Fatores de privacidade e confiança em websites

Sérgio Donizetti Zorzo, Paulo Roberto Massa Cereda

RCCT

Vol. I
Nº. 1

Resumo

Mecanismos de coleta são utilizados por websites para capturar informações acerca dos usuