles são necessários para mensurar fatores subjetivos que envolvem a proteção de privacidade.

Os resultados do desenvolvimento do sistema e dos testes aplicados são apresentados no capítulo 11. Esses resultados evidenciam o aumento de segurança das informações e de garantias de privacidade, bem como provam o aumento da percepção do usuário de confiança no acesso a serviços da *Web*.

No capítulo 12, são discutidos os resultados e as implicações das modificações efetuadas no servidor de mascaramento e no P3P. As conclusões do trabalho concluído e os trabalhos futuros a serem realizados são apresentados.

2 Personalização

O termo “personalização” possui diversas acepções que variam de acordo com situações e contextos. De forma geral, conforme *Cambridge Dictionary*[1], uma dessas acepções é apresentada como “fazer algo apropriado para as necessidades de uma pessoa particular” (*to make something suitable for the needs of a particular person*). Assim, nesse contexto, personalizar é tornar algo mais próximo às características de alguém, adaptando-o e adequando-o a suas vontades, necessidades e preferências. Personalização é apresentar algo de forma diferente a cada pessoa, pois cada uma tem um gosto definido, um perfil formado [8].

A personalização possui uma aplicação muito vasta por ser dependente das características e dos interesses humanos. Ela pode estar presente em objetos pessoais para identificá-los como pertencentes a alguém, como uma forma de demarcação e defesa das posses de um indivíduo ou para torná-los mais agradáveis por apresentarem características mais próximas de um determinado perfil.

Um produto ou serviço pode atender as necessidades fundamentais de uma pessoa por suas funcionalidades e características primárias. Além disso, um serviço, através da personalização, pode possuir determinadas características que o torna mais parecido com um indivíduo. Essas qualidades secundárias são consideradas tão importantes que, em muitos casos, a escolha do produto ou serviço é regida somente através delas.

Nesse sentido, para aplicar personalização é preciso identificar e conhecer as necessidades e preferências das pessoas. Essas necessidades podem ser vistas como sendo a falta e a procura de algum produto ou serviço de significativa importância para algum indivíduo. Além disso, deve ser delimitado o modo como esse conhecimento é aplicado para gerar uma caracterização que atenda melhor as necessidades identificadas. Os sistemas computacionais existentes nas empresas apresentam divergência entre as informações apresentadas por eles e as procuradas por um indivíduo [9].

No contexto de caracterização do usuário, personalizar para a *Web* significa modificar a experiência do usuário no acesso a conteúdo de sites de acordo com suas preferências [10]. As preferências são identificadas através da análise de comportamento do usuário e são modeladas para modificar a maneira como o usuário interage com o sistema. Em uma outra visão, Goderis [11] se refere à personalização como políticas e regras que integram

[1] http://dictionary.cambridge.org/

determinado modelo de usuário à aplicação sendo construída. Essas regras evidenciam o modo como as características do usuário influenciam a aplicação. Considerando a usabilidade, essas regras devem levar em conta fatores cognitivos do usuário. Dessa forma, a preocupação em adaptar a forma e o conteúdo de um *site* da *Web* deve ser parte integrante de seus requisitos de desenvolvimento centrados no usuário [12].

O comportamento humano não é estático. Com o passar das experiências vividas, as características individuais sofrem alterações, o que deve ser tratado como um alvo móvel para um possível mapeamento ou medição dos interesses de um indivíduo. O contexto ou ambiente onde o indivíduo está inserido deve ser considerado, devido à influência que ele exerce sobre as características individuais. PVA, serviços de agentes e 1:1Pro são exemplos de sistemas de modelagem computacional do usuário [13].

Na visão da organização de conteúdo, a personalização é um meio de os empreendimentos modificarem seus produtos e serviços para uma maior aprovação das pessoas. Sae-Tang e Esichaikul [14] afirmam que personalização é a ferramenta utilizada por comerciantes para adaptar os *sites* da *Web* e anunciar seus produtos. Porém, não é suficiente que o conteúdo de um *site* da *Web* esteja organizado e bem estruturado hierarquicamente, já que experiências obtidas mostram que isso não atende inteiramente ao usuário [15].

A organização do conteúdo tem grande importância na navegação dos usuários pela *Web*. McGarry [16] defende que o conteúdo é encontrado pelos indivíduos a partir do significado. Esse significado é relacional, envolve outros elementos de conteúdo e depende do contexto em que ele está presente. Como exemplo, existe o hipertexto que foi criado para permitir a associação de elementos de dados por seu conteúdo, e não mais por sua estrutura [17].

Existem definições de personalização que procuram mesclar as características do usuário às informações disponíveis em um sistema. O objetivo mais importante de personalizar a interface de um sistema computacional é trazer ao usuário informações relevantes com os seus interesses, como prover a ele o que ele quer, da maneira que quer [18] e no tempo certo [19]. Quando um usuário exercita seu comportamento na busca de informação, normalmente, ele deseja resolver algum problema ou alcançar um objetivo para o qual seu atual nível de conhecimento é inadequado ou insuficiente [20].

O modelo de usuário, o modelo de conteúdo e a integração usuário-conteúdo são fatores determinantes de personalização [21]. Existem diversos trabalhos para essa integração: estabelecimento de regras e algoritmos para extrair informações do conteúdo a ser direcionado, para depois cruzar essas informações com perfis de usuário [22], estudos

matemáticos de como relacionar as informações encontradas de forma caótica e arranjá-las de forma lógica, de acordo com os requisitos do usuário [23] e discussão da organização do conteúdo através das técnicas de personalização [24].

Sistemas que aplicam personalização podem ser classificados como adaptados ou adaptativos. Os sistemas adaptados são criados segundo requisitos de características de usuários. Essa customização é definida como a capacidade do usuário ou do sistema de modificar a interface conforme a necessidade [25]. Ao longo do tempo, eles não são capazes de se alterar e atender diferentes exigências e preferências. Por outro lado, os sistemas adaptativos têm seus parâmetros modificados em tempo de interação com o usuário [26]. Essas adaptações da interface são realizadas pelo próprio sistema [25].

Entre todas as situações em que a personalização pode ser aplicada, o ambiente comercial é o que mais a emprega. Boar [27] defende que é preciso utilizar a personalização do conteúdo aos clientes para que a empresa seja capaz de obter diferencial competitivo e conseqüente vantagem. Nesse meio, a personalização é utilizada no atendimento, em propagandas, em produtos e em serviços que estão à venda. Através da identificação do perfil de clientes, os interesses e objetivos dos usuários podem ser delimitados. Com isso, empreendimentos são capazes de melhorar o atendimento e diferenciá-lo, de aumentar a interatividade com o usuário. A personalização é muito utilizada na *Web*, como o comércio eletrônico e os portais corporativos.

As técnicas de personalização, que são voltadas para a identificação e caracterização do usuário, são implementadas seguindo três tarefas: coleta de dados, análise e uso de dados e recomendação e apresentação [13]. De acordo com Koch [28], para usar a personalização, é necessário que algumas informações sejam obtidas de forma a serem utilizadas com fins de divulgação e de cálculos de estatísticas. Dessa forma, a preferência do usuário pode ser obtida de um modo geral. Segundo a estrutura apresentada, a obtenção de informações pessoais é essencial para a viabilização de personalização, visto que sem essa coleta não é possível identificar as características de uma pessoa.

Em decorrência da personalização, a informação vem se apresentando como um fator fundamental para as organizações que desejam se manter no mercado de forma competitiva e inovadora. Aqueles que detêm informações sobre seus negócios podem utilizá-las para determinar o perfil e preferências de seus usuários, definir estratégias de marketing e reduzir riscos ao ingressar em novas áreas de negócios [28]. O acesso às informações relevantes, como preferências de usuários, torna-se imprescindível para que os dirigentes possam decidir

a melhor maneira de administrar os negócios e fazer com que a organização atinja seus objetivos mercadológicos [29].

A personalização pode ser aplicada sem a necessidade de conhecer o indivíduo detalhadamente. Ela pode ser guiada por características obtidas estatisticamente sobre um determinado grupo. Essas características podem ser obtidas através de pesquisa de opinião, de análise comportamental de uma população e de outros métodos.

Dessa forma, o nível de personalização, que pode ser mais específico ou mais abrangente, depende da quantidade de atributos de um indivíduo alvo. Assim, quanto mais específica é a personalização aplicada, mais informação é necessária. A obtenção e a armazenagem de informação são tarefas difíceis de serem realizadas, principalmente quando se manipula um volume grande de dados. Além disso, a obtenção dessas informações implica em grandes responsabilidades, pois elas são dados pessoais e com grande importância individual.

A personalização é aplicada amplamente em todos os ambientes. Na *Web*, isso não é diferente. Pela facilidade do tráfego e da manipulação de informação através do uso de técnicas para automatização de processos, o emprego da personalização se torna fácil e flexível. A necessidade de personalizar os serviços oferecidos vem fazendo com que técnicas e ferramentas sejam criadas e aprimoradas com o intuito de facilitar a tarefa de personalização [6].

A aplicação de personalização, realizada pelo marketing, é uma das mais fortes vantagens competitivas no mercado *on-line*. Ela tem se tornado uma ferramenta crítica para o progresso de serviços e negócios *on-line*. O recebimento de serviços personalizados de *sites Web* visitados é bastante atrativo, além de muito bem aceito pelos usuários. De acordo com Kobsa [30], clientes precisam sentir que possuem um relacionamento pessoal e único com a empresa. Para prová-lo, é apresentada uma pesquisa que mostra que *sites* que oferecem serviços personalizados conseguiram um aumento de 47% no número de clientes.

A coleta de dados na *Web* pode ser realizada de forma implícita ou explícita [29]. No primeiro caso, a obtenção sucede sem a ciência ou o consentimento do usuário. Essa falta de ciência de um indivíduo do que ocorre com seus dados pode levar a uma perda de controle de suas informações. Para realizar essa coleta, os *sites* utilizam basicamente *cookies*, *Web bugs* e observação de requisições de páginas, conhecida como *clickstream*. A observação da navegação do usuário captura dados de controle, os quais são relacionados aos protocolos de comunicação utilizados para transmissão de informações na *Web*. Assim, através da análise dos dados coletados, é possível delimitar um perfil para aqueles que freqüentam um *site*.

As informações obtidas pelo método implícito são fundamentais para a interação do usuário com o *e-commerce*. Ela possibilita a criação de perfis do usuário baseados em seus interesses, padrões de navegação, preferências e outros. Devido a sua importância, esse tipo de coleta de dados não deve ser impedida por mecanismos de navegação anônima.

Na coleta explícita, é necessário que o usuário exponha suas informações de forma evidente. Os dados coletados são os de conteúdo, eles são relacionados às informações pessoais de um indivíduo, as quais não podem ser obtidas através de um mecanismo automático. Nesse tipo de obtenção de dados, o visitante está ciente da existência dele e tem a opção de consenti-lo. Normalmente, os dados são enviados por formulários apresentados em páginas da *Web*. Dessa forma, é mais difícil de ocorrer algum tipo de agressão na privacidade segundo esse processo de coleta. Porém, uma invasão pode suceder em razão da atitude do *site* no destino dado às informações coletadas, no método de armazenagem, na segurança utilizada, na divulgação para terceiros, na promoção de marketing e em outros.

Os dados dos usuários, após serem coletados, passam por processos de filtragem de informação. Técnicas como de mineração de dados são utilizadas com a finalidade de separar as informações mais interessantes ou de colocar essas informações em certo formato padrão específico para serem trabalhadas posteriormente. Normalmente, os dados coletados não estão em um formato apropriado para aplicar personalização, principalmente quando são dados obtidos a partir da observação da navegação do usuário. Por esse motivo, é necessário se fazer uma análise das informações obtidas.

Ao reunir todas essas informações, é preciso encontrar uma forma de armazenagem organizada para que, no futuro, elas possam ser recuperadas a fim de serem usadas como fonte para aplicar personalização. Isso pode ser feito de diferentes formas. Armazenam-se os dados em um banco de dados, *data warehouse*, no mesmo servidor em que se encontra a página ou em um servidor de terceiros. Da mesma forma, as informações podem ser armazenadas no computador do próprio usuário através de *cookies*, evitando-se, desse modo, o trabalho de gerenciar um banco de dados.

Quando um usuário faz acesso ao *site*, deve haver um modo de o sistema identificar e relacionar as informações armazenadas com um determinado usuário, depois processar essas informações e produzir algo personalizado. Nesse caso, *cookies* de processos de *login* são utilizados para relacionar a identidade do usuário com o computador que ele utiliza para se comunicar com o servidor do *site*.

Ao ser identificada a pessoa que vai receber o resultado da personalização, resta somente ao *site* em questão aplicar regras e políticas específicas para promover o resultado

personalizado para esse usuário. Esse resultado pode ser, em termos gerais, um produto ou serviço, uma visualização ou uma propaganda personalizada.

Através da especificação da personalização na Internet, torna-se evidente a grande funcionalidade que ela possibilita para os *sites* de *e-commerce*. Por ela, pode-se apresentar o *site* de diversas maneiras, cada uma específica para cada usuário ou para um grupo de usuários. Assim, há um grande benefício promovido para a disposição de conteúdo, mostrando-se os produtos e serviços mais interessantes de acordo com os interesses de cada um. Os produtos e serviços podem possuir características específicas de cada usuário, e as propagandas são beneficiadas pela facilidade de se personalizá-las e enviá-las pela Internet.

3 Mecanismos de personalização existentes

A Internet comporta diversos meios de se oferecer serviços aos usuários. O meio mais comum é a *Web*. Esse meio é o mais utilizado por usuários finais. Para fornecer esse serviço, é necessário que páginas *Web* tenham um formato uniforme para que possam ser disponibilizadas ou apresentadas para a visualização correta. Essa uniformidade é dada pela descrição das páginas em linguagem de marcação e hipertexto (HTML – *Hypertext Markup Language*). Tal linguagem de marcação delimita os parâmetros que determinam o formato do contexto apresentado na *Web*.

A transferência de páginas de um servidor para um computador do usuário é fornecida através do protocolo de transferência de hipertexto, HTTP – *Hypertext Transfer Protocol*. HTTP é um protocolo que age no nível de aplicação e é caracterizado por não identificar estados de navegação [31].

De forma geral, os mecanismos de análise de navegabilidade do usuário, baseados no protocolo HTTP e no formato HTML, são apresentados como ferramentas utilizadas para a aplicação de personalização. Para coleta de informações, são mais comuns o uso da análise da entrada de dados em formulários e a análise de navegação do usuário. Na interpretação das informações coletadas, são utilizadas, de forma mais abrangente, técnicas de mineração de dados para atuarem com um repositório com grande quantidade de informações.

3.1 Web Bugs

Web bugs são mecanismos utilizados para identificar o acesso do usuário a determinado serviço da *Web*. Eles são pequenas imagens inseridas em mensagens de correio eletrônico ou em páginas *Web* com o objetivo de monitorar o acesso dos usuários a alguma informação. Por exemplo, um *Web Bug* pode ser usado para a validação de e-mails [32].

Esse mecanismo de monitoração é chamado de *Web Bug*, de *Web Beacon* e de outros nomes, por ser um pequeno *eavesdropper* (espião); ou de *invisible* GIFS, de *Clear Gif* e de *Tracker Gif*, por ser uma imagem transparente de tamanho 1 pixel X 1 pixel, do tipo *Graphics Interchange Format* (GIF) [33].

Devido às suas características, um *Web Bug* é imperceptível aos olhos humanos. Dessa forma, os usuários requisitam e visualizam páginas e mensagens da *Web* sem estar cientes de

que sua navegação pode estar sendo monitorada. Para poder identificá-los, é necessário observar *tags* de imagens no código HTML (*Hypertext Markup Language*) da página ou da mensagem que os contém.

Geralmente, um *Web bug* pertence a um servidor distinto daquele que é acessado pelo usuário. Esse servidor fornece serviços de coleta de dados para aplicação de personalização ou para criação de dados estatísticos sobre os visitantes de um determinado *site* ou página. Os *Web Bugs* podem trabalhar em conjunto com *cookies*, monitorando quais *sites* cada usuário visita, exibindo, assim, anúncios específicos para cada usuário [34].

Esses mecanismos são baseados na premissa de que o navegador busca os elementos formadores de uma página HTML nos respectivos servidores que os comportam. Para montar a página HTML com a imagem do *Web Bug*, o navegador faz uma requisição para o servidor que contém o *Web Bug*, como apresentado na figura 1. Esse servidor, ao receber uma requisição do navegador do usuário, obtém nos cabeçalhos http e TCP/IP informações que identificam o computador do visitante da página. Essas informações são o IP, a porta, o navegador e o sistema operacional utilizados, o instante em que a imagem foi carregada, a URL do *site* que está sendo visitado pelo usuário, etc. [32] [35].

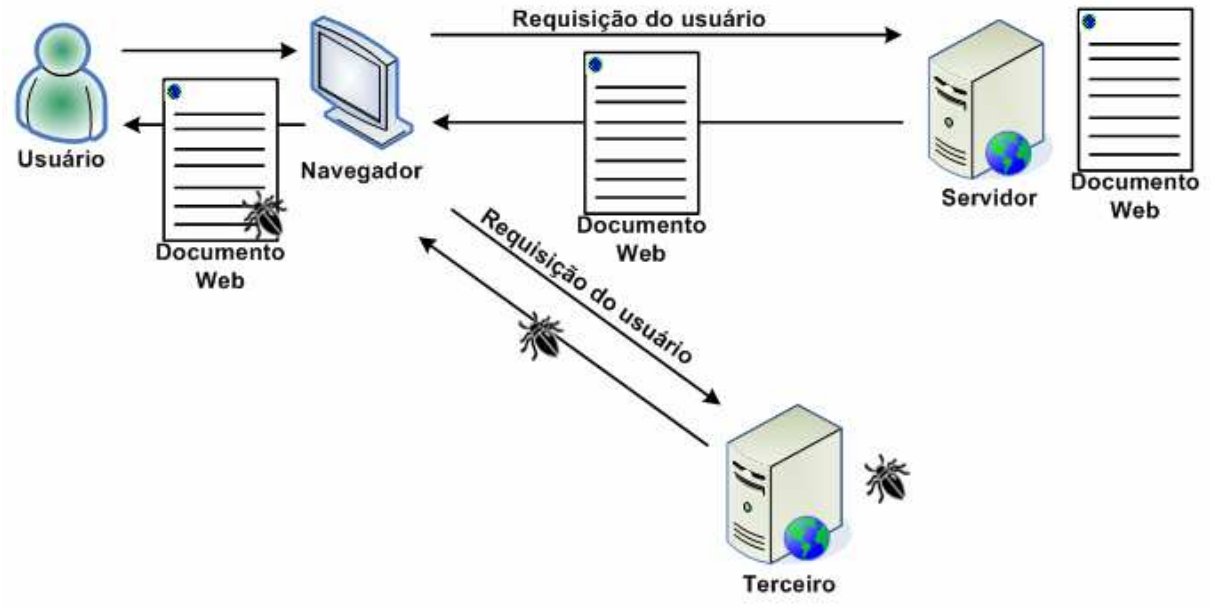

Figura 1. Identificação do Computador do usuário por um Web Bug em uma requisição de página.

O uso de *Web bugs* obedece às seguintes etapas: criação de uma imagem do *Web Bug* com um identificador de mensagem ou da página a ser monitorada; hospedagem da imagem em um servidor no qual seja possível ter acesso às informações de *log*; criação de uma

mensagem de e-mail ou uma página no formato HTML, que tenha em seu conteúdo a URL completa da imagem correspondente ao *Web Bug*; envio da mensagem para o endereço de e-mail a ser validado ou publicação da página a ser monitorada. Quando o usuário abre a mensagem em seu navegador, o *Web Bug* é acessado e a confirmação de que o e-mail é válido ou a página é acessada é gerada.

O código HTML das páginas e das mensagens visualizadas pode ser filtrado para a eliminação desses espiões. Para isso, devem-se procurar as *tags* de imagens que delimitam o tamanho de uma imagem de *Web bug* e eliminá-las. Esse processo é realizado por mecanismos para detecção de *Web bugs*, como *Bugnosis* [2] e *SaferSurf* [3].

Alguns defendem a idéia de que *Web Bugs* permitem que empresas melhorem a qualidade de seus serviços, com os dados estatísticos disponibilizados. Entretanto, por mais que os *Web bugs* sejam utilizados de forma inofensiva, eles podem gerar uma invasão de privacidade.

3.2 Cookies

Quando alguém acessa um servidor na *Web*, o navegador da máquina do usuário envia uma requisição de informação para o computador que contém o *site*, chamado de servidor *Web*. Esse servidor responderá à requisição transmitindo a informação solicitada para o computador do usuário. Finalmente, o navegador do usuário mostrará a informação recebida na tela do computador.

Entretanto, a característica do protocolo HTTP de não possuir estados torna cada transação distinta de outra, não apresentando uma memória para identificar alguém que está retornando ao *site*. Os *cookies* foram criados para estender o protocolo HTTP e introduzir uma memória nele. Através deles, é possível armazenar o último estado que se encontrava antes de terminar a conexão com um determinado *site* e permitir o retorno a partir daquele estado em que se encontrava em um próximo acesso. Assim, o emprego dos *cookies* permitiu acrescentar estados ao protocolo HTTP durante a navegação do usuário.

Os *cookies* originalmente foram criados com a intenção de serem um simples mecanismo para facilitar a navegação para os usuários no acesso a *sites Web*, sem ter que realizar um longo processo de identificação em toda visita. Dessa forma, é requisitada a

[2] http://www.bugnosis.org/
[3] http://www.safesurf.org/

informação pessoal somente da primeira vez que se acessa um *site*, e os *cookies* armazenam informações para serem usadas em acessos futuros.

Em termos mais específicos, *cookie* é um pequeno fragmento de informação referente a um determinado usuário, sendo gerado através de algum mecanismo específico em certo servidor *Web*. Ele é representado por uma cadeia de caracteres contida na memória do navegador do usuário. O tempo de vida desse valor é maior que o tempo que o usuário gasta para acessar um determinado *site*; dessa forma, esse valor é salvo em um arquivo para uma referência futura.

Os *cookies* são embutidos no fluxo de resposta HTTP da informação HTML e são enviados do servidor *Web* para o usuário. Eles podem ser criados para viabilizar a personalização das informações da *Web*. Um uso comum dos *cookies* é o armazenamento dos itens de listas de compras que um determinado usuário seleciona enquanto navega por uma loja virtual.

Essencialmente, os *cookies* fazem uso de informação específica de um indivíduo. Essa informação é transmitida pelo servidor *Web* ao computador do usuário para que ela possa estar disponível em um acesso posterior por esse mesmo servidor ou por outros do mesmo domínio. Na maioria dos casos, a armazenagem de informação pessoal e também o acesso a ela são feitos sem serem percebidos pelo usuário.

Os servidores *Web* podem automaticamente ganhar acesso a *cookies* relevantes a qualquer momento em que o usuário estabeleça uma conexão com eles, normalmente na forma de requisições *Web.* A implantação de *cookies* para um resgate futuro possibilita a construção de perfis detalhados dos usuários, interesses, hábitos de consumo e estilo de vida. Pode ser preocupante a maneira como o conhecimento das preferências pessoais e atividades privadas de um usuário possa ser usado para determiná-lo como membro de um grupo particular e, com isso, aplicar práticas direcionadas a ele.

Cookies são baseados em processos de dois estágios [36], o que pode ser visualizado na figura 2. No primeiro estágio, é feita a armazenagem do *cookie* com determinadas informações no computador do usuário sem seu consentimento ou conhecimento. Por exemplo, com mecanismos de busca *Web* personalizáveis como *My Yahoo!*, um usuário seleciona categorias de interesse de uma página *Web*. O servidor *Web*, então, cria um *cookie* específico e o transmite para o computador do usuário. O navegador do usuário recebe o *cookie* e o armazena em um arquivo especial chamado lista de *cookies*. Como resultado, as informações armazenadas são as preferências de categorias do usuário, que definem e moldam perfis.

No segundo estágio, o *cookie* é automaticamente transferido do computador do usuário para um servidor *Web*. Quando o usuário faz uma requisição de seu navegador *Web* para visualizar certa página de um determinado servidor, o navegador, sem o conhecimento do usuário, transmite o *cookie* contendo informações pessoais para esse servidor.

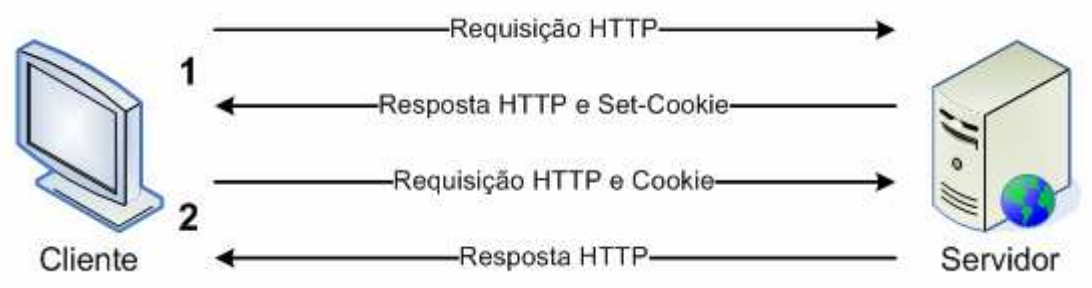

Figura 2. Etapas de criação e recuperação de cookies.

A criação de um *cookie* geralmente envolve a inserção de um cabeçalho de *cookie* nos pacotes de comunicação do protocolo HTTP, de forma que o navegador receba e armazene o par "nome e valor" do *cookie* na memória. Algumas linguagens constróem exatamente um cabeçalho HTTP para ser enviado, enquanto outras usam funções embutidas para ajudar no processo de construção. Uma vez que ele é criado, ele deve fluir facilmente de servidor para o cliente e o inverso via os cabeçalhos HTTP.

De acordo com a sua especificação pela RFC 2965 [3], um *cookie* possui propriedades sintáticas envolvendo pares de atributo e valor. Em uma comunicação comum, o servidor do *site* inicia uma sessão[4] retornando um cabeçalho extra ao cliente pela diretiva *Set-cookie*. O navegador do usuário retorna, junto à requisição HTTP, o cabeçalho do *cookie* requisitado, se ele possui algum que foi anteriormente armazenado no computador do usuário. O servidor, ao receber esse cabeçalho, pode decidir usá-lo ou não para determinar o estado corrente da sessão em que o usuário se encontrava num acesso anterior. Após isso, o servidor pode enviar para o navegador um cabeçalho ou múltiplos cabeçalhos de resposta de *cookies*. O servidor também pode terminar a sessão em que o usuário se encontrava enviando um cabeçalho de *cookie* com *Max-Age* = 0.

Como dito anteriormente, o cabeçalho de resposta de um *cookie* é constituído basicamente por pares de atributo e valor e possui sintaxe de acordo com sua especificação [3], apresentada na figura 3.

[4] Uma sessão é iniciada quando o servidor armazena algum cookie no computador do usuário. Essa sessão existe enquanto não terminar o tempo de vida do cookie.

```
set-cookie      =  "Set-Cookie2:" cookies
cookies         =  1#cookie
cookie          =  NAME "=" VALUE *(";" set-cookie-av)
NAME            =  Attr
VALUE           =  Value
set-cookie-av   =  "Comment" "=" value
                |  "CommentURL" "=" <"> http_URL <">
                |  "Discard"
                |  "Domain" "=" value
                |  "Max-Age" "=" value
                |  "Path" "=" value
                |  "Port" [ "=" <"> portlist <"> ]
                |  "Secure"
                |  "Version" "=" 1*DIGIT
portlist        =  1#portnum
portnum         =  1*DIGIT
```

Figura 3. Gramática da sintaxe do cabeçalho cookie de resposta.

Pela sintaxe, o cabeçalho de resposta pode conter um ou mais *cookies*. Cada *cookie* deve apresentar seu nome (*NAME*) e valor (*VALUE*), que não são considerados um par atributo-valor. Esses dois campos são seguidos por zero ou mais pares atributo-valor. Cada um desses pares é detalhado logo abaixo.

O campo "*NAME=VALUE*" delimita o nome do *cookie* por *NAME* para futuros acessos, o qual irá identificá-lo entre outros num mesmo domínio. *VALUE* contém a informação que o *cookie* vai armazenar. Esse conteúdo é de interesse do *site* que o criou e ele não possui nenhum significado relevante para o navegador que está acessando-o.

O par "*Comment=value*" é usado para informar o usuário sobre as intenções do *site* no uso do *cookie*. O valor dessa informação contém uma *string* em linguagem natural para que o usuário possa visualizar e decidir iniciar ou continuar uma sessão por esse *cookie*.

O par "*CommentURL=http_URL*" é usado para informar o usuário sobre as intenções do *site* de um forma mais completa e detalhada. O usuário pode acessar a informação identificada pela URL para decidir em iniciar ou continuar uma sessão por esse *cookie*.

Os dois pares delimitados podem ser usados para que o usuário tenha um maior controle sobre as informações que o *site* pode estar armazenando em *cookies*; esses podem ser usados para receber ou armazenar informações privadas de um usuário.

O campo "*Discard*" informa para o navegador descartar o *cookie* incondicionalmente quando ele terminar.

O par “*Domain=value*” especifica o domínio para o qual o *cookie* é válido. O par “*Path=value*” especifica o caminho para um subconjunto de endereços do *site* aos quais esse *cookie* se aplica.

O par “*Max-Age=value*” é um valor decimal, inteiro e maior ou igual a zero que se refere ao tempo de vida global do *cookie* em segundos. Se esse tempo de vida possui valor menor que o tempo corrente ou esse seu valor de tempo é igual a zero, ele é imediatamente descartado.

O par “*Port [=”portlist”] ”* restringe a porta a que um *cookie* pode ser retornado em um cabeçalho de requisição do *cookie*.

O campo “*Secure*” direciona o navegador do usuário a somente usar meios seguros para retornar o *cookie* para o seu *site* de origem. Com isso, a confidencialidade e a autenticidade da informação contida no *cookie* são protegidas. Esse campo pode ser considerado como um aviso do *site* ao navegador, que indica a relevância das informações trocadas nessa sessão.

O par “*Version=value*” identifica a versão da especificação que o *cookie* obedece; o seu valor é um número decimal e inteiro.

São obrigatórios somente os campos “*NAME=VALUE*” e “*Version=value*” no cabeçalho do *cookie*. O resto das informações é opcional e pode ser omitida. Na da falta dos seguintes valores opcionais o navegador pode inserir valores padrões.

O comportamento do campo “*Discard*” é relacionado à presença ou ausência de um atributo *Max-Age*.

Para o valor de domínio (*Domain*), o padrão é ser o próprio *host* requisitado pelo navegador; no caso do valor de *Path*, o padrão é o próprio caminho da URL respectiva à resposta do *cookie*.

O atributo *Max-Age* possui comportamento padrão de descartar o *cookie* quando o navegador termina.

No comportamento padrão do campo *Port*, o *cookie* pode ser retornado a qualquer porta de requisição.

O campo *Secure* estando ausente no cabeçalho, o *cookie* pode ser transmitido por um canal inseguro.

O navegador pode rejeitar ou não armazenar o *cookie* caso haja problemas de má formação do cabeçalho. Isso pode ocorrer por não haver o campo *Version*, pelos campos *Domain* e *Path* não corresponderem ao *host* do *site* que está respondendo o cabeçalho do *cookie* ou por o atributo *Port* conter um valor que não corresponda à porta de requisição.

Para retornar o *cookie* ao seu *site* origem, é necessário colocá-lo em um cabeçalho de requisição de *cookie* no envio de uma requisição HTTP pelo navegador para o *site*. A sintaxe desse cabeçalho [3] é apresentada pela figura 4.

```
cookie          =  "Cookie:" cookie-version 1*((";" | ",") cookie-
                   value)
cookie-value    =  NAME "=" VALUE [";" path] [";" domain] [";" port]
cookie-version  =  "$Version" "=" value
NAME            =  Attr
VALUE           =  Value
path            =  Path" "=" value
domain          =  "$Domain" "=" value
port            =  "$Port" [ "=" <"> value <"> ]
```

Figura 4. Gramática da sintaxe do cabeçalho de requisição do cookie.

Os elementos dessa sintaxe são os mesmos elementos do cabeçalho de resposta analisado anteriormente. Nem todos os campos armazenados pelo navegador são retornados ao *site* origem.

Antes de o navegador iniciar uma requisição a um *site*, são selecionados determinados *cookies* dentre todos os que estão armazenados por certos valores obtidos pelo critério a seguir. O atributo de domínio (*Domain*) dos *cookies* deve condizer com o domínio do *site* em que está sendo feita a requisição, e o atributo de caminho (*Path*) dos *cookies* deve condizer com a URL requisitada.

Uma seleção de porta também é feita da seguinte maneira. A porta padrão é utilizada quando o atributo *Port* não existe; o *cookie* pode ser selecionado para ser enviado independentemente da porta de envio. Se o atributo está presente e não contém valor, o *cookie* é selecionado para a porta de requisição. Caso haja algum valor na lista de portas, o *cookie* é selecionado se a porta de requisição estiver nessa lista.

Os *cookies* que possuem *Max-Age*, que denota os seus tempos de vida, são descartados se possuírem um tempo expirado. Desse modo, eles não são enviados para o *site* de origem.

Assim, segundo o tempo de expiração dos *cookies*, eles podem ser classificados em persistentes e de nível de sessão [37]. O *cookie* persistente possui data de expiração prolongada. Ele é utilizado para identificar um usuário. O *cookie* de sessão possui tempo de vida limitado, geralmente não possuindo data de expiração. Dessa forma, ele não é retido no computador do usuário.

No caso de múltiplos *cookies* satisfazerem os critérios acima, eles são ordenados no cabeçalho do *cookie* de modo que a precedência é dos que possuem o caminho mais específico para aqueles são menos específicos.

Para um maior controle de espaço, alguns limites são impostos na implementação. Uma especificação mínima [3] é apresentada para a sua construção no navegador: um número máximo de 300 *cookies* em geral, no máximo, 4096 bytes por *cookie* e, no máximo, 20 *cookies* para um único host ou nome de domínio.

A recuperação de um valor armazenado por um *cookie* é feita pela maioria das linguagens, que lêem esse cabeçalho de *cookie* do protocolo HTTP e o disponibilizam em uma variável ou objeto para ser manipulado. Os *cookies* podem ser lidos no lado do navegador ou no lado do servidor, quando enviados para ele, e o que vai determinar o modo de leitura no lado servidor será a linguagem utilizada.

Os *cookies* podem ser implementados em qualquer linguagem de programação *Web*. Essas linguagens, ao gerarem as páginas dinamicamente, produzem o cabeçalho *cookie*. Java por *Servlets* ou JSP, CGI e PHP são exemplos dessas linguagens.

Ilustraremos a criação e utilização de *cookies* na linguagem PHP [38]. Nessa linguagem, "*$*_COOKIE" é um *array* associativo de variáveis e com escopo global.

Com a função *setcookie(name, value, expire, path, domain, secure)* pode ser criado um novo *cookie*. O parâmetro "*name*" indica o nome para um determinado *cookie* e está relacionado com a variável *"$_COOKIE["name"]"*; *"value"* é o conteúdo a ser armazenado pelo *cookie*; *"expire"* é a data final para expiração do *cookie*; *"path"* é o caminho no servidor no qual o *cookie* estará disponível; *"domain"* é o domínio na Internet onde se encontra o servidor do respectivo *cookie*; e "*secure*" especifica se deve ser transmitido por um canal de transmissão seguro, no caso, por uma conexão HTTPS.

A função deve ser utilizada antes da *tag "<HTML>"*, pois não deve haver nenhuma saída antes de sua chamada.

O código apresentado na figura 5 exemplifica o uso de *cookies* para a linguagem PHP.

No exemplo, cria-se um *cookie* que irá contar o número de vezes que alguém acessa a página. A manipulação do *cookie* pode ser feita de qualquer maneira por quem o constrói. Informações pessoais, como nome de uma pessoa, endereço, senha e outros dados, podem ser armazenadas. Para que isso seja feito, é necessário que o usuário tenha acessado e disponibilizado essas informações anteriormente nesse *site*.

Os *cookies* também podem ser classificados como *first-party* ou *third-party* [34]. Um *first-party cookie* está associado com o *site* que o usuário requisitou. Um *third-party cookie* está associado a uma imagem, a um anúncio, a um frame, ou a qualquer outro conteúdo proveniente de outro domínio, que esteja incorporado ao *site* que o usuário requisitou.

```
<?php
if ( $_COOKIE["TestCookie"] == '' )
      /* primeira vez, o cookie não existe nesse computador */
 {
   setcookie("TestCookie", 0 , time()+3600); /* expira em 1 hora   */
 }
else
{     /* deleta cookie tempo é anterior ao atual */
  setcookie("TestCookie", "", time()-3600); setcookie("TestCookie",
  $_COOKIE["TestCookie"] + 1 , time()+3600);
      /* expira em 1 hora */
}
?>
<HTML>
<HEAD>
TESTE DE COOKIES<br><br>
</HEAD>
<BODY>
<?php
 if ( $_COOKIE["TestCookie"] == 0 )
       echo "Essa página foi visitada nenhuma vez por você!<br/>";
 else if ( $_COOKIE["TestCookie"] == 1 )
       echo "Essa página foi visitada " . $_COOKIE["TestCookie"] .
            " vez por você!<br/>";
      else
       echo "Essa página foi visitada " . $_COOKIE["TestCookie"] .
            " vezes por você!<br/>";

?>
</BODY>
</HTML>
```

Figura 5. Exemplo de um código PHP de uma página que registra um cookie no computador do usuário.

Um dado computador pode conter os *cookies* de páginas que nunca foram acessadas pelo usuário. Isso ocorre devido ao fato de existirem *sites* que são assinantes de algum serviço de personalização de alguma organização. Esses *sites* podem possuir em sua página principal uma requisição de *cookie* para os servidores dessa organização. Por exemplo, o usuário pode encontrar diversos *cookies* do *site* da *DoubleClick* [5] sem tê-lo visitado, mas alguns dos *sites* que foram visitados são assinantes do serviço da *DoubleClick*.

Quando se acessa um *site* semelhante a esse descrito anteriormente, ele requisita o *cookie* armazenado no computador do usuário. Através dessa requisição, ele obtém as informações para saber quem está visitando-o e qualquer outra informação contida no arquivo do *cookie*. Então, é enviada uma requisição para um outro servidor, o qual promove esse serviço de personalização, informações de propagandas a respeito de quem visitou o *site*. Com a resposta dessa requisição o *site* pode promover propagandas personalizadas para um usuário específico.

O *cookie* é um mecanismo inofensivo, mas é a maneira como alguns *sites* encontraram para utilizá-lo que pode causar problemas de invasão de privacidade. Com a preocupação de

[5] http://www.doubleclick.com

privacidade, navegadores são construídos com opções de mostrar um alerta antes de aceitar um *cookie* ou de barrá-lo completamente.

As informações contidas em *cookies* podem delimitar perfis de usuário genéricos ou específicos. Para os *sites* que trabalham com tais informações, os indivíduos são considerados anônimos a não ser que se identifiquem voluntariamente. Entretanto, não há nada que impeça que algum *site* armazene informações pessoais em *cookies*.

Os *cookies* são relacionados a um determinado computador e não a um determinado usuário. Um mesmo computador pode ser utilizado por duas ou mais pessoas. A aplicação de personalização com dados desses *cookies* pode não funcionar da forma mais correta. Porém, existem sistemas operacionais que permitem o uso da identificação de usuário no sistema para o armazenamento das informações de perfil de usuário. Mas o problema persiste se houver mais de um usuário utilizando um mesmo *login* no Sistema.

Assim, para alguns, a presença dos *cookies* e a maneira como alguns *sites* os utilizam pode causar uma invasão de privacidade. Por esse motivo, uma coalizão de advogados de privacidade está trabalhando para mudar esse protocolo [38] e sugerir uma nova proposta para a IETF[6]. Nessa proposta, a persistência de *cookies* é limitada e o usuário possui uma maior escolha de onde e de quais *cookies* permitir. Essa nova especificação será integrada em todos os navegadores.

Essa mesma proposta requer que os navegadores notifiquem a aceitação de *cookies*. Dessa forma, os *cookies* seriam menos transparentes para novos usuários. Esta proposta é apoiada por organizações americanas, como *Center for Media Education*, *Computer Professionals for Social Responsibility*, *Consumer Project on Technology*, *Electronic Frontier Foundation*, e *Electronic Privacy Information Center* (EPIC) [38].

A limitação ou exclusão das requisições a *cookies* de servidores de terceiros faria com que firmas que fazem o serviço de marketing direcionado deixassem de funcionar. Muitos *sites* atualmente usam os serviços dessas companhias para suas propagandas. Exemplos de companhias que fazem tal serviço são: *Doubleclick*, *Focalink*, *Globaltrack* e *ADSmart*. Todas essas companhias usam *cookies* para direcionar anúncios aos usuários.

Além disso, há programas que permitem o controle ou a exclusão de *cookies*. Existem alguns exemplos de softwares para a plataforma Windows, como o *Cookie Pal*, que é um gerenciador de *cookies*. *ZDNet's CookieMaster* é uma ferramenta para monitorar os *cookies*. *Cookie Crusher* é uma ferramenta para rejeitar todo tipo de *cookie*. *Cookie Cutter* PC permite

[6] A IETF (Internet Engineering Task Force) é uma organização sem fins lucrativos com milhares de membros e atualmente possui grande influência nas decisões que irão guiar o futuro da Web, estando ativa desde outubro de 1996.

apagar qualquer *cookie* que seja indesejado pelo usuário. *Cookie Cruncher* gerencia os *cookies* e permite visualizar seus conteúdos. *IEClean* é um gerenciador de *cookies* para o *Internet Explorer*. *NSClean* gerencia os *cookies*. *HistoryKill* remove automaticamente os *cookies* do computador.

Os *cookies* não são capazes de danificar dados e de tornar inconsistente um computador. Os *cookies* são armazenados como arquivos texto. Eles não são executáveis, e, mesmo que fossem, seria necessário que alguém os acessasse. Dessa forma, eles não podem ser ou conter vírus ou cavalos de tróia que poderiam causar algum dano.

Assim, espera-se qualquer uso dos *cookies*, até mesmo o prejuízo à privacidade dos usuários através de um uso malicioso, como a identificação do perfil do usuário. Nesse caso, torna-se preocupante a presença de *cookies*, pelo uso malicioso que pode ser dado a eles.

A principal preocupação é a execução dessas tarefas sem o conhecimento do usuário. Para alguns usuários, a obtenção de informação pode aparentar invasão de privacidade. Mas o uso desse tipo de informação é inofensivo para aqueles que estão cientes das limitações da rede, do receptor das informações e do propósito da coleta de dados. Por outro lado, ninguém tem o direito de coletar informações sobre um indivíduo e fazer uso delas sem o seu conhecimento ou consentimento; dessa forma, pode ser violado o seu direito à privacidade.

3.3 Clickstream

O termo *clickstream*, também conhecido como *clickpaths*, denota o caminho ou rota que um visitante realiza através de um ou mais *sites* na *Web* [4]. O trajeto reflete uma série de escolhas feitas dentro de um *site*. Esse caminho é uma lista de todas as páginas vistas por um visitante; elas são apresentadas na ordem em que foram vistas. O caminho pode ser também definido como uma sucessão de cliques de mouse que cada visitante faz.

Os dados de *clickstream* é um produto natural da navegação na *Web*, que é gerado automaticamente sem ser necessária a interação com o usuário. Além disso, essa interação poderia alterar seu comportamento de navegação.

O motivo primordial para a coleta de dados de *clickstream* é a monitoração da janela do navegador do usuário e o registro de quando e o que está em foco nessa janela pela identificação de qual URL está atualmente sendo requisitada.

Como estatísticas agregadas, os dados de *clickstream* dizem, em média, quanto tempo as pessoas gastam no *site*, quão frequentemente elas retornam e quais páginas são as mais

freqüentemente visitadas. Para lojas *on-line* que pretendem atingir visitantes de suas páginas, essa informação de *clickstream* é muito mais rica que as informações convencionais por apresentar características demográficas e de perfil do usuário.

Além disso, dados de caminho percorrido por um usuário na *Web* podem conter informações sobre seu objetivo, conhecimento e interesses. Segundo essas informações, pela perspectiva de marketing, há grande vantagem em minerar esses dados para melhorar o entendimento e predição do comportamento de escolha do usuário, até mesmo em predizer uma compra.

Há também um amplo interesse da parte do marketing na grande interatividade, presente na *Web*, entre *sites* e usuários. A interatividade se restringe à monitoração e resposta das ações de cada consumidor. Com os dados de *clickstream*, é possível construir sistemas de propaganda de forma dinâmica, enquanto o usuário interage com o sistema.

Isso é possível, pois, segundo Montgomery [4], as escolhas de navegação do usuário em um *site* envolvem o número de páginas vistas, o tempo gasto na visualização de uma página ou o *site* por completo, a decisão de ficar ou sair do *site* em uma dada página e a escolha de qual *link* seguir e/ou quais páginas ver. Essas decisões podem refletir na atratividade do *site Web* e podem também influenciar a habilidade do *site* em aumentar o rendimento através de propagandas com *banners* ou outros veículos relacionados (*pop-ups*, *pop-unders*, etc).

Muitos visitantes não têm em mente um objetivo de compra quando acessam um *site*. Para ajudar esse tipo de usuário, dados de *clickstream* podem ser usados a fim de inferir individualmente o seu objetivo. Tal uso atribuído a esses dados é de grande importância para esse tipo de usuário.

Essa inferência de objetivos é realizada por sistemas de recomendação [4]. A descoberta de objetivos fornece ao usuário orientação sobre quais produtos selecionar através do uso de uma estrutura de preferências do usuário individualmente e das preferências e escolhas de outros visitantes do *site*. Conseqüentemente, é possível definir o aspecto da navegação que é mais próxima de uma compra *on-line*. A *DoubleClick* pode ser capaz de usar o histórico de visita do *site* de um indivíduo a fim de predizer com precisão o gênero do usuário e o que é útil para selecionar um material de propaganda direcionado a futuras visitas a esses *sites Web*.

Esse novo tipo de dados de *clickstream* é vasto em tamanho e potencialmente muito complexo. Métodos estatísticos e de mineração de dados são necessários para tratar esse novo tipo de dados.

Alguns termos para descrever a navegação na *Web* são essenciais para análise de *clickstream* e para sua compreensão: requisição de página, visão de página e sessão [39]. Uma requisição de página refere-se a um usuário que está requisitando uma URL através de seu programa de navegação. Requisição de página é uma marcação de requisição no arquivo de *log* do servidor. Um usuário pode pressionar o botão de voltar em sua janela do navegador para rever uma página. Essa gerará uma nova visão de página, mas não uma nova requisição de página. Dessa forma, o programa de navegação retornará uma cópia da página previamente armazenada. Nesse caso, uma sessão é definida como um período de navegação *Web* assistida pelo usuário ou uma seqüência de visões de página. Se um usuário não tem visto qualquer página por 20 minutos, assume-se que a sessão tenha terminado.

Dados de *clickstream* sem refinamento são coletados das seguintes maneiras: arquivos de *log* de servidores, dados de painel e provedor de serviço de Internet.

A mais popular é através de arquivos de *log* de servidores dos *sites* que estão sendo visitados. Eles identificam informações como endereço IP, última URL e tipo do navegador, informações que são registradas nos *logs* do servidor [40]. Esses arquivos são mantidos em benefício de um proprietário de *site Web* e contêm todas as requisições e informações transferidas entre o computador do usuário e o servidor durante uma visita a um *site Web*. Isso ocorre porque os arquivos de *log* de servidor são capazes de gravar informação no identificador do *cookie* do visitante, o que possibilita identificar usuários individualmente e suas respectivas visitas de retorno.

Os dados de painel são fornecidos por ferramentas como *ComScore*, *NetRatings* e *MediaMetrix*. Eles são relatórios de acesso de usuários a determinado serviço *Web*. Essas ferramentas capturam as URL’s de todas as páginas requisitadas durante a navegação na *Web* e transmitem informações do computador do usuário para esses fornecedores de dados de painel.

Os dados de *clickstream* podem também ser coletados por um Provedor de Serviço de Internet (ISP). O provedor de Internet nem sempre encaminha a requisição; ele pode satisfazer a requisição usando um *cache* armazenado localmente, como um meio de diminuir o tráfego na Internet. Como o ISP processa todas essas requisições, ele pode também registrá-las para capturar o *clickstream* do usuário. Se a requisição é passada para o servidor destino, esse servidor pode também registrar o computador que gerou a requisição. Isso fornece dois

repositórios de dados de *clickstream*, mas a gerada pelo provedor é uma fonte mais completa de dados.

Dados de painel de Internet não são tão ricos quanto os arquivos de *log* de servidores, que fornecem um registro de todas as informações requisitadas do servidor para uma dada visão de página. Dados de painel de Internet somente registram a URL do *site* visitado, o que torna difícil reconstruir o que o usuário atualmente viu quando estava visitando a página *Web* ou dificulta o entendimento da interação que ocorre entre as páginas de um *site*.

As fontes de dados de painel ou de ISP podem também ser capazes de associar informação demográfica detalhada do local da máquina com cada indivíduo. Mas, por outro lado, faltam algumas informações específicas sobre a interação entre usuário e o servidor, as quais são coletadas por *log* de servidores *Web*. Ambas as fontes tipicamente incluem informação do endereço IP do visitante, o tipo de navegador usado, um *timestamp*, e a URL visitada anteriormente. Embora muita informação de potencial interesse não é capturada por essas origens, dados de *clickstream* fornecem um nível razoável de detalhamento nas informações coletadas.

Há uma abundância de informação a ser analisada. Também é possível serem examinados os *clickstreams* do visitante em combinação com qualquer informação fornecida por um programa estatístico, como duração de visitas, termos de busca, ISP's, países dos visitantes, navegadores, etc. O processo apresenta um relatório sobre o público que acessa um determinado serviço.

Esses programas de análise geralmente possuem filtros e rotinas para classificar o perfil ou definir um comportamento de cada visitante, usando conceitos de inteligência artificial e as informações coletadas pelos métodos anteriormente descritos.

3.4 Data Mining

A Mineração de Dados (*Data Mining*) é um conjunto de técnicas automatizadas usadas para extrair ou minerar conhecimento previamente desconhecido ou oculto de grandes quantidades de dados em algum tipo de banco de dados. O significado para esse processo pode receber outras terminologias; a que mais a representa é a descoberta de conhecimento em banco de dados [41]. No contexto deste trabalho, a mineração de dados é empregada para quatro propósitos principais: melhorar a aquisição e retenção do consumidor, reduzir fraude,

identificar ineficiências internas, consertar operações e mapear áreas inexploradas da Internet [7].

Por definição, a mineração de dados produz informação que o usuário não conhecia ou da qual possuía somente hipóteses [42]. Utilizam-se abordagens baseadas em descoberta nas quais o casamento de padrões e outros algoritmos são usados para descobrir relacionamentos chave nos dados. O uso do mecanismo para mineração de informação traz grandes benefícios à extração de conhecimento, pois a assimilação de tais dados não é sempre intuitiva. Com uma mineração bem sucedida, é possível encontrar padrões e relacionamentos e, então, usar essa informação para realizar decisões de negócios de conhecimento dirigido [7].

A introdução da mineração de dados ocasionou uma mudança da manufatura, publicidade e marketing em massa, que possuem suas origens na revolução industrial, para manufatura, publicidade e marketing personalizados e direcionados a seguimentos específicos da população [7].

Com o aumento da globalização econômica e evolução da tecnologia da informação, dados financeiros são gerados e acumulados a uma taxa sem precedentes. Dessa forma, a necessidade iminente por tornar tais dados em informação e conhecimentos úteis atraiu a atenção na indústria de informação nos anos recentes. Claramente, há benefícios comerciais potenciais do emprego de mineração de dados. Tais benefícios podem ser exemplificados por Zhang [43], que discute a importância de mineração de dados para aplicações financeiras especificas e compara as diferentes técnicas de mineração a partir das perspectivas técnicas e de aplicação.

A descoberta de padrões de dados escondidos contribui para estratégias de negócios, para bases de conhecimento, para pesquisa médica e científica e para predizer tendências e comportamentos futuros em mercados financeiros. Ela cria oportunidades para companhias realizarem decisões pró-ativas e dirigidas a conhecimento para ganhar uma vantagem competitiva [43].

Em princípio, a mineração de dados deveria ser aplicada a qualquer tipo de repositório de informação. Isso inclui bancos de dados relacionais, armazéns de dados, bancos de dados transacionais, sistemas de banco de dados avançado, arquivos simplórios e a Internet [41]. Os desafios e técnicas de mineração são diferentes em cada sistema de repositórios.

A mineração de dados é aplicada nos mais diversos ramos. No setor bancário, é utilizada para identificar clientes fiéis, comportamento de uso de cartão de crédito e correlações escondidas entre diferentes indicadores financeiros. Na área de Marketing e comércio, ela é aplicada para se determinar a aceitação de um novo produto, padrões de

comportamento dos consumidores, probabilidade de compra de determinado produto e abrangência das campanhas de marketing. No setor de planos de saúde e de seguros, ela é usada para identificar clientes de risco e de comportamento fraudulento. No campo da medicina, ela é utilizada para inferir a probabilidade de um paciente contrair certa doença, analisar prontuários, identificar terapias para doenças e realizar seqüenciamento genético. No ambiente *Web*, é usada para determinar perfis de usuário.

Devido às necessidades da indústria de informação, uma arquitetura de banco de dados emergiu recentemente, o chamado armazém de dados (*Data Warehouse*). Ele é um repositório de múltiplas origens de dados heterogêneos organizado sob um esquema unificado em uma localização central para facilitar a realização de decisão de gerenciamento. Esse repositório realiza armazenagem e entrega de quantidades massivas de dados. A vantagem do *Data Warehousing* é a consolidação e o gerenciamento de informações de bancos de dados díspares por um único banco [7].

Assim, quando dados sobre processos de organizações se tornam prontamente disponíveis e de fácil acesso, se torna economicamente viável realizar a mineração neles [7]. Para analisar esses bancos de dados, é preciso realizar uma análise em profundidade, como classificação de dados, agrupamento e caracterização de mudanças de dados ao longo do tempo [41].

Na realidade, o conceito “mineração de dados” atribuído a esses tipos de sistemas pode ser mais bem expresso como “descoberta de conhecimento”, sendo que a mineração de dados é somente uma tarefa dentro de todo o processo para se descobrir informações úteis dentro de um conjunto de dados.

O processo de descoberta de conhecimento consiste em uma seqüência iterativa de passos efetuados nos dados. Os passos executados são: a limpeza, que corresponde à remoção de ruído e inconsistências; a integração, que consiste na combinação de múltiplas origens; a seleção, que recupera informações do banco de dados relevantes para a tarefa de análise; a transformação, que consolida os dados em formas apropriadas para mineração por efetuar operações de sumarização ou agregação; a mineração, que consiste no processo de aplicação de métodos para extração de padrões de dados; a avaliação de padrão, que corresponde à identificação dos padrões que realmente representam conhecimento baseado em algumas medições de importância; e a apresentação do conhecimento, que consiste em técnicas de visualização que são usadas para apresentar conhecimento minerado para o usuário.

A arquitetura simplificada de um sistema de descoberta de conhecimento pode possuir os seguintes componentes: banco de dados e servidor de banco de dados; base de

conhecimento, que contém o domínio de conhecimento; máquina de mineração de dados; módulo de avaliação de padrão e interface gráfica do usuário. A classificação desses sistemas pode ser realizada de acordo com os tipos de bancos de dados minerados, de conhecimento minerado, de técnicas utilizadas e de aplicações adaptadas [41].

A máquina de mineração de dados pode utilizar diversas técnicas, com abordagens diferentes, para encontrar padrões no repositório de dados. Muitas dessas técnicas foram originadas nas pesquisas de inteligência artificial dos anos 80 e 90 [7]. Através delas, os sistemas de mineração de dados são capazes de descobrir padrões em várias granularidades e em diferentes níveis de abstração. Essas técnicas são utilizadas para descrever características ou predizer eventos e comportamentos.

As técnicas utilizadas pela mineração de dados são: a caracterização de dados é o resumo das características gerais ou aspectos de uma classe alvo de dados; a discriminação de dados é uma comparação dos aspectos gerais de objetos de dados da classe alvo com aspectos gerais de objetos de uma classe ou de um conjunto de classes contrastantes; a análise de associação é o descobrimento de regras de associação mostrando condições atributo-valor que ocorrem freqüentemente em um certo conjunto de dados; a classificação é o processo de busca de um conjunto de modelos ou funções que descrevem e distinguem classes ou conceitos de dados; a análise de agrupamento (*clustering*) agrupa os objetos em conjuntos de afinidade, baseando-se no princípio de maximizar as similaridades internas das classes e minimizar as similaridades entre as classes [41]; a análise marginal ou de exceções é um processo de busca de objetos que não consentem com o comportamento ou modelo geral dos dados, ao passo que testes estatísticos auxiliam na técnica para se encontrar comportamentos gerais; e a análise de evolução descreve e modela regularidades ou tendências para objetos que possuem comportamento que muda ao longo do tempo.

A classificação é provavelmente a mais comum atividade na atualidade, ela é capaz de usar o modelo para predizer a classe de objetos cujo rótulo é desconhecido. Essa técnica auxilia a descobrir as características dos consumidores e fornecer um modelo para predizer quem eles são.

As técnicas de mineração utilizam ferramentas computacionais como redes neurais, algoritmos genéticos, inferências estatísticas, indução de regras e visualização de dados. Essas técnicas geralmente produzem cinco tipos de informação: associações, seqüências, classificações, agrupamentos e previsão, que levam a aplicações que podem gerar saídas, como características demográficas de consumidores, predições, padrões de uso fraudulento de

cartão de crédito, identificação de lealdade de consumidores, padrões de comportamento de consumidores (definições de perfis de usuário) e outros [7].

4 Privacidade

Privacidade não possui uma definição bem delimitada e objetiva, já que é fortemente relacionada com informações pessoais e íntimas de um indivíduo. Essa subjetividade em sua definição causa a aparição de diversas outras, pois cada um interpreta privacidade de acordo com as características que mais lhe convém. Fernandes [44] define que o significado de privacidade para uma pessoa pode diferir por completo do significado de privacidade para outra, mesmo dentro de um mesmo grupo étnico-cultural.

Dessa forma, existem inúmeras definições para privacidade, sendo cada definição pertencente a um determinado contexto mais geral ou específico. Por exemplo, apresentam-se duas definições para privacidade. Uma delas, apresentada em dicionários, é o estado de estar só, que é um conceito muito genérico e pode causar inúmeras controvérsias quando aplicado para uma área mais específica. Outro mostra que, para ter privacidade, uma pessoa precisa ter controle sobre as informações existentes sobre si mesma e exercer este controle de forma consistente com seus interesses e valores pessoais [45] [46], o que é um conceito mais específico para o contexto que abrange um ambiente regido pelas informações.

Atualmente, as informações têm grande importância em todas as áreas. No caso, para as pessoas estarem inseridas na sociedade, elas precisam ter um nome, um endereço e outras informações pessoais que contribuam para identificar o ser humano/pessoa com sua presença física. Desse modo, para um indivíduo ser reconhecido e provar sua existência para os gestores governamentais, ele deve possuir documentos com números de identificação presentes no sistema de identificação do governo. Da mesma maneira, e principalmente no campo virtual, que não deixa de ser uma representação do físico, a existência de um indivíduo é mantida somente por suas informações.

Porém, mesmo a privacidade sendo delimitada por uma definição específica, as pessoas divergem sobre a maneira como a aplicam. Por ser subjetiva, ela não pode ser moldada por um padrão, pois é posta em prática de maneira individual para atender necessidades individuais.

O anonimato pode ser considerado uma ação extrema e segura para se manter a privacidade. Sem divulgação alguma de informação, o controle sobre as informações é completo, a não ser que elas sejam obtidas através de algum ataque malicioso. Porém, ao se

aplicar o anonimato, uma pessoa pode deixar de acessar serviços e de se comunicar, já que algum dado deve ser divulgado para uma pessoa ser identificada, e serviços requerem informações para poderem ser oferecidos.

O anonimato é vantajoso quando realmente alguém não quer ser identificado de maneira alguma para a prática de ações sigilosas. Por exemplo, para ver e analisar algum produto em uma loja, uma pessoa não precisa divulgar nenhuma informação pessoal. Entretanto, para se enviar uma carta pelos correios, certas informações devem ser adicionadas ao envelope para que o envio da correspondência possa ser realizado efetivamente e para que o destinatário reconheça o remetente da carta.

Assim, certas informações devem ser disponibilizadas para que uma pessoa participe de meios sociais e acesse certos serviços. Para isso, é necessário analisar a importância de cada informação pessoal divulgada e a segurança que o receptor pode oferecer. Dessa maneira, o nome de alguém pode ser divulgado abertamente, enquanto a senha de uma conta em um banco somente será fornecida com a autenticação da organização receptora e com segurança nos meios de transmissão de informações. A grande dificuldade para este caso é saber quais informações são realmente necessárias para serem disponibilizadas e confiar em quem vai recebê-las.

Portanto, a relevância de privacidade é proporcional à importância e à quantidade de informações relacionadas. Esse grau de importância é subjetivo, dependente de cada indivíduo. Um outro fator relevante são as ações tomadas por quem pode receber as informações pessoais, a maneira como ele obtém as informações e o que faz com elas.

No campo eletrônico, a confiança se torna muito mais difícil de ser proporcionada. Devido à imersão das pessoas no contexto físico, elas estão habituadas a confiar mais nos relacionamentos desse meio. O discernimento de confiança baseia-se em um longo histórico de aprendizagem durante a vida de um indivíduo. Quanto ao mundo virtual, essas mesmas pessoas não possuem ou não conhecem mecanismos para assegurar confiança de suas informações por estarem inseridas em um ambiente novo e desconhecido, o qual pode estar prejudicando-as de alguma maneira diferente da usual. Na *Web*, o fato de as pessoas não saberem ao certo o motivo da coleta e a quantidade de seus dados que são coletados representa um grande risco à privacidade dos usuários que utilizam seus serviços [47].

Da mesma forma, essa desconfiança pode ser fortalecida através do reconhecimento de que o mundo virtual não possui um nível de segurança razoável e estável. Esse reconhecimento é feito através dos fatos e notícias sobre ocorrências de falha de segurança e de privacidade que se dão diariamente no meio eletrônico. A privacidade envolve a maneira

como alguns tratam as informações pessoais, como o uso delas para propagandas e marketing indesejados ou com vistas à divulgação para terceiros. Assim, a confiança é prejudicada, pois ela baseia-se na segurança fornecida.

Um indivíduo sempre tem necessidade de privacidade, independentemente do meio onde ele está inserido, a não ser que ele a desconheça ou não se importe com ela. A privacidade, quando necessária, pode se apresentar em diferentes níveis. Esses níveis variam de acordo com os indivíduos e com as situações.

Como dito anteriormente, a privacidade encontra uma barreira para a sua existência no mundo virtual, na facilidade da transmissão da informação. Quando conectado à rede mundial, um computador pode estar vulnerável a todo tipo de ataque externo. As informações contidas nesse computador podem ser acessadas e divulgadas para o resto da rede.

A privacidade é um termo abrangente que envolve e utiliza a segurança da informação. Nesse caso, essa segurança é necessária para os computadores e para os meios de transmissão de informação. No mundo virtual, a segurança relaciona-se estritamente com os dados nele armazenados. Portanto, a segurança é utilizada para garantir nesse meio certos aspectos que são apresentados no ambiente físico, como confidencialidade, autenticação, integridade, não-repúdio, controle de acesso e disponibilidade [48].

Entretanto, é difícil fornecer esses mesmos aspectos de segurança no mundo virtual devido à flexibilidade com que as informações podem ser manipuladas. A fim de se promover segurança para a informação nesse meio, existe um desenvolvimento contínuo de mecanismos e ferramentas, como *Firewalls* [7], antivírus, *IpSec* [8], VPN [9] (*Virtual Private Network*) e outros. Entretanto, todos esses mecanismos não oferecem segurança completa.

A única maneira de se estar livre dos problemas de segurança e, conseqüentemente, de invasão de privacidade é estar desconectado da rede. Com isso, não há maneira de ocorrer qualquer tentativa de ataque ao computador ou à transmissão de dados vindos da rede. Mas um ataque pode acontecer através do acesso físico de algum estranho ao computador. Desse modo, a segurança total é muito difícil de ser alcançada; o que ocorre normalmente é a obtenção de um determinado nível de segurança.

A invasão de privacidade pode ocorrer no meio de comunicação e nas partes comunicantes. A confidencialidade da comunicação é perdida pela observação por terceiros

[7] Firewall é um regulador do tráfego de informação entre redes distintas; ele impede a transmissão de dados nocivos ou não autorizados entre elas.
[8] IpSec é um padrão seguro de comunicação, utilizado no tráfego de redes ou de pacotes de comunicação nas redes. Esse padrão é de extrema utilidade para a conexão segura com redes confidenciais
[9] VPN é uma rede privada virtual. Ela é construída sobre uma rede de comunicação pública e utiliza criptografia para tornar o tráfego de dados privados.

do tráfego de dados que transcorre pela rede. Criptografia é vastamente utilizada para manter a privacidade das informações nessa comunicação.

O computador do usuário pode ser invadido por algum ataque malicioso. Esse ataque malicioso permite que terceiros acessem informações confidenciais dos usuários. Ele pode ser evitado através do uso de antivírus, de *firewalls* e outros mecanismos que impedem o acesso indesejado aos dados do usuário. No caso da navegação *Web*, o servidor do *site* pode utilizar as mesmas técnicas que o usuário para evitar que as informações nele contidas sejam acessadas ou comprometidas.

Entretanto, mesmo com segurança das informações, a privacidade do usuário pode ser prejudicada. *Sites Web* precisam obter certas informações pessoais para possibilitar a oferta de determinados serviços. A privacidade pode ser comprometida com relação às atitudes dos *sites*. A maneira como eles obtêm as informações e como eles as manipulam quando as têm em sua posse pode caracterizar uma invasão da privacidade do usuário.

A coleta de dados na *Web* pode ser realizada de forma implícita ou explícita [29]. No primeiro caso, a obtenção sucede sem a ciência ou o consentimento do usuário. A falta de ciência de um indivíduo do que ocorre com seus dados pode levar a uma perda de controle de suas informações. Para realizar essa coleta, os *sites* utilizam basicamente *cookies* e observação de requisições de páginas, conhecida como *clickstream*. Essa observação da navegação captura dados de controle, os quais são relacionados aos protocolos de comunicação utilizados. Assim, através da análise dos dados coletados, é possível delimitar um perfil para aqueles que freqüentam um *site*.

Na coleta explícita, é necessário que o usuário exponha suas informações de forma evidente. Os dados coletados são os de conteúdo; eles são relacionados às informações pessoais de um indivíduo, as quais não podem ser obtidas através de um mecanismo automático. Nesse tipo de obtenção de dados, o visitante está ciente da existência dele e tem a opção de consenti-lo. Dessa forma, nesse processo de coleta é mais difícil de ocorrer algum tipo de agressão da privacidade [49]. Porém, uma invasão pode suceder em razão da atitude do *site* no destino dado às informações coletadas, no método de armazenagem, na segurança utilizada, na divulgação para terceiros, na promoção de marketing e em outros.

No contexto do tema desse trabalho, as informações pessoais são consideradas como todas as informações pertencentes a uma pessoa que identificam sua identidade real com a adição de novos dados de identificação dessa pessoa no ambiente eletrônico. Esses dados podem ser informações técnicas como o IP do computador do usuário e as informações de navegação do usuário obtidas através da análise dos acessos que ele fez nos *sites* e outras.

A coleta de dados do usuário pode auxiliar no marketing de *sites* de *e-commerce*, como também pode atrapalhá-lo. Esses dados coletados são necessários para a implementação da personalização. Porém, a maneira como é realizado o processo de coleta e de análise dos dados dos usuários pode apresentar uma invasão de privacidade.

Essa falta de privacidade leva a uma perda de confiança do usuário, que deixa de acessar certos serviços por temer que suas informações pessoais sejam divulgadas ou tenham um uso indesejado. Constata-se em pesquisa [50] que 64% dos usuários da *Web* deixaram de acessar alguma vez um *site Web*, ou não compraram algo por não saberem como a suas informações seriam utilizadas. Jutla afirma que 53% dos usuários não confiam em *sites Web* comerciais de coleta de dados, 66% não se registram em *sites on-line* temendo que suas informações sejam usadas inapropriadamente e 40% falsificam dados quando se registram *on-line* [51].

Além disso, a privacidade é tida como intrinsecamente relacionada com o controle que um indivíduo possui sobre determinada informação [45]. Desse modo, ela deve ser inerente em transações confiáveis, de outra maneira uma falta de privacidade contribuirá para causar uma falha do modelo de negócio do comércio eletrônico. Vale ressaltar que o usuário é uma peça importante nos negócios, já que ele é o maior envolvido em termos de adoção difundida de qualquer modelo de negócio de privacidade eletrônica.

O aumento da percepção de controle do usuário sobre seus dados reflete no crescimento da adoção de serviços e produtos baseados na *Web*. Esse aumento da percepção do usuário é proporcionado pela explicação do uso que é feito dos dados coletados ou pela divulgação de informação sobre o recebimento de algo de valor em troca. Jutla [51] reporta que 51% dos usuários da *Web* desejariam divulgar dados pessoais para receber algo de valor, e Teltzrow [50] mostra que 90% dos usuários querem que se peça permissão antes que suas informações sejam usadas.

Portanto, o desenvolvimento ou expansão do *e-commerce* depende muito da segurança de informação que pode ser oferecida e da confiabilidade que os *sites* podem adquirir através da maneira como atuam. O recebimento de informações pessoais dos usuários é essencial para a existência do comércio eletrônico. Dessa forma, é necessário procurar uma forma de promover segurança e confiabilidade. Sem isso, pode haver uma perda no número de usuários que se dispõem a fornecer suas informações pessoais em troca de serviços dos *sites* de *e-commerce*.

A privacidade visa proteger as informações do usuário contra alguma forma de obtenção indevida delas, e a personalização procura se abastecer dessas informações para se

tornar mais característica ao usuário. Essa situação evidencia o problema da coexistência de privacidade e personalização. As duas funcionam de maneira antagônicas. Enquanto que para a aplicação de personalização procura-se coletar uma maior quantidade de informações, para a privacidade ser mantida é preciso que o mínimo de informação seja divulgado para não ocorrer uma perda de controle sobre ela.

O cuidado com a privacidade do usuário é pouco observado nos sistemas *Web*. Apesar de existirem muitos *sites* que promovem serviços personalizados, apenas um número limitado deles fornece alguma privacidade para os seus usuários através de políticas de privacidade. São aqueles que realmente precisam que seja promovida alguma segurança das informações pessoais, como os *sites* de bancos que necessitam da confiança de seus clientes e manipulam informações de grande importância.

Além disso, segundo Fernando [44], mesmo havendo políticas de privacidade declaradas explicitamente, os usuários deixam de atentar para elas ou não as compreendem.

Uma pesquisa publicada por Joseph Turow, no relatório "*Americans and Online Privacy: The System is Broken*" mostrou alguns dados curiosos:

> Concluiu-se que os norte-americanos compreendem mal a finalidade das políticas de privacidade, embora a maioria dos entrevistados possuam grau elevado de escolaridade. Mesmo aqueles que têm consciência que sua navegação está sendo rastreada, e que informações pessoais estão sendo fornecidas, não se preocupam como essas informações podem vir a ser utilizadas. Na verdade, quando informados que os sites costumam guardar informações sobre seus clientes, dizem que isto é um fato inaceitável.
>
> - 57% dos entrevistados acreditam incorretamente que quando um site possui uma política de privacidade, este site não irá compartilhar as informações dos usuários com outros sites e organizações.
> - Embora 47% dos norte-americanos afirmam que as políticas de privacidade são de fácil compreensão, 67% desses 47% também acreditam (erroneamente) que sites com políticas de privacidade não irão compartilhar os dados dos usuários.
> - 85% dos adultos norte-americanos que acessam a internet de casa não concordam que seus dados sejam colhidos pelos sites, nem mesmo os que oferecem serviços pagos. 54% dos pesquisados responderam que preferiam pagar para continuar com acesso anônimo ou até obter a informação em outro lugar, fora da Web.
> - Dentre esses 85% que disseram não concordar com as políticas, mais da metade confidenciou já ter informado em sites pagos seus nomes e endereços eletrônicos reais.
> - Embora toda essa preocupação com a privacidade on-line, 64% deles afirmaram nunca ter procurado informações sobre como proteger suas informações na rede. Apenas 9% dos entrevistados disseram saber como evitar que os sites coletem suas informações pessoais.
> - 86% dos entrevistados acreditam que leis que obriguem as políticas de privacidade dos sites a possuírem um formato padrão ajudarão os usuários a se protegerem melhor contra essa coleta de informações pessoais.

Deve-se harmonizar a personalização com a privacidade, viabilizar obtenção de informações do usuário sem infringir sua privacidade [52]. Entretanto, encontrar essa

harmonia é difícil por ser a noção de privacidade subjetiva: cada pessoa possui seu critério de privacidade ou nível de privacidade. Conseqüentemente, a coleta de informações feita pelos *sites* é guiada de acordo com uma visão subjetiva, pois cada indivíduo permite as divulgações de informações que melhor lhe parecem.

Portanto, a coleta de dados para a aplicação de personalização deve ser controlada e regulamentada, e seguir princípios que visam a manutenção de privacidade. Para isso, políticas de privacidade podem ser utilizadas para aumentar a ciência do usuário e seu consentimento. Para evitar abusos, mecanismos de navegação anônima apresentam segurança das informações com relação à coleta implícita de dados, e leis de proteção de privacidade e certificados de privacidade oferecem garantias de privacidade através da regulamentação da coleta de informação.

5 Camadas de proteção de privacidade

O conceito de camadas de proteção de privacidade foi introduzido por Ishitani [53] [6]. Foi definido que o uso do termo “camadas” é o mais correto para esse contexto de privacidade. Para que não restrinja a existência de uma determinada camada, não é necessária a presença das respectivas camadas inferiores. Como a privacidade é um termo subjetivo, o uso do conceito de proteção de privacidade permite uma divisão em diferentes camadas para cada método diferente de proteção.

O uso da divisão em camadas é justificado pela presença de diversas abordagens para proteção de privacidade. Cada uma dessas abordagens é classificada em uma camada. Essa divisão diferencia e classifica os diversos mecanismos de proteção de privacidade. Uma determinada abordagem pode ser relacionada a ferramentas que possuem uma determinada lógica. Uma ferramenta pode abordar de diferentes maneiras a proteção de privacidade e ela pode se enquadrar em mais de uma camada.

Segundo Ishitani, a proteção de privacidade pode ser dividida em seis camadas que possuem responsabilidade ou controle transitório entre o usuário e a sociedade, o que pode ser visualizado na figura 6 [53]. Os mecanismos nas camadas iniciais estão presentes no computador do usuário; nas camadas subseqüentes, eles se apresentam em *proxies*, depois, em cada *site* e, finalmente, no controle de toda a sociedade regida por leis criadas pelos governos.

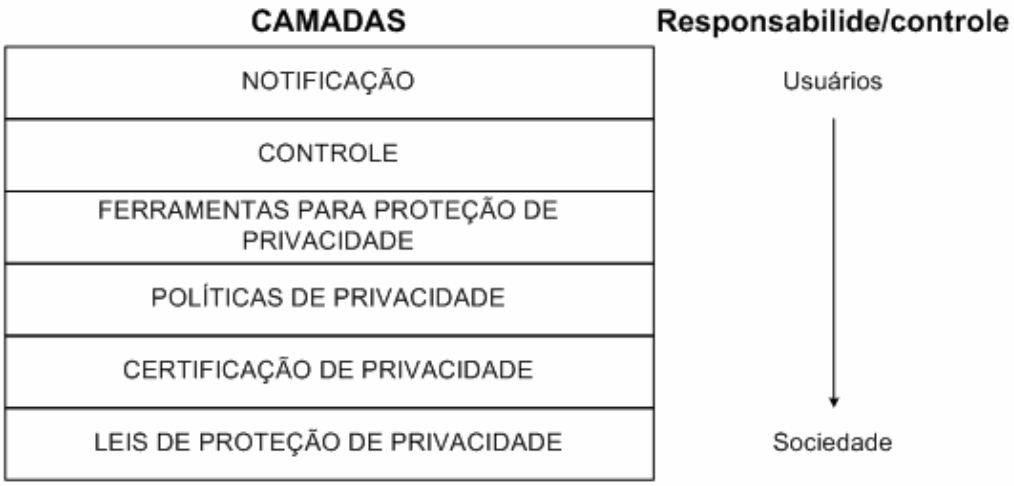

Figura 6. Camadas de proteção de privacidade.

Por essa linha de transitoriedade é apresentada cada uma das camadas de proteção de privacidade.

A primeira é a camada de notificação. A maioria dos usuários que navega pela *Web* é leiga em noções de infra-estrutura e segurança na rede de comunicação. Dessa forma, eles não estão cientes da existência da invasão de privacidade e muito menos conhecem o tipo de informação que pode ser derivada a partir de sua interação com um *site* [53]. Por isso, essa camada de proteção é necessária. Ela tem o papel de estar informando os usuários sobre as situações de risco durante sua navegação pela *Web*. Como exemplo desses tipos de mecanismos, há o *Privacy Critics* [54].

A segunda é a camada de controle. Essa camada inclui mecanismos que permitem os usuários controlarem melhor suas informações. Ela evita mecanismos não identificáveis de obtenção de dados, os quais são tentativas explícitas de violação de privacidade. Como exemplo desses tipos de ferramentas, há os *cookies* de terceiros e *Web bugs* [6]. Entretanto, deve-se ressaltar a necessidade de ferramentas da camada de notificação, que podem informar os usuários sobre a necessidade de alguns *cookies* e a presença de outros que são maliciosos.

A tecnologia chave nessa camada é o navegador *Web* ou *plugins* adicionados a ele, que filtram o fluxo de informações da navegação do usuário buscando métodos indesejáveis de coleta de dados. Mas somente a presença desses filtros não garante a privacidade do usuário; são necessários outros mecanismos para atingir esse objetivo.

A terceira camada é a que comporta as ferramentas para proteção de privacidade. A maioria dos mecanismos de proteção de privacidade se apresenta nessa camada. A característica principal desses mecanismos é a tentativa de mascarar ou esconder a identidade virtual ou real do usuário. Para realizar tal tarefa, eles utilizam técnicas de anonimato e de pseudônimos, que inserem um servidor intermediário entre o computador do usuário e o *site*. Tais ferramentas serão apresentadas no sétimo capítulo.

A quarta camada é a de políticas de privacidade. O uso de políticas de privacidade pode aumentar a confiança do usuário no acesso a um determinado *site*. A política é uma descrição do *site* sobre suas práticas de privacidade; ela informa o usuário para que ele, ao concordar com ela, realize o acesso aos serviços do *site*. Como exemplo de ferramenta dessa camada, há o P3P, uma plataforma para leitura e análise automática de políticas de privacidade. Entretanto, a política de privacidade não é nada além de um documento que informa o usuário. Ela necessita da confiança do usuário em *sites*, que devem cumprir suas políticas. Para melhorar essa confiança, leis e selos de privacidade podem ser utilizados.

A quinta camada é a de certificação de privacidade. Essa camada compreende os serviços de regulamentação de práticas de privacidade do *sites*. Esses serviços são prestados por entidades que realizam auditoria, análise de práticas de privacidade e supervisões

periódicas em *sites Web*. De acordo com o resultado dessas análises e supervisões, os *sites* podem ganhar selos ou certificados de garantia de privacidade. Essas análises e supervisões são baseadas na classificação proposta por Wang [55] para preocupações do usuário: acesso impróprio, coleta imprópria, monitoramento impróprio, análise imprópria, transferência imprópria, transmissão não desejada e armazenamento impróprio.

A sexta camada é a de leis que regulamentam a proteção de privacidade. Esse é um controle para os abusos de privacidade que incentivam o respeito e a proteção da privacidade dos usuários. Sem leis para moderar, alguns mecanismos não trabalham apropriadamente, não são capazes de fornecer garantias de seus serviços de proteção de privacidade. Um exemplo disso são as políticas de privacidade, que necessitam de entidades legislativas para garantir que elas sejam seguidas por seus respectivos *sites*.

Os governos devem tentar regular os códigos ou o funcionamento de aplicações *Web*, pois não é possível controlar o comportamento da *Web* [6]. Mesmo assim, há uma dificuldade em conseguir um consenso internacional para criação de leis, pois a privacidade é dependente de questões políticas e culturais. Apesar dessas dificuldades, existe um conjunto de atividades comuns que estas leis devem regular [53], como: notificar os usuários sobre os dados que serão coletados e sobre o objetivo do processamento destes; coletar dados para um uso específico e pertinente ao objetivo para os quais serão usados; armazenar dados por um tempo limite de armazenamento; não repassar dados para terceiros; em caso de venda da empresa, obrigar os novos proprietários de uma determinada coleção de dados a respeitar a política de privacidade da antiga empresa; permitir o acesso dos usuários aos dados coletados e possibilitar atualizá-los ou removê-los e obedecer às restrições de coleta de dados do país do usuário, apesar de os *sites* serem registrados em um país diferente do do usuário.

No contexto da coleta de informações, essas seis camadas podem ser divididas em duas categorias: a coleta implícita e a coleta explícita. Segundo ela, as três primeiras camadas (notificação, controle e ferramentas para proteção da privacidade) são atribuídas basicamente à coleta implícita, e as outras três (políticas de privacidade, certificação de privacidade e leis de proteção à privacidade), à coleta explícita.

Atualmente não existe mecanismo que abranja toda a taxonomia em camadas de proteção de privacidade [53]. Há uma grande dificuldade de se unir diferentes abordagens para solução do mesmo problema em níveis diferentes. Por exemplo, o anonimato exclui todo tipo de dado identificador do usuário e, assim, não permite o acesso a certos tipos de serviço. Na construção de sistemas de proteção de privacidade, os interesses dos usuários têm grande

relevância. Segundo Clarke [56], a proteção de privacidade é o processo de encontrar o balanceamento apropriado entre privacidade e múltiplos interesses competitivos.

6 Certificados ou selos de privacidade

Certificados de privacidade são utilizados para incrementar a visão de confiança dos usuários nos *sites Web* visitados. Os selos de privacidade são marcas de privacidade de confiança mostradas nas páginas principais dos sites que informam aos visitantes as suas práticas de segurança [57]. O progresso do *e-commerce* é dependente do número de adeptos que se sentem seguros em realizar suas transações de negócios via *Web*. No caso, engendrar confiança em consumidores *on-line* é visto como um componente crítico de qualquer estratégia de *e-business* de sucesso [58].

Um exemplo da importância da confiança foi a especulação apresentada por *Jupiter Research*, que diz que preocupações de privacidade causarão bilhões de perdas de vendas para negócios relacionados à Internet [59]. Embora todas as transações comerciais e de negócios possuam um elemento de confiança inerente, a porção de dados requeridos para completar uma transação *on-line* e o potencial para ocorrer uma fraude levam continuadamente a aumentar as preocupações de privacidade [60] [55]. Apesar de tudo isso, as vendas de *e-commerce* nos Estados Unidos totalizaram $56 bilhões em 2003, um aumento de 26,4% sobre o ano de 2002 [61].

A confiança do usuário na *Web* é um crédito atribuído a um site ou serviço através da segurança do usuário, bom conceito ou tradição que são oferecidos no acesso. Essas características se apresentam pelos aspectos de boa conduta e de segurança da informação, ou seja, privacidade, presentes no comportamento de sites. Assim, a privacidade está relacionada com a confiança, pois ela oferece para o usuário um controle de suas informações, o que resulta em um aumento de sua credibilidade no *site* que acessa.

A visão de garantia de privacidade é apresentada por entidades certificadoras de confiança, as quais entregam selos de privacidade a partir de uma análise realizada no *site* e em suas práticas de coleta e de uso de informações pessoais dos usuários. O selo é um sinal de que o *site* que os usuários visitam foi analisado e passou por critérios de privacidade que garantem a confiança deles para continuar a realizar suas transações. Esses *sites* concordam que a confiança pode ser cultivada para melhorar as vidas pessoais e sociais e aumentar capital social das pessoas [62].

A criação dessas entidades que distribuem selos de privacidade *on-line* tem início com a preocupação da legislação americana, que vem crescendo desde a década de 90. Essa legislação observava que incentivos substancialmente maiores eram necessários para

estimular a auto-regulamentação e assegurar a implementação difundida de princípios básicos de privacidade [63]. Com isso, a indústria de *e-commerce* nos Estados Unidos tomou providências de uma política de auto-regulamentação que centra o uso de selos de privacidade. O selo é produzido para introduzir confiança no consumidor *on-line* por verificar que o *site Web* tem uma política sobre sua coleta e uso de informação pessoal [64].

O anonimato pode ser utilizado para fornecer privacidade ao usuário. Entretanto, altos graus de anonimato fornecem desafios significantes para responsabilidade ou para a justificação de ações [65]. Essa responsabilidade diz respeito à identificação do usuário, necessária para interação e realização de alguns serviços, que pode levar a uma falta de anonimato. Atenção cuidadosa é necessária para focar no equilíbrio entre anonimato e responsabilidade. Deve-se buscar engendrar esforços para construção de uma confiança *on-line* sem ignorar outros valores humanos importantes.

Os selos de privacidade, criados por TRUSTe, WebTrust e BBBOnline, adotam um sensato conjunto de princípios de privacidade e empenham-se em assegurar que *sites* estejam em concordância com tais princípios. Esses selos foram desenvolvidos pela indústria de *e-commerce* para informar consumidores *Web* que um *site* particular pode ser confiável [66]. Entretanto, geralmente eles exigem o pagamento de taxas pelos *sites* para receber o selo ou certificado de privacidade, após passarem por um processo de auditoria e avaliação e estarem sujeitos a supervisões.

Segundo Moores [66], o processo de aquisição do selo tipicamente envolve a escrita de uma política de privacidade e a construção de um questionário auto-avaliado em práticas de negócios. A partir disso, o material é então submetido a uma organização de confiança para revisão e aprovação. Além disso, outra exigência é a segurança das informações, pois todas essas organizações reconhecem que dados de privacidade são intimamente relacionados à segurança. Um dado que não seja seguro não pode ser considerado por ser privado [64].

TRUSTe[10] foi quem desenvolveu o primeiro dos principais selos de confiança da *Web*; ele foi lançado em junho de 1997. Ela é uma organização sem fins lucrativos que possui a missão de construir a confiança do usuário na Internet promovendo o princípio de divulgação. O programa requer que *sites Web* o adotem e cumpram com uma divulgação justa de suas práticas. Ela trabalha como um coordenador de *sites* para assegurar que a declaração de privacidade seja escrita corretamente [64].

[10] http://www.truste.org

Ao seguir uma requisição de uma organização *on-line*, a TRUSTe realiza auditoria para o *site Web* da organização. Ela permite o *site* mostrar o seu selo de aprovação e mantém a certificação se a organização apresenta certos critérios mínimos [62]. Para um *site Web* ser concordante com os critérios da TRUSTe, ele deve adotar e implementar uma política de privacidade para o seu *site* e estar sujeito a procedimentos de resolução de concordância e supervisão da TRUSTe. Licenças são monitoradas por uma revisão inicial e supervisões periódicas [64]. Essas licenças concordam com os requisitos de Notificação, Escolha, Segurança, Acesso e Qualidade de Dados e Verificação e Supervisão [57].

Na notificação, o *site Web* deve divulgar uma declaração de privacidade ligada à página principal, a qual inclui divulgação sobre a coleta de informações realizada pelo *site* e as práticas de disseminação. A TRUSTe trabalha com o *site Web* para desenvolver declarações de privacidade compreensíveis que são fáceis de ler e entender.

Segundo o requisito de escolha, o *site Web* deve fornecer aos usuários pelo menos a opção de não possuírem suas informações pessoais usadas por terceiros e para propósitos secundários.

Para atender o requisito de segurança, o *site Web* deve implementar procedimentos razoáveis para proteger informações pessoais de perda, mau uso ou alteração não autorizada.

No acesso e na qualidade de dados, o *site Web* deve fornecer um mecanismo para consumidores corrigirem imprecisões em suas informações.

Verificação e supervisão estão relacionadas à imposição da TRUSTe de que *sites* sigam suas práticas de privacidade declaradas através de revisões iniciais e periódicas. O *site Web* concorda em cooperar com esse processo de supervisão para receber o selo de privacidade.

Para a TRUSTe, as declarações de privacidade devem divulgar qual tipo de informação pessoal de identificação está sendo coletada, quem a está coletando, como ela será usada e com quem ela será divida.

Para ser concedido o selo da WebTrust[11], lançada em setembro de 1997, o site *Web* deve ser examinado para assegurar concordância com os seus princípios vigentes, semelhantes à política da TRUSTe. É necessário também que o *site* publique suas práticas de privacidade. Assim como a TRUSTe, o selo é representado por uma imagem que pode levar a uma função de verificação ou autenticação, administrada por Verisign.com.

[11] http://www.cpawebtrust.org

A BBBOnline[12] é a mais recente das três organizações, lançada em março de 1999 [64]. Para um *site* receber o selo, ele deve escrever uma declaração de privacidade que seja fácil de ler e que liste todas as divulgações em um único documento. Assim como TRUSTe e WebTrust, esse documento precisa descrever todos os tipos de informações pessoais que podem ser coletadas, como são coletadas e quais são os usos dados a elas.

Como exemplo do uso aplicado desses certificados de privacidade, são apresentados os receptores mais eminentes dos selos concedidos por essas três entidades: América Online, AT&T, Bell Canadá, IBM, Intel, Microsoft e Hewlett-Packard [66].

Dentre as centenas de milhares de *sites* comerciais que existem, somente poucos milhares possuem alguma declaração de privacidade. Além disso, entre aqueles *sites* que adotam esse sistema, existem casos em que houve abusos do uso dos selos de confiança. Em outubro de 2000, a TRUSTe processou dois *sites Web*, *American-Politics.com* e *SurfAssured.com*, por uso ilegal de seu selo [64].

Usuários não entendem completamente a forma ou função dos selos de privacidade, poucos podem reconhecer um selo como verdadeiro e poucos deles reconhecem como uma ferramenta importante na decisão para confiar em *sites Web*. Moores observa que para mudar essa visão que o usuário possui dos selos de privacidade é necessário colocá-los em lugares mais proeminentes para um reconhecimento e entendimento mais fácil [66].

Apesar de todos esses dados negativos, os usuários da *Web* estão começando a reconhecer os selos de confiança e o que eles significam. *Cheskin Research* reportou que 69% dos usuários *Web* reconheceram o selo da TRUSTe e 37% o selo da BBBOnline [67]. O selo da TRUSTe aumentou a confiança em um *site Web* para 55% segundo essa mesma pesquisa.

Os fatos e resultados de pesquisas mostram que muito deve ser explorado nesse contexto de selos de privacidade para melhorar algumas características e levar a uma maior adoção. Mas também outros resultados mostram um maior reconhecimento dos usuários da importância desses certificados. O que é um aspecto relevante que não deve ser ignorado na construção de um sistema de privacidade de informações dos usuários.

No Brasil, a certificação de privacidade não foi implantada. Não existem entidades certificadoras que garantem que sites sigam princípios para não prejudicarem a privacidade do usuário. Por enquanto, somente o governo é responsável por assegurar os direitos de privacidade do usuário na *Web*, que são observados como direitos do consumidor e do cidadão, presentes na constituição brasileira.

[12] http:// www.bbbonline.org

7 Leis de proteção de privacidade

Segundo Bittar [68], a privacidade é um dos direitos essenciais, vitalícios e intransmissíveis, que protegem valores inatos ou originários da pessoa humana, como a vida, a honra, a identidade, o segredo e a liberdade. Segundo o autor, o direito de expor ou não a sua própria identidade nada mais é do que um dos componentes da Liberdade. No contexto deste trabalho, a privacidade está envolvida com o conhecimento da coleta de informação e com o consentimento dela, o que resulta no controle que um usuário exerce sobre suas informações.

A *Web* não possui uma legislação que a regulamenta. Cada nação procura aplicar suas leis com relação à privacidade de seus cidadãos e também aos usuários da Internet. Essas aplicações podem parecer semelhantes, mas possuem diferentes características referentes à diversidade cultural ao redor do mundo.

A abordagem para regulamentação é dependente da funcionalidade empregada na rede mundial: correio eletrônico, Debates Eletrônicos (*Newsgroups*), comércio eletrônico, divulgação de material informativo científico, votação eletrônica e outras.

A falta de uma regulamentação única para a Internet é pertinente a sua característica fundamental: a de não pertencer a ninguém, não ser financiada por instituições, governos ou organizações internacionais e não ser um veículo comercial. Os únicos órgãos que desenvolvem a função de direção, controle e funcionamento da rede são a ISOC (*Internet Society* [13]) e a IETF (*Internet Engineering Task Force* [14]) [69].

Os países sempre discutem leis para regulamentar e restringir a divulgação de informações para manter a privacidade de seus cidadãos na *Web*. A *Privacy International* apresenta anualmente um relatório que aborda a legislação e a situação de privacidade em vários países do mundo. No relatório de 2004[15], foram listados 65 paises, dentre eles o Brasil, os Estados Unidos, a França, a Alemanha, o Japão e a Austrália.

A *Privacy International* [16] é uma organização independente e não governamental que trabalha com direitos humanos; foi formada em 1990 como um vigia de governos e corporações. Ela monitora legislações de países segundo a diretiva de proteção de dados da

[13] http://www.isoc.org
[14] http://www.ietf.org
[15] http://www.privacyinternational.org/survey/phr2004
[16] http://www.privacyinternational.org

União Européia; exerce ações legais contra companhias que violam regras de privacidade e monitora o desenvolvimento de regulamentações de privacidade ao redor do mundo.

Ela foi criada em resposta a um número crescente de ameaças de privacidade e é uma organização mundial para a proteção de privacidade formada por mais de centenas de especialistas de privacidade de direções e organizações de Direitos Humanos de 40 países.

A formação da *Privacy International* é o primeiro sucesso na tentativa de estabelecimento de uma estrutura mundial focada nessa área crucial de direitos humanos.

Diversas outras instituições existem com a finalidade de dar suporte e amparo à privacidade dos indivíduos, visto que a maioria delas tem base em ações americanas e européias. As instituições mais importantes nesse contexto são apresentadas a seguir.

A EPIC (*Electronic Privacy Information Center* [17]) é um centro americano de pesquisa de interesse público. Ela foi estabelecida em 1994 para focar a atenção pública em edições de liberdades civis emergentes e para proteger a privacidade, a Primeira Emenda Americana, e valores constitucionais.

A *Privacy.org*[18] é um *site* para divulgação de notícias diárias e informações e para desenvolvimento de iniciativas em privacidade. Esse *site Web* é um projeto conjunto da EPIC e *Privacy International*.

A PRC (*Privacy Rights Clearinghouse*[19]) é uma organização americana direcionada ao consumidor, sem fins lucrativos e com a missão de informar e amparar o consumidor. Ela foi estabelecida em 1992. Ela é principalmente mantida por doações e serve indivíduos em âmbito nacional. Eis os objetivos da PRC: aumentar a ciência dos consumidores de como a tecnologia afeta a privacidade pessoal; capacitar consumidores a tomar ações para controlarem suas próprias informações pessoais através do fornecimento de dicas práticas de proteção de privacidade; responder a reclamações específicas dos consumidores relacionadas à privacidade; interceder a seu interesse, e, no momento apropriado, direcionar esses consumidores a organizações para assistência adicional; documentar a natureza das reclamações dos consumidores e as questões sobre privacidade em relatórios, testemunhos e discursos (essas informações são disponibilizadas a construtores de políticas, a representantes de indústrias, a advogados de consumidores e à mídia); e amparar os direitos de privacidade do consumidor para procedimentos de políticas públicas federais, o que inclui a prova e o

[17] http://www.epic.org/
[18] http://www.privacy.org
[19] http://www.privacyrights.org

testemunho legislativo, a audição de agências reguladoras, as forças de tarefa e comissões de estudo, como conferências e *workshops*.

Nesse contexto, a FTC (*Federal Trade Comission*[20]) e a OECD (*Organization for Economic Co-operation and Development*[21]) são as instituições que mais se destacam no intuito de regulamentar coleta de informações pela *Web* e proteger a privacidade dos usuários.

A OECD apresenta oito princípios que especificam a forma como os dados devem ser protegidos segundo as práticas de privacidade dos *sites* na *Web* [7] [70]: o limite de colete, a qualidade dos dados, a especificação do objetivo, a limitação do uso, a segurança, a transparência, a participação individual e a responsabilidade.

O princípio do Limite da Coleta estabelece um limite para a coleta de dados dos usuários. Esse limite deve ser regido por meios legais, uma vez que o proprietário dos dados deve estar ciente dele e consenti-lo.

No princípio da Qualidade dos Dados, a coleta de informações pessoais deve se limitar à relevância dos objetivos de seu uso, visto que elas devem ser precisas, completas e mantidas atualizadas para não ocorrer nenhuma incoerência.

Segundo o princípio da Especificação do Objetivo, antes de ser realizada a coleta de dados, os objetivos da coleta devem ser delimitados e especificados, e o uso desses dados deve se restringir a esses objetivos.

No princípio da Limitação do Uso, os dados pessoais não podem ser divulgados, disponibilizados ou usados para outros propósitos além dos especificados. Exceto quando há consentimento do proprietário ou quando é exigido por uma autoridade legal.

O princípio da Segurança leva o *site* ser responsável por utilizar mecanismos de segurança razoáveis para proteger os dados pessoais coletados e armazenados.

Segundo o princípio da Transparência, as práticas do *site* com dados pessoais devem ser divulgadas para os usuários de forma transparente através de políticas ou declarações de privacidade.

O princípio da Participação Individual diz que o acesso e a obtenção de informações próprias devem ser direitos de todo usuário.

No princípio da Responsabilidade, um controlador de dados deve ser responsável por cumprir todos os princípios acima.

A privacidade é um elemento central da missão de proteção do consumidor da FTC. Ela é uma instituição que trabalha para zelar pela vida econômica dos cidadãos americanos.

[20] http://www.ftc.gov
[21] http://oecd.org

Nos anos recentes, avanços na tecnologia da computação tem tornado possível o detalhamento de informação sobre pessoas para serem tratadas e divulgadas mais facilmente. Isso tem produzido muitos benefícios para sociedade como um todo ou individualmente, mas também potencializa os riscos de privacidade.

Desse modo, a FTC auxilia no reforço de leis para vasculhar criminosos, para que bancos previnam fraude e para que consumidores aprendam sobre novos produtos e serviços, o que permite a tomada de decisões de compra com mais segurança. Ao mesmo tempo, conforme a informação pessoal se torna mais acessível, companhias, associações, agências governamentais e consumidores devem tomar precauções para se protegerem contra o mau uso das informações.

A FTC objetiva educar consumidores e negócios sobre a importância da privacidade da informação pessoal e sobre a segurança de informações. A FTC luta contra a deslealdade e a decepção, reforçando promessas de privacidade de companhias sobre como elas coletam, usam e asseguram informação pessoal dos consumidores.

Princípios de práticas justas de privacidade também são apresentados pela FTC[22] [71] [72]; esses princípios são um resumo dos oito que são especificados pela OECD. Essas práticas justas de informação são comumente responsabilidades que governam a coleta, o acesso e o controle sobre informações pessoais. Essas práticas incluem: notificação e ciência; escolha e consentimento; acesso e participação; integridade e segurança; reforço e retificação.

Segundo a Notificação e a Ciência, os usuários devem receber notificações sobre práticas de informação da entidade antes que qualquer informação pessoal seja coletada. Sem a notificação, o usuário é incapaz de decidir sobre a divulgação de suas informações. Os outros princípios só possuem significado quando o usuário está ciente das políticas de privacidade do *site*.

A escolha e o consentimento significam dar aos usuários opções de como suas informações pessoais podem ser utilizadas, por relatar os usos que estão além dos necessários para realizar uma transação na *Web*.

O acesso e a participação se referem à capacidade de um indivíduo de acessar os dados sobre ele, contestar a precisão e a integridade desses dados, corrigi-los e atualizá-los.

Na integridade e na segurança, os dados devem ser precisos e estarem seguros. *Sites* devem assegurar a integridade das informações coletadas.

[22] http://www.ftc.gov/reports/privacy3/fairinfo.htm

O reforço e a retificação são mecanismos utilizados para reforçar os princípios anteriores de proteção de dados e torná-los efetivos.

No Brasil, existem leis e projetos de leis como uma tentativa de manter a privacidade dos cidadãos para qualquer contexto em que ele esteja inserido [69]. Entretanto, não há nada relacionada diretamente à privacidade do usuário no *Web*. O artigo 220 da Constituição Federal do Brasil dá liberdade de informação para os cidadãos e permite que eles usufruam de instrumentos informáticos para instruir e para se instruírem. O artigo quinto fornece ferramentas para julgar crimes na *Web*. Entretanto, nenhuma lei menciona explicitamente privacidade ou trata dos crimes virtuais.

Art. 220. “A manifestação do pensamento, a criação, a expressão e a informação, sob qualquer forma, processo ou veículo não sofrerão qualquer restrição, observado o disposto nesta Constituição.”

Em outro trecho da constituição brasileira, art 5º. – IV e IX, são assegurados a livre manifestação do pensamento, a liberdade de expressão e o direito à informação.

Art. 5º. – IV – “é livre a manifestação do pensamento, sendo vedado o anonimato”; complementando o inciso X.

Art. 5º. – IX – “é livre a expressão da atividade intelectual, artística, científica e de comunicação, independentemente de censura ou licença.”

Em 1995, foi criado o Comitê Gestor da Internet, que conta com a participação do Ministério das Comunicações e do Ministério da Tecnologia, de representantes de provedores de acesso ou de informações, de representantes de usuários, da comunidade acadêmica e de entidades operadoras e gestoras de *backbones* [69]. Esse comitê é responsável por fomentar o desenvolvimento de serviços de Internet no Brasil, recomendar padrões e procedimentos técnico-operacionais, coordenar atribuição de endereços IP, providenciar o registro de nomes de domínios, coletar, organizar e disseminar informações sobre serviços Internet e outros.

Em 1996, foi criada a Lei 9.296 que pune o indivíduo que realiza interceptação de comunicações em sistemas informáticos.

O artigo 5º. – X da Constituição Federal do Brasil de 1988 resguarda a vida privada e a intimidade.

Art. 5º. – X – “são invioláveis a intimidade, a vida privada, a honra e a imagem das pessoas, assegurado o direito a indenização pelo dano material ou moral decorrente de sua violação”.

Além disso, o Projeto de Lei 3.360/00 está tramitando no Congresso Nacional [73]. Ele foi aprovado pela Comissão de Ciência e Tecnologia, Comunicações e Informática da

Câmara dos Deputados. Essa lei se refere aos crimes contra a inviolabilidade de dados e de comunicações através de computadores. Ela não delimita especificamente os abusos à privacidade, mas poderá contribuir para uma punição mais adequada aos que a violam na *Web*.

Esse Projeto de Lei no Brasil é um passo importante para a definição de uma legislação específica e que esteja coerente com os aspectos de privacidade para o meio virtual. Para isso é necessário delimitar claramente privacidade e os direitos relacionados a ela, apresentados pelos princípios de privacidade.

A privacidade constitui um limite natural ao direito da informação, mas as informações pessoais poderão ser divulgadas quando houver consentimento da pessoa, com consentimento implícito ou demonstração de interesse em divulgar, e contra a sua vontade, se houver relevância pública [69].

8 Mecanismos de privacidade

Como formas de enfrentar o problema de privacidade na *Web*, existem, na literatura, muitas propostas, as quais podem ser divididas em duas formas básicas de abordagem. Uma das formas visa a introdução de uma arquitetura ou mecanismo que tenta manter o anonimato do usuário ou dificultar a identificação do mesmo. Entretanto, mecanismos dessa linha de abordagem podem impossibilitar que a coleta de informação do usuário seja realizada. Ferramentas de anonimato, criação de pseudônimos e MASKS [53], seguem essa lógica de abordagem.

A outra linha de abordagem introduz um método de policiamento dos *sites* ou de informação do usuário sobre as políticas de privacidade que são adotadas por eles. As políticas comunicam as informações que são coletadas e advertem sobre irregularidades. Os mecanismos P3P [74] e a ferramenta de fornecimento de informação relacionada a contexto [30] se encaixam segundo esse critério.

8.1 Anonimato

Existem muitas razões para uma pessoa esconder a sua identidade real quando usa a Internet. Ela pode querer se proteger de governos opressivos ou publicar mensagens em grupo de notícias ou de discussão sem se identificar para todos como o autor.

Mudar ou apagar o campo que corresponde a uma identificação real não impossibilita que alguém descubra quem foi o responsável pelo acesso a alguma página ou pelo envio de alguma mensagem. É possível investigar o caminho de volta (*back tracking*) de uma mensagem enviada para o *site* original dela, e o administrador desse *site* pode facilmente ver o nome real do transmissor dela.

Um melhoramento para esse caso da mensagem é usar um *site* que a envie de forma anônima e apague toda a informação referente ao usuário antes de transmiti-la. Esse *sites* também fornecem um endereço anônimo, ao qual outras pessoas podem enviar correspondência eletrônica. No final, essa correspondência é transmitida para o endereço real. Esses sistemas são algumas vezes chamados de “servidores de pseudônimos”. Desse modo, a mensagem é enviada sem deixar qualquer sinal do nome ou endereço do transmissor.

A falta ou a sensação de falta de privacidade das transações na *Web* está presente em uma grande parte de seus usuários. Por esse motivo, um número crescente de pessoas começa a procurar serviços que forneçam uma navegação anônima na Internet, como *anonymizer.com* ou *the-cloak.com* [75].

Em uma busca no *site* "*google.com"* por esse serviço de anonimato, uma longa lista de *sites* que fornecem serviços diferenciados aparece como resposta. Esses *sites* podem apresentar serviços gratuitos ou pagos e com variações de funcionalidades e desempenhos. Entretanto, todos eles afirmam que podem tornar a navegação na *Web* anônima. Todos esses serviços são baseados nas mesmas suposições, no mesmo modelo de abordagem e mesmo projeto.

Além desses *proxies* de anonimato, existem também sistemas mais complexos, necessários para proteger um usuário de um agressor. Entretanto, para isso é preciso realizar a instalação de software extra e um maior conhecimento do usuário para efetuar a instalação desse software, que apresenta obrigações adicionais. Desses sistemas complexos, existem somente dois que hoje estão publicamente disponíveis: *Crows* e JAP [75].

O anonimato é definido como o estado de um indivíduo não identificável entre um conjunto de sujeitos [76]. Segundo essa significação, existe o anonimato do transmissor, que é a não identificação de uma mensagem do mesmo; o anonimato do receptor, que é a não identificação do recebimento de uma mensagem; e a não ligação entre o transmissor e o receptor, que é a não identificação por um terceiro da existência da mensagem por si própria.

Segundo essa definição, é necessário considerar também dois conceitos: o agressor contra quem se quer manter o anonimato e o grau de anonimato, ou seja, a certeza com a qual o agressor pode localizar o transmissor, o receptor ou a ligação entre os dois. Os agressores podem ser passivos, podem estar somente observando o tráfego, ou podem ser ativos, inserindo suas próprias informações em uma comunicação pré-existente.

Novas propostas são criadas para solucionar problemas dos mecanismos antigos e para possibilitar um anonimato mais seguro. Um exemplo é o anonimato que usa memória a *cache* de um *proxy*. Essa solução visa acabar com os problemas com relação ao agressor global que é capaz de observar por um longo período. A proposta é direcionar para um *proxy* as requisições dos usuários. Esse *proxy* encaminha a requisição de página para o servidor e a armazena em seu próprio *cache*, e as requisições subseqüentes dos usuários são atendidas através de acesso à memória [75].

Resumindo, de forma geral, um *proxy* de anonimato mascara todas as requisições dos usuários como se fossem as suas, através da modificação dos IPs dos pacotes encaminhados.

Entretanto, essa camuflagem da requisição do usuário torna impossível para os *sites* a identificação do perfil de um usuário determinado. Conseqüentemente, esse anonimato acarreta o fornecimento de serviços personalizados.

Essas práticas de anonimato falham quando o usuário explicitamente envia suas informações pessoais para algum receptor. Com isso, o computador do usuário não é identificado, mas a identidade do usuário pode ser determinada facilmente, e ele pode estar correndo o risco de perder confidencialidade de suas informações pessoais.

Segundo essa linha, um *proxy*, para realizar a navegação anônima, é classificado em duas abordagens: *proxies* de Anonimato de único nó e *proxies* de Anonimato de vários nós.

Além disso, existe o sistema *Crowds* [77], como outra abordagem para o problema das transações *Web* anônimas. Ele introduz o conceito de camuflagem para esconder a origem de uma requisição de um usuário em uma multidão ou conjunto de usuários. *Crowds* tenta fornecer anonimato ao transmissor e ao receptor, mas atua diferentemente das redes de *proxies* de vários nós.

8.1.1 Proxies de Anonimato de único nó (Anonymizer)

A maioria dos serviços de anonimato é construída através de *proxies* de único nó que encaminham requisições HTTP para o usuário. O usuário envia a URL da página que deseja requisitar para o *proxy* de anonimato que imediatamente emite uma requisição HTTP para o servidor relacionado à URL recebida. A requisição recebida pelo *site* ou servidor de destino aparenta ser proveniente do *proxy* e não do computador do usuário. Assim que o servidor do *site* responde, o *proxy* recebe e envia para o usuário o documento HTML com todos os *links* reescritos, para que eles direcionarem a passagem da requisição pelo *proxy* e não diretamente para os *sites* que eles originalmente estavam direcionados. Isso significa que, quando se utiliza um serviço gratuito de navegação anônima, como o *anonymizer.com*, um link para *http://slashdot.org* é reescrito e se torna *http://anon.free.anonymizer.com/http://slashdot.org* [75].

Além disso, para impossibilitar a identificação da ligação entre o transmissor e o receptor da comunicação, a conexão do usuário com o *proxy* pode ser criptografada. Alguns serviços, como *the-cloak.com*, fornecem esse serviço de criptografia. Outros, como *anonymizer.com*, provêm gratuitamente somente a navegação sem criptografia e cobra-se pela navegação criptografada [75].

Além de ocultar o IP do usuário e possivelmente permitir criptografia do tráfego de dados, outras funcionalidades extras podem ser incluídas nesses *proxies* de único nó. Essas funcionalidades adicionais podem ser a filtragem ou manipulação de *cookies*, a filtragem ou reedição de códigos JavaScript, Java e outros códigos de conteúdo ativo, a filtragem de anúncios e *banners*, o bloqueio de pacotes HTTP e a falsificação do campo *http_user_agent* e *http_referer* do cabeçalho HTTP para não se revelar informações sobre os sistemas operacionais do usuário, sobre o navegador e sobre o *site* que foi previamente visitado [75].

A figura 7 [75] exemplifica o funcionamento do *proxy* de um único nó. O usuário faz uma requisição de página para o *site* *http://www.slashdot.org* da forma *http://ww.anonymizer.com/http://www.slashdot.*org em (1). O *proxy*, que aplica seus mecanismos de anonimato, repassa a requisição para o *site* destino (2). O *site* retorna a página requisitada para o *proxy* com um link na página para *http://www.something.org* em (3). Por fim, em (4), o *proxy* repassa ao usuário a página com modificações em seus *links* *http://ww.anonymizer.com/http://www. something.org*.

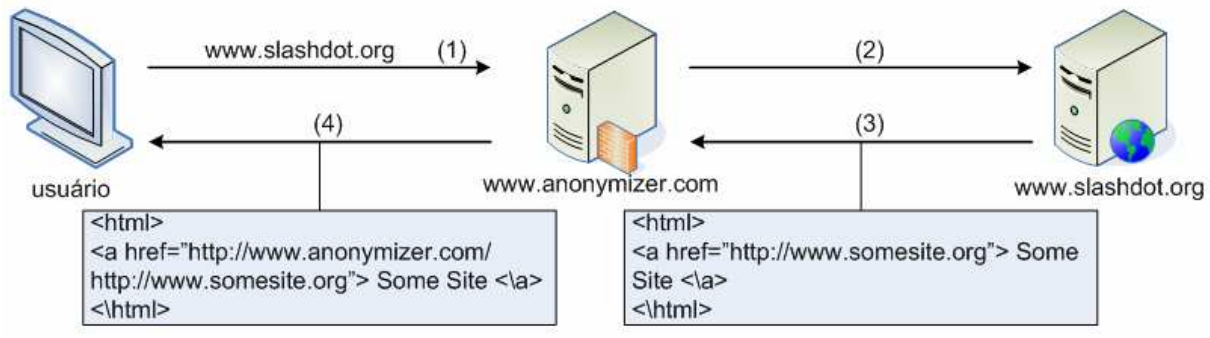

Figura 7. Exemplo de requisição de página através de um Proxy de anonimato de único nó.

A tabela 1 [75] lista alguns serviços de navegação anônima de acesso privado e público.

Tabela 1. Listagem de alguns serviços presentes na Web de navegação anônima.

Nome	URL	Encriptação
Anonymizer	http://www.anonymizer.com	Somente versão paga
the-Cloak	http://www.the-cloak.com	Sim
ProxyWeb.net	http://www.proxyWeb.net	Somente versão paga
SnoopBlocker.com	http://www.snoopblocker.com	Somente versão paga
Proxify.com	http://proxify.com	Não
Anonymouse	http://anonymouse.ws	Não
Web Warper	http://Webwarper.net	Não
Anonymization	http://www.anonymization.net	Não
PurePrivacy	http://www.pureprivacy.com	Não

A figura 8 [78] apresenta, como exemplo, a arquitetura do *Anonymizer*. Na figura há dois fluxos de dados, com e sem criptografia, para o acesso do usuário a ele. Esse acesso criptografado é construído através de uma *Virtual Private Network*.

Proxies de único nó não abrangem o modelo em que o agressor é capaz de observar um tráfego entre o usuário e o *proxy*. Nos casos em que a comunicação entre eles não é criptografada, qualquer um é capaz de analisar o tráfego de comunicação e de obter uma visão total do que está sendo transmitido.

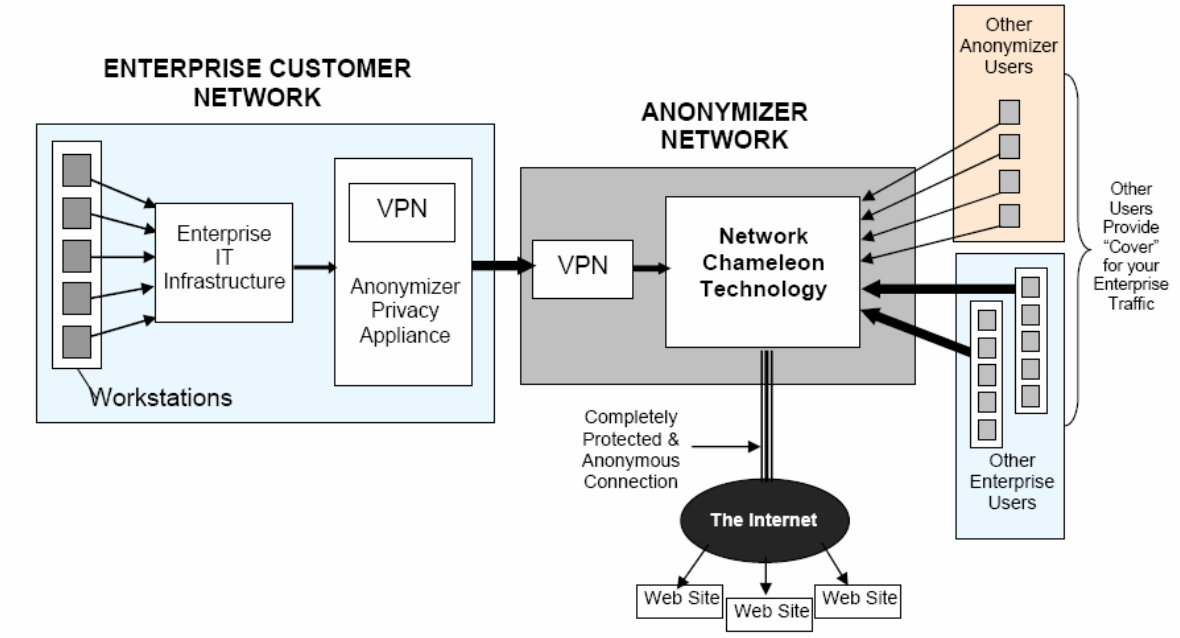

Figura 8. Arquitetura de funcionamento de Anonymizer.

Mesmo que o tráfego entre o usuário e o *proxy* seja criptografado, é possível para um agressor global ver quais dados o usuário requisita e os obtém fazendo análise de tráfego do *proxy* através do fluxo de dados de saída e de entrada, visto que esse agressor é capaz de observar ambos os caminhos de comunicação, do usuário para o *proxy* e do *proxy* para o destino.

8.1.2 Proxies de Anonimato de vários nós (Onion Routing)

Em 1981, Chaum [79] introduziu o conceito de Mix-nets. Mix-nets são grupos de servidores ou *proxies* que promovem anonimato por conduzir o tráfego do usuário através de nós chamados Mixes, os quais podem atrasar, reordenar, recriptografar, adicionar dados inúteis e passar adiante o tráfego. Para fornecer anonimato ao transmissor, uma rede de Mix-

nets tenta desvincular nas comunicações que passam por ela o aspecto de ligação entre o transmissor e o receptor.

Entre os exemplos de Mix-nets, existem *Onion Routing*, o qual está temporariamente fora de ação para testes de sua segunda geração, *Zero Knowledge Systems' Freedom Network* [80], que está permanentemente fora do ar, *Web MIXes* [81], e *Tarzan* [82]. Como exemplo dessa classe de abordagem para o problema de privacidade, a rede *Onion Routing* será apresentada, dado que se trata de uma arquitetura representativa do funcionamento de *proxies* de anonimato de vários nós.

Onion Routing opera por conexões anônimas construídas dinamicamente com uma rede de Mixes de tempo real [79], que pode ser visualizada na figura 9. Cada Mix dessa rede é um dispositivo que armazena e passa adiante informação aceitando um número de mensagens com tamanho fixo de numerosas e variadas origens. Cada nó aplica transformações criptográficas nas mensagens e, então, passa as mensagens adiante para o próximo destino em uma ordem randômica. A dificuldade de se determinar a comunicação entre dois *hosts* na rede é proporcional à quantidade de nós de Mix que estão executando o roteamento de mensagens.

A rede *Onion Routing* é constituída por vários roteadores *onion* (Mix) e tem como característica principal a distribuição. Conseqüentemente, ela é tolerante a falhas e está sob o controle de domínios de administração múltipla. Assim, a perda de um único roteador *onion* não implica no comprometimento de toda a rede ou na perda da privacidade do usuário. Entretanto, a distribuição dessa rede traz também as mesmas dificuldades de gerenciamento que uma rede distribuída possui.

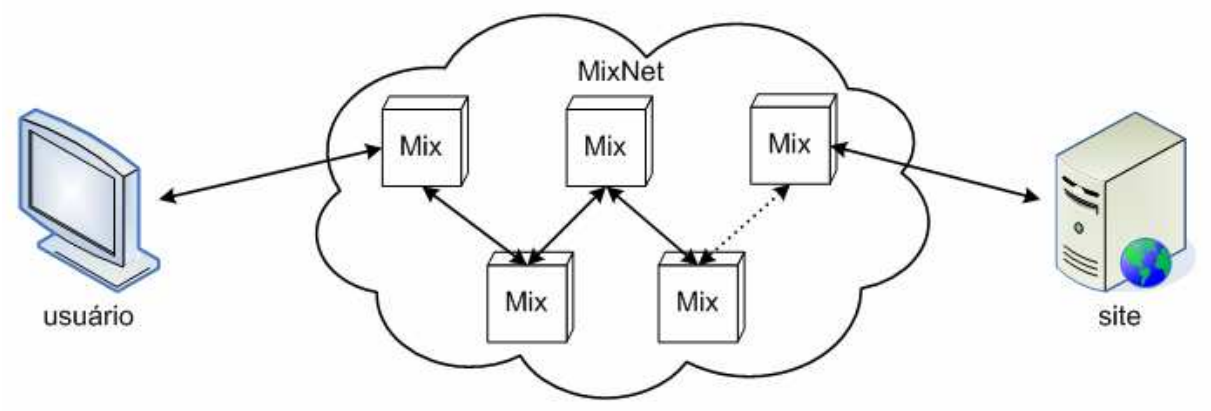

Figura 9. Ilustração da formação de uma rede de Mixes.

Uma *Onion Routing* pode ser usada com aplicações que não são cientes de sua existência; não é necessário realizar modificações nelas. Atualmente, os seguintes protocolos

são suportados pela arquitetura: HTTP, FTP, SMTP, *rlogin*, *telnet*, NNTP, *finger*, *whois*, e *sockets raw*. *Proxies* estão em desenvolvimento para Socks5, DNS, NFS, IRC, HTTPS, SSH, e *Virtual Private Networks* (VPNs) [77].

Um *proxy* ou uma mix-net tem três camadas lógicas: um filtro de privacidade específico da aplicação e que limpa os fluxos de dados, um *proxy* específico da aplicação, que traduz os fluxos de dados em um formato independente de aplicação aceito pela rede *Onion Routing* e um *proxy onion* que constrói e gerencia as conexões anônimas.

Por construir e gerenciar essas conexões, o *proxy onion* deve ser o componente mais confiável no sistema. Para construir esses *onions* e então definir rotas, o *proxy onion* deve conhecer necessariamente a topologia, o estado de ligação da rede, os certificados públicos de nós na rede e as políticas de saída de nós da rede [77]. Essa informação é distribuída automaticamente e de forma segura pela rede conforme novos nós se tornam ativos ou conforme ocorrem alterações de informações.

Segundo Reiter [77], as conexões anônimas de *Onion Routing* possuem protocolos independentes incorporados em três fases: configuração de conexão, movimentação de dados, e destruição da conexão.

A configuração começa quando o iniciador cria um *onion*, que define o caminho da conexão através da rede, visualizada na figura 10. Um *onion* é uma estrutura de dados criada em camadas que especifica propriedades de conexão a cada ponto da rota, informação de controle de criptografia, como os seus diferentes algoritmos simétricos e suas diferentes chaves públicas usadas durante a fase de movimentação de dados. Cada roteador *onion* da rota determinada usa sua chave pública para descriptografar completamente o *onion* que ele recebe. Essa operação extrai a informação de controle de criptografia, a identidade do próximo roteador de *onion* e o próprio *onion* embutido. Depois, o roteador do *onion* o envia para o próximo roteador. Assim que a conexão é estabelecida, dados podem ser enviados em ambas as direções.

Dados do iniciador são em seguida pré-criptografados repetidamente usando os algoritmos e chaves que foram especificadas no onion. Conforme os dados se movem através das conexões anônimas, cada roteador onion remove uma camada de criptografia como definida pela informação de controle de criptografia quando o onion define a rota. Quando o dado chega ao receptor, ele está na forma de texto. Essa construção de camadas também ocorre na ordem reversa para os dados que se movimentam pelo caminho de volta e utiliza diferentes algoritmos e chaves.

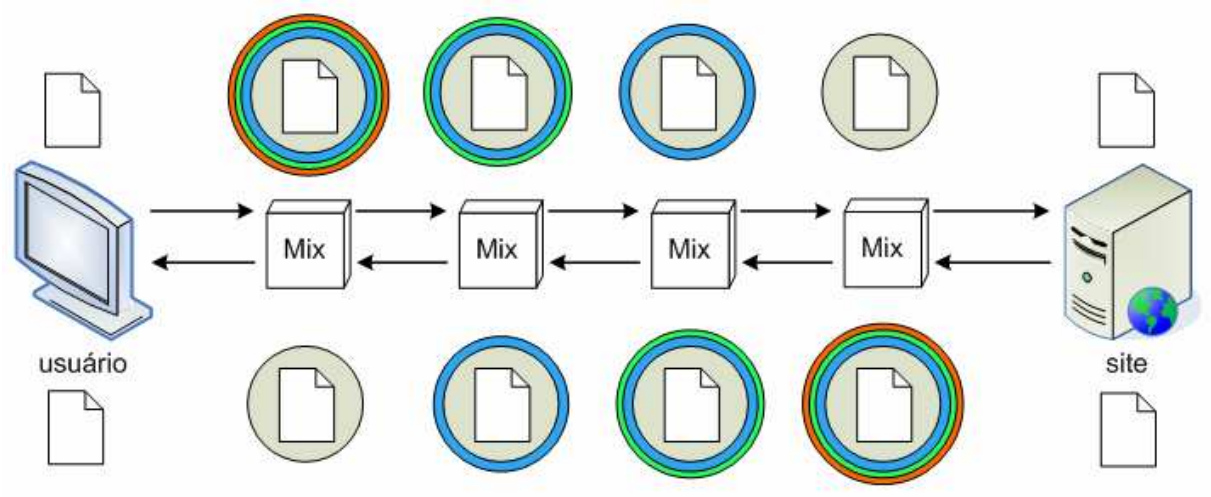

Figura 10. Ilustração da formação do caminho de comunicação de Mixes e aspecto do Onion.

A destruição de conexão pode ser iniciada pelas extremidades ou por roteadores internos, se for preciso.

Todas as informações são enviadas através da rede *Onion Routing* em células de tamanho uniforme. Todas as células que chegam a um roteador *onion* com um intervalo de tempo fixo são misturadas para reduzir a correlação entre elas para observadores internos de rede. Um *onion* aparenta diferente para cada roteador *onion* ao longo de uma conexão por ser esta estendida em camadas de criptografia de chave pública ou de criptografica simétrica. Esse plano resiste à análise de tráfego com mais efetividade que qualquer outro mecanismo desenvolvido para a comunicação na Internet [77].

8.1.3 Crowds

Uma *crowd* ou multidão pode ser vista como uma coleção de usuários que trabalham de forma cooperativa para auxiliar cada membro dela a se manter anônimo. Um usuário é representado em uma *crowd* por um processo em seu computador chamado “jondo” [77], que vem da expressão “*John Doe*”, que conduz a representação de um participante desconhecido ou anônimo.

Quando um jondo é iniciado por um usuário, ele contata um servidor chamado *blender*, que é um misturador, para requisitar a admissão a uma *crowd*. Se admitido, o *blender* reporta para esse jondo a atual sociedade da *crowd* e a informação que habilite esse jondo a participar na *crowd*.

Para poder utilizar o *blender*, é necessário que o usuário estabeleça uma conta com ele e crie um nome de conta e senha que o *blender* deve armazenar para poder autenticar a comunicação com os *jondos*. Após essa autenticação e comunicação, o *blender* adiciona o novo jondo (seu endereço IP, número de porta, e nome de conta) para sua lista de membros e reporta essa lista de volta para o jondo. O *blender* gera e reporta de volta uma lista de chaves comuns, cada uma das quais pode ser usada para autenticar outro membro da *crowd*. No final do processo, todos os membros estão equipados com os dados necessários para que o novo membro participe da *crowd*. Mas, para se proteger de ataques, o novo membro priva-se de fazer conexão até o momento em que ele receba uma mensagem de aceitação do *blender*.

Cada membro mantém sua própria lista da sociedade de *crowd*. Essa lista é iniciada quando esse jondo se junta a *crowd*, e é atualizada quando o jondo recebe avisos do *blender* de membros novos ou excluídos. Ao ser detectado por um jondo que outro está falho, ele pode removê-lo de sua lista de membros. Isso permite que cada lista de jondo se diferencie de outras, se jondos diferentes detectam diferentes falhas de elementos na *crowd*.

Uma desvantagem dessa abordagem para manutenção da sociedade é que o *blender* é um terceiro na comunicação que necessita ser confiável para a distribuição de chaves e para a realização de relatório da sociedade atual para um novo *jondo*. O *blender* pode ser replicado em vários computadores como forma de segurança. Mesmo assim, a comunicação HTTP do usuário não é roteado através do *blender*, e assim, um ataque passivo ao *blender* não revelará imediatamente as transações dos usuários. Com isso, diminui a responsabilidade do *blender*, que tem somente a função de distribuir chaves públicas de *Diffie-Hellman* para os membros.

Segundo Reiter [77], o usuário seleciona um jondo como seu *proxy Web* e especifica o nome do *host* e número de porta em seu navegador *Web* como o *proxy* para todos o serviços. Assim, qualquer requisição vinda do navegador é enviada diretamente para o jondo.

Quando o navegador recebe a primeira requisição do usuário, o jondo inicia o estabelecimento de um caminho randômico de jondos que comportará as comunicações entre os usuários e os servidores *Web*. O jondo escolhe um outro jondo da sua lista da *crowd* randomicamente para encaminhar a sua requisição; nesse processo ele mesmo pode ser escolhido. Quando esse jondo recebe a requisição, ele sorteia um número aleatório para determinar o encaminhamento ou não da requisição para outro jondo. Esse número indica a probabilidade do envio, visto que, dependendo do resultado, o jondo encaminha para um outro ou submete a requisição para o servidor final, ao qual a requisição foi destinada inicialmente. Então, cada requisição parte do navegador do usuário, passa por um ou mais jondos e finalmente chega ao servidor final. Exemplos de possíveis caminhos são mostrados

na figura 11 [77]. Nesta figura, os caminhos são 1 → 5 → servidor; 2 → 6 → 2 → servidor; 3 → 1 → 6 → servidor; 4 → 4 → servidor; 5 → 4 → 6 → servidor; e 6 → 3 → servidor. Requisições subseqüentes iniciadas pelo mesmo jondo seguem o mesmo caminho, porém com destinos diferentes.

Para cada caminho instaurado existe um identificador que muda para cada jondo, uma vez que esse identificador deve ser único para manter a integridade de cada comunicação de um jondo. Toda a comunicação entre qualquer dois jondos é criptografada com uso de uma chave conhecida somente pelos dois. Essas chaves de criptografia são estabelecidas e listadas assim quando os jondos se unem à *crowd* [77].

Para não comprometer a rede Crowd, é necessário controlar a quantidade e a freqüência de admissão de jondos. Para isso, o *blender* organiza as uniões no momento em que elas ocorrem segundo uma programação. Esses são eventos discretos chamados *commits*. O *blender* informa todos os membros da *crowd* sobre o *commit* de união. Por este ponto, todos os membros novos são habilitados a participar da *crowd* e todos os membros velhos recompõe seus caminhos.

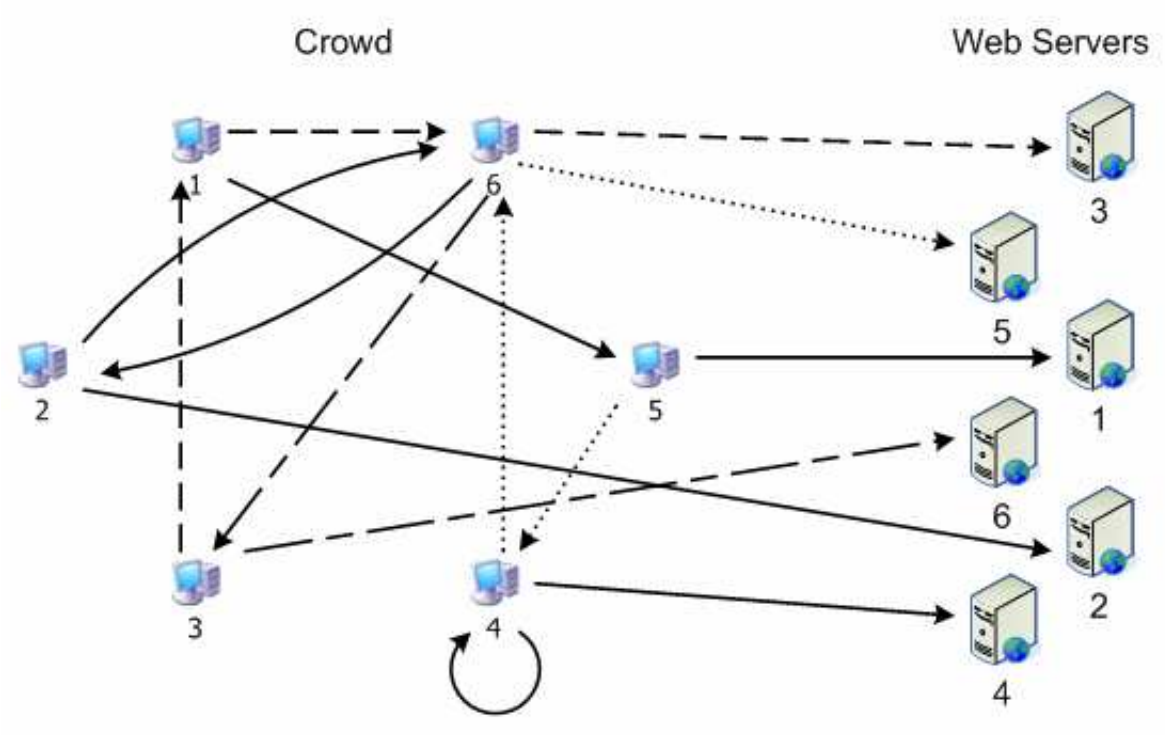

Figura 11. Exemplos de caminhos de comunicação em uma Crowd.

8.2 Pseudônimos (LPWA)

Uma outra abordagem para permitir privacidade ou anonimato e prover certo nível de personalização é o uso de pseudônimos. Esse método consiste basicamente em criar para os usuários nomes fictícios que poderiam disfarçar a identificação real do usuário, o que permite que *sites* possam disponibilizar um serviço personalizado.

Os *sites Web* de *e-commerce* desempenham um papel crescente na Internet na área de comércio eletrônico, o qual possui um grande potencial a ser aproveitado. Entretanto, para que esses *sites* tenham crescimento é necessário que seus clientes os acessem de maneira fácil, segura, anônima e pessoal. Para isso, é necessário que sejam usadas algumas ferramentas que beneficiem ambos, que permitam que os proprietários dos *sites Web* construam seus *sites* de uma forma que possam melhorar suas vendas *on-line*, sem que os usuários tenham sua privacidade prejudicada.

O ideal para a navegação do usuário é a conciliação entre os seus interesses, representados pela privacidade de suas informações, e os interesses dos *sites*, como a coleta de informações para aplicação de personalização e de outras técnicas. Introduzir uma combinação que alie esses dois interesses, navegação anônima e coleta de informação, foi denominado “navegação *Web* personalizada anônima” [8].

O anonimato na navegação *Web* se apresenta como dois aspectos importantes, anonimato de conteúdo de dados e anonimato de conexão. Anonimato de conteúdo de dados significa que a identidade do usuário não é revelada pelo conteúdo de qualquer fluxo de dados entre usuários e *sites Web*, sem deixar de obedecer ao protocolo HTTP. Anonimato de conexão significa que não é possível identificar os usuários pelos aspectos que envolvem suas conexões; a análise desses aspectos pode ser efetuada por *sites* Web, por bisbilhoteiros e por outros intrusos [83].

Atualmente apresentam-se diversas ferramentas que implementam esse mecanismo. Mas *Janus Personalized Web Anonymizer* [8] será apresentado por ser o primeiro sistema existente que permite aos usuários navegarem na *Web* com personalização e anonimato. Esse sistema foi desenvolvido e é conhecido atualmente por *Lucent Personalized Web Anonymizer* [84]. Nessa versão atual, o *proxy* Janus é dividido em três componentes: gerador de pseudônimos, *proxy* de navegação e encaminhador de e-mail. A ferramenta original é detalhada, pois outros sistemas da mesma abordagem seguem a mesma base de sua estrutura e sua arquitetura.

Janus atua como uma entidade intermediária, um *proxy*, entre o usuário e um *site Web*. A privacidade e identificação na navegação dos usuários são viabilizadas pela geração automática de apelidos por esse *proxy*. Tais apelidos permitem ao usuário fazer o *login* em sua conta através de um pseudônimo que esconde a sua verdadeira identidade [8]. Normalmente, esse apelido é um nome de usuário, uma senha e um endereço de e-mail. Cada usuário possui um apelido diferente de qualquer outro usuário, e esse apelido é apresentado toda vez que um determinado usuário deseja acessar um *site* particular.

O usuário, sem nenhuma ferramenta de anonimato e privacidade em sua navegação, cria nos *sites* em que acessa contas com nome de usuário e senha e e-mails que não expressam realmente a sua identidade. Além disso, todas essas contas e e-mails são diferentes uns dos outros. O mecanismo Janus retira do usuário todo esse trabalho de invenção e memorização desses nomes de usuário e senhas seguras para os *sites*. Ele também realiza o trabalho de inserção de tais nomes de usuário e senhas a toda vez que o usuário retorna a algum *site Web*.

Para o usuário poder navegar através do sistema, é necessário que ele inicie uma sessão de navegação no Janus. Para isso, ele deve fornecer uma vez uma única identificação de endereço de e-mail e um segredo. Através dessa entrada do usuário, o sistema Janus gera todos os apelidos durante aquela sessão. O segredo pode ser considerado como a senha universal do usuário para todas as contas de *site Web* do usuário, e não é necessário que esse segredo seja único globalmente para todos os usuários de Janus.

O sistema, por um critério de privacidade, não mantém qualquer informação sobre os seus usuários. Não é possível identificar quem atualmente está em uma sessão de navegação. Apelidos para endereços de e-mail também são fornecidos por Janus. Esses apelidos se apresentam da forma "*alias-email@hostname*", em que "*hostname*" é uma máquina intermediária confiável e que faz parte desse sistema, e "*alias-email*" é uma *string* que esconde a identidade do usuário e permite ao sistema obter o endereço de e-mail real do usuário. Essa troca de e-mail anônima entre um usuário e um *site Web* é realizada pelos mecanismos que Janus fornece [8].

Para assegurar que nenhuma informação comprometedora da privacidade seja transferida, o *proxy* Janus filtra o fluxo de dados do navegador do usuário [8]. Um *site Web* desconhece que um usuário está acessando-o por meio de um *proxy* como esse.

A tradução de nomes é o centro da navegação *Web* personalizada anônima. O endereço de e-mail do usuário e o seu segredo são transformados em um apelido que possa cumprir um número de propriedades, como anonimato, consistência, discrição, exclusividade de apelido e proteção da criação de dossiês (fichas). Uma função criptográfica é usada para

manter o anonimato, e a função Janus é utilizada para realizar o processo de interação personalizada [8].

O *proxy* Janus se localiza em cada máquina de usuário. Esse *proxy* realiza a função de Janus e é responsável também pela filtragem do fluxo de dados do usuário que se comunica com um *site Web*. A filtragem é efetuada nas mensagens HTTP e possivelmente nos *cookies* enviados pelo navegador do usuário. Um navegador do usuário deve ser configurado para conectar ao *proxy* Janus e passar todas as suas mensagens de comunicação através desse *proxy*. A figura 12 [8] exemplifica o funcionamento do sistema Janus, demonstrando uma situação com dois usuários, *Anne Miller* e *John Smith*, e dois *sites Web*, jornal *Wall Street* e *New York Times*.

Acompanhando a exemplificação da figura 12, pode-se verificar que um usuário, quando realiza um primeiro acesso a uma página *Web* que necessita de autenticação, terá sua comunicação intermediada pelo sistema Janus, o qual automaticamente reconhece a situação de *login*. Desse modo, ele responde mostrando no próprio navegador do usuário um formulário de autenticação do usuário no sistema Janus, com campos de nome de usuário e de segredo.

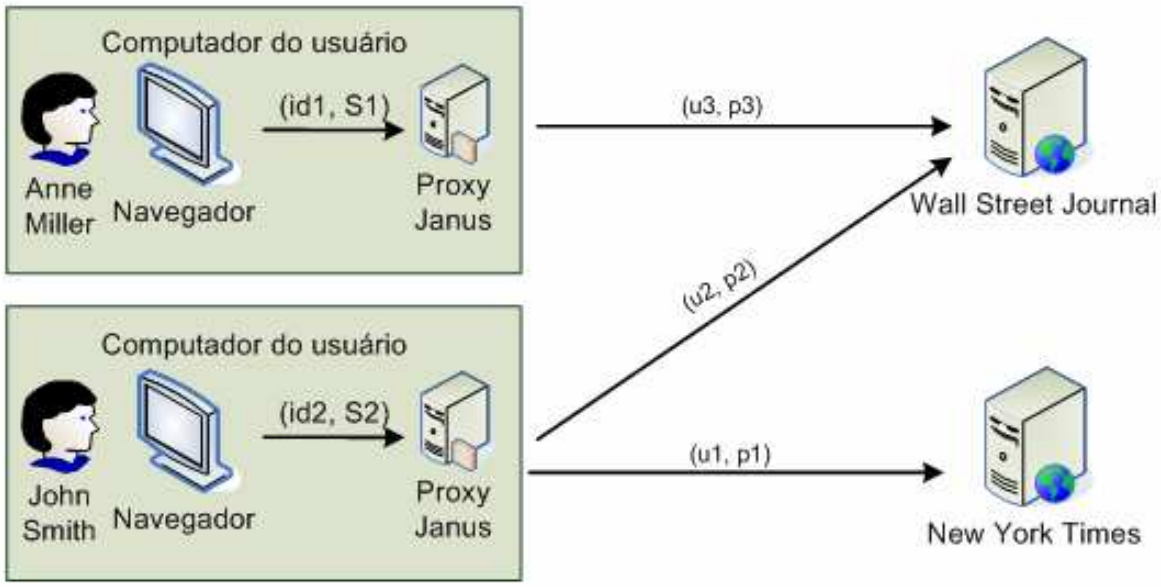

Figura 12. Exemplo de funcionamento do sistema Janus, criação e autenticação automática de contas de usuário.

A autenticação no sistema pelo usuário John com id2 e S2 é efetuada através do envio de dados pelo navegador do usuário para o *proxy* Janus. Visto que essa sessão é iniciada e finalizada pelo usuário, somente lhe são mostradas as páginas requisitadas explicitamente.

Depois de realizar a autenticação do usuário, ele pode continuar a navegação normalmente através do protocolo HTTP, encaminhando a página inicialmente requisitada. Uma requisição de usuário poderá conter informações pessoais no cabeçalho HTTP e *cookies* de suas mensagens, o que leva o sistema Janus a realizar uma filtragem desses pacotes antes de encaminhá-los para os *sites*.

Quando o usuário acessa pela primeira vez um *site* que necessita de autenticação, o *site* disponibiliza uma página de criação de conta, um formulário com nome de usuário, senha e um endereço de e-mail. Desse ponto, o usuário somente insere *strings* de fuga compreensíveis para Janus, como “\U” para nome de usuário, “\P” para senha e “\@” para endereço de e-mail. Com o nome de usuário (id2), o segredo (S2) e a parte do nome do domínio do *site*, o sistema gera um pseudônimo de nome de usuário e pseudônimo de senha para o *site*. No caso de *John Smith*, u1, p1 e o *site New York Times*.

Em visitas subseqüentes de um usuário a *sites* que passaram por esse processo, o usuário responderá sempre as páginas de autenticação com as *strings* de fuga “\U” e ”\P”. Desse modo, o sistema Janus gera e envia automaticamente u1 e p1, no caso de John.

O sistema Janus é construído com o objetivo de fornecer anonimato de conteúdo de dados na navegação *Web*. Se a rede permite rastrear as conexões, então o sistema Janus não tem utilidade. Assim, sugere-se que a comunicação entre um usuário e um *site Web* se faça sobre uma rede de comunicação anônima. Esforços de pesquisa e implementação para tais redes anônimas estão sendo realizados a fim de possibilitar uma maior segurança para a privacidade do usuário [83].

A figura 13 [8] descreve uma configuração do *proxy* Janus. Nela, ele está localizado fora do próprio computador do usuário. Na figura, um conjunto de usuários está localizado em uma *Intranet* confiável atrás de um *firewall* e o *proxy* Janus se localiza no *firewall*. Se o número de usuários atrás do *firewall* é suficientemente grande, então o *proxy* Janus pode se tornar um alvo potencial de ataque de bisbilhoteiros e outros usuários maliciosos.

Na conexão entre um *proxy* Janus e um *site Web*, são transmitidos os apelidos dos usuários. Um bisbilhoteiro que observa essa conexão pode obter um apelido de usuário e então, subseqüentemente, usá-lo para acessar um *site Web* particular e assumir o caráter de outra pessoa.

Assim, é evidente a necessidade de segurança nesse *proxy*, pois a sua corrupção possibilitaria o acesso a uma identidade de usuário ou às identidades dos usuários e seus respectivos pseudônimos.

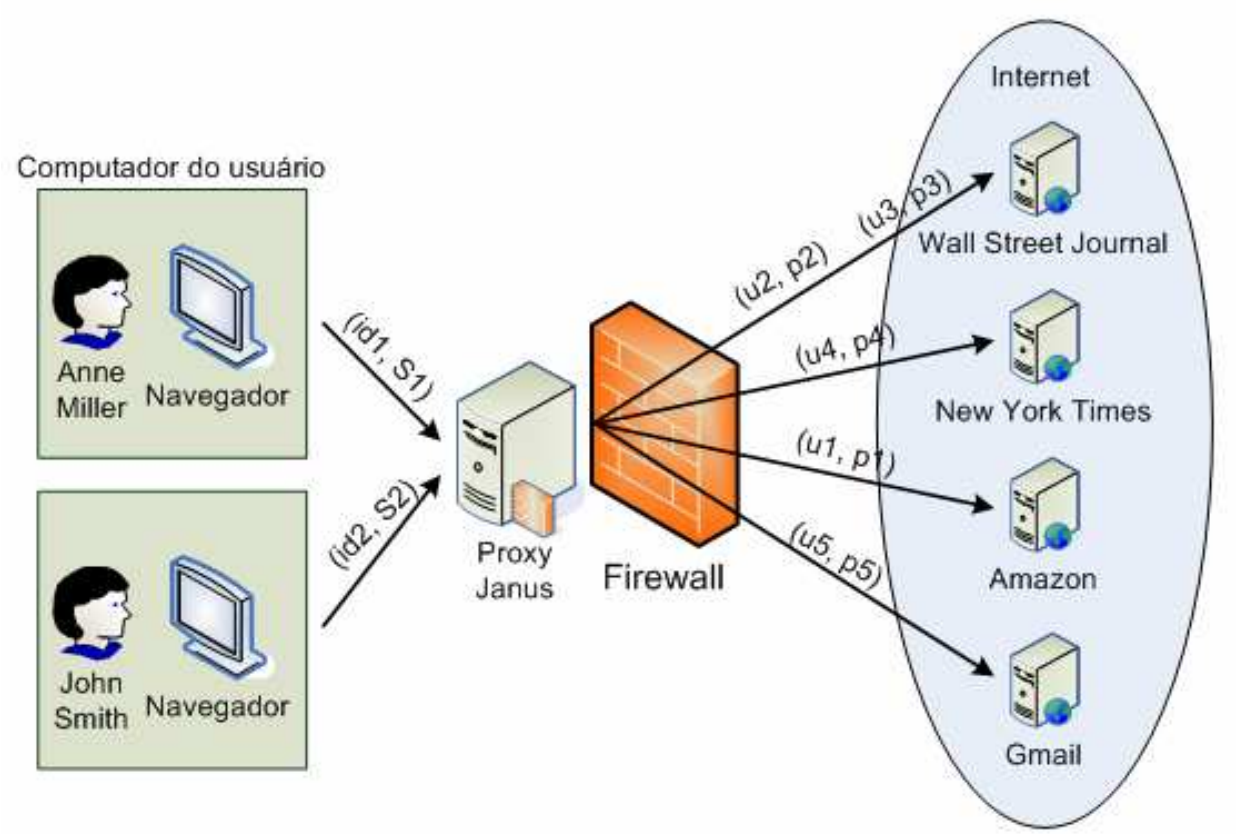

Figura 13. Configuração do sistema Janus com proxy para uma Intranet.

8.3 MASKS

MASKS (*Managing Anonymity while Sharing Knowledge to Servers*) é um mecanismo que introduz uma arquitetura que visa trazer para a navegação do usuário as seguintes funcionalidades: garantir privacidade com anonimato sem deixar de permitir personalização; evitar a armazenagem de informação (o servidor poderia se tornar um alvo para obtenção de informações de seus usuários); tornar flexível a quantidade de informação que o usuário deseja divulgar; adaptar-se aos interesses do usuário a todo momento; permitir a existência dos *cookies*; não inserir nenhuma modificação nos protocolos *Web* existentes, HTTP, TCP, IP e outros (pois isso atrasaria o seu desenvolvimento na *Web*) [35].

Essa arquitetura fornece aos usuários privacidade através da criação de máscaras ou pseudônimos. Um usuário é caracterizado por um perfil de certo grupo de acordo com os seus interesses apresentados durante a sua navegação na *Web*. Esse perfil é representado por uma máscara que é associada a esse usuário.

Detectar o perfil do usuário durante a sua navegação pela *Web* é muito difícil. É impraticável a predição do comportamento do usuário; ele pode alterar os seus interesses

enquanto navega. Assim, o usuário não é designado para um determinado perfil, mas cada requisição do usuário é encaminhada para um determinado grupo que melhor represente os interesses do usuário. Essa atribuição de perfis para cada requisição de página evita que o usuário envie informações suas para identificação no sistema.

A requisição representa o tipo de informação que o usuário está procurando em determinado momento. Somente uma porção da navegação do usuário vai ser mascarada em cada grupo através do destino e da classificação de sua requisição. Desse modo, o usuário não precisa divulgar nenhum tipo de informação pessoal. Os dados privados não são mantidos no servidor do *site*, e um *site* não é capaz de perceber se o acesso a ele é feito por um grupo de interesses similares em vez de um único usuário [53].

O sistema MASKS realiza as seguintes tarefas para atribuir perfis às requisições: a criação de um grupo, a associação das requisições dos usuários a um determinado grupo e a criação das requisições do grupo que substituirão as requisições dos usuários [6].

Cada grupo de interesse pode possuir diversas máscaras, uma para cada acesso ou requisição a um *site*. Os *sites* identificam as requisições que chegam a eles como as de usuários comuns, sem suspeitar que elas estejam caracterizadas como um membro de um grupo de um *proxy* de anonimato. A figura 14 exemplifica o processo de atribuição de grupos de perfil às requisições dos usuários e de atualização de máscaras de grupo no retorno das respostas das requisições. Dessa forma, *sites Web* têm acesso ao padrão de navegação do grupo e são capazes de oferecer serviços personalizados para cada grupo, mas não podem traçar o perfil dos usuários porque eles não têm acesso a todas as requisições de um usuário.

Por outro lado, suponha-se a situação seguinte: o *site* W2 da figura 14 é um portal que oferece diferentes classes de informação, tal como turismo e investimentos; Maria faz uma requisição por informações de investimento (A1) e, depois de algum tempo, faz outra requisição por serviços de turismo (A2). A partir disso, W2 determinaria as duas requisições de Maria (M (A1) e M (A2)) como as de dois usuários distintos, os quais são representados por dois grupos. Nas respostas das requisições, o *site* atualiza os *cookies* para cada perfil identificado e responde para o servidor MASKS, o qual atualiza as máscaras com as novas informações no *cookies* e repassa a página requisitada para o respectivo usuário.

O processo de mascarar os usuários é feito por um pequeno programa chamado PSA (*Privacy and Security Agent*), que atua junto com o navegador de cada usuário. Ele mantém os usuários informados sobre as máscaras que foram escolhidas para eles, permite que eles classifiquem suas próprias requisições e possibilita a interação direta dos usuários com os *sites* [35].

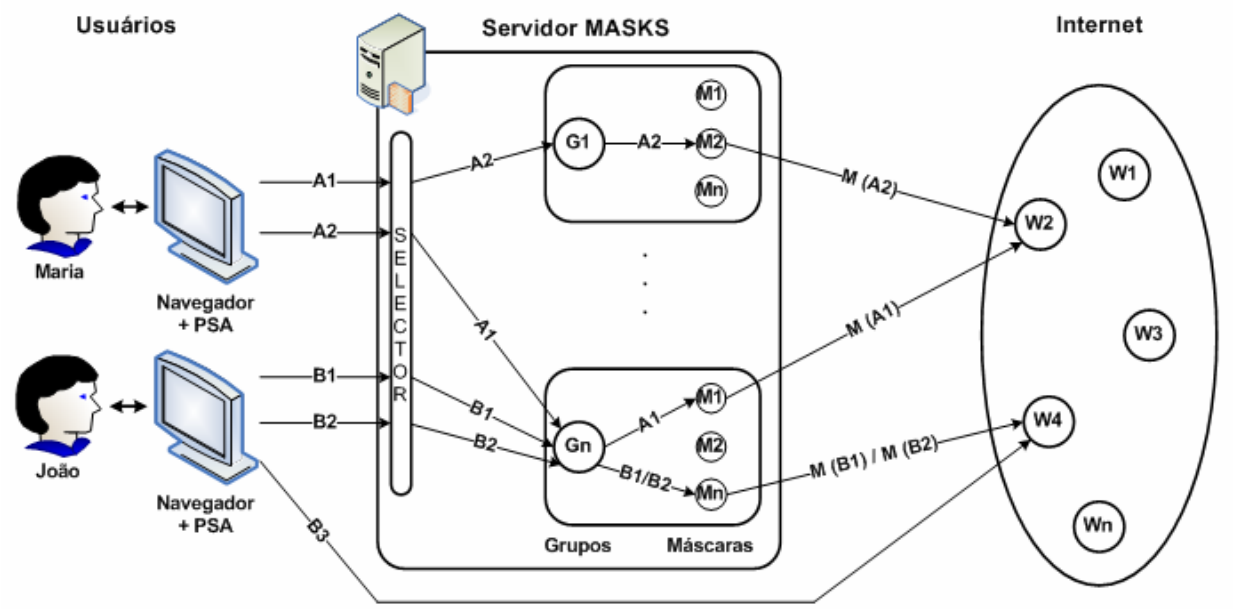

Figura 14. Arquitetura de funcionamento do Sistema MASKS.

Entre o usuário e os *sites Web* há um *proxy* de anonimato, o servidor MASKS. Esse servidor MASKS é o elemento responsável pelo processo de criação de grupos de usuários e de fornecimento de anonimato. Ele apresenta um componente chamado Seletor, o qual é responsável pela seleção de grupos que correspondam de forma mais próxima aos interesses dos usuários em suas requisições [53]. Assim, o servidor MASKS encaminha as requisições de usuários através de uma máscara de grupo apropriada, e na volta, as respostas dos *sites* seguem o mesmo caminho que passa pelos grupos.

Em virtude de a máscara de grupo ser ligada a uma requisição de usuário ou a um interesse de usuário, o usuário poderá possuir várias máscaras, uma para cada interesse que ele apresentar durante a sua navegação na *Web*.

O sistema pode comportar diferentes situações relativas ao processo de mascarar. Requisições que são relacionadas a diferentes *sites* podem ser atribuídas a diferentes máscaras de um mesmo grupo. *Sites* diferentes podem possuir tópicos semelhantes. Geralmente, um usuário possui várias máscaras durante a interação com um *site Web*, cada *site* possui uma máscara específica a um de seus tópicos de interesse. Desse modo, requisições a um mesmo *site* podem ser encaminhadas por diferentes grupos.

Para permitir condições de personalização dos *sites*, os *cookies* são aceitos por MASKS. Além disso, eles são usados no processo de mascaramento das requisições dos usuários e agrupados de acordo com cada interesse em particular.

O sistema MASKS não fornece nenhuma garantia de privacidade quando o usuário deseja ou precisa divulgar explicitamente suas informações pessoais. Com isso, essas informações não podem ser aplicadas a todos os indivíduos do grupo. Entretanto, quando um usuário tem o desejo ou a necessidade de divulgar suas informações pessoais para algum *site*, o que é necessário para muitos *sites* de *e-commerce*, ele não pode considerar invasão de privacidade o comportamento do *site* de coletar seus dados, pois essa sua atitude é realizada de forma consentida.

Entretanto, grande parte da navegação dos usuários se resume em buscas e acessos a documentos. Existem poucos momentos em que eles enviam informações. Dessa forma, nessa grande parte da navegação, MASKS pode atuar e fornecer anonimato ao usuário.

O algoritmo de designação de grupo é a parte mais importante da arquitetura de MASKS. As sessões formadas pelas requisições dos usuários podem ser usadas como base para uma variedade de técnicas de personalização, já que há um risco pequeno de os resultados recuperados de uma base de dados do tipo oferecerem informações errôneas dos interesses do usuário, pois cada grupo é atribuído somente a páginas correlacionadas [35].

O algoritmo de seleção possui a característica de separar as requisições dos usuários de acordo com suas semânticas. A extração tradicional de dados ou os algoritmos de reunião de informação não são apropriados, pois ambos os métodos requerem alguma informação inicial do usuário. Entretanto, o uso da classificação de assunto feita por humanos é vantajoso, pois a classificação semântica de objetos é uma tarefa muito difícil de ser aplicada a técnicas automáticas [53]. Para isso, a árvore de categorias definida por *Open Directory Project* [85] é usada como base do processo de determinação de grupo.

Essa árvore de categorias contém uma listagem de *sites Web* que é organizada em categorias e constantemente atualizada e revisada por editores voluntários por todo o mundo [85]. O projeto de criação dessa árvore é fundamentado nas características do movimento de código aberto[23], e seu uso é totalmente livre. Ela representa um ponto inicial para definir os grupos e seus relacionamentos, exemplificados pela figura 15 [35] em a).

Um nó da árvore representa uma categoria semântica ou um grupo. O nó é formado por um conjunto de páginas relacionadas e um conjunto de termos que caracteriza o tópico referente a um determinado conjunto de máscaras [53]. Assim como um nó de uma árvore, um grupo pode também ter filhos e ligações. Os filhos são especializações semânticas de um grupo e as ligações são especializações que se referem a um grupo pré-existente. Um grupo

[23] http://www.opensource.org

vai possuir um ou mais caminhos, os quais correspondem às seqüências de nós que devem ser visitados na árvore para alcançar esse grupo. Um grupo também possuirá um caminho extra para cada ligação que se refere diretamente a ele [35]. A parte a) da figura 15 ilustra o relacionamento entre os grupos; as setas contínuas indicam filhos, e as pontilhadas indicam ligações.

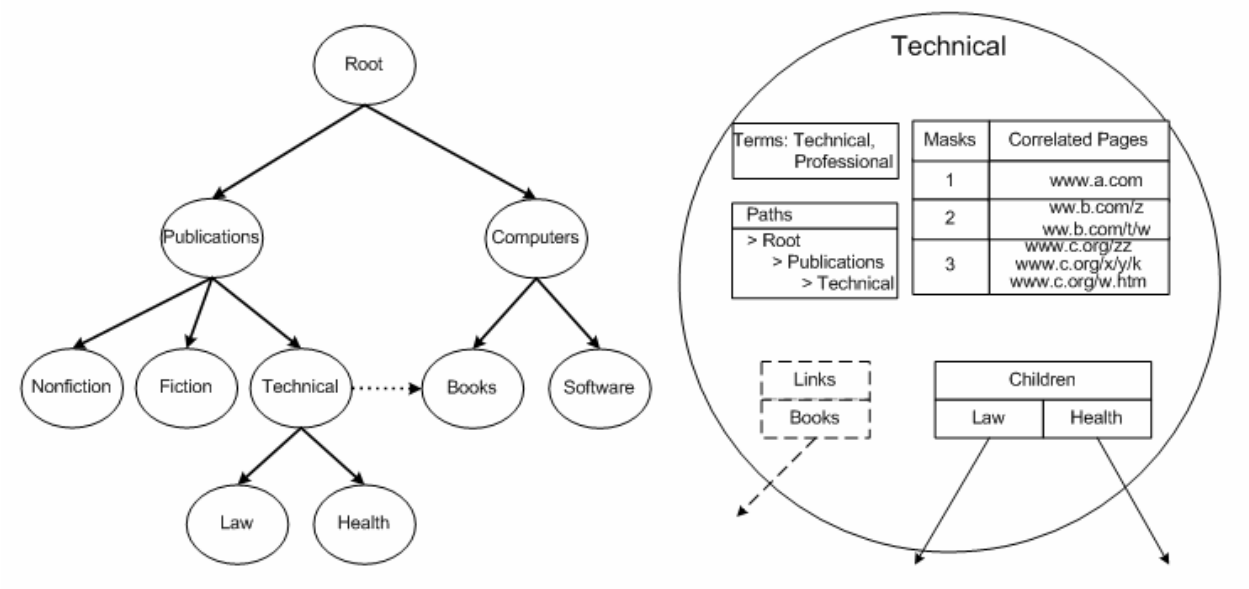

Figura 15. a) Um exemplo de árvore de categoria. b) Os detalhes do nó *Technical*.

Na parte b) da figura, o nó *Technical* é analisado mais detalhadamente. Ele é uma especialização de *Publications* e possui três especializações: *Health* e *Law* em forma de filhos e uma ligação para *Root -> Computers -> Books*. Os termos que identificam o tópico representado por esse nó são "*Technical*" e "*Professional*". Há seis páginas correlacionadas para esse nó e três máscaras distintas, uma para cada *site Web* onde essas páginas são hospedadas [35].

A partir dessa árvore de categorias, podem ser obtidas duas outras estruturas: a tabela de termo e a tabela de conteúdo. A tabela de conteúdo faz a correlação entre as URLs presentes na árvore e seus respectivos grupos. Ela visa simplificar e agilizar o processo de consulta de grupos e endereços de páginas da *Web*. A tabela de termos contém palavras discriminadas no campo "Termos" dos grupos para determinar mais rapidamente o grupo de uma requisição segundo seus termos de consulta [53].

O algoritmo utilizado para escolher o grupo que mascara uma requisição não requer um conhecimento prévio do usuário. Para escolher um grupo, é necessário conhecer a requisição corrente e o grupo para o qual o usuário foi designado, de acordo a última

requisição. Assim, ao empregar essa abordagem, todo usuário pode ter seus interesses associados a um grupo sem a necessidade de armazenar qualquer informação pessoal [35].

Basicamente, o algoritmo escolhe o grupo que representa o tópico da requisição do usuário para designar uma máscara a ela. A análise da URL e a escolha de grupo consistem em procurar primeiramente o grupo da árvore semântica que indexa a URL; se não encontrado, o algoritmo determina o grupo pelos termos de consulta da URL; em caso negativo, indica o grupo de acordo com os termos da URL; por fim, seleciona o grupo raiz. [53].

O mecanismo de MASKS traz anonimato para as informações pessoais dos usuários somente pela divulgação implícita de dados. Além disso, se o algoritmo de atribuição de grupos não estiver bem formulado, a requisição pode não ser designada para o seu respectivo grupo corretamente ou ser sempre designada para o grupo raiz. Desse modo, o usuário não poder receber serviços personalizados. Por esse *proxy* ou servidor MASKS ser o ponto onde se encontra o mecanismo principal de MASKS, ele se torna um ponto de risco. Com isso, é necessário o uso de métodos para aumentar a segurança da comunicação, evitar a invasão de sistemas para não comprometer o servidor e a integridade da privacidade dos usuários e armazenar históricos de requisições de usuários para uso de análise forense de crimes virtuais.

O sistema MASKS apresenta alto desempenho, pois utiliza a árvore semântica para determinar perfis e atribuir máscaras às requisições. Entretanto, o desvio da comunicação do computador do usuário para o servidor de mascaramento e o processamento para gerenciar *cookies* e atribuir perfis acrescentam um tempo nas respostas das requisições que deve ser considerado. Esse tempo acrescido a cada requisição, por ser pequeno e imperceptível para o usuário, não prejudicou sua navegação.

Avaliações empíricas foram realizadas para identificar a quantidade de tempo gasta para realizar o processo de mascaramento e para verificar algum atraso para a navegação do usuário. Nos testes empíricos, foram feitas requisições de página para um *site* criado para testes e que possuía 1 *cookie*. Nas avaliações, o servidor MASKS original, o qual não realiza gerenciamento de *cookies*, acrescentou, em média, para cada requisição, 10,847 milissegundos. Ademais, o servidor MASKS modificado para realizar gerenciamento de *cookies* apresentou um acréscimo de 14,6423 milissegundos para cada requisição de página.

8.4 P3P

O P3P, Projeto de Plataforma para Preferências de Privacidade - (*Plataform for Privacy Preferences Project*), desenvolvido pelo Consórcio da *World Wide Web*[24], propõe fornecer ao usuário uma maneira de ele não ter sua privacidade prejudicada ao acessar algum serviço disponibilizado por algum *site* na *Web*. Esse mecanismo provê maior controle do usuário sobre suas informações por torná-lo consciente das ações do *site*, e, conseqüentemente, capaz de consentir a coleta de seus dados.

Essa plataforma de privacidade visa disponibilizar aos usuários preferências de privacidade, para que o comportamento de um agente de usuário seja moldado de acordo com as necessidades de privacidade de cada usuário, sem prejudicar os serviços prestados pelos *sites*. Pela introdução dessas preferências de privacidade do usuário, a agente de usuário P3P se adapta melhor à subjetividade do conceito de privacidade.

Em uma navegação sem a presença do P3P, todo usuário que se preocupa com a sua privacidade deve procurar as políticas de privacidade de cada *site* que ele visita e deve observar as práticas de privacidade do *site* com relação às suas próprias preferências. O trabalho que todo usuário teria em analisar as políticas de privacidade de cada *site* ou de cada página seria muito grande e competiria com os propósitos de navegação do usuário. Por isso, o P3P visa tornar automático esse mecanismo de leitura e de aceitação das políticas de privacidade dos *sites* [86].

O objetivo do P3P é permitir aos *sites Web* apresentarem suas práticas de coleta de dados de uma maneira padronizada, fácil de ser localizada e capaz de ser lida pelo computador. Assim, ele visa permitir que usuários entendam qual informação será coletada pelos *sites* que eles visitam e como essa informação será usada.

O P3P foi desenvolvido para dar um formato padrão de leitura às políticas de privacidade e para inserir um protocolo que capacite os navegadores da *Web* a processarem essas mesmas políticas de forma automática. Ele permite que as políticas de privacidade possam ser encontradas automaticamente por ferramentas de agente de usuário e possibilita aos agentes utilizar essa plataforma para informar os usuários através de símbolos ou tomar outras ações apropriadas.

Embora o P3P disponibilize um mecanismo técnico para assegurar que os usuários possam ser informados sobre políticas de privacidade antes que eles disponibilizem alguma

[24] http://www.w3.org/

informação pessoal, ele não fornece um método que assegure que o comportamento dos *sites* esteja de acordo com suas políticas, que mascare requisições ou que garanta anonimato. Ferramentas podem implementar essa especificação apresentando alguma assistência para isso. Nessa mesma linha, o P3P pode ser considerado como um complemento e um mecanismo de reforço a leis e programas de auto-regulamentação.

O protocolo introduzido pelo Projeto 3P é projetado em um formato XML, conhecido como política P3P de privacidade. Ele tem o papel de informar os usuários sobre as práticas de coleta de dados dos *sites* da *Web* e sobre o uso dado a essas informações obtidas. A especificação do P3P define: um esquema padrão para dados que *sites Web* podem coletar; um conjunto de padrões de uso, "receptores", categoria de dados e outras divulgações de privacidade; um formato XML para expressar uma política de privacidade; uma maneira de associar políticas de privacidade com páginas *Web* e *cookies*; e um mecanismo de transporte de políticas P3P sobre o protocolo HTTP [74].

O P3P pode ser implementado pelos *sites Web* em seus servidores através da tradução de suas políticas de privacidade escritas numa linguagem humana para a sintaxe P3P. No final, cria-se um ou mais arquivos de texto que contém suas políticas de privacidade traduzidas para essa sintaxe no formato XML [86]. Depois de publicar esses arquivos resultantes, um arquivo de referência da política é colocado junto com as políticas para indicar para quais partes do *site* que elas serão aplicadas. Para auxiliar operadores a desempenhar essa tradução das políticas de privacidade para um formato padrão, existem diversas ferramentas automáticas.

A implementação do P3P pode ser feita em servidores *Web* que estão compatíveis com o protocolo HTTP/1.1 existente, sem requerer software adicional ou alguma atualização [74]. Servidores podem publicar seus arquivos de referência da política em conteúdo HTML/XHTML através do uso de um *link tag* para indicar algum arquivo de referência. Alternativamente, servidores compatíveis podem ser configurados para inserir um cabeçalho de extensão P3P em todas as respostas HTTP e indicar a localização do arquivo de referência da política P3P de um *site* [74].

Proprietários de *sites Web* têm certa flexibilidade no modo de usar P3P. Eles podem optar por uma política P3P para todo o *site* ou podem designar diferentes políticas para diferentes partes de seus *sites*. Uma política P3P deve cobrir toda informação gerada ou trocada como parte de uma interação HTTP de um *site* com os visitantes.

Para que esse arquivo que contém as políticas do *site* possa ser lido e construído, é necessário que ele esteja sob uma notação formal. No caso, a linguagem XML, uma ABNF

[87] é criada e é utilizada. Ela é uma gramática representativa usada para acrescentar e melhorar a capacidade de representação proporcionada pela sintaxe XML. Todas as sintaxes XML definidas pela especificação do P3P devem seguir o Schema XML para o P3P, o qual, junto com outras restrições expressas em linguagem natural em sua especificação, constitui uma definição normativa [74].

8.4.1 Arquivo de Referências de Políticas P3P de privacidade

Localizar uma política P3P é um dos primeiros passos na operação do protocolo P3P. Serviços usam referências de política para relatar quais políticas se aplicam a um específico endereço ou conjunto de endereços de algum recurso, como páginas, figuras e outros elementos. Agentes de usuário usam referências de política para localizar a política de privacidade que se aplica a um recurso *Web*, pois eles podem processar a política em benefício de seus usuários.

Referências de políticas são usadas extensamente como uma otimização de desempenho. Elas facilitam a busca de políticas P3P, a transferência de dados e o processamento de análise das políticas [74]. Essas referências também reduzem a necessidade de computação; as políticas podem ser unicamente associadas com endereços, e o agente de usuário somente analisa e processa uma política uma vez, em vez de invés de processá-la para todo o documento que a política cobre. O arquivo de referências é um centralizador para a análise das políticas P3P.

Um arquivo de referência de política é usado para associar políticas P3P a certas regiões de endereçamento. O arquivo de referência de política é um arquivo XML com nomes espaçados que pode especificar a política para um único documento *Web*, para porções de um *site Web* ou para um *site* inteiro [74]. O arquivo de referência pode referir a uma ou mais políticas P3P; isso permite que um único arquivo possa cobrir um *site* inteiro, mesmo que diferentes políticas P3P se apliquem a diferentes porções do *site*.

Essas referências delimitam a abrangência das políticas P3P através de indicações do endereço onde a política P3P se encontra, o endereço ou regiões de espaço de endereçamento do *site* coberto por uma política, o endereço ou regiões de espaço de endereçamento do *site* não coberto pela política, as regiões de espaço de endereçamento para conteúdo embutido em outros servidores que estão cobertos pela política, os *cookies* que estão e que não estão

cobertos pela política, os métodos de acesso para os quais essa política é aplicável e o período de tempo para o qual essas declarações são consideradas válidas [74].

Entretanto, é necessário que o agente do usuário seja informado sobre a localização desse arquivo de referências. Dessa forma, a especificação da plataforma de privacidade apresenta basicamente três maneiras de localizar esse arquivo [74].

O arquivo de referências pode estar em uma localização bem conhecida ou pré-definida. No caso, é sugerido que o arquivo de referência de política esteja disponível no *site* pelo caminho */w3c/p3p.xml* [74]. Desse modo, as políticas P3P estarão acessíveis aos agentes de usuários antes que qualquer outro recurso seja requerido do *site*. O fluxo de dados de obtenção do arquivo de referências é exemplificado pela figura 16.

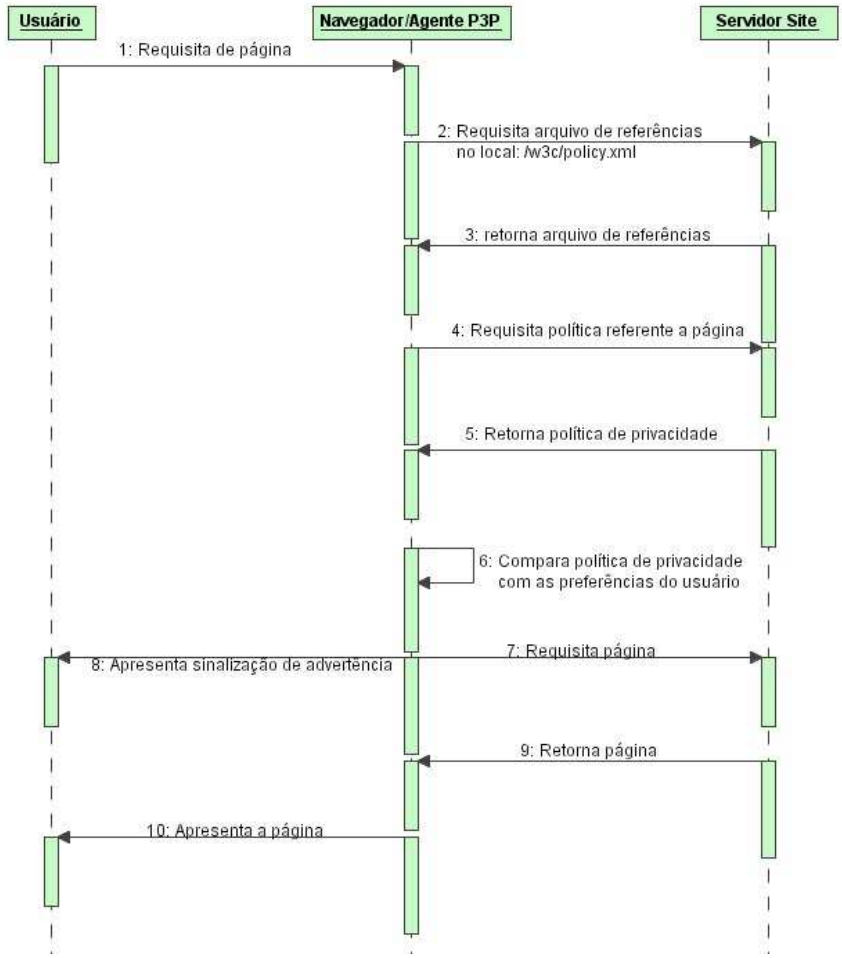

Figura 16. Diagrama de Seqüência com busca do arquivo de referências por um lugar bem conhecido.

O arquivo de referências de política pode ser indicado por qualquer documento trazido pelo protocolo HTTP através da construção de um novo cabeçalho de resposta, o cabeçalho P3P. Se um *site* está usando cabeçalhos P3P, ele deve incluir isso na resposta para todos os

métodos apropriados de requisição. Esse cabeçalho contém uma ou mais diretivas separadas por vírgulas para designar o local onde se encontra o arquivo. Numa consulta à diretiva *policyref* do cabeçalho P3P, obtém-se o endereço que especifica a localização de um arquivo de referência de política.

Por exemplo [74]:

O cliente faz uma requisição GET para o *site catalog.example.com*:

```
GET /index.html HTTP/1.1
Host: catalog.example.com
Accept: */*
Accept-Language: de, en
User-Agent: WonderBrowser/5.2 (RT-11)
```

O servidor retorna o conteúdo e o cabeçalho P3P apontando para a política do recurso.

```
HTTP/1.1 200 OK
P3P: policyref="http://catalog.example.com/P3P/PolicyReferences.xml"
Content-Type: text/html
Content-Length: 7413
Server: CC-Galaxy/1.3.18
```

A maneira como o arquivo de referências de políticas de privacidade pode ser obtido através desse mecanismo é exemplificada na figura 17.

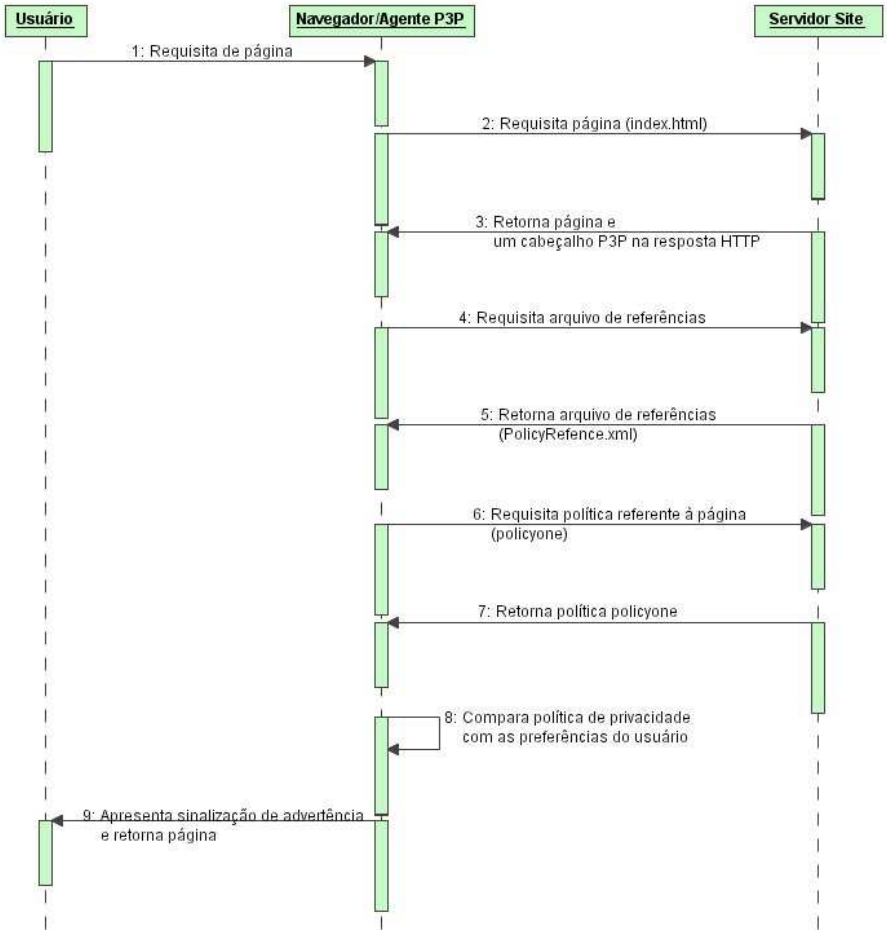

Figura 17. Diagrama de Seqüência com busca do arquivo de referências pelo cabeçalho P3P.

Outra forma de se encontrar o arquivo de referência é fazendo-se indicações dentro do código HTML. Servidores podem servir conteúdo HTML com *link tags* embutidos, os quais informam a localização do arquivo de referências de política P3P relevante. Esse uso da P3P não requer qualquer mudança no comportamento do servidor. O *link tag* codifica a informação de referência de política que poderia ser expressa usando o cabeçalho P3P. O atributo *href* do *link tag* delimita o endereço onde se encontra o arquivo de referências. Esse atributo embutido em um documento HTML apresenta a seguinte codificação:

<link rel="P3Pv1" href="http://catalog.example.com/P3P/PolicyReferences.xml">

O fluxo de dados para esse mecanismo de obtenção do arquivo de referências de políticas de privacidade é exemplificado na figura 18.

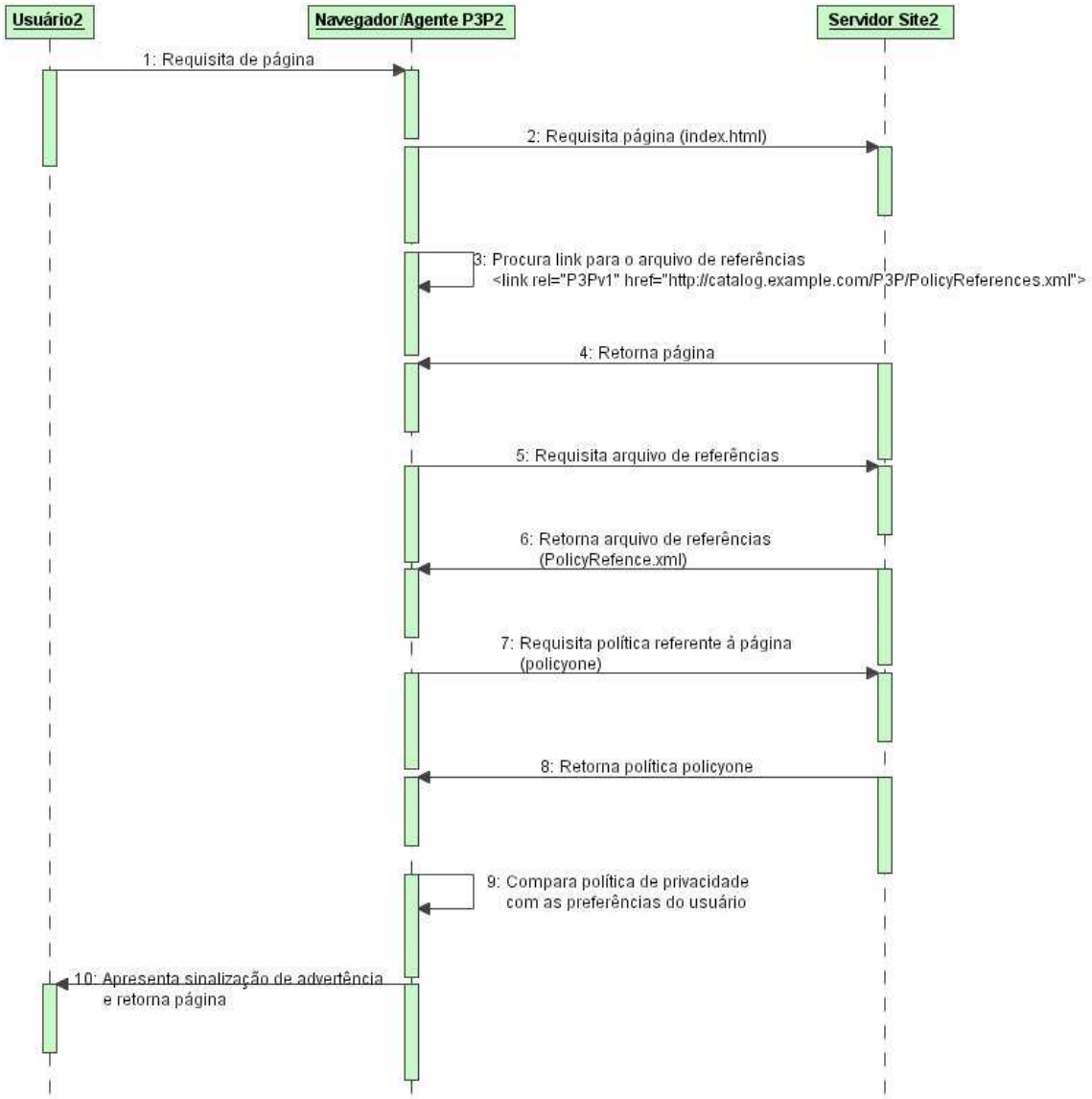

Figura 18. Diagrama de Seqüência com busca do arquivo de referências pela *link tag* de um código HTML.

Os mecanismos descritos acima são usados para transações do protocolo HTTP sobre qualquer protocolo do nível de transporte, o que inclui as conexões TCP/IP, SSL ou de qualquer outro protocolo que possa ser projetado.

Assim como as políticas de privacidade, as referências também possuem uma estrutura para a sua construção; elas utilizam certa sintaxe e semântica para designar as políticas para as devidas partes do *site*. Para isso, diretivas são definidas e apresentadas na tabela 2.

Tabela 2. Diretivas para construção do arquivo de referências.

<META>	identifica o arquivo de referência de políticas
<POLICY-REFERENCES>	contém um ou mais elementos de referência e pode conter também um tempo de expiração delas
<EXPIRE>	delimita o tempo, absoluto ou relativo, de duração das referências de políticas
<POLICY-REF>	descreve atributos, localização da política e área de cobertura, para uma ou mais políticas P3P
<INCLUDE> e <EXCLUDE>	usados para especificar a porção do *site* Web que é coberto por uma política referenciada pela diretiva <POLICY-REF>
<HINT>	usado para identificar outras referências de políticas além das contidas no arquivo através dos atributos de escopo (escope) e caminho (path)
<COOKIE-NCLUDE> e <COOKIE-EXCLUDE>	usados para associar políticas aos cookies
<METHOD>	usado para relatar que a aplicação da política será feita somente quando o acesso ao recurso é feito através de algum determinado método

A figura 19 contém um exemplo [74] de arquivo de referências de política:

8.4.2 Políticas P3P de privacidade

As políticas na P3P consistem em indicações feitas com base no vocabulário da P3P para expressar práticas de privacidade de cada *site*. Essas indicações são feitas por uma codificação XML com nomes espaçados para fornecer informação de contato à entidade legal, fazer a representação das práticas de privacidade em uma política, enumerar os tipos de dados ou elementos de dados coletados e explicar o uso que será designado à informação coletada.

A especificação inclui um mecanismo para definição de novos elementos de informação e conjunto de informação e, assim, permite extensões do vocabulário P3P. Porém, ela não é a única especificação para a construção de políticas de privacidade em uma forma padronizada. A linguagem formal EPAL [88] foi criada pela IBM como uma linguagem paralela para especificação das políticas. Além disso, Stufflebeam et al. [88] afirmam que,

nessa linguagem, as políticas são reforçadas por uma máquina de coação que assegura a coleta de informação.

```
<META xmlns="http://www.w3.org/2002/01/P3Pv1">
 <POLICY-REFERENCES>
  <EXPIRY max-age="172800"/>

    <POLICY-REF about="/P3P/Policies.xml#first">
      <INCLUDE>/*</INCLUDE>
      <EXCLUDE>/catalog/*</EXCLUDE>
      <EXCLUDE>/cgi-bin/*</EXCLUDE>
      <EXCLUDE>/servlet/*</EXCLUDE>
    </POLICY-REF>

    <POLICY-REF about="/P3P/Policies.xml#second">
      <INCLUDE>/catalog/*</INCLUDE>
    </POLICY-REF>

    <POLICY-REF about="/P3P/Policies.xml#third">
      <INCLUDE>/cgi-bin/*</INCLUDE>
      <INCLUDE>/servlet/*</INCLUDE>
      <EXCLUDE>/servlet/unknown</EXCLUDE>
    </POLICY-REF>

    <POLICY-REF about="/P3P/Policies.xml#first">
       <COOKIE-INCLUDE name="*" value="*" domain="*" path="*"/>
       <COOKIE-EXCLUDE name="obnoxious-cookie" value="*"
domain=".example.com" path="/"/>
    </POLICY-REF>

    <POLICY-REF about="/P3P/Policies.xml#first">
      <INCLUDE>/docs/*</INCLUDE>
      <METHOD>GET</METHOD>
      <METHOD>HEAD</METHOD>
    </POLICY-REF>

    <HINT scope="http://www.example.org" path="/mypolicy/p3.xml" />
    <HINT scope="http://www.example.net:81" path="/w3c/prf.xml" />

 </POLICY-REFERENCES>
</META>
```

Figura 19. Exemplo de um arquivo de referências de políticas P3P de privacidade.

Essas políticas identificam os receptores de dados e fazem uma variedade de outras divulgações, o que inclui os dados sobre a resolução de discussões e sobre o endereço para um texto de uma política de privacidade do *site* para a leitura dos usuários. As políticas P3P devem cobrir todos os elementos de informação relevantes e as atitudes do *sites* respectivas a eles [74]. Declarações P3P devem ser utilizadas pelos *sites* para expressarem suas atitudes com as informações de seus usuários.

O vocabulário do projeto 3P não é planejado para indicar obediência a uma lei particular ou a um código de conduta. Apesar disso, as políticas de privacidade podem ser

construídas para serem analisadas por agentes de usuário desenvolvidos para testar se as práticas de um *site* estão de acordo com leis e códigos específicos.

Em casos em que o vocabulário P3P não é preciso o bastante para descrever as práticas de privacidade, ele deve ser utilizado para melhor representar o comportamento do *site* Web, sem fazer declarações falsas ou equivocadas.

Mais de uma política de privacidade podem ser utilizadas para delimitar as práticas de privacidade de um *site*, visto que em cada política especificam-se: o órgão responsável pelo *site* que está obtendo as informações; o órgão que garante a política de privacidade em caso de discussões; a delimitação das informações que estão sendo coletadas pelo *site*; o lugar onde as informações coletadas vão ser armazenadas; o propósito da obtenção das informações através do atributo *purpose*, como para propagandas, para uso próprio no gerenciamento e para auxiliar na compra de mercadorias feita pelo usuário; o uso que vai ser dado à informação coletada, através da especificação de uma a doze maneiras predefinidas de usos da informação; quem usará os dados pessoais obtidos em seis tipos diferentes de companhias ou pessoas; e o direito de acesso à informação coletada do usuário.

Quando um agente de usuário resgata essas políticas, ele precisa analisá-las para advertir o usuário. Para ser possível analisar as políticas e identificar cada um de seus itens, é necessário que elas sejam definidas por um vocabulário que segue uma sintaxe e uma semântica padronizadas. Assim, são apresentadas na tabela 3 algumas diretivas definidas para possibilitar essa tradução das políticas [74].

Tabela 3. Diretivas para construção das políticas P3P de privacidade.

<POLICIES>	reúne uma ou mais políticas identificadas por um nome e podem conter um atributo xml:lang, um elemento <EXPIRY> relacionado à referência que foi feita ou um elemento <DATASCHEMA>
<POLICY>	contém uma definição completa de uma política com os atributos nome (name), endereço da política em linguagem natural (discuri), endereço para instruções de quando o dados do usuário foram usados por algum determinado propósito (optcuri) e a linguagem em que a política foi escrita (xml:lang)
<TEST>	usado para propósitos de teste
<ENTITY>	informa precisamente a entidade legal com o nome, endereço postal, número de telefone, endereço de email, endereço virtual
<ACCESS>	indica se o *site* provê acesso a vários tipos de informação aos usuários (<nonident/>, <all/>, <contact-and-other/>, <ident-contact/>, <other-ident/>, <none/>)
<DISPUTES>	descreve procedimentos de resolução de discussão que podem com relação a práticas de privacidade de serviços ou violação de protocolo (resolution-type, service, verification, short-description, <LONG-DESCRIPTION>, <IMG> , src, width, height, alt)

<REMEDIES>	especifica as possíveis ações no caso de ocorrer uma infração de política.(<correct/>, <money/> , <law/>)
<STATEMENT>	descreve práticas de privacidade que são aplicadas a tipos particulares de dados e que são constituídas pelos elementos <PURPOSE>, <RECIPIENT>, <RETENTION>, <DATA-GROUP> e <CONSEQUENCE>
<CONSEQUENCE>	define uma explicação do *site* para o uso da informação
<NON-IDENTIFIABLE/>	define se não há dado coletado por esse <STATEMENT>
<PURPOSE>	define um ou mais propósitos para o processamento ou a coleta dos dados
<RECIPIENT>	contém um ou mais recipientes (entidades legais ou domínios) dos dados coletados
<RETENTION>	indica o tipo de retenção que se aplica aos dados definidos na declaração
<DATA-GROUP>	descreve os dados transferidos ou inferidos contendo um ou mais elementos <DATA> (são usados para descrever o tipo de dados que o *site* coleta)
<CATEGORIES>	usados nos elementos de dados para prover dicas para usuários e seus agentes como usos pretendidos dos dados
<EXTENSION>	descreve uma extensão para a sintaxe P3P

A figura 20 contém um exemplo [74] de política P3P de privacidade.

Além disso, existem políticas de privacidade compactas. Elas são políticas P3P resumidas que provêem pistas para os agentes de usuários realizarem decisões rápidas e sincronizadas sobre a aplicação da política. Essas políticas possuem de forma compacta as mesmas diretivas que as políticas P3P normais, sem detalhar algumas informações. Para cada diretiva normal, apresentam-se somente os atributos abreviados para formar as diretivas das políticas compactas. No caso de <ACCESS>, haverá as diretivas NOI, ALL, CAO, IDC, OTI, NON que são os seus atributos.

As políticas de privacidade são obtidas e analisadas por agentes P3P de usuários que atuam junto com os navegadores.

8.4.3 Agente P3P de usuário (AT&T Privacy Bird)

Os agentes de usuário do P3P podem ser construídos nos navegadores de *Web*, como *plug-ins* de navegadores ou como servidores *proxy*. Eles podem ser implementados como *Java aplets* ou *Javascript* ou outras ferramentas de gerenciamento de dados do usuário [74]. Os agentes procuram referências à política P3P em lugar bem conhecido, em cabeçalhos P3P de respostas HTTP e em links P3P colocados em conteúdo HTML. Eles obtêm a política no

lugar indicado, interpretam-na e disponibilizam o resultado da análise das práticas de privacidade P3P do *site* em forma de sinais e símbolos.

```
<POLICIES xmlns="http://www.w3.org/2002/01/P3Pv1">
 <POLICY name="forBrowsers"

discuri="http://www.catalog.example.com/PrivacyPracticeBrowsing.html"
     xml:lang="en">
  <ENTITY>
   <DATA-GROUP>
    <DATA ref="#business.name">CatalogExample</DATA>
    <DATA ref="#business.contact-info.postal.street">4000 Lincoln
Ave.</DATA>
    <DATA ref="#business.contact-info.postal.city">Birmingham</DATA>
    <DATA ref="#business.contact-info.postal.stateprov">MI</DATA>
    <DATA ref="#business.contact-info.postal.postalcode">48009</DATA>
    <DATA ref="#business.contact-info.postal.country">USA</DATA>
    <DATA ref="#business.contact-
info.online.email">catalog@example.com</DATA>
    <DATA ref="#business.contact-
info.telecom.telephone.intcode">1</DATA>
    <DATA ref="#business.contact-
info.telecom.telephone.loccode">248</DATA>
    <DATA ref="#business.contact-
info.telecom.telephone.number">3926753</DATA>
   </DATA-GROUP>
  </ENTITY>
  <ACCESS><nonident/></ACCESS>
  <DISPUTES-GROUP>
   <DISPUTES resolution-type="independent"
     service="http://www.PrivacySeal.example.org"
     short-description="PrivacySeal.example.org">
    <IMG src="http://www.PrivacySeal.example.org/Logo.gif"
alt="PrivacySeal's logo"/>
    <REMEDIES><correct/></REMEDIES>
   </DISPUTES>
  </DISPUTES-GROUP>
  <STATEMENT>
   <PURPOSE><admin/><develop/></PURPOSE>
   <RECIPIENT><ours/></RECIPIENT>
   <RETENTION><stated-purpose/></RETENTION> <!-- Note also that the
site's human-readable privacy policy MUST mention that data is purged
every two weeks, or provide a link to this information. -->
   <DATA-GROUP>
    <DATA ref="#dynamic.clickstream"/>
    <DATA ref="#dynamic.http"/>
   </DATA-GROUP>
  </STATEMENT>
 </POLICY>
</POLICIES>
```

Figura 20. Exemplo de uma política P3P de privacidade.

Através da política de privacidade, é realizada a comparação entre as práticas de privacidade e as preferências feitas pelo usuário. De acordo com o resultado obtido, é possível tomar ações apropriadas. O agente do usuário P3P vai autorizar a liberação de dados somente se a política é consistente com as preferências do usuário e se a requisição de transferência de

dados é consistente com a política. Se alguma dessas condições não é seguida, o usuário pode ser informado sobre a discrepância e pode opinar em autorizar a troca de informação.

A interface do agente do usuário é especificada com poucos requisitos. Aqueles que a implementarem podem fazer suas próprias escolhas sobre quais palavras e símbolos apresentar para os usuários ao divulgar informações sobre uma política de privacidade de um *site Web*. Dessa forma, a implementação do agente deve obedecer à especificação da plataforma para preferências de privacidade, mas ela permite que se insiram novas funcionalidades, como as de segurança.

Para o agente de usuário possuir um funcionamento correto, é necessário que o usuário o configure de acordo com suas preferências de privacidade. O agente é regido segundo essas preferências de privacidade. Para todo *site* visitado, procura-se um arquivo que contenha as referências de suas políticas de privacidade; encontrando-o, sabe-se onde se encontram as políticas de privacidade e a quais partes do *site* cada política se aplica. Ao comparar essas políticas com as preferências do usuário, pode-se permitir a entrega dos dados pessoais ao *site* ou deixar para o usuário resolver por conta própria.

Um exemplo comum de execução de um agente pode ser observado na figura 21. O usuário faz uma requisição de uma página de um determinado *site*. Para essa requisição, o agente do usuário solicitará os arquivos de políticas de privacidade da página. A resposta com os arquivos é enviada, e com eles faz-se uma verificação da política da página requisitada com relação às preferências do usuário. Ao final, a requisição inicial do usuário é atendida encaminhando-a para o respectivo servidor.

Entretanto, para acessar a política de privacidade, o agente de usuário deve realizar um acesso no servidor que contém a política P3P. Esse acesso é independente de como é informado sobre a localização do arquivo de referências de políticas, o que pode ser registrado pelo servidor e utilizado para gerar alguma informação. Como não é possível evitar o registro desse acesso antes da apuração da política de privacidade pelo agente e pelo usuário, então é necessário que o *site* não considere o acesso às referências e políticas P3P para não prejudicar a privacidade do usuário.

Como exemplo de caso de uso do P3P, é usado um plugin criado pela AT&T denominado *Privacy Bird*[25]. Ele está disponível para *download* na página *www.privacybird.com*, onde também pode ser encontrado seu código fonte. Entretanto, esse *plugin* é uma versão beta para testes. Ele tem enfoque inicial de pesquisa para conhecer a

[25] http://www.privacybird.com

aceitação do usuário com relação ao seu uso. Assim, algumas funcionalidades mais específicas não se apresentam.

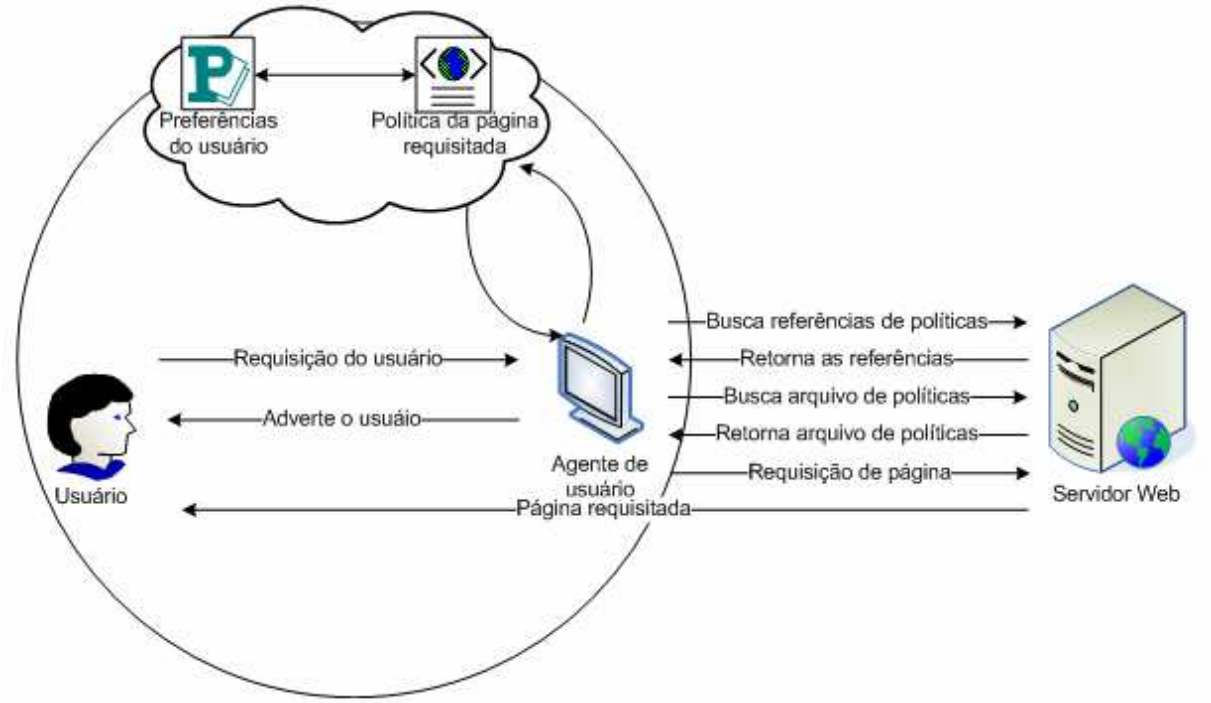

Figura 21. Exemplo de uma requisição de usuário coberta pela plataforma P3P.

Segundo a especificação do agente AT&T *Privacy Bird*, ele foi desenvolvido como um agente de usuário P3P que pode comparar políticas P3P com as preferências do usuário. Desde que P3P foi adotado como uma recomendação W3C em Abril de 2002, pouco trabalho foi feito para estudar como ele está sendo usado e, especialmente, seu impacto nos usuários. Descobriu-se que uma grande proporção de usuários de AT&T *Pribacy Bird* começou a ler políticas de privacidade mais freqüentemente e a ser mais ativa com relação à proteção de suas privacidades como um resultado do uso desse software.

A utilidade dos agentes de usuário P3P é severamente limitada pelo número de *sites Web* que têm implementado P3P. Os resultados de pesquisas também sugerem que, ao facilitar a comparação de política de privacidade através de *sites Web* de *e-commerce*, um grupo significante de consumidores usaria essa informação em suas decisões de compra [89].

O Projeto 3P apresenta limitações que interferem na confiança do usuário [53]. A especificação do P3P somente insere um contrato de privacidade para análise do agente de usuário. Ela não oferece nenhuma garantia com relação a práticas maliciosas de alguns *sites*. Um *site* pode delimitar políticas P3P de privacidade que divergem do comportamento real do

site. Dessa forma, o uso de certificados de privacidade incrementaria a confiança nas políticas ao assegurar que elas estão corretas de acordo com princípios de privacidade.

Kobsa evidencia outras falhas que o P3P apresenta [90]. Ele requer que usuários façam decisões de privacidade antecipadamente e sem relacioná-las a circunstâncias específicas de um determinado contexto. O sistema não informa sobre os benefícios de fornecer os dados requisitados. Ele não acrescenta entendimento do usuário sobre colocações básicas de privacidade.

8.5 Método de fornecimento de informação relacionada a contexto para E-Business

Uma outra abordagem na linha de notificação do usuário sobre as práticas de privacidade de um *site* é um método que visa fornecer informação relacionada a contexto sobre opções de privacidade e personalização [50]. Esse método aconselha os *sites* a construírem um sistema de suporte de navegação do usuário, o qual exibe mensagens sobre as práticas de privacidade e os benefícios de personalização em cada situação.

Conceitos de privacidades são importantes para os usuários tomarem decisões sobre a divulgação de informações pessoais ou sobre o acesso a determinados serviços. Em muitos casos, os usuários podem deixar de acessar certos serviços, devido ao fato de *sites* não apresentarem com clareza suas práticas de privacidade.

Cerca de 90% dos usuários querem que os serviços acessados procurem permissões antes que eles usem suas informações pessoais para o marketing [91]. Uma outra pesquisa mostrou que 76% dos usuários acreditam que as políticas de privacidade são muito importantes [92] e 55% deles declaram que uma política de privacidade torna mais reconfortante a divulgação de informações [93]. Entretanto, segundo essas pesquisas, as políticas de privacidade são escritas de uma maneira complexa para serem entendidas. Garantias de proteção de privacidade são vantajosas para *sites Web* de personalização, pois essa personalização requer informações detalhadas dos usuários, o que leva a altos riscos de privacidade [94].

Devido à complexidade das políticas de privacidade, apresenta-se uma necessidade de criação de métodos mais avançados para comunicar aos usuários sobre as práticas de privacidade e os benefícios de personalização.

A atual abordagem predominante para comunicação de práticas de privacidade de visitantes de *sites Web* é o Projeto 3P. A atual taxa de adoção dele se estagnou nos 30% dos 100 maiores *sites Web*. Isso é devido à baixa adoção por apresentar problemas com implicações legais e com a falta de suporte aos usuários na avaliação das políticas P3P [95].

A partir dos princípios da Diretiva Européia de Proteção de Dados, quatro diretrizes podem ser derivadas para criação de um projeto efetivo para interface de privacidade: compreensão, consciência, controle e consentimento [90]. Ao seguir essas quatro diretrizes, são apresentados padrões de projeto de interface de usuário que comunicam as práticas de privacidade de um *site* em níveis global e local. Kobsa [50] insere diretivas para desenvolvedores com vistas a um projeto eficiente e efetivo de interfaces destinados a usuários.

Essa declaração contextualizada pode ser dividida em uma comunicação local e global para o usuário. A comunicação global de práticas de privacidade divulga declarações de privacidade na página principal de uma companhia ou em todas suas páginas *Web*. Declarações de privacidade na *Web* estão ligadas legalmente a muitas jurisdições. A *Federal Trade Comission* e vários estados americanos têm processado de forma crescente companhias que não se comportam de acordo com as suas políticas de privacidade divulgadas, por práticas de negócios injustas e enganosas [90]. É necessário manter as declarações de privacidade para referências e para proteções legais. A divulgação desse tipo de informação acrescenta maior segurança para a confiança do usuário na divulgação das práticas de privacidade e nos benefícios em contexto local.

Explicações de práticas de privacidade e benefícios de personalização adaptadas no contexto local podem ser endereçadas para as preocupações dos usuários muito melhor que divulgações globais, que não têm contexto. Tal abordagem divide as políticas de privacidade longas em menores e em partes mais compreensíveis. Elas se referem de forma mais concreta ao atual contexto e permitem aos usuários decidirem de forma mais apropriada com relação à coleta de informações [90].

As práticas de privacidade e os benefícios de personalização podem ser explicados em diversos níveis: referidos aos campos de entrada, resumidos para uma página ou para várias páginas consecutivas. A escolha de um ou outro nível de divulgação para o usuário deve seguir as seguintes considerações [90] de fechamento, de separação, de diferenças de sensibilidade e de diferenças legais.

“Fechamento” significa que seqüências de entrada devem ser projetadas de tal maneira que suas conclusões levem ao completo entendimento. O nível mais grosso para o qual o fechamento deveria ser alcançando é o nível da página.

“Separação” quer dizer que, dentro de uma página, freqüentemente existem sub-contextos; supõe-se que eles sejam visualmente separados uns dos outros. A conclusão de cada sub-contexto deve levar a seu fechamento.

Nas diferenças de sensibilidade, os usuários possuem diferentes níveis de disposição para enviar dados pessoais; isso depende do tipo de dados e se os dados relacionam-se a eles. Mesmo em um sub-contexto, usuários possuem diferentes níveis de conforto que levam a tratar cada campo de entrada separadamente.

De acordo com as diferenças legais, nem todos os dados podem ser semelhantes. Deve haver uma explanação separada de práticas de privacidade e benefícios de personalização de dados que são diferentes de um ponto de vista legal, os quais necessitam do consentimento explícito do usuário. Esses dados, possivelmente, devem ser combinados em um acordo durante a navegação do usuário.

Segundo essas considerações, a estratégia mais segura é comunicar práticas de privacidade e benefícios de personalização no nível de cada campo de entrada para dados pessoais individualmente.

Um experimento foi conduzido para verificar empiricamente os méritos desse padrão de projeto de interface. Nessa pesquisa, equiparou-se a divulgação global tradicional e outra contextualizada. A pesquisa de Kobsa [90] revelou que para todos os termos mensurados, como percentagens de questões respondidas, taxa de compra realizada, noção da importância da privacidade, o *site* que possui explanações contextuais apresentou uma maior taxa. Isso mostra que a explanação contextualizada apresenta um efeito positivo e significante no comportamento da divulgação de dados dos usuários e suas percepções de práticas de privacidade.

Declarações de privacidade constituem atualmente documentos legais importantes e compreensíveis. Necessariamente uma política de privacidade de um *site Web* deve ser amigável para o usuário, a fim de melhorar a sua leitura e compreensão.

9 Sistema de Integração de Técnicas de Proteção de Privacidade

O mecanismo foi desenvolvido para gerar um sistema que integra diferentes técnicas de proteção de privacidade para oferecer garantias de segurança das informações pessoais na divulgação implícita e explícita. O aumento de garantias de privacidade tem como resultado maior confiança no acesso a serviços da *Web* pelo usuário. Como requisito, a personalização não deve ser prejudicada pelos cuidados de privacidade. Assim, todas as diferentes técnicas que compõem o sistema permitem que *sites* coletem informações do usuário.

O objetivo do sistema é proteger a privacidade do usuário em toda coleta de dados que ocorre durante a sua navegação. Ele é composto por três módulos que correspondem ao sistema MASKS estendido, à Plataforma de Preferências de Privacidade estendida e à certificação de privacidade. A combinação desses mecanismos concilia as qualidades de cada um e elimina suas limitações. Dessa forma, são oferecidas garantias de privacidade na coleta implícita de dados pela navegação anônima e segurança na coleta explícita de informações através dos contratos e dos selos de privacidade.

Segundo a estrutura do sistema desenvolvido, ele contempla três aspectos de uma nova abordagem de proteção de privacidade. A taxonomia apresentada por Ishitani [6] é contemplada e sintetizada nessa abordagem. A introdução dessa nova interpretação para o contexto de privacidade e personalização na *Web* permite a definição de uma nova classificação para proteção de privacidade, que é discutida a seguir.

Assim como a taxonomia de Ishitani, essa síntese de classificação é composta por camadas de proteção de privacidade. A equivalência entre a taxonomia de Ishitani e a classificação nova é apresentada na figura 22. Essa divisão torna as camadas independentes, por não exigir a presença de uma camada para a existência de outras. Apesar da certificação de privacidade estar intimamente relacionada à camada de proteção de privacidade para a identificação explícita do usuário, ela não exige a existência dessa camada.

Essa classificação sintetizada é criada em virtude de os usuários, no contexto da navegação Web, apresentarem duas situações: a navegação para visualização de páginas e o envio explícito de suas informações pessoais para acessar algum serviço. As técnicas de proteção de privacidade podem ser arranjadas por uma divisão em dois grupos que

representam essas duas situações de navegação, pois esses dois casos apresentam abordagens distintas. A navegação anônima não é capaz de oferecer privacidade para o usuário se ele se identifica explicitamente; da mesma forma, as políticas de privacidade não oferecem nenhuma garantia de privacidade para a coleta implícita de dados, pois ela não pode ser identificada.

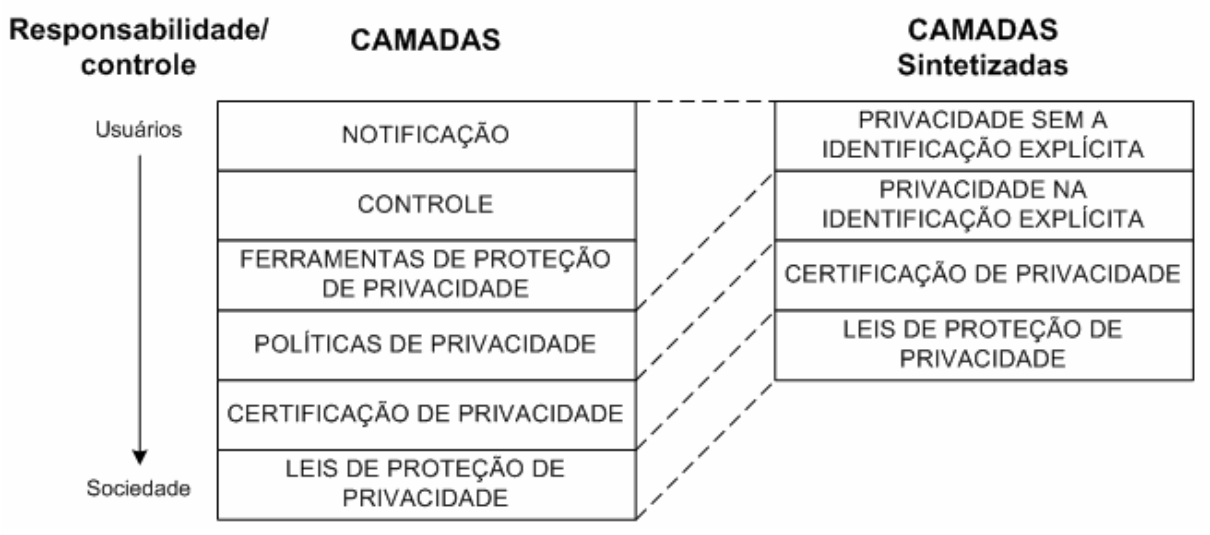

Figura 22. Equivalência entre a classificação de Ishitani e a classificação sintetizada.

Portanto, a proteção de privacidade considera os aspectos da proteção na coleta implícita de dados, na segurança de privacidade da identificação explícita do usuário, na certificação de privacidade e nas leis de privacidade, o que pode ser visualizado na figura 22. A proteção da coleta implícita de dados engloba as camadas de notificação de riscos, de controle e de ferramentas de proteção de privacidade, segundo a taxonomia de Ishitani. A segurança de privacidade da identificação do usuário é equivalente à camada de políticas de privacidade; ela adverte o usuário sobre as práticas de privacidade de um *site*. Essas advertências são necessárias quando o usuário se identifica explicitamente, porque, a partir dessa identificação, o seu anonimato não pode ser mantido somente com a navegação anônima. Dessas abordagens, as três primeiras são utilizadas na construção do sistema.

O mecanismo construído possui um agente de usuário e um servidor de mascaramento de requisições. O agente redireciona a navegação do usuário para o servidor de anonimato, retira informações dos cabeçalhos HTTP e procura políticas P3P estendidas e selos de privacidade para sinalização ao usuário. O servidor mascara as requisições do usuário para oferecer anonimato e permitir a coleta implícita de dados para personalização de serviços. A arquitetura de funcionamento de todo o sistema desenvolvido pode ser visualizado na figura

23. Na figura, o usuário navega pela *Web* através do navegador e é auxiliado pelo agente de usuário do sistema, o qual está incorporado ao navegador, para garantir sua privacidade.

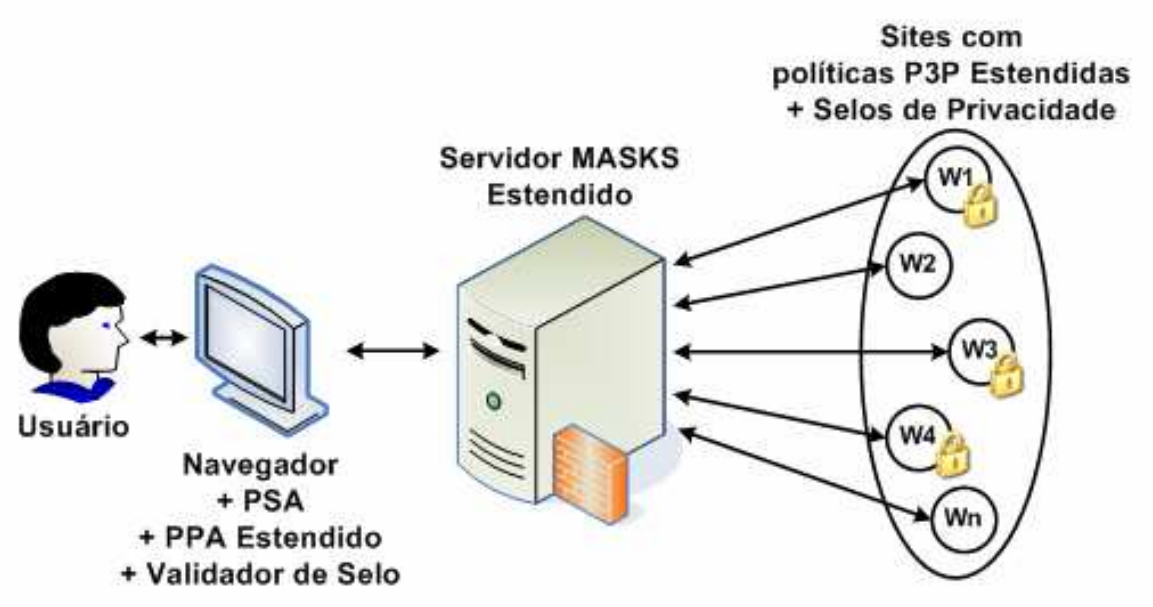

Figura 23. Arquitetura de funcionamento do Sistema de Integração.

9.1 Agente de Usuário

O agente de usuário é um *plugin* adicionado ao navegador. Ele é composto por três mecanismos diferentes: o PSA (*Privacy and Security Agent*), o PPA (*Privacy Policy Agent*) estendido e o validador de selos. O primeiro é o agente do sistema MASKS, responsável por configurar a navegação anônima do usuário. O PPA estendido é um agente de usuário P3P que foi expandido para ser capaz de verificar as políticas P3P de privacidade estendidas [96]. O validador de selos realiza o processo de obtenção, de análise e de autenticação dos selos de privacidade.

O módulo de divulgação de políticas de privacidade e de benefícios de personalização equivale ao Projeto 3P estendido. A extensão incorpora na plataforma de privacidade as informações sobre os benefícios do usuário ao divulgar seus dados pessoais [96]. Para cada propósito de coleta de dados, as vantagens do usuário devem ser explicitadas, considerando-se os interesses dos usuários em fornecer seus dados para receber algum benefício em troca [51].

Uma infra-estrutura de autenticação compõe o módulo de certificação de privacidade. O processo de validação de selos de privacidade utiliza a arquitetura de certificados digitais [97]. O uso dessa arquitetura visa fornecer confiabilidade para o certificado de privacidade e autenticidade para o *site* que o possui e para a entidade que o gera.

A representação esquemática das partes que compõem o agente do usuário é apresentada na figura 24. No esquema, os três mecanismos apresentam independência em seu funcionamento. Assim como o modelo em camadas introduzido, esses mecanismos não necessitam dos outros para estarem auxiliando a navegação do usuário. Toda a comunicação do usuário, bem como a comunicação do PPA e do mecanismo de certificação de privacidade, é encaminhada pelo PSA, o qual decide sobre o uso de mascaramento de navegação através do *proxy* de anonimato.

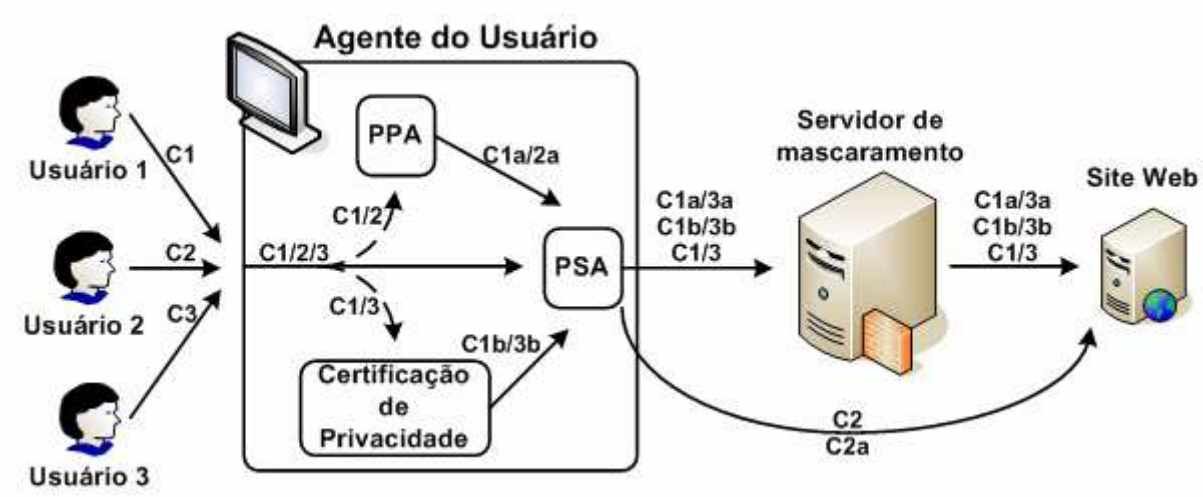

Figura 24. Arquitetura geral do Agente de usuário.

A independência de funcionamento dos mecanismos que compõem o agente de usuário pode ser visualizada através das setas da figura 24. As setas tracejadas não representam o fluxo de comunicação, mas a observação do fluxo pelo mecanismo para obter e analisar políticas de privacidade ou selos de privacidade. As letras destacam o fluxo de comunicação dos mecanismos; o PPA é representado por “a”, e o mecanismo de certificação de privacidade é representado por “b”.

No esquema da figura 24, todas as comunicações, C1, C2 e C3, trafegam pelo PSA, mas somente algumas são observadas pelo PPA e pelo mecanismo de certificação de privacidade, como as requisições C1/2 e C1/3. Assim, o usuário que deseja navegar anonimamente tem sua comunicação redirecionada para o servidor de mascaramento pelo

PSA; aquele que deseja observar as políticas de privacidade tem suas requisições observadas para análise das políticas P3P e selos de privacidade; e aquele que deseja receber todos os serviços de privacidade do sistema tem sua comunicação redirecionada e observada para análise de políticas. No caso, como exemplificado pela seta C2a, o usuário 2 não utilizou o serviço de navegação anônima, mas, pela figura, ele utiliza o PPA para observar as políticas P3P de privacidade do *site*.

9.1.1 Privacy and Security Agent (PSA)

O PSA é parte integrante do sistema MASKS, agindo como agente de usuário nesse sistema. Esse agente é essencial para o funcionamento do sistema, pois redireciona o fluxo de comunicação de navegação do usuário para o servidor de mascaramento e configura esse servidor [53]. Para a extensão do sistema MASKS, não são realizadas modificações estruturais nesse mecanismo. Essencialmente, da mesma maneira como ele foi construído, ele foi utilizado na construção do novo agente de usuário.

As modificações realizadas no PSA têm a finalidade de permitir a comunicação desse mecanismo com o PPA e com o mecanismo de certificação de privacidade. Para isso, são adicionados à arquitetura do PSA outros dois fluxos de comunicação, um para cada mecanismo pertencente ao agente de usuário do sistema de integração, o que pode ser visualizado na figura 25.

O PSA deve estar ciente da existência desses dois mecanismos, pois toda a comunicação do navegador do usuário é encaminhada pelo PSA, mesmo se nenhum método de segurança de privacidade é aplicado nas requisições do usuário. Dessa forma, os fluxos de comunicação dos outros mecanismos do agente, que são paralelos ao fluxo de comunicação do navegador, devem ser mascarados, assim como as requisições do usuário. Esse mascaramento adicional ocorre para evitar que o computador do usuário seja identificado pelos dois outros mecanismos.

De acordo com Ishitani [53], o PSA apresenta as seguintes funções básicas, que não são alteradas para a incorporação dele no agente de usuário do sistema de integração desenvolvido. Esse mecanismo criptografa as URL’s que trafegam entre o computador do usuário e o servidor MASKS. Ele mantém o usuário ciente dos riscos de navegação e informa sobre as máscaras que são inseridas em suas requisições. Ele bloqueia os métodos conhecidos de invasão de privacidade e remove informações que permitam identificar o navegador do

usuário, bem como o próprio usuário. Além disso, ele permite que os usuários desliguem o processo de mascaramento.

Para que as funcionalidades do PSA sejam devidamente atendidas, o agente deve apresentar uma interface simples de ser utilizada, em que o usuário seja capaz de configurar as opções de privacidade para sua navegação [53]. A interface tem um papel fundamental para um melhor esclarecimento do usuário sobre o controle de sua privacidade, já que ela é o meio em que as informações do sistema são divulgadas para o usuário.

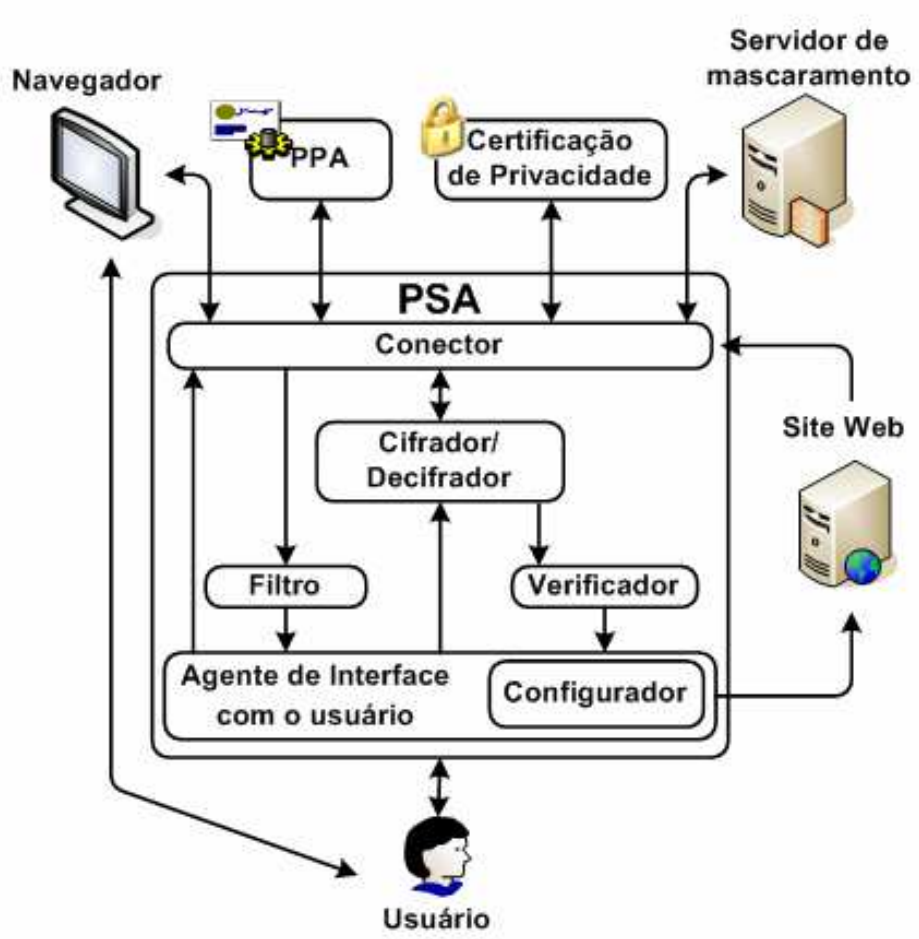

Figura 25. Arquitetura do PSA.

A figura 25 apresenta a arquitetura original do PSA com adição de dois outros fluxos de comunicação para o PPA e o mecanismo de certificação de privacidade. Na figura, a estrutura é divida em módulos, que são o conector, o cifrador/decifrador, o filtro, o verificador, o agente de interface do usuário e o configurador.

O conector realiza a conexão entre o agente de usuário e o servidor MASKS estendido. Ele age como mediador entre o navegador e o *proxy* de mascaramento,

interceptando as requisições do navegador para processá-las e repassando as respostas do *proxy*.

O cifrador/decifrador é responsável por criptografar e descriptografar as URL's enviadas e recebidas no tráfego de comunicação entre o servidor MASKS estendido e o agente de usuário.

O filtro é utilizado para retirar as informações do cabeçalho HTTP que podem identificar o navegador, o sistema operacional e a navegação do usuário.

O verificado é responsável por identificar mecanismos de invasão de privacidade, como os *Web bugs* e os códigos maliciosos que advêm nos pacotes de comunicação HTTP. Esses mecanismos, por essência, podem trazer risco à privacidade do usuário. Dessa forma, eles podem ser eliminados da navegação. Entretanto, a funcionalidade do verificador de reconhecer a presença de formulários no código HTML das páginas navegadas não é utilizada, pois a extensão do servidor MASKS possibilita que toda comunicação do usuário seja mascarada.

O agente de interface do usuário realiza a comunicação com o usuário. Ele os informa sobre os riscos de invasão de privacidade, visto que a notificação do usuário é uma função necessária e deve ser realizada de forma discreta e simples.

O configurador possibilita que o usuário configure o PSA através da seleção das opções de filtro, verificação e mascaramento do mecanismo.

Portanto, segundo os componentes do PSA, apresentados na figura 25, a requisição do usuário e sua respectiva resposta são processadas da seguinte forma: a requisição é filtrada retirando-se as informações do cabeçalho HTTP, e o usuário é informado caso ocorra algum erro na filtragem; antes da requisição ser encaminhada para o servidor MASKS estendido, ela é criptografada. A resposta da requisição vinda do servidor é descriptografada, e o verificador retira dela os mecanismos de invasão de privacidade e adverte o usuário sobre eles. Por fim, a resposta é encaminha para o navegador.

Além desses passos, o PSA recebe requisições do PPA e do mecanismo de certificação de privacidade. Os fluxos de comunicação desses dois módulos do agente de usuário recebem o mesmo serviço de mascaramento que as requisições do usuário. Esse processo evita a ocorrência de invasão de privacidade pela incorporação dessas novas funcionalidades.

9.1.2 Extensão da Plataforma de Preferências de Privacidade (PPA)

A extensão expande a plataforma para melhorar o entendimento do usuário sobre a divulgação de suas informações em transações via *Web* [96]. O usuário acessa serviços da *Web* para receber em troca algum benefício de seu interesse. O interesse do usuário em acessar algum serviço é dependente do conhecimento das vantagens que ele pode receber. Assim como a política de privacidade instrui sobre as práticas de privacidade, ela pode informar sobre os benefícios recebidos através do acesso a um serviço.

O projeto de desenvolvimento apresentado por Kobsa [90] é aplicado para orientar a construção de políticas de privacidade e para incrementar as informações divulgadas. A política P3P de privacidade é construída de forma contextualizada e específica para uma página ou para um conjunto de páginas. Dessa forma, o usuário visualiza somente as políticas que são pertinentes à região do *site* que acessa. Além disso, a especificação de políticas P3P é estendida para comportar dados sobre os benefícios dos usuários.

Construção Hierárquica das Políticas P3P de Privacidade

Para divulgar práticas de privacidade de forma global e contextualizada, as políticas P3P construídas apresentam uma organização hierárquica. Declarações de privacidade comuns são herdadas entre as políticas. Políticas gerais para um conjunto de páginas têm suas declarações “herdadas” por políticas mais específicas, o que é exemplificado na figura 26. Essa divisão hierárquica especifica o escopo de cada política de privacidade de um *site*, o que possibilita a existência de várias políticas específicas para certa região de um *site*. Dessa forma, contextos gerais e específicos podem ser delimitados para todo um *site*.

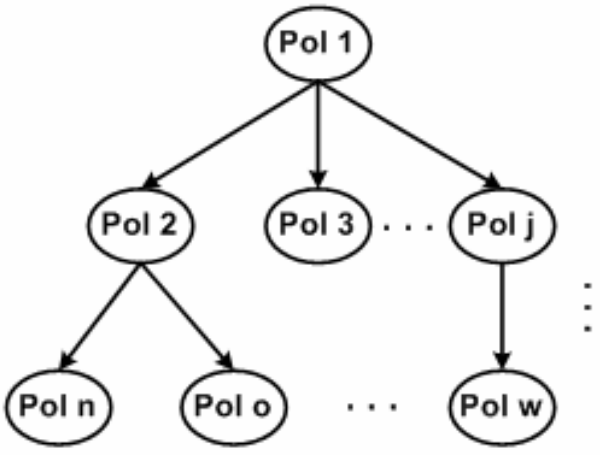

Figura 26. Hierarquia de delimitação de Políticas P3P de Privacidade.

A divisão hierárquica das políticas P3P permite que o usuário receba informações sobre as práticas de privacidade do *site* de acordo com o contexto da página que ele acessa. Essa construção das políticas tem resultado no modo como o agente P3P apresenta o resultado da checagem das políticas de privacidade para o usuário. Com isso, a disposição hierárquica das políticas P3P de privacidade resulta em uma maior facilidade de visualização e de entendimento das práticas de privacidade.

Divulgação dos Benefícios na Entrega de Informações

A divulgação de informações sobre as práticas de privacidade de um *site* auxilia o usuário a tomar decisões sobre o acesso a certo serviço *Web*. Informações adicionais podem ser inseridas nessa divulgação para melhorar a decisão do usuário. O processo de decisão do usuário é facilitado quando são apresentados os benefícios que ele poderá receber ao divulgar seus dados pessoais.

Para as políticas P3P de privacidade informarem sobre os benefícios do usuário, são delimitados campos adicionais para cada elemento "*PURPOSE*" das políticas. O usuário tem interesse em divulgar suas informações pessoais quando ele pode receber algo vantajoso em troca. Assim, para justificar ao usuário o envio de suas informações ao *site*, o propósito de coleta de dados deve apresentar algum beneficio. Além disso, ao serem explicitadas as vantagens do usuário, o propósito é definido mais claramente e pode ser melhor compreendido para se tomar decisões mais conscientes sobre a divulgação de informação para o *site* [90].

Os benefícios são definidos de acordo com o contexto em que o usuário possa se encontrar. Dado que eles são específicos para cada aplicação, a elaboração de uma classificação de forma padronizada para a divulgação dessas vantagens se torna complexa. Dessa forma, cada *site* deve descrever as vantagens do usuário de forma contextualizada para seus serviços.

A diretiva "*PURPOSE*" é delimitada no contêiner "*statement*". Os elementos dessa diretiva são expandidos para comportar uma descrição dos benefícios do usuário. Esses elementos são modificados para admitir uma especificação do propósito de coleta. Os propósitos são modificados para a seguinte forma:

<current></current>,
<admin></admin>,
<develop></develop>,

<tailoring></tailoring>,
<pseudo-analysis></pseudo-analysis>,
<pseudo-decision></pseudo-decision>,
<individual-analysis></individual-analysis>,
<individual-decision></individual-decision>,
<contact></contact>,
<historical></historical>,
<telemarketing></telemarketing>,
<other-purpose value=””></other-purpose>.

Um novo elemento é criado a fim de descrever as vantagens do usuário para cada propósito: <user-benefits>PCDATA</user-benefits>. Esse elemento é opcional para especificar as vantagens do usuário em cada propósito apresentado anteriormente.

Devido às modificações realizadas na sintaxe das políticas P3P, o agente de usuário é estendido para ser capaz de analisar a nova especificação desses benefícios. Para cada elemento de “*PURPOSE*”, o agente deve procurar a descrição das respectivas vantagens dos usuários em consentir a obtenção de seus dados. No final, ele deve apresentar, junto com a advertência das práticas de privacidade do *site*, as informações que esse novo elemento descreve. Dessa forma, o usuário está ciente da coleta de dados, da política do *site* e dos benefícios que pode receber.

PPA

O mecanismo PPA (*Privacy Policy Agent*) é desenvolvido para auxiliar o usuário a manter sua privacidade durante a navegação. Esse mecanismo realiza o processo de análise das políticas P3P de privacidade estendidas e informa o resultado para que o usuário fique ciente das práticas de privacidade do *site* que ele acessa. Como um módulo do agente de usuário do sistema de integração, o fluxo de comunicação do PPA é encaminhado pelo PSA.

A figura 27 apresenta a arquitetura do PPA, a qual é dividida em componentes e em interações entre eles. O mecanismo observa a comunicação do navegador e não a intercepta, como apresentado na figura 27 e na figura 24 pelas setas tracejadas. No final do processo, o resultado da verificação da política P3P de privacidade estendida é apresentado ao usuário através de sinalizações e de um relatório da checagem realizada.

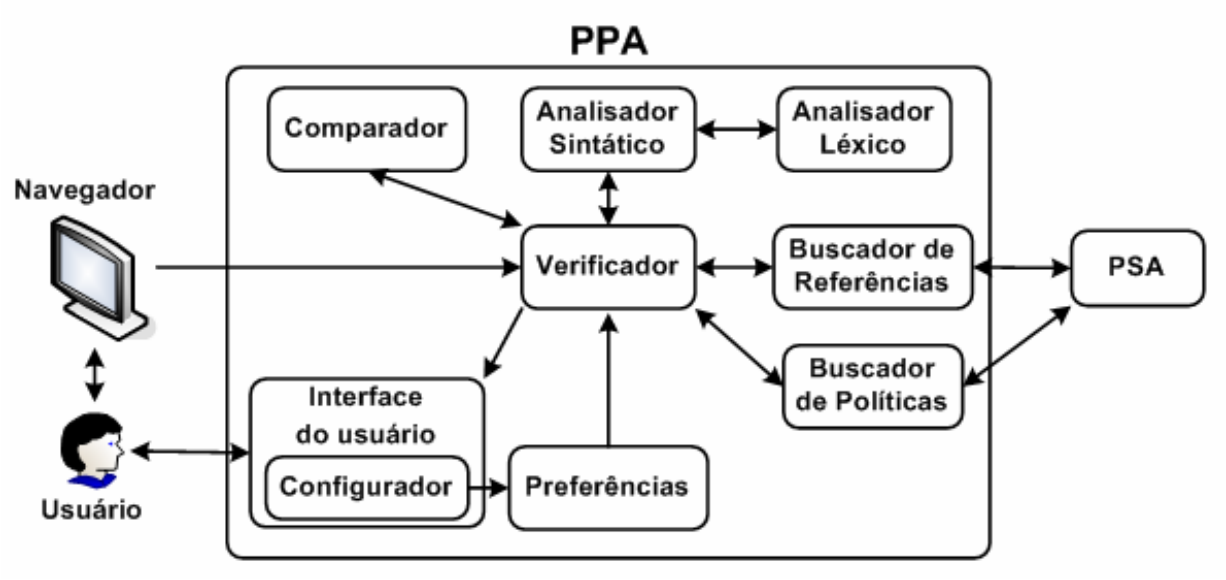

Figura 27. Arquitetura simplificada do PPA.

A interface do usuário é responsável pela comunicação com o usuário. Ela permite a configuração das preferências de privacidade, que são necessárias para modelar o mecanismo de acordo com as necessidades do usuário. Além disso, ela notifica o usuário sobre as práticas de privacidade de um *site* que não são compatíveis com suas preferências.

Os buscadores de políticas de privacidade e de referências são utilizados para obter, respectivamente, o arquivo de políticas de privacidade e o arquivo de referências de política. No desenvolvimento da ferramenta, para encontrar as referências de políticas de privacidade, o lugar bem conhecido (*/w3c/p3p.xml*) foi adotado.

O analisador sintático realiza o processo de análise das referências de políticas e das políticas de privacidade. Essa análise possibilita conhecer o local onde se encontram as políticas P3P de privacidade e permite extrair informações delas para inferir sobre as preferências de privacidade que foram delimitadas.

O comparador é responsável por encontrar as divergências entre as políticas de privacidade analisadas e as preferências de privacidade. O final do processo resulta em advertências que devem ser expostas ao usuário, no caso de as políticas apresentarem algo que o usuário não aceite.

O verificador integra todos os componentes; ele, através das informações contidas na URL requisitada pelo usuário, realiza o processo de análise, utilizando cada componente em uma determinada ordem, e gera uma notificação que é encaminhada para o usuário através da interface. Ele é centralizador no processo e seqüencializa todos os passos para análise de políticas P3P de privacidade.

Na figura 28 são apresentados os passos da avaliação de política de privacidade para uma requisição do usuário pelo PPA. De acordo com o diagrama da figura, o processo se inicia com a requisição do usuário, a qual fornece informações para a obtenção das políticas de privacidade. Após a análise da política e a comparação de seus elementos com as preferências de privacidade do usuário, o resultado é sinalizado para o usuário.

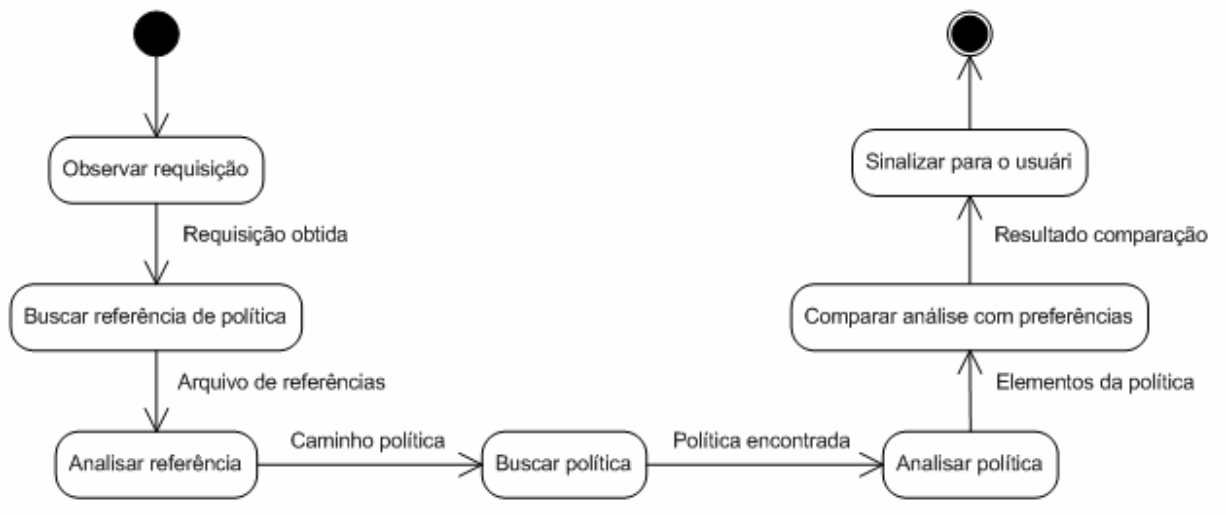

Figura 28. Diagrama de estados de verificação de política P3P de privacidade.

9.1.3 Certificação de Privacidade

Os certificados ou selos de privacidade são sinalizações que fornecem garantias de privacidade aos visitantes de um *site*. Essas sinalizações são utilizadas para comprovar a validade das políticas de privacidade. Os selos evidenciam a existência de uma entidade de confiança que supervisiona as atitudes do *site*. As políticas de privacidade são declarações expressas sobre essas atitudes. Entidades de privacidade podem realizar auditoria nas políticas delimitadas no formato do P3P para garantir que são confiáveis. Elas seguem leis e princípios de privacidade que visam a segurança das informações dos usuários. Com isso, o selo de privacidade fornece meios para auxiliar os mecanismos de regulamentação de leis de privacidade.

Aplicações de criptografia são utilizadas para aumentar a confiança no selo de privacidade. Sem um mecanismo que mantenha a autenticidade do selo, ele pode ser maliciosamente copiado de um *site*. Isso leva a perder todas as garantias de privacidade fornecidas pela entidade do certificado [64]. Assim, a autenticidade e a unicidade do selo podem ser fornecidas através do uso das propriedades da criptografia assimétrica, evitando-se,

dessa forma, a cópia indevida do certificado. Para isso, os selos de privacidade são delimitados no formato de certificados digitais [97]. A recomendação X.509 da ITU-T define um *framework* para o fornecimento de serviços de autenticação [48].

Certificação digital

O padrão X.509 é baseado no uso de criptografia assimétrica e de assinaturas digitais [48]. Ele recomenda o algoritmo RSA para criptografia assimétrica. A função *hash* é utilizada para manter o esquema de assinatura digital. Alguma autoridade confiável de certificação cria o certificado e o coloca em lugar preestabelecido para que possa ser obtido para a verificação. O certificado delimita várias informações para a verificação de autenticidade, entre elas, identificadores da autoridade que gerou o certificado e do sujeito que o recebeu. Um código *hash* é gerado com essas informações e, depois, ele é criptografado para assinar o certificado. Esse processo de assinatura é apresentado na figura 29.

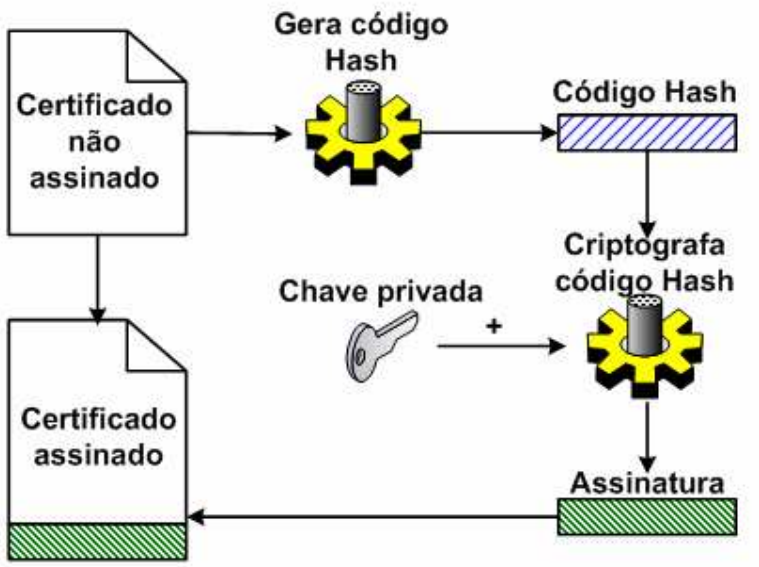

Figura 29. Processo de criação de assinatura em um certificado digital.

Para certificar a autenticidade do certificado digital e das informações contidas nele, a assinatura do certificado é analisada. Essa análise é realizada pela comparação do código *hash* da assinatura por outro que é gerado no momento da verificação com as informações do certificado. Nesse caso, o código da assinatura, para ser comparado na análise, deve ser descriptografado com a chave pública da entidade que gerou a assinatura digital. Esse mecanismo de validação do certificado digital pode ser visualizado na figura 30.

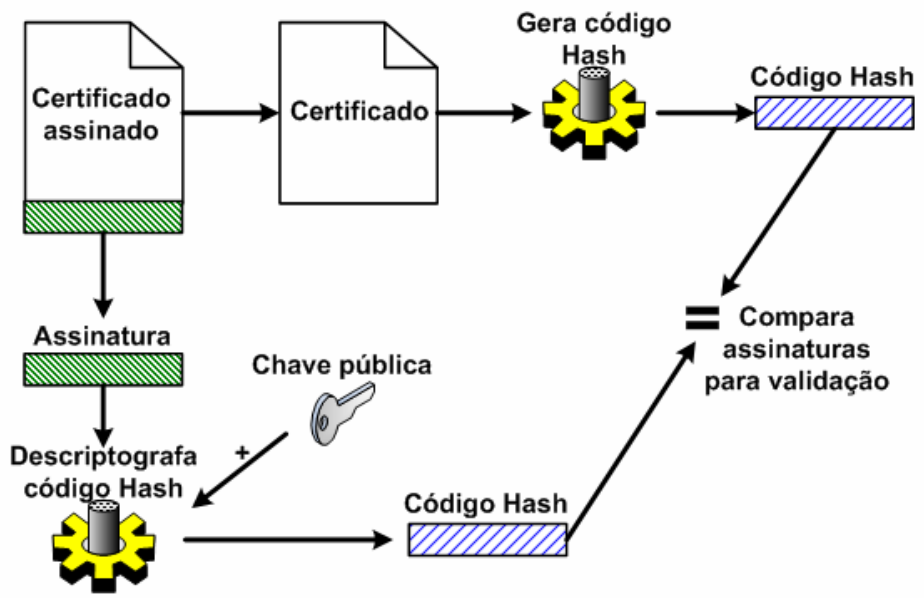

Figura 30. Processo de validação de um certificado digital por meio de sua assinatura.

Devido às propriedades de criptografia assimétrica, o valor *hash*, criptografado por uma chave primária, somente pode ser descriptografado pela respectiva chave pública. Essa propriedade garante que a única entidade capaz de gerar o valor criptografado seja a detentora da chave privada. Assim, se a chave privada da entidade não é divulgada para ninguém, esse processo garante autenticidade na criação de certificados digitais e nas informações contidas nele.

Mecanismo de Certificação de Privacidade

O módulo de certificação de privacidade do sistema desenvolvido delimita o selo de privacidade no padrão X.509 de certificados digitais. Esse padrão de certificados é utilizado para comportar as informações que identificam o *site* visitado, a entidade que fornece o certificado e o próprio certificado. Em um acesso a um *site*, o módulo realiza uma verificação do selo obtido através da comparação das informações existentes nele com identificações do *site*. Dessa forma, essa estrutura de certificação de privacidade garante privacidade através da auditoria da entidade de privacidade, garante autenticidade do selo com o uso da certificação digital e evita a cópia do selo de privacidade ao verificar as informações do *site* que o possui.

Esse módulo de certificação objetiva automatizar a verificação de existência de selos de privacidade em *sites* da *Web*. Após a avaliação do selo de privacidade, a ferramenta informa o usuário sobre as garantias de privacidade oferecidas pela presença do selo da entidade no *site*. Para realizar essa avaliação, informações de identificação do *site* são

inseridas no selo, no caso, o endereço do *site* na *Web*. Essas informações são necessárias para realizar comparações que determinam a autenticidade do selo.

A entidade de privacidade avalia as práticas de privacidade do *site*. Se elas são condizentes com os princípios de privacidade, ele recebe um selo da entidade. A entidade gera esse selo com informações que identificam o *site*. O selo é entregue ao *site* para ser colocado em um lugar preestabelecido para que possa ser obtido automaticamente pela ferramenta; esse local conhecido foi utilizado como */privacySeal/privacySeal.pem*. O usuário, no acesso ao *site*, é informado pela ferramenta sobre as políticas de privacidade do *site*, e, além disso, essas políticas são garantidas pelo selo de privacidade existente no *site*. A ferramenta valida a confiança e autenticidade do selo através da criptografia e das informações contidas nele sobre o *site*. A figura 31 apresenta a arquitetura de autenticação de selos de privacidade desenvolvida nesse trabalho.

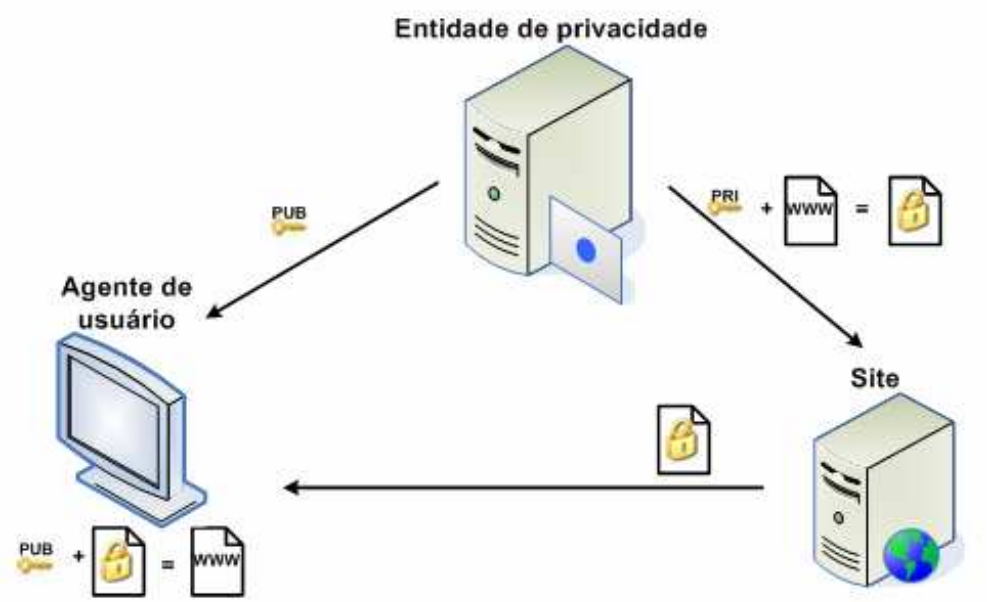

Figura 31. Arquitetura de autenticação de selos de privacidade.

O mecanismo de certificação de privacidade é adicionado ao agente de usuário do sistema para trazer garantias às políticas P3P de privacidade. O mecanismo reconhece a autenticidade do certificado através da validação de sua assinatura digital. O certificado de privacidade válido oferece garantias de que as práticas de privacidade de um *site* são corretas. Como resultado da verificação do certificado, são geradas sinalizações para o usuário estar ciente desse acréscimo de garantia de privacidade.

A figura 32 apresenta a arquitetura do mecanismo que realiza a validação dos certificados de privacidade. O processo é iniciado por uma requisição de usuário para o navegador. Através dessa requisição, obtêm-se informações para encontrar o selo de privacidade e validá-lo apropriadamente. O mecanismo é composto por componentes, os quais são apresentados a seguir.

O decifrador é responsável por descriptografar a assinatura digital de cada selo de privacidade que é verificado.

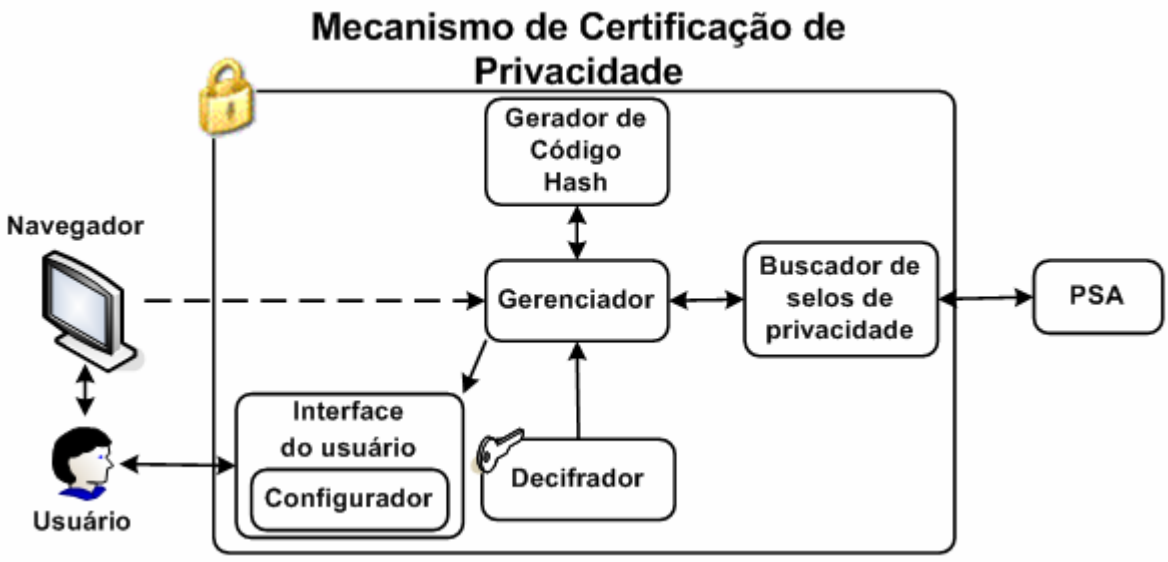

Figura 32. Arquitetura do Mecanismo de Certificação de Privacidade.

O buscador de selos de privacidade é utilizado para obter o certificado de privacidade. Esse componente procura o selo no local pré-estabelecido.

O gerador de código *hash* é utilizado para produzir um código *hash* com as informações contidas no certificado de privacidade.

A interface do usuário é o meio pelo qual o usuário interage com o mecanismo. Ela apresenta sinalização de existência de certificados de privacidade e permite ao usuário configurar o mecanismo.

O gerenciador é responsável por centralizar todo o processo e por seqüencializar os passos; ele observa as requisições encaminhadas pelo navegador, realiza a comparação entre o código *hash* descriptografado e o código *hash* gerado e valida o selo de privacidade através das informações do site contidas nele. Ao final do processo, o gerenciador gera um resultado que é apresentado ao usuário através da interface.

Na figura 33, são apresentados os passos do processo de validação de selo de privacidade para uma requisição do usuário. De acordo com o diagrama da figura, o processo se inicia com a requisição do usuário, a qual fornece informações para a obtenção do certificado de privacidade. Após a análise do selo e a comparação das informações contidas nele com as do site, um resultado é enviado para o usuário.

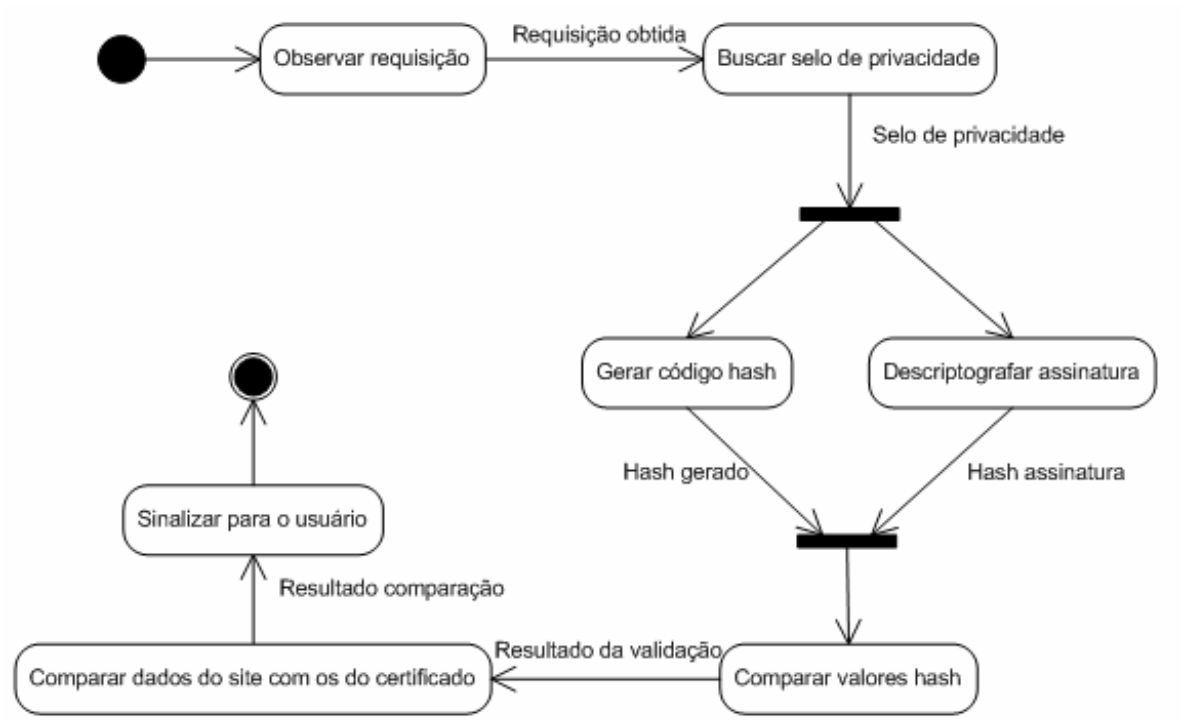

Figura 33. Diagrama de estados de validação de selos de privacidade.

Dessa forma, no sistema desenvolvido, os certificados de privacidade são aliados às políticas P3P de privacidade para assegurar o cumprimento delas. As políticas são contratos de privacidade com assinaturas legítimas dos *sites*. Essa legitimidade garante ao usuário que, ao ser firmado o contrato, ele será seguido corretamente. O sistema desenvolvido garante ao usuário alguma segurança de suas informações, quando ele se identifica explicitamente enquanto navega pela *Web*.

9.1.4 Implementação

Em busca de facilitar o desenvolvimento, a implementação do agente de usuário foi baseada nas mesmas linguagens e arquiteturas para a construção do PSA. A linguagem utilizada é específica para a criação de *plugin's* para o navegador Mozilla. Essa linguagem é

conhecida como XUL (*XML-based User Interface Language*), ela utiliza *scripts* em *JavaScript*, e componentes e interfaces de uma arquitetura conhecida como XPCOM. Esses componentes são utilizados para gerar as interfaces de comunicação com o usuário, para interceptar as requisições do navegador, para filtrar as informações dos pacotes de comunicação e para descriptografar assinaturas digitais.

Na implementação do PPA no agente de usuário, foi utilizada a especificação APPEL 1.0 (*A P3P Preference Exchange Language* [26]) para armazenagem das preferências de privacidade do usuário. A especificação expressa as preferências do usuário em um conjunto de regras de preferência, que podem ser usadas pelo agente de usuário para realizar decisões automáticas ou semi-automáticas relacionadas à avaliação de políticas P3P de privacidade habilitadas em sites da *Web*.

9.2 Extensão do servidor de mascaramento

O módulo de mascaramento de requisições corresponde ao sistema MASKS estendido. A extensão implementada no sistema objetiva introduzir sessões no *proxy* de mascaramento para possibilitar a criação de sessões de navegação, as quais são necessárias para o acesso aos serviços da *Web* e para a segurança das informações divulgadas explicitamente [98]. Além disso, essa extensão viabiliza a cooperação entre os três mecanismos que compõem o agente de usuário, o PSA, o PPA e o Mecanismo de Certificação de Privacidade.

A extensão no servidor de mascaramento restabelece a memória dos *cookies* para a navegação do usuário e mantém o anonimato de sua navegação. Para isso, sessões são inseridas no servidor MASKS. Essas sessões permitem a identificação do estado de navegação de um usuário, o qual é perdido no término delas. Essa perda é decorrente da mistura de *cookies* entre os usuários. No sistema estendido, todos os usuários têm acesso aos *cookies* de um grupo. Entretanto, os *cookies* relacionados à sessão de navegação de um usuário em um *site* não são acessíveis a todos os usuários: eles são protegidos para evitar a perda de segurança de informações.

Uma sessão pode ser definida como o período em que se realiza determinada atividade ou parte dela. Na *Web*, ela pode ser considerada como o período em que ocorre uma transação entre o usuário e um *site*, e *cookies* são utilizados para identificá-la e determinar sua

[26] http://www.w3.org/TR/P3P-preferences/

existência. Assim, a validade de um *cookie* determina o período de existência de uma sessão, que pode ser de alguns minutos, dias, meses ou até mesmo anos, caracterizando um *cookie* como persistente ou de nível de sessão.

De modo geral, os *cookies* persistentes ou de longo prazo são utilizados para identificar o perfil de acesso de um usuário. Através deles, é possível rastrear os acessos do usuário a determinadas páginas em certo período de tempo. As escolhas de acesso delimitam um perfil de navegação e de interesses. Os *cookies* persistentes, por suas características, representam um conjunto de usuários que apresentam um mesmo perfil. Esses *cookies* podem ser usados para caracterizar o acesso de um grupo de usuário de mesmo interesse. Esse processo não prejudica a personalização porque permite a identificação de perfis de usuário através de grupos de máscaras e nem invade a privacidade dos usuários porque não identifica diretamente um perfil de usuário.

Os *cookies* de nível de sessão ou de curto prazo são usados para realizar *login* em sistemas, compras e outros serviços em que há grande interação do usuário. Esses *cookies* delimitam sessões de navegação, que são essenciais para existência de certos serviços *Web*. Sem elas não é possível manter o estado de navegação do usuário e nem armazenar de forma acumulativa as informações adquiridas durante o acesso às páginas.

Para manter essas sessões de navegação de curta duração, *sites* da *Web* normalmente utilizam *cookies* sem validade, sem o campo "*expires*". Sem a delimitação da validade, a sessão é encerrada quando o navegador utilizado pelo usuário é desligado. Esse mecanismo evita que sessões de navegação sejam mantidas, mesmo quando o usuário não realiza o *logout* delas. Dessa forma, sessões de usuário são inseridas no servidor de mascaramento para possibilitar a existência de *cookies* de nível de sessão.

No servidor, as sessões são criadas no nível de grupo de mascaramento e no nível de usuário. A sessão de nível de usuário comporta todas as sessões de que o usuário participa durante a sua navegação; ela possui uma ou mais sessões de nível de grupo. Essa sessão é terminada quando o usuário deixa de acessar o sistema por certo tempo.

Nas sessões de nível de grupo, cada grupo recebe uma validade para permanecer ativo em uma área de *cache* específica de um usuário. A área de *cache* do usuário contém cópias ativas de grupos de mascaramento e um repositório que armazena todos os *cookies* ativos do usuário. Ao iniciar uma sessão de nível de grupo no servidor de mascaramento, o grupo selecionado é copiado para uma lista de grupos ativos do usuário, e seus *cookies* são adicionados a uma *cache* onde são mantidas todas as máscaras de navegação do usuário. Enquanto o grupo está ativo, ele recebe novas máscaras ou suas máscaras são atualizadas

conforme o usuário requisita as páginas da *Web*. Ao expirar a validade da sessão do grupo, essa sessão é terminada e a cópia do grupo na área de *cache* é copiada em seu respectivo grupo original da árvore semântica para atualizar as máscaras existentes para que um acesso futuro de outro usuário seja mascarado por elas.

O processo de gerenciamento de sessões de grupo no sistema permite atribuir especificamente máscaras de navegação para cada usuário. O usuário não perde sua privacidade, pois os *cookies* de sessão são eliminados no final do processo e os *cookies* persistentes são embaralhados com outros *cookies* de mesmo grupo de interesse do sistema de mascaramento. Além disso, em decorrência da atualização que ocorre com a navegação do usuário, as máscaras dos grupos da árvore semântica permanecem atuais com os interesses dos usuários. A atualização da árvore semântica ocorre quando há perda de interesse do usuário por algum tópico, o que é caracterizado pelo término do tempo de validade da sessão do grupo na *cache*.

Para estender o sistema MASKS, foram adicionados a sua arquitetura um gerenciador de sessão, um buscador de máscaras e uma área de *cache* [98]. Esses três módulos são responsáveis por manter as sessões de grupo criadas no sistema. A área de *cache* comporta as sessões de usuário no servidor de mascaramento. Essas *caches* são criadas, atualizadas e eliminadas pelo gerenciador de sessão. Para atualizar a *cache,* é preciso que o buscador de máscaras procure máscaras existentes para toda requisição do usuário. Esses módulos novos atuam com o agente de usuário, o seletor e o gerenciador de grupos sem a necessidade de modificar as partes originais. A arquitetura do sistema MASKS estendido é apresentada na figura 34.

9.2.1 Área de Cache

A área de *cache* contém repositórios de sessão de máscaras e cópias dos grupos. O comportamento e o conteúdo da área de *cache* são apresentados na figura 35. Essa área é dividida em sessões de usuário. Cada sessão simula o armazenamento e o gerenciamento de *cookies* do navegador de um usuário.

As cópias dos grupos são usadas para manter referência com os grupos originais e auxiliar na manutenção dos grupos de perfil e na identificação de máscaras que já pertencem à sessão de navegação do usuário. Elas podem ser consideradas como instâncias de grupos, em virtude de, ao longo do tempo, elas se modificarem de acordo com as requisições do usuário.

Ao terminar a validade de sessão da cópia, ela conterá valores, ou máscaras, diferentes daqueles armazenados no grupo original. Essa diferença é importante para adicionar características e manter o grupo original atual com as necessidades dos usuários, quando a validade da cópia do grupo expira.

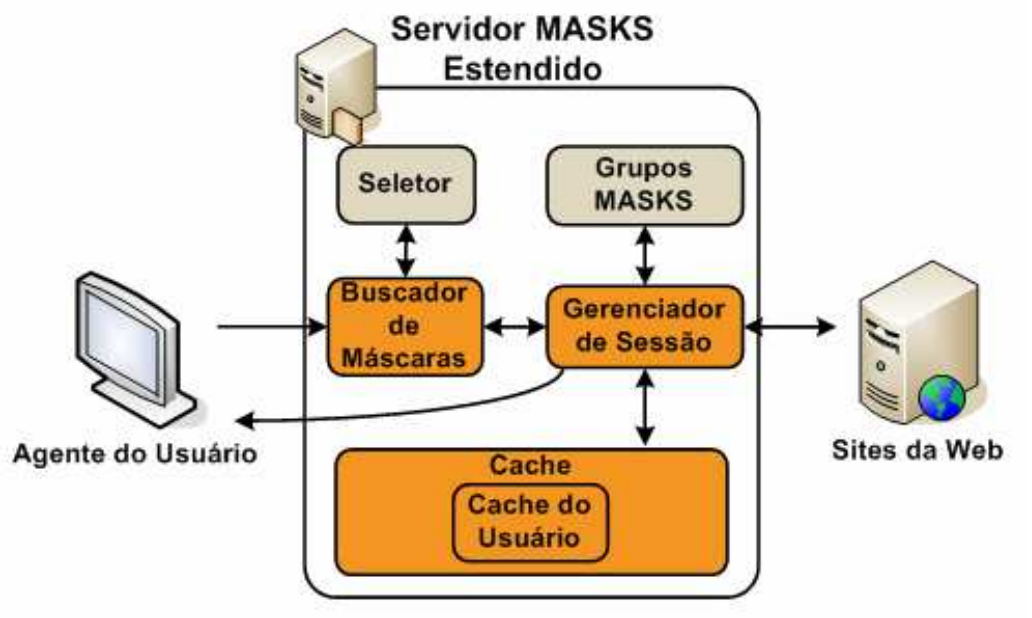

Figura 34. Arquitetura do Sistema MASKS estendido.

Na sessão de um usuário, além de haver cópias de todos os grupos ativados por sua navegação, um repositório de máscaras é criado nela. Esse repositório abriga todos os *cookies* utilizados pelo usuário em sua navegação. Ele simula a armazenagem de *cookies* que um navegador dispõe. Essa armazenagem objetiva auxiliar a manutenção de sessões de navegação; sem ela, não é possível gerenciar e organizar os *cookies* de sessão, pois é preciso reuni-los em uma única área para que sejam administrados por uma ordem cronológica.

O repositório mantém cópias das máscaras dos grupos da árvore semântica para que elas sejam atualizadas sob certa organização conforme o usuário navega. Esse processo é apresentado na figura 35. Ele tem a finalidade de não interromper a sessão de navegação do usuário e de possibilitar a aplicação dos critérios de gerenciamento de *cookies* segundo a especificação deles [3]. A ordem temporal de sessão é mantida no repositório através da substituição dos *cookies* presentes pelos respectivos *cookies* mais novos. Isso ocorre de maneira global na sessão do *cache* e específica nas cópias de grupos da sessão do usuário para mantê-las atuais. Assim, o repositório contém todos os *cookies* da sessão de cada usuário,

mantém os grupos originais atuais através da atualização das respectivas cópias e é responsável por manter o estado atual de todas as sessões de um usuário.

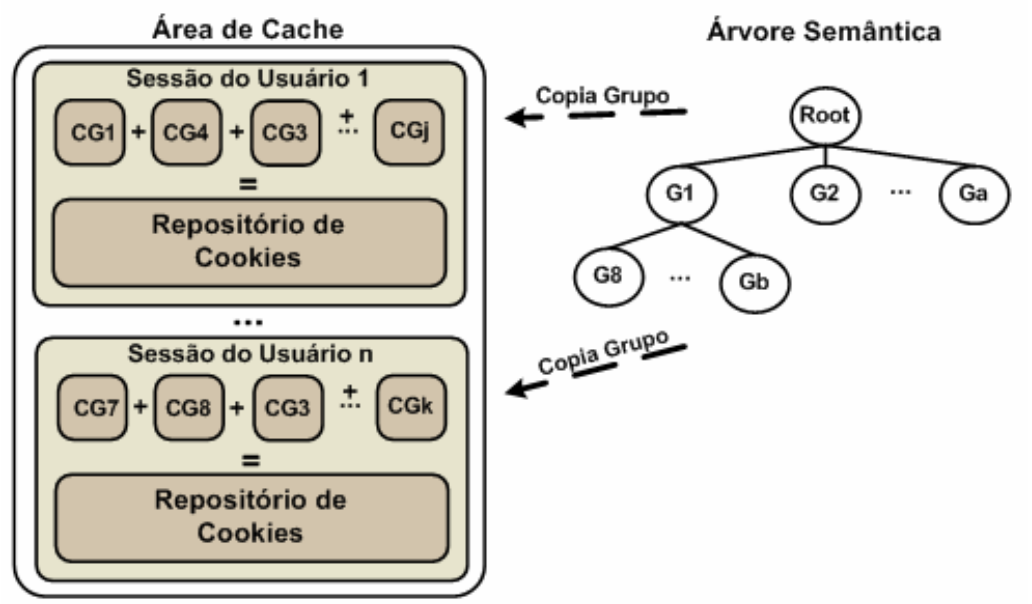

Figura 35. Área de cache e suas sessões de usuário.

9.2.2 Buscador de Máscaras

O buscador de máscaras é responsável por identificar a existência de máscaras relacionadas à requisição do usuário. As máscaras pertencentes à sessão de navegação do usuário devem sempre estar presentes na sessão do usuário no servidor de mascaramento. A presença das máscaras na sessão é sinalizada pela existência de cópias de grupos de perfil na área de *cache* do usuário. Assim, a atuação do buscador é essencial para efetuar a atualização de uma sessão de usuário no servidor com as máscaras de grupos de perfil existentes.

A figura 36 apresenta a seqüência de passos que o buscador realiza para encontrar as respectivas máscaras de uma requisição de página. Ele, inicialmente, precisa identificar o usuário para atribuir sua requisição a alguma sessão no servidor. Depois dessa identificação, com a URL requisitada, o buscador encontra o grupo de interesse através do seletor, o que pode ser visualizado na figura 34. Após isso, com o auxílio do gerenciador de sessão, ele procura a cópia do grupo na sessão do usuário, como apresentado na figura 34. Caso a sessão não contenha uma cópia, cria-se nela uma nova e copia-se o grupo original para a sessão do usuário. Se existirem *cookies* no grupo copiado, eles são transferidos ao repositório para a

inserção das novas máscaras de perfil. Finalmente, a requisição recebe os respectivos *cookies* e é encaminhada ao *site*.

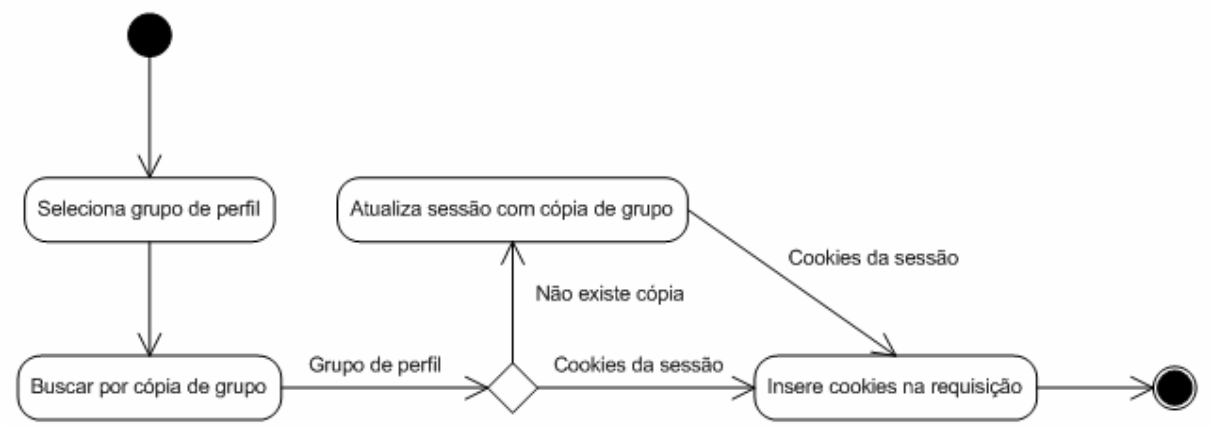

Figura 36. Diagrama de estados de busca de máscaras.

A ordem em que a busca de máscaras é realizada mantém coesas as sessões de navegação do usuário. Necessariamente, antes de qualquer atualização no repositório de máscaras, a busca deve ser realizada na sessão do usuário. Com isso, as máscaras existentes em uma sessão no servidor de mascaramento, que são atualizadas pela navegação do usuário, não são substituídas por outras que pertencem aos grupos de perfil e que não participam da sessão de navegação.

9.2.3 Gerenciador de Sessão

O gerenciador de sessão é responsável pela criação, manutenção e eliminação das sessões de usuários e de grupos no *proxy* de mascaramento. A sessão de um usuário contém uma ou mais sessões de grupos. A sessão do usuário é gerada quando o usuário faz sua primeira requisição através do servidor de mascaramento; ela contém somente a sessão de grupo respectiva à requisição do usuário. Sessões de usuários são apresentadas na figura 35.

O usuário, ao navegar, pode ter sua requisição atribuída a uma sessão de grupo nova ou a uma que havia expirado. Quando um grupo não existe na sessão de um usuário, uma sessão de grupo é gerada. O grupo é copiado para a área de *cache* do usuário e seus *cookies* são inseridos no repositório de máscaras. Essa inserção ocorre com a cópia de *cookies* que não existem no repositório. Essa forma de cópia contribui para manter a consistência das sessões dos *sites*, já que os *cookies* mais atuais não são substituídos. Dessa forma, o sistema estendido

utiliza os *cookies* existentes para inserir características de um grupo de interesse no início de comunicação com um *site*, sem prejudicar as sessões de navegação existentes.

A manutenção do gerenciador é realizada na validade das sessões de grupos e na atualização dos *cookies* do repositório e de cada cópia de grupo. A validade da sessão de um grupo é atualizada toda a vez que uma requisição é encaminhada para ele. A atualização dos *cookies* ocorre em todas as respostas das requisições. Essa atualização ocorre de forma geral no repositório de máscaras de usuário e de forma específica nas máscaras da cópia de um grupo na área de *cache* do usuário.

Ao finalizar uma sessão de grupo, o gerenciador a elimina da *cache* de sessão do usuário e insere as máscaras da cópia do grupo no respectivo grupo original. Essa inserção seleciona os *cookies* que apresentam validade definida, *cookies* de persistência. Como medida de segurança e de privacidade de informação, os *cookies* de sessão existentes no grupo não são copiados para evitar que outros usuários acessem as informações pertencentes às sessões de navegação de um usuário. A sessão do usuário deve ser terminada quando o tempo de validade da sessão do *cookie* expira, ou quando o *cookie* não tem validade e o navegador é fechado [3]. Essa cópia é responsável por manter os *cookies* atualizados para o acesso de outros usuários e por desvincular os *cookies* à identidade de um indivíduo.

O gerenciador encerra uma sessão de usuário quando ela não possui mais nenhuma sessão de grupo. Ele elimina a sessão do usuário da área de *cache*, bem como seu repositório de *cookies*.

A figura 37 apresenta a estrutura do Gerenciador de Sessão, que realiza a manutenção de sessões do usuário. O processo de atribuição de *cookies* à requisição é iniciado pelo buscador de máscaras e é finalizado com a resposta do site Web. O gerenciador é composto por componentes, apresentados a seguir.

O manipulador de requisições insere *cookies* nas requisições do usuário e os retira delas. Ele é responsável por iniciar o processo de gerência de *cookies* do repositório, de grupos e de suas cópias nas sessões de usuários.

O gerenciador de grupos atualiza as cópias de grupos com *cookies* novos através do gerenciador de *cookies* e é responsável por gerenciar a validade dessas cópias. Ele possui um temporizador para determinar o tempo de expiração das cópias, o qual inicia o processo de atualização dos grupos de perfil do sistema.

O gerenciador de repositório administra os *cookies* de sessão do usuário. Ele é importante para o processo de manutenção de sessões pelo sistema por organizar a atualização de *cookies* de sessão de navegação do usuário.

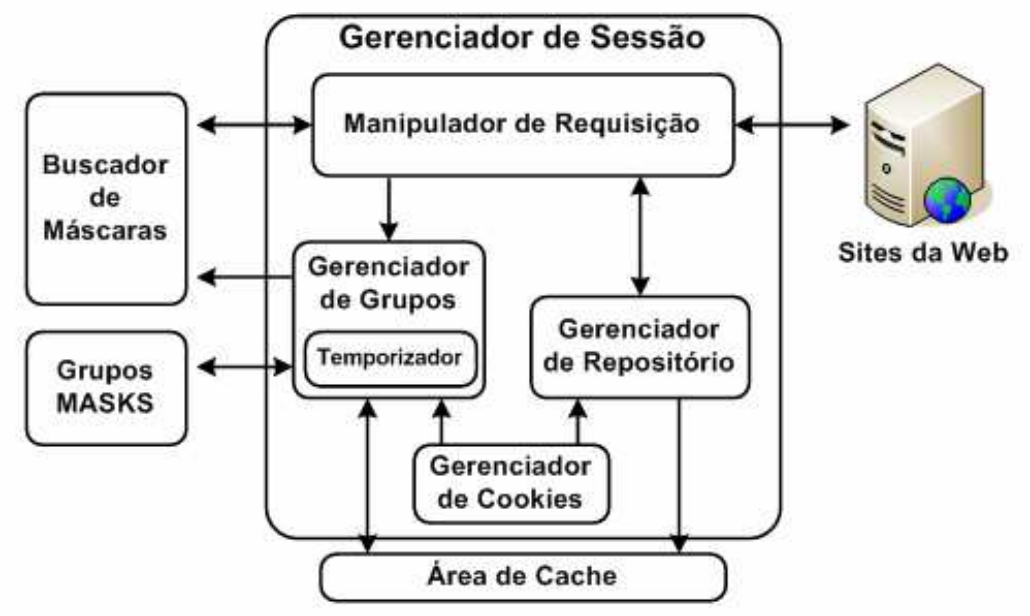

Figura 37. Componentes do Gerenciador de Sessão.

Como ilustrado na figura 35, temos um usuário que fez uma requisição da página que é atribuída ao grupo G3. O buscador de máscaras, com a delimitação da requisição por esse grupo, procura-o na seção do usuário. Se o grupo G3 não é encontrado, o seu original é copiado para a área de *cache*. Os *cookies* dessa cópia que não estão no repositório de máscaras do usuário são inseridos no grupo G3. Quando o usuário pára de acessar páginas da *Web* pertinentes a esse grupo G3 por um tempo maior que a validade da sessão do grupo G3, o conteúdo da cópia do grupo é copiado para o grupo original e a cópia é removida da sessão do usuário. Essa cópia somente estará novamente na sessão do usuário, se ele fizer, novamente, uma requisição de página relacionada ao grupo G3.

Assim como em MASKS, a extensão adicionada continua impossibilitando a identificação do computador do usuário. Porém, a manutenção da sessão do usuário no servidor de mascaramento possibilita a criação de sessões de navegação. Além disso, a sessão no servidor permite uma personalização mais próxima do usuário para a sessão criada, pois a seqüência de páginas requisitadas pelo usuário não é fragmentada.

9.2.4 Implementação

Modificações foram realizadas no *proxy* do Sistema MASKS para criar um novo servidor de mascaramento que comporte as funcionalidades especificadas nesse trabalho.

Novos componentes, como estruturas de armazenagem de dados e procedimentos de tratamento de informações, foram adicionados para incluir sessões de usuários no Sistema. Para o desenvolvimento, foi utilizada a linguagem C em virtude de o servidor do sistema MASKS ser implementado sobre o Squid [27], um *proxy* HTTP.

[27] http://www.squid-cache.org

10 Avaliação Qualitativa do Sistema

Foram realizados testes qualitativos com 15 pessoas, visando-se avaliar a usabilidade e confiabilidade da ferramenta desenvolvida. O sistema desenvolvido introduz maior privacidade das informações pessoais através da navegação anônima e das garantias de privacidade. Devido às diferenças existentes entre a disposição de um conteúdo e a interpretação de um usuário, é preciso que a interface do sistema implementado seja avaliada por pessoas.

10.1 Metodologia

A subjetividade da privacidade e a interpretação das informações por um usuário são critérios essenciais para o desenvolvimento do sistema de proteção de privacidade. Conceitos de usabilidade devem ser utilizados para medir o ganho de proteção de privacidade apresentado pelo mecanismo. A usabilidade pode ser entendida como a qualidade de um sistema de ser usado fácil e efetivamente por um determinado grupo de usuários dentro de um conjunto de cenários específicos [99]. O estudo de caso é utilizado como método de pesquisa para avaliar a ferramenta desenvolvida por medir com maior precisão a eficiência do sistema em informar o usuário sobre sua privacidade.

10.1.1 Método de Pesquisa

O método de estudo de caso é um método específico de pesquisa de campo. Estudos de campo são investigações de fenômenos e de suas ocorrências, sem qualquer interferência significativa do pesquisador, já que a interferência do investigador influenciaria a observação. O objetivo do estudo de caso é compreender o evento em estudo e, ao mesmo tempo, desenvolver teorias mais genéricas a respeito dos aspectos característicos do fenômeno observado [100].

O estudo de caso visa investigar um fenômeno contemporâneo dentro de seu contexto, especialmente quando não existe uma definição clara dos limites entre o fenômeno e o

contexto [101]. O fenômeno explorado não está isolado de seu contexto. Além disso, a pesquisa realizada objetiva delimitar a relação entre o fenômeno observado e o contexto em que ocorre. Além de ser um método de pesquisa propriamente dito, ele é considerado por alguns autores como uma estratégia de pesquisa [102].

Segundo Dias [103], o estudo de caso consiste em uma investigação detalhada de uma ou mais organizações, ou grupos dentro de uma organização, com vistas a oferecer uma análise do contexto e dos processos envolvidos no fenômeno em estudo. O fenômeno, assim como as pesquisas de laboratório, não é isolado de seu contexto, já que o interesse do pesquisador é justamente essa relação entre o fenômeno e seu contexto. A abordagem de estudo de caso não é um método propriamente dito, mas uma estratégia de pesquisa; ele não é uma metodologia específica, mas uma forma de organizar dados através da preservação do caráter único do objeto social em estudo [104].

Portanto, o estudo de caso é um exame empírico que investiga um fenômeno contemporâneo dentro de um contexto da vida real, no qual os comportamentos relevantes não podem ser manipulados, mas onde é possível se fazer observações diretas e entrevistas sistemáticas [101]. Ele se caracteriza pela capacidade de lidar com uma completa variedade de evidências, como documentos, artefatos, entrevistas e observações. Além disso, o estudo de caso pode ter caráter interpretativo; as informações coletadas pelos métodos adotados oferecem base para determinar resultados e inferir conclusões. Sendo assim, o estudo de caso, como outras estratégias de pesquisa, representa uma maneira de se investigar um tópico empírico seguindo-se um conjunto de procedimentos pré-estabelecidos.

O uso desse método de investigação cientifica proposto por Jones citado por Dias [103] é justificado por suas características básicas, que são apresentadas a seguir. O fenômeno em estudo é observado em seu ambiente natural, e uma ou mais entidades, como pessoa, grupo ou organização, são examinadas pelo método. A pesquisa é dirigida aos estágios de exploração, classificação e desenvolvimento de hipóteses do processo de construção do conhecimento. Para não haver interferência nos resultados das análises, não são utilizados controles experimentais ou manipulações. Os resultados dependem fortemente da capacidade do pesquisador integrar as informações coletadas e inferir um resultado. A investigação é envolvida com questões relacionadas ao modo e à razão dos fenômenos, em vez de freqüências ou incidências.

Segundo Lee citado por Dias [103], as deduções lógicas por meio de proposições verbais são tão válidas quanto aquelas derivadas de proposições matemáticas. Dessa forma, as

deduções determinadas através do uso de técnicas de pesquisa qualitativa possuem validade para delimitar conclusões.

Segundo Brewer & Hunter citado por Coutinho [104], os estudos de caso podem ser utilizados para observar e analisar indivíduos, atributos de indivíduos, ações e interações, atos de comportamento, ambientes, incidentes e acontecimentos e, ainda, coletividades. Além disso, Bell citado por Dias [103] afirma que esse estudo é particularmente apropriado para pesquisadores individuais, pois oferece oportunidades para que um aspecto de um problema seja estudado em profundidade dentro de um período de tempo limitado.

Como um método de pesquisa, o estudo de caso é empregado para investigação de fenômenos que apresentam uma grande variedade de fatores e relacionamentos, que podem ser diretamente observados, sem, entretanto, existirem leis básicas para determinar quais deles são importantes [100]. Esses estudos são úteis para auxiliar a compreensão dos processos sociais em seu contexto organizacional ou ambiental e para realizar comparações científicas nas quais é essencial compreender os comportamentos e as concepções das pessoas em diferentes localidades ou organizações [102]. Dessa forma, Bell citado por Dias [103] declara que esse método de pesquisa permite que o pesquisador concentre-se em um aspecto ou em uma situação específica e identifique, ou tente identificar, os diversos processos que interagem no contexto estudado.

A técnica de coleta de dados utilizada no método de pesquisa foi a entrevista de participantes. Essa técnica faz parte de um conjunto de outras técnicas para fornecer informação à investigação. Cada uma dessas estratégias utilizadas em estudos de caso apresenta vantagens e desvantagens próprias, as quais são dependentes basicamente de condições como o tipo de questão de pesquisa e o controle que o pesquisador possui sobre os eventos comportamentais efetivos. Assim, de acordo com Bell citado por Dias [103], no estudo de caso, os métodos de coleta de dados são escolhidos de acordo com a tarefa a ser cumprida. Hamel citado por Dias [103] afirma que existem diversos métodos que podem ser utilizados para coleta de informações, sendo os mais comuns as entrevistas, a observação do participante e os estudos de campo.

Dentro da ampla estratégia de pesquisa do estudo de caso, os vários métodos empregados são classificados como qualitativos, quantitativos ou ambos. Em função dos tipos de problemas que geralmente são associados e melhor compreendidos por meio de estudos de caso, há uma ênfase para empregar métodos qualitativos. Entretanto, métodos qualitativos podem utilizar critérios quantitativos para inferir conclusões [100], como análises estatísticas e comparações numéricas.

Os métodos qualitativos mais utilizados são: a observação, a observação do participante, as entrevistas semi-estruturadas ou não estruturadas e análise dos dados. Para coletar informações, a entrevista é aplicada de forma intensa, mas as técnicas empregadas para realizá-la não são quantitativas [100]. Questionários podem ser utilizados também para auxiliar na obtenção de dados ou para complementar os dados obtidos a partir de observação e de entrevistas [102].

Enquanto as pesquisas quantitativas são utilizadas para generalizar dados de uma amostra em relação à população, as pesquisas qualitativas e os estudos de caso se mostram envolvidas com a generalização de proposições teóricas, comparando-as com outros casos na literatura existente [102]. No contexto do trabalho, a proposição teórica abordada é o uso de navegação anônima, de políticas de privacidade e de selos de privacidade para adicionar maiores garantias de segurança das informações pessoais de um usuário no acesso aos serviços da *Web*.

Portanto, o estudo de caso empregado no trabalho é utilizado como uma ferramenta de pesquisa qualitativa. Ele é empregado para analisar indivíduos, bem como suas opiniões sobre a confiança na navegação utilizando as ferramentas desenvolvidas. Essa pesquisa é devidamente sustentada por uma investigação bibliográfica, documental e exploratória. A fundamentação teórica busca demonstrar as atuais bases bibliográficas sobre o tema, possibilitando, dessa forma, a realização de uma comparação em relação aos resultados obtidos com outros estudos.

A ferramenta estudada nesse trabalho não é avaliada para averiguar a aceitação percentual de uma população, mas é observada a eficiência com que o mecanismo provê privacidade na navegação do usuário. Para se chegar a uma conclusão, são utilizadas comparações controladas. Essas comparações envolvem os resultados obtidos pelas entrevistas realizadas com os participantes. Os casos são examinados para fornecer introspecção sobre um assunto, esclarecer uma teoria, proporcionar conhecimento sobre algo que não é exclusivamente o caso em si. Assim, segundo Stake citado por Coutinho [104], o estudo do caso funciona como um instrumento para compreender outros fenômenos.

10.1.2 Local de Estudo e População

O local onde foram realizados os testes do estudo de caso foi o Laboratório do Grupo de Sistemas Distribuídos e Redes do Departamento de Computação da Universidade Federal

de São Carlos. Devido à necessidade de instalação de *plugins* em navegadores de diferentes sistemas operacionais, as simulações de navegação com auxílio de cada ferramenta de proteção de privacidade não puderam ser realizadas remotamente.

A população participante do estudo de caso é composta por pessoas familiarizadas com navegação *Web*. Assim, alunos de graduação e de pós-graduação do Departamento de Computação foram selecionados para avaliar as ferramentas estudadas.

Todos os convidados a participar do estudo de caso assinaram um termo de consentimento livre e esclarecido. O uso desse termo é necessário para tornar os participantes cientes de que a avaliação não oferece nenhum risco para eles e garantir-lhes que as informações coletadas não os prejudicam.

10.1.3 Coleta de Dados

Para realizar a coleta de informações do estudo de caso, a técnica de entrevista foi utilizada. Os usuários, após realizarem a simulação de navegação, observando o auxílio da ferramenta desenvolvida, responderam um questionário. Através dele, os usuários estimaram o ganho de confiança que a ferramenta proporcionou, em comparação com a confiança proporcionada por outras ferramentas existentes.

10.1.4 Planejamento

O método de estudo de caso é aplicado para provar que o sistema desenvolvido no projeto de mestrado é eficiente em fornecer confiança para o usuário final no acesso a serviços da *Web*. Esse método de pesquisa de campo é necessário para observar a visão que um usuário possui ao utilizar um sistema que protege sua privacidade no acesso a serviços de *sites*.

A avaliação comprova o acréscimo na confiança do usuário no uso de uma ferramenta que introduz anonimato em sua navegação, e disponibiliza informação sobre seus benefícios em divulgar seus dados pessoais e sobre a existência de uma entidade de privacidade. O entendimento do usuário das políticas de privacidade e a segurança em navegar e disponibilizar suas informações foram os critérios usados na análise dos testes.

Uma avaliação comparativa é realizada entre o agente P3P de usuário e sua versão estendida, e entre a divulgação de políticas de privacidade global e contextualizada. Os efeitos da introdução de selos de privacidade são averiguados para medir o aumento da credibilidade do usuário nas políticas de privacidade. Além disso, é observado o impacto na confiança do usuário no uso de um *proxy* de anonimato para navegar pelos *sites Web*.

Os participantes do teste utilizam três agentes de usuários, os quais informam sobre as políticas de privacidade de um *site*. Eles simulam a navegação por um mesmo *site* com três versões diferentes de políticas P3P de privacidade construídas, uma estendida, uma contextualizada e outra global. No final das três navegações, esses participantes respondem um questionário comparativo que informa quais métodos trouxeram maior segurança para eles na navegação e no envio de informações.

A avaliação com usuários foi realizada através do uso de três agentes P3P de privacidade: *Privacy Bird*, PPA (*Privacy Police Agent*) e PPA estendido. Para realizar o teste de forma padronizada, a interface de interação com usuário do agente *Privacy Bird* foi utilizada como base para construção do PPA e do PPA estendido. O PPA está de acordo com a especificação do P3P, e o PPA estendido é criado em conformidade com as modificações do P3P. Adicionalmente, o agente estendido tem sua aparência modificada para comportar as extensões do P3P apresentadas nesse trabalho, para analisar e informar o usuário sobre os certificados de privacidade e para avisar sobre o uso do *proxy* de mascaramento de navegação.

A interface de configuração das preferências de privacidade do usuário do *Privacy Bird* é utilizada como base para construção dos outros dois agentes. Para o teste, as preferências do usuário são configuradas para o nível máximo de privacidade. Dessa forma, o usuário visualiza de forma padronizada o relatório de checagem e tem acesso a todas as possíveis mensagens de aviso.

Um *site* é construído para a realização da avaliação de navegação. Esse *site* possui outras duas réplicas modificadas. Essas réplicas contêm políticas P3P estendidas contextualizadas e políticas P3P estendidas globais. O participante do teste realiza quatro simulações de navegação por esses três *sites* similares. As duas primeiras simulações apresentam checagens das políticas P3P de privacidade sem modificações, em uma simulação é utilizado o *Privacy Bird* e em outra é usado o PPA. São realizadas duas simulações para o usuário compreender o que é informado para ele.

As outras duas simulações de navegação avaliam a extensão inserida no P3P, a divulgação de políticas de privacidade de forma contextualizada, o uso de selos de

privacidade e navegação anônima. Nas duas navegações pelo *site*, é utilizado o PPA estendido. Na primeira simulação, as políticas P3P estendidas estão delimitadas de forma contextualizada. Na segunda, existe somente uma política que especifica as práticas de privacidade de todo o *site*.

Na parte prática da avaliação, o usuário é familiarizado com cada agente que ele utilizará. Essa familiarização consiste em apresentar o funcionamento básico da ferramenta e a localização das informações de privacidade relacionadas ao *site*. O participante da avaliação acessa uma seqüência de páginas do *site* construído para o teste. Após cada acesso de página, o usuário deve consultar o relatório de checagem que o agente disponibiliza. No final, uma quarta navegação é realizada em uma versão do *site* que possui políticas de privacidade globais. Nessa navegação, todas as checagens resultam em um mesmo relatório.

Depois da simulação de navegação, um questionário comparativo é apresentado ao usuário. Nesse conjunto de perguntas, o usuário informa a ferramenta utilizada que apresenta maior confiança em seus relatórios de privacidade. O participante indica a construção de política P3P estendida que mantêm o usuário melhor informado. Ademais, o usuário informa o seu interesse e sua segurança em utilizar uma ferramenta de navegação anônima que permite a coleta implícita de dados. Por fim, ele evidencia o aumento de segurança no acesso a serviços *Web* que apresentam políticas P3P de privacidade garantidas por selos de privacidade.

10.2 Análise dos Dados

A análise dos dados foi aplicada de forma qualitativa. Em um conjunto de 15 participantes, a respostas do questionário revelaram percentuais de aprovação do sistema em cada critério evidenciado por cada questão. As questões exigiam respostas objetivas “sim” ou “não”, ou a escolha entre um conjunto de possibilidades (ferramenta 1, 2 ou 3). A justificativa era opcional, mas ela era considerada para análise quando introduzida nas respostas.

10.3 Riscos e Benefícios dos Participantes da Avaliação

Devido ao estudo de caso contemplar todos os requisitos da Resolução CONEP 196/96, a coleta de dados na pesquisa não apresenta nenhum risco aos participantes. Além

disso, o planejamento do estudo de caso foi submetido ao Comitê de Ética em Pesquisa em Seres Humanos para aprovação. A resposta positiva desse comitê comprova que os testes realizados não prejudicam os indivíduos que participaram.

A participação na avaliação da ferramenta não apresenta diretamente benefícios aos entrevistados. Eles contribuíram para um avanço tecnológico que beneficiará toda a comunidade de usuários da *Web*. Portanto, indiretamente e futuramente, eles poderão usufruir desses avanços como usuários finais da ferramenta desenvolvida e estudada nesse trabalho.

10.4 Respostas do Questionário Aplicado

As respostas do questionário utilizado nos testes são apresentadas na tabela 4. Elas evidenciam a opinião subjetiva, satisfação ou conforto do usuário em utilizar o sistema desenvolvido. Para analisar comparativamente o questionário, inicialmente, procurava-se saber a opinião do participante sobre a navegação anônima e sobre a personalização. Com isso, a avaliação é direcionada para aqueles que se preocupam com a privacidade e com o acesso a serviços *Web*.

Como apresentado anteriormente, as 4 primeiras questões são gerais no contexto de privacidade e personalização na Web, e suas respostas se limitam a “sim” ou “não”. Os questionamentos seguintes das perguntas 5, 6 e 7 comparam o sistema desenvolvido com um agente de usuário P3P, e as respostas das questões 5 e 6 era a escolha entre as opções 1, 2 e 3. Nas questões 8 e 9, o participante respondia sobre sua opinião com relação à existência de certificados e entidades de privacidade. Na última questão, a contextualização das políticas de privacidade é avaliada para averiguar se, realmente, auxilia o usuário no entendimento das práticas de privacidade de um *site*.

Tabela 4. Respostas das perguntas do questionário de avaliação do Sistema de Integração.

Pergunta	Resposta
Utilizar navegação anônima é vantajoso?	13 - Sim
Receber personalização é vantajoso?	12 - Sim
Uso de ferramenta de navegação anônima traz mais segurança?	15 - Sim
Receber personalização e navegar anonimamente é vantajoso?	13 - Sim
Qual ferramenta utilizada que ofereceu mais confiança na navegação?	15 - 3
Qual ferramenta usada que tornou mais confortável a divulgação de dados?	15 - 3

Não enviaria dados usando as 2 primeiras ferramentas, mas seria favorável no envio utilizando a terceira?	5 - Sim
Entidade de privacidade traz mais segurança no acesso a serviços *Web*?	15 - Sim
Existência de um selo privacidade aumenta a confiança em um *site*?	15 - Sim
Políticas de privacidade contextualizadas são melhores que as globais?	14 - Sim

As informações contidas na tabela de respostas são mais bem detalhadas no capítulo que discute os resultados. Os resultados da avaliação qualitativa são formulados pela análise e pela interpretação das informações obtidas através das respostas objetivas do questionário e das respostas discursivas às quais os participantes acrescentavam opiniões.

11 Resultados

Os resultados desse trabalho podem ser descritos em dois contextos diferentes. Em um deles, são apresentadas as informações obtidas através da aplicação do testes comparativos. Esses dados subjetivos evidenciam a opinião do usuário sobre o uso do sistema introduzido nesse trabalho, a melhora de entendimento das políticas de privacidade, o aumento de segurança e de confiança no uso de selos e de práticas de privacidade. No outro contexto, o resultado do acréscimo de sessões no *proxy* de MASKS é descrito. A eficiência em manter sessões durante a navegação anônima é demonstrada.

De forma geral, o sistema desenvolvido adiciona um custo de comunicação na navegação do usuário, o que gera um aumento de tempo de resposta das requisições de página do usuário. O tempo de resposta é maior devido ao processamento gasto para mascarar requisições, o que envolve a criação e o gerenciamento de sessões no servidor. Apesar da execução dessas tarefas não ser custosa, há um acréscimo pequeno no tempo de resposta das requisições, o que é tolerável, dada a vantagem oferecida pelo Sistema de Integração.

O agente de usuário do sistema acrescenta procedimentos aplicados a todas as requisições do navegador. Porém, esse custo adicional é imperceptível para o usuário. Os procedimentos acrescentados apresentam um processamento mínimo, que é facilmente realizado pelo computador do usuário.

11.1 Avaliação da eficiência do sistema MASKS estendido

A inserção de sessões no servidor de mascaramento permitiu que o usuário desfrutasse dos serviços da *Web* que utilizam sessões de navegação. Da mesma maneira como o sistema MASKS original, a extensão implementada impede que qualquer informação pertinente aos protocolos HTTP e TCP/IP seja identificada pelos *sites* da *Web*. Assim, o sistema estendido oferece privacidade através da navegação anônima e permite descoberta de conhecimento e sessões de navegação pelos *cookies*.

A coleta de dados implícita foi beneficiada com a sessão no servidor. Porém, ela não tem capacidade de identificar um usuário, a não ser que ele envie explicitamente suas informações. A navegação de um usuário pode ser identificada através dos *cookies* da sessão

de navegação, mas o término da sessão no servidor traz o anonimato novamente para o usuário. Assim, o sistema estendido fornece uma identificação momentânea da navegação do usuário, a qual dura enquanto a sessão no sistema é válida.

O servidor MASKS, sem as modificações para inserção de sessões, retirava a funcionalidade de memorização dos *cookies*. Um usuário que navegava por um *site* recebia uma personalização semi-estática, presa a um grupo de interesse. Além disso, ele não permitia serviços que utilizavam sessões de navegação, e apresentava insegurança nas sessões de navegação dos usuários, elas eram de domínio público no sistema.

Como exemplo de uma suposta navegação, o usuário requisita uma seqüência de páginas, p1→p2→p3→p4→p5→p6. Nessa seqüência, as páginas P1 e P2 são de esportes, P3 e P4 são de finanças, P5 e P6 são páginas de turismo, e o *site* as observa como uma seqüência de um único usuário.

Ao utilizar o sistema MASKS para mascarar as mesmas requisições do usuário, a seqüência é dividida em 3 grupos de interesse no servidor, p1→p2, p3→p4 e p5→p6. Através dos *cookies* inseridos de acordo com a semântica de cada requisição, o *site* identifica por eles 3 seqüências diferentes como p1→p2, p3→p4 e p5→p6. Segundo os *cookies*, não há correlação entre requisições de grupos diferentes, o que limita a personalização e impossibilita a criação de sessões de navegação. Se essa seqüência pertencesse a uma sessão de compras, o usuário perderia os produtos já selecionados a partir da página p3.

A extensão do sistema organiza a seqüência de navegação do usuário. A seqüência de requisições de páginas, apresentada anteriormente, executada no sistema estendido não é fragmentada e nem perde sua ordem. O *site* identifica a seqüência como a de um único usuário, e ele é capaz de identificar a transição entre duas páginas de grupos de interesses diferentes.

Ao criar sessões no servidor, a navegação de um usuário não invade a sessão de navegação de outro. Com o término de uma sessão, a identidade do usuário não é mantida, pois os *cookies* de identificação, não relacionados a sua sessão, podem ser utilizados por outros. Desse modo, a personalização é melhor produzida, dado que a coleta implícita é dinâmica no nível de sessão no *proxy* de MASKS estendido, e a privacidade do usuário é mantida.

Em testes, o servidor MASKS estendido distinguiu requisições de diferentes usuários. Três computadores foram utilizados para avaliar a capacidade do servidor de manter sessões. Em cada computador, a comunicação do navegador utilizado foi redirecionada para o servidor de mascaramento. Por motivos de análise, um site foi criado para avaliar a manutenção de

sessões de navegação nos acessos dos usuários. As sessões foram mantidas corretamente em todos os computadores, o que provou a eficácia da extensão incorporada ao servidor MASKS.

Nas avaliações empíricas, o servidor MASKS estendido apresentou pouco acréscimo no tempo de resposta das requisições de página feitas para um *site* criado para testes e que possuía 1 *cookie*. Em média, houve um acréscimo de 16,5659087 milissegundos para cada requisição, uma adição de 1,9236087 milissegundos em relação ao servidor MASKS com gerenciamento de *cookies*. Dessa forma, o gerenciamento de sessões de navegação no servidor MASKS compromete muito pouco a navegação do usuário, e o processo para criação e para manutenção de sessões apresentou pouca diferença de tempo em relação ao processamento do servidor de mascaramento com gerenciamento de *cookies*.

11.2 Resultados dos Testes Comparativos

Nos testes realizados, todos os participantes concordaram que o *proxy* de anonimato traz maior segurança em sua navegação. Porém, 84% deles acreditam ser conveniente não ter sua navegação identificada. Dessa forma, apesar do anonimato não ser um requisito necessário para todos, ele traz mais segurança na manutenção das informações de navegação.

Na avaliação, 84% dos usuários consideraram agradável receber serviços personalizados enquanto navegam anonimamente. Os outros 16% afirmam que não gostariam de acessar esses serviços nesse contexto e também não apreciariam receber personalização de forma geral. Assim, para aqueles que têm interesse em receber serviços personalizados, o sistema MASKS estendido é vantajoso, pois adiciona maior confiança na navegação e não prejudica o fornecimento de serviços.

Nos testes, o agente estendido foi o que transmitiu mais confiança para o usuário e satisfação em divulgar suas informações [96]. Todos os participantes do teste concluíram que ele detalha mais o propósito de coleta de dados, apresenta mais informações e as coloca em um contexto próximo do entendimento do usuário. Um exemplo desse caso é o propósito de gerar dados sobre os interesses do usuário sem identificá-lo com seu respectivo benefício de receber páginas e serviços personalizados.

Entretanto, para todas as ferramentas testadas, houve diversos comentários negativos sobre a interface de interação. As informações apresentadas para o usuário devem ser direcionadas a sua visão. O objetivo principal dos agentes de usuários é informar o usuário

sobre os riscos de abuso de privacidade que podem ocorrer. Dessa forma, o modo como as informações são dispostas influencia o entendimento do usuário.

Na situação de envio explícito de informações, 34% dos que participaram da pesquisa somente aceitariam enviar seus dados se soubessem dos benefícios que receberiam. Os outros 66% rejeitariam ou aceitariam divulgar suas informações pessoais de qualquer forma, independentemente dos relatórios apresentados. Entretanto, todos eles concluíram que o agente estendido torna mais clara a vantagem e os torna mais cientes.

Todos os participantes concluíram que os selos de privacidade oferecem mais confiança no acesso a serviços da *Web*. A existência de um selo indica a presença de uma entidade confiável que regulamenta e supervisiona as práticas de privacidade dos *sites*.

A contextualização de políticas P3P de privacidade foi a preferida por 92% dos participantes do teste. Eles concluíram que ela é mais objetiva, não compete com a navegação e é pertinente ao contexto da página acessada. Outros 8% escolheram a construção global das políticas porque ela permite ao usuário fazer uma verificação de todo o *site* de uma só vez.

12 Conclusões

O sistema apresentado fornece privacidade para a navegação do usuário na coleta implícita e na coleta explícita de informações. A extensão do *proxy* de mascaramento de requisições garante o anonimato sem prejudicar a obtenção de dados da forma implícita. O P3P estendido insere contratos de privacidade que asseguram maior confiança no envio explícito de informações. Os selos de privacidade garantem que as práticas do *sites* não prejudiquem a privacidade do usuário.

A extensão do servidor MASKS é eficiente em fornecer privacidade para a navegação do usuário, em permitir descoberta de informações para a aplicação de personalização e em manter sessões de navegação do usuário [98]. A sessão no servidor de mascaramento simula o gerenciamento do navegador do usuário, o qual é executado no tratamento dos *cookies* adquiridos durante o acesso a páginas da *Web*. Desse modo, é possível manter sessões de navegação sem que elas sejam ligadas diretamente a um navegador.

A identidade real do usuário somente pode ser identificada através da coleta explícita de dados. Pelo sistema estendido, um usuário, ao conectar-se em um *site*, está sujeito a ter sua navegação observada pelos *cookies*. Sua navegação pode ser registrada em sua conta no *site*. Porém, esse processo não pode ser impedido, pois, ao realizar o processo de *login*, o usuário se identifica explicitamente e concorda com as políticas de privacidade do *site* que acessou.

A inserção de sessões no servidor permitiu aumentar o acesso do usuário a serviços da *Web* que utilizavam a funcionalidade de memória dos *cookies*. As sessões de navegação de um usuário somente são finalizadas pelo sistema se elas utilizarem *cookies* sem validade definida. Sem isso, o usuário pode ter sua sessão invadida por outros.

Através dos testes aplicados, a inserção de benefícios do usuário provou aumentar o conhecimento do usuário para tomar decisões. Ele, ao estar ciente das vantagens que receberá na divulgação de suas informações, está mais apto a escolher em acessar um determinado serviço. Além disso, a divulgação dessas informações permitiu focalizar no contexto da aplicação e direcionar o aviso para um linguajar mais próximo do usuário, sem o uso de termos técnicos.

A validação de selos de privacidade incorpora garantias das práticas de privacidade do *sites*. Os selos têm total aprovação dos usuários nos testes realizados. O uso de certificados

digitais forneceu confiabilidade e autenticidade dos selos de privacidade. Devido às propriedades da criptografia assimétrica e à arquitetura de certificação, os selos não podem ser copiados, e são autenticadas a identidade dos *sites* e a entidade que emitiu o selo.

A forma contextualizada de construção de políticas de privacidade foi a mais preferida pelos que participaram da avaliação. Ela traz mais objetividade à divulgação para o usuário e enfoca as mensagens de advertência no cenário de navegação em que o usuário está inserido. Dessa forma, as práticas de privacidade do *site* são apresentadas de forma mais compreensível. Entretanto, uma política de privacidade global para o *site* deve ser oferecida para observação do usuário de todas as práticas do *site*.

O uso do sistema MASKS estendido oferece confiança para o usuário em sua navegação e permite a coleta de dados implícita. Segundo a pesquisa realizada, os usuários consideram que a presença de um sistema que mascara requisições e permite coleta de dados aumenta a confiança deles no acesso a serviços da *Web*. Através do anonimato, o usuário tem controle total sobre suas informações e não precisa se preocupar com a divulgação delas.

Como trabalho futuro, é necessário melhorar a interface de acesso do agente de usuário do sistema. Essa melhoria deve considerar critérios de usabilidade para fornecer ao usuário uma navegabilidade mais intuitiva pelo sistema e um entendimento melhor das informações apresentadas por ele. Além disso, para trazer maiores garantias de privacidade às políticas P3P estendidas, um mecanismo de análise semântica de política pode ser adicionado ao sistema [105]. Esse mecanismo avalia as políticas P3P de privacidade com o comportamento do *site* em uma dada página. Dessa forma, cada campo de entrada explícita de informação de uma página pode ser avaliado, e a respectiva política para esse campo pode ser apresentada na página visualizada.

A arquitetura de autenticação de certificados de privacidade garante autenticidade e unicidade do selo. Para garantir a confiabilidade da entidade que gera o selo de privacidade, seu selo deve ser autenticado por uma autoridade certificadora confiável. Assim, a infra-estrutura de chaves públicas será utilizada para autenticação dos selos de privacidade. Eles possuirão um valor jurídico com a assinatura de autoridades certificadoras.

No servidor de MASKS estendido, foi utilizada uma validade de 20 minutos para a sessão de grupo, a qual expira nesse tempo após nenhuma requisição do usuário referir-se a ela. No entanto, é necessário realizar mais testes para adequar a validade da sessão dos grupos a um valor mais coerente com a navegação do usuário. A adequação desse tempo é crucial para a manutenção de sessões de navegação do usuário e para a segurança de suas informações.

Referências Bibliográficas

[1] XING, S.; PARIS, B. P. Mapping the Growth of the Internet. In: INTERNATIONAL CONFERENCE ON COMPUTER COMMUNICATIONS AND NETWORKS (ICCCN), oct. 2003, Dallas, Texas, USA. **Proceedings...** IEEE Press, 2003. p. 199-204. Disponível em: < http://ieeexplore.ieee.org/iel5/9031/28660/01284170.pdf?tp=&isnumber=&arnumber=1284170>. Acesso em: 15 set. 2006.

[2] INTERNET WORLD STATS. **Internet growth statistics**. Disponível em: <http://www.internetworldstats.com/emarketing.htm>. Acesso em: 13 set. 2006.

[3] KRISTOL, D.; MONTULLI L. **HTTP state management mechanism**. Bell Laboratories, Lucent Technologies, oct. 2000. RFC 2965. Disponível em: <http://www.ietf.org/rfc/rfc2965.txt>. Acesso em: 20 mar. 2005.

[4] BUCKLIN R. E. et al. Choice and the Internet: from clickstream to research stream. **Marketing Letters**, v. 13, n. 3, p. 245-258, aug. 2002.

[5] MONTGOMERY, A. L. et al. Modeling online browsing and path analysis using clickstream data. **Marketing Science**, v. 23, n. 4, p. 579-595. 2004.

[6] ISHITANI, L.; ALMEIDA, V.; MEIRA Jr., W. Masks: bringing anonymity and personalization together. **IEEE Security & Privacy Magazine**, v.1, n. 3, p. 18–23, may/jun., 2003.

[7] CAVOUKIAN, A. **Data mining:** staking a claim on your privacy. Technical report, Information and Privacy Commissioner. Toronto, Ontario, jan., 1998. Disponível em: <http://www.ipc.on.ca/english/pubpres/papers/datamine.htm>. Acesso em: 21 jun. 2005.

[8] GABBER, E. et al. (1997) How to make personalized web browsing simple, secure, and anonymous. In: International Conference on Financial Cryptography, 1, 1997. **Proceedings…** London, UK: Springer-Verlag, 1997. p. 17-32. (Lecture Notes in Computer Science, v. 1318) Disponível em: < http://www.pittsburgh.intel-research.net/people/gibbons/papers/fc97.pdf >. Acesso em: 13 maio 2004.

[9] LAUDON, K. C.; LAUDON, J.P. **Gerenciamento de sistemas de informação**. 3. ed. Rio de Janeiro: LTC, 2001. 434 p.

[10] MOBASHER, B. et al. Integrating web usage and content mining for more effective personalization. In: INTERNATIONAL CONFERENCE ON ELECTRONIC COMMERCE AND WEB TECHNOLOGIES, 1, 2000. **Proceedings...** London, UK: Springer-Verlag, 2000. p.165-76. (Lecture Notes In Computer Science, v. 1875).

[11] GODERIS, S. et al. Combining meta-level and logic-based constructs in web personalization. In: INTERNATIONAL COMPUTER SCIENCE CONFERENCE ON ACTIVE MEDIA TECHNOLOGY, 6, 2001. **Proceedings...** London, UK: Springer-Verlag, 2001, p.57-64. (Lecture Notes in Computer Science, v. 2252).

[12] KRAMER, J.; NORONHA, S.; VERGO, J. A user-centered design approach to personalization. **Communications of the ACM**, v. 43, n. 8, p.45-8, 2000.

[13] ALVES, C. R. C. **Conceitos e aplicação de personalização na navegação em ambientes web – Sistema Argo**. 2005. 260 p. Dissertação (Mestrado). Escola Politécnica, Universidade de São Paulo, São Paulo, 2005.

[14] SAE-TANG, S., ESICHAIKUL, V. Web personalization techniques for e-commerce. In: INTERNATIONAL COMPUTER SCIENCE CONFERENCE ON ACTIVE MEDIA TECHNOLOGY, 6, 2001. **Proceedings...** London, UK: Springer-Verlag, 2001. p. 36-44. (Lecture Notes In Computer Science, v. 2252).

[15] PERKOWITZ, M.; ETZIONI, O. Towards adaptive web sites: conceptual framework and case study. **Computer Networks**, Toronto, Canadá, v.31, n. 11-24, p.1245-1258, 1999.

[16] MCGARRY, K. **O contexto dinâmico da informação**: uma análise introdutória. Brasília: Briquet de Lemos livros, 1999. 189p.

[17] KREITZBERG, C.B. Designing the electronic book: human psychology and information structures for hypermedia. In: INTERNATIONAL CONFERENCE ON HUMAN-COMPUTER INTERACTION ON DESIGNING AND USING HUMAN-COMPUTER INTERFACES AND KNOWLEDGE BASED SYSTEMS, 3, 1989, Amsterdã, Holanda. **Proceedings...** New York, NY, USA: Elsevier Science, 1989. p.457-464.

[18] MURUGESAN, S.; RAMANATHAN, A. Web Personalisation: an overview. In: INTERNATIONAL COMPUTER SCIENCE CONFERENCE ON ACTIVE MEDIA TECHNOLOGY, 6, 2001. **Proceedings...** London, UK: Springer-Verlag, 2001. p. 65-76.

[19] SMYTH, B.; COTTER, P. A personalized television listings service. **Communications of the ACM**, New York, NY, USA, v. 43, n. 8, p.107-111, aug. 2000.

[20] BELKIN, N.J. Helping people find what they don´t know. **Communications of the ACM**, New York, NY, USA, v. 43, n. 8, p. 58-61, 2000.

[21] ALVES, C.R.C.; FILGUEIRAS, L.V.L. Fatores para personalização de navegação na Internet. In: LATIN AMERICAN CONFERENCE ON HUMAN COMPUTER INTERACTION (CLIHC), 2003. **Anais...** Rio de Janeiro, RJ: Brasil, 2003. p. 267.

[22] YU, P.S. Data mining and personalization technologies. In: INTERNATIONAL CONFERENCE ON ADVANCED SYSTEMS FOR ADVANCED APPLICATIONS, 6, 1999. **Proceedings…** Washington, DC, USA: IEEE Computer Society, 1999. p.6-13.

[23] SUNDARSEN, N.; YI, J. Mining the web for relations. **Computer Networks: The International Journal of Computer and Telecommunications Networking**, Amsterdam, The Netherlands, v. 33, n. 1-6, p. 699-711, 2000.

[24] AGGARWAL, C.C.; YU, P.S. On text mining techniques for personalization. In: INTERNATIONAL WORKSHOP ON NEW DIRECTIONS IN ROUGH SETS, DATA MINING, AND GRANULAR-SOFT COMPUTING, 1999. **Proceedings...** London, UK: Springer-Verlag, 1999. p.12-8. (Lecture notes in artificial intelligence, v.1711).

[25] DIX, A. et al. **Human-computer interaction**. 2. ed. New York, NY, USA: Prentice Hall, 1997. 572 p.

[26] BROWNE, D.; TOTTERDELL, P.; NORMAN, M. **Adaptive user interfaces**. Londres: Academic Press, dec. 1990. 240 p.

[27] BOAR, B. Tecnologia da informação. **A arte do planejamento estratégico**. 2. ed. São Paulo, SP, Brasil: Berkeley, 2002. 339 p.

[28] KOCH, M.; MÖSLEIN, K. User representation in eCommerce and collaboration applications. In: Bled eCommerce Conference eTransformation, 16, Bled, Slovenia, 2003. **Proceedings…** Bled, Slovenia: 2003. p. 649-661.

[29] CRANOR, L. F.'I didn't buy it for myself' privacy and ecommerce personalization. In: THE 2003 ACM WORKSHOP ON PRIVACY IN THE ELECTRONIC SOCIETY (WPES), Washington, DC, USA, 2003. **Proceedings...** New York, NY, USA: ACM Press, 2003. p. 111-117.

[30] KOBSA, A. Tailoring privacy to users' needs. In: INTERNATIONAL CONFERENCE IN USER MODELING, 8, 2001. **Proceedings…** London, UK: Springer-Verlag, 2001. p. 303-313.

[31] FIELDING, R. el al. **Hypertext Transfer Protocol (HTTP/1.1)**. Network working group: UC Irvine, Compaq, Xerox, Microsoft, W3C e MIT. jun. 1999. RFC 2616. Disponível em: <http://www.w3.org/Protocols/rfc2616/rfc2616.html>. Acesso em: 20 ago 2004.

[32] CERT.BR. **Cartilha de segurança para a Internet, parte VI**: spam. Versão 3.0. Comitê Gestor da Internet no Brasil, 2005. Disponível em: <http://cartilha.cert.br>. Acesso em: 01 jun. 2006.

[33] MARTIN, D. **Detecting web bugs with bugnosis**: privacy advocacy through education. Boston University Computer Science Department, 2003. Disponível em: <

http://www.bugnosis.org/faq.html> Acesso em: 12 nov. 2005.

[34] CAVOUKIAN, A.; HAMILTON, T., J. **The privacy payoff**: how successful business build customer trust. Ohio, USA: McGraw-Hill Tyerson Limited, 2002. 332 p.

[35] ROCHA, B. G. et al. Disclosing users' data in an environment that preserves privacy. In: WORKSHOP ON PRIVACY IN THE ELECTRONIC SOCIETY, Washington, DC, USA, 2002. **Proceedings…** New York, NY, USA: ACM Press, 2002. p. 71-80.

[36] SHIFLETT, C. **Essential PHP Security**. Sebastopol, CA, USA: O'Reilly Media, 01 nov. 2005. 124 p.

[37] KIMBALL, R.; MERZ, R. **Data Webhouse**: construindo o data warehouse para a web. Rio de Janeiro, RJ, Brasil: Campus, 2000. 384 p.

[38] COOKIE CENTRAL. **Cookies**. Disponível em: <http://www.cookiecentral.com/>. Acesso em: 06 set. 2005.

[39] MONTGOMERY, A. L. **Using clickstream data to predict WWW usage**. Tepper School of Business, Carnegie Mellon University, Pittsburgh, PA, USA. Working Papers Series: Marketing, 2000. Disponível em: < http://www.andrew.cmu.edu/user/alm3/papers/predicting%20www%20usage.pdf>. Acesso em: 03 set. 2005

[40] MONTGOMERY, A. L.; FALOUTSOS, C. **Using clickstream data to identify world Wide Web browsing trends**. Tepper School of Business, Carnegie Mellon University, Pittsburgh, PA, USA. Working Papers Series: Marketing, 2000. Disponível em: < http://www.andrew.cmu.edu/user/alm3/papers/web%20trends.pdf>. Acesso em: 10 set. 2005

[41] HAN, J.; KAMBER, M. **Data mining**: concepts and techniques. 2. ed. USA: Morgan Kaufmann, 2001. 500 p.

[42] ROBINSON, N.; SHAPCOTT, M. Data mining information visualisation: beyond charts and graphs. In: International Conference on Information Visualisation, 6, 2002, London, England. **Proceedings…** Los Alaminos: IEEE Computer Society Press, 2002. p. 577.

[43] ZHANG, D.; ZHOU, L. Discovering golden nuggets: data mining in financial application. **IEEE Transactions on Systems, Man and Cybernetics, Part C:** Applications and Reviews, v. 34, n. 4, p. 513-522, nov. 2004.

[44] FERNANDES, C. H. **A privacidade na sociedade da informação**. Disponível em: <http://www.linux.ime.usp.br/~carloshf/0302-mac339/fase2/node2.html>. Acesso em: 08 set. 2005.

[45] PRIVACILLA.ORG. **Source for privacy policy from a free-market, pro-technology perspective**. Disponível em: <http://www.privacilla.org>. Acesso em: jul. 2005.

[46] CHARLES, F. Privacy [a moral analysis]. In: SCHOEMAN, F. D. **Philosophical dimensions of privacy:** an anthology. Cambridge, England: Cambridge University Press, 1984. p. 203-222.

[47] SPIEKERMANN, S.; GROSSKLAGS, J.; BERENDT, B. E-privacy in 2nd generation e-Commerce: privacy preferences versus actual behavior. In: THE ACM CONFERENCE ON ELECTRONIC COMMERCE, 3, 2001. Tampa, Florida, USA. **Proceedings…** New York, NY, USA: ACM Press, oct. 2001. p. 38-47.

[48] STALLINGS, W. **Network security essentials:** applications and standards. Upper Saddle River, NJ , USA: Prentice-Hall, 2000. 366 p.

[49] WARREN, S. D.; BRANDEIS, L. D. The right to privacy. Boston, USA: **Harvard Law Review**, v. 4, n. 5, 15 dec. 15, 1890. Disponível em: < http://www-swiss.ai.mit.edu/6805/articles/privacy/Privacy_brand_warr2.html >. Acesso em: 20 jul. 2006.

[50] TELTZROW, M.; KOBSA, A. Communication of privacy and personalization in e-business. In: THE WORKSHOP WHOLES: A MULTIPLE VIEW OF INDIVIDUAL PRIVACY IN A NETWORKED WORLD, 1, 2004, Stockholm, Sweden. **Proceedings…**

[51] JUTLA, D.; BODORIK P. A client-side business model for electronic privacy. In: BLED ECOMMERCE CONFERENCE AND TRANSFORMATION, 16, Bled, Slovenia, 2003. **Proceedings…** Bled, Slovenia: 2003. p. 463-479

[52] EARP, J. B.; BAUMER, D. Innovative web use to learn about consumer behavior and online privacy. **Communications of ACM,** New York, NY. USA, v. 46, n. 4, p. 81-83, apr. 2003.

[53] ISHITANI, L. **Uma arquitetura para controle de privacidade na web**. 2003. 92 p. Tese (Doutorado em Ciência da Computação). Departamento de Ciência da Computação, Universidade Federal de Minas Gerais, Belo Horizonte, MG, Brasil, dez. 2003.

[54] ACKERMAN, M. S.; CRANOR, L. F. **Privacy critics**: safeguarding users' personal data. Web Techniques, set. 1999. Disponível em: <http://www.webtechniques.com/archives/1999/09/ackerman>. Acesso em: 13 jun 2004.

[55] WANG, H.; LEE, M. K. O.; WANG, C. Consumer privacy concerns about Internet marketing. **Communications of the ACM**, New York, NY, USA, v. 41, n. 3, p. 63-70, mar. 1998.

[56] CLARKE, R. The digital persona and its application to data surveillance. **The Information Society**, London, England, v. 10, n. 2, p. 77-92, jun. 1994. Disponível em: < http://www.anu.edu.au/people/Roger.Clarke/DV/DigPersona.html>. Acesso em: 22 abr. 2005.

[57] BENASSI, P. TRUSTe: an online privacy seal program. **Communications of the ACM**, New York, NY, USA, v. 42, n. 2, p. 56-59, fev. 1999. Disponível em: < http://portal.acm.org/ft_gateway.cfm?id=293461&type=pdf&coll=GUIDE&dl=GUIDE &CFID=19047425&CFTOKEN=24364791 >. Acesso em: 20 nov. 2005.

[58] URBAN, G.L.; SULTAN, F.; QUALLS, W.J. Placing trust at the center of your Internet strategy. **MIT Sloan Management Review**, Massachusetts, USA, v. 42, n. 1, p. 39-48. 2000.

[59] DOHERTY, S. Keeping data private. **Network Computing**, Manhasset, NY, USA, v. 12, n. 13, p. 83–91, jun. 2001.

[60] CULNAN, M. J. How did you get my name? An exploratory investigation of consumer attitudes toward secondary information use. **MIS Quarterly**, v. 17, n. 3, p. 341–363, 1993.

[61] UNITED STATES OF AMERICA. Department of Commerce, U. S. Census Bureau News. **Quarterly retail e-commerce sales**: 2nd Quarter 2005. Disponível em: <http://www.census.gov/mrts/www/data/html/05Q2.html>. Acesso em: 23 jun. 2005.

[62] FRIEDMAN, B.; KHAN Jr., P. H.; HOWE, D. C. Trust online. **Communications of the ACM**, New York, NY, USA, v. 43, n. 12, p. 34-40, dec. 2000.

[63] UNITED STATES OF AMERICA, Department of Commerce. **Discussion draft; elements of effective self-regulation for protection of privacy**. 1998. Disponível em: <www.ecommerce.gov/staff.htm>. Acesso em: 11 mar. 2004.

[64] MOORES, T. T.; DHILLON, G.. Do privacy seals in e-commerce really work?. **Communications of the ACM**, New York, NY, USA, v. 46, n. 12, p. 265-271, dec. 2003.

[65] NISSENBAUM, H. Accountability in a computerized society. In: HUMAN VALUES AND THE DESIGN OF COMPUTER TECHNOLOGY. **Proceedings…** Stanford, CA, USA: Center for the Study of Language and Information, 1997. p. 41-64.

[66] MOORES, T. Do consumers understand the role of privacy seals in e-commerce?. **Communications of the ACM**, New York, NY, USA, v. 48, n. 3, p. 86-91, mar. 2005.

[67] CHESKIN RESEARCH. **Trust in the wired Americas**. jul. 2000. Disponível em: <http://www.cheskin.com /think/studies/trustIIrpt.pdf>. Acesso em: 12 mar. 2006.

[68] BITTAR, C. A. **Os direitos da personalidade**. 2 ed. Rio de Janeiro, RJ, Brasil: Forense Universitária, 1995. 64 p.

[69] PAESANI, L. M. **Direito e Internet: liberdade de informação, privacidade e responsabilidade civil**. 2. ed. São Paulo: Atlas, 2003. 141 p.

[70] KOBSA, A. Personalized Hypermedia and International Privacy. **Communications of the ACM**, New York, NY, USA, v. 45, n. 5, p. 64-67, maio 2002.

[71] UNITED STATES OF AMERICA. Federal Trade Commission. **Privacy online**: fair information practices in the electronic marketplace. Federal Trade Commission's Report to Congress. 30 jul., 2004. Disponível em: < http://www.cdt.org/testimony/000525berman.shtml >. Acesso em: 21 abr. 2005.

[72] UNITED STATES OF AMERICA. Federal Trade Commission. **The fair credit reporting act**. 13 out. 2006. Disponível em: <http://www.ftc.gov/os/statutes/fcradoc.pdf>. Acesso em: 21 abr. 2005.

[73] AGUIAR, A. Proteção na rede. **Revista Consultor Jurídico**. jul. 2006. Disponível em: <http://conjur.estadao.com.br/static/text/46139,1>. Acesso em: 13 ago. 2006.

[74] CRANOR, L. et al. **Platform for privacy preferences project 1.0 (P3P1.0) specification**. World Wide Web Consortium recommendation, abr. 2002. Disponível em: <http://www.w3.org/TR/P3P/>. Acesso em: 8 set. 2005.

[75] SHUBINA, A. M.; SMITH, S. W. Using caching for browsing anonymity. In: ACM SIGECOM EXCHANGES, 2003. **Proccedings…** New York, NY, USA: ACM Press, v. 4, n. 2, jun. 2003. p. 11-20.

[76] PFITZMANN, A.; KÖHNTOPP, M. Anonymity, unobservability, and pseudonymity - a proposal for terminology. In: INTERNATIONAL WORKSHOP ON THE DESIGN ISSUES IN ANONYMITY AND OBSERVABILITY (DESIGNING PRIVACY ENHANCING TECHNOLOGIES), 2000, Berkeley, CA, USA. **Proceedings…** Berlin, Alemanha: Springer-Verlag, v. 2009, jul. 2000. p. 1-9.

[77] REITER, M. K.; RUBIN, A. D. Crowds: anonymity for Web transactions. **ACM Transactions on Information and System Security (TISSEC)**, v. 1, n. 1, p. 66-92, nov. 1998.

[78] ANONYMIZER. **Anonymizer enterprise network privacy/security appliance.** Technology Overview, 2004. Disponível em: <www.anonymizer.com. Acesso em: 12 maio 2004.

[79] CHAUM, D. Untraceable electronic mail, return addresses and digital pseudonyms. **Communications of the ACM**, New York, NY, USA, v. 24, n. 2, p. 84-88, fev. 1981.

[80] GOLDBERG, I.; SHOSTACK, A. **Freedom network 1.0 architecture and protocols**. out. 2001. Disponível em: <http://www.homeport.org/~adam/zeroknowledgewhitepapers/arch-tech.pdf>. Acesso em: 06 ago. 2004.

[81] BERTHOLD, O.; FEDERRATH, H.; KOPSELL, S. Web MIXes: a system for anonymous and unobservable Internet access. In: INTERNATIONAL WORKSHOP ON DESIGNING PRIVACY ENHANCING TECHNOLOGIES: DESIGN ISSUES IN ANONYMITY AND UNOBSERVABILITY, 2001, Berkeley, California, USA. **Proceedings...** New York, NY, USA: Springer-Verlag, 2001. 2001. p. 115-129.

[82] FREEDMAN, M. J.; MORRIS, R. Tarzan: A peer-to-peer anonymizing network layer. In: ACM CONFERENCE ON COMPUTER AND COMMUNICATIONS SECURITY (CCS), 9, 2002, Washington, DC, USA. **Proceedings…** New York, NY, USA: ACM Press, 2002. p. 193-206.

[83] GOLDBERG, I.; WAGNER, D.; BREWER, E. Privacy-enhancing technologies for the Internet. In: IEEE INTERNATIONAL COMPUTER CONFERENCE (COMPCON), 42, fev. 1997, San Jose. **Proceedings…** Washington, DC, USA: IEEE Computer Society, 1997. p. 103.

[84] GABBER, E. et al. Consistent, yet anonymous, Web access with LPWA. **Communications of the ACM**, New York, NY, USA, v. 42, n. 2, p. 42-47, fev. 1999.

[85] NETSCAPE COMMUNICATIONS CORPORATION. **Open directory project**. Disponível em: <http://dmoz.org/ about.html>. Acesso em: 20 ago. 2005.

[86] CRANOR, L. F.; BYERS, S.; KORMANN, D. **An analysis of P3P deployment on commercial, government, and children's Web Sites as of May 2003**. Technical Report prepared for the 14 May 2003 Federal Trade Commission. In: WORKSHOP ON TECHNOLOGIES FOR PROTECTING PERSONAL INFORMATION. Disponível em: <http://lorrie.cranor.org/pubs/p3p-census-may03.pdf>. Acesso em: 23 set 2005.

[87] CROCKER, D.; OVERELL P. **Augmented BNF for syntax specifications: ABNF**. Internet Mail Consortium, Demon Internet, nov. 1997. Disponível em: <http://www.ietf.org/rfc/rfc2234.txt>. Acesso em: 23 ago 2005.

[88] STUFFLEBEAM, W. et al. Specifying Privacy Policies with P3P and EPAL: Lessons Learned. In: WORKSHOP ON PRIVACY IN THE ELECTRONIC SOCIETY (WPES), 3, jun. 2004, Washington DC, USA. **Proceedings…** New York, NY, USA: ACM Press, 2004. p. 35.

[89] CRANOR, L. F.; ARJULA, M.; GUDURU, P. Use of a P3P User Agent by Early Adopters. In: ACM WORKSHOP ON PRIVACY IN THE ELECTRONIC SOCIETY (WPES), 21 nov. 2002, Washington, DC, USA. **Proceedings…** New York, NY, USA: ACM Press, 2002. p. 1-10.

[90] KOBSA, A.; TELTZROW, M. Contextualized communication of privacy practices and personalization benefits: Impacts on users' data sharing and purchase behavior. In: INTERNATIONAL WORKSHOP PRIVACY ENHANCING TECHNOLOGIES, 4, 2004 Toronto, Canada. **Proceedings…** p. 329-343. (Springer LNCS v. 3424).

[91] ROY MORGAN RESEARCH. **Privacy and the community**. Prepared for the Office of the Federal Privacy Commissioner, jul. 2001. Disponível em: <http://www.privacy.gov.au/publications/rcommunity.html>. Acesso em: 23 jan. 2006.

[92] UNITED KINGDOM, Department for Trade and Industry**. Informing consumers about e-Commerce:** quantitative survey report. Conducted by MORI, London. 2001. Disponível em: <http://www.mori.com/polls/2001/pdf/dti-ecommerce.pdf>. Acesso em: 03 fev. 2006.

[93] BEHRENS, L. Gartner G2. **Privacy and security**: the hidden growth strategy. 2001. Disponível em: <http://www.gartner.com/Init>. Acesso em: 12 abr. 2004.

[94] TELTZROW, M.; KOBSA, A. Impacts of user privacy preferences on personalized systems - a comparative study. In: WORKSHOP: DESIGNING PERSONALIZED USER EXPERIENCES FOR ECOMMERCE, 2004, Netherlands. **Anais…** Norwell, MA, USA: Kluwer Academic Publishers, 2003. p. 315-332.

[95] CRANOR, L. F.; REIDENBERG, J. R. Can user agents accurately represent privacy notices?. In: RESEARCH CONFERENCE ON COMMUNICATION, INFORMATION AND INTERNET POLICY, 30, 2002, Alexandria, VA, USA. **Anais…** Disponível em: <http://intel.si.umich.edu/tprc/archive-searchabstract.cfm?PaperID=65>. Acesso em: 13 ago. 2004.

[96] GRANDE, R. E.; ZORZO, S. D. Privacy protection without impairing personalization by using the extended system MASKS and the extended and contextualized P3P privacy policies. In: SIMPÓSIO BRASILEIRO DE SISTEMAS MULTIMÍDIA E WEB (WebMedia), 12, 2006, Natal, RN, Brasil. **Proceedings...** New York, NY: ACM Press, 2006. p. 89-98.

[97] GOTARDO, R. A. et al. Garantia de políticas de privacidade utilizando-se certificação digital. In: INTERNATIONAL CONFERENCE OF FORENSIC COMPUTER SCIENCE, 3, nov. 2006, Brasília, DF, Brasil. **Proceedings...** Brasília, Brasil: Departamento de Polícia Federal, v. 1, 2006. p. 82-88.

[98] GRANDE, R. E. e ZORZO, S. D. Extensão do sistema MASKS para permitir sessões de navegação. In: INTERNATIONAL SYMPOSIUM ON SYSTEM AND INFORMATION SECURITY, 8, nov. 2006, São José dos Campos, SP, Brasil. **Proceedings...** São José dos Campos, SP, Brasil: Fundação Biblioteca Nacional, 2006. p. 1-10.

[99] DILLON, A. **Designing usable electronic text:** ergonomic aspects of information

usage. Bristol, PA, USA: Taylor & Francis, 1994. 195 p.

[100] FIDEL., R. Quality methods in information retrieval research. **Library and Information Science Research**, v. 15, n. 3, p. 219-247, 1993. Disponível em: <http://www.ischool.washington.edu/fidelr/RayaPubs/QualitativeMethodsinInformation RetrievalResearch.pdf>. Acesso em: 20 ago. 2006.

[101] YIN, R. K. **Case study research**: design and methods. 3. ed. USA: Sage Publications, 2004. 200 p.

[102] HARTLEY, J. F. Case studies in organizational research. In: CASSELL, C.; SYMON, G. **Qualitative methods in organizational research:** a practical guide. London: Sage Publications, 1994. p. 208-229.

[103] DIAS, C. **Estudo de caso**: idéias importantes e referências. Disponível em: <http://www.geocities.com/claudiaad/case_study.pdf> . Acesso em: 19 ago. 2006.

[104] COUTINHO, C. P.; CHAVES, J. H. O estudo de caso na investigação em tecnologia educativa em Portugal. **Revista Portuguesa de Educação,** Universidade do Minho, Portugal, v. 15, n. 1, p. 221-243, 2002.. Disponível em: <https://repositorium.sdum.uminho.pt/retrieve/940/ClaraCoutinho.pdf>. Acesso em: 30 set. 2006.

[105] GRANDE, R. E.; ZORZO, S. D. P3P semantic checker of site behaviours. In: NETWORK CONTROL AND ENGINEERING FOR QOS, SECURITY, AND MOBILITY OF 19TH WORLD COMPUTER CONGRESS (WCC), 5, 2006, Santiago, Chile. **Proceedings…** Boston: Springer, 2006. v. 213. p. 41-53.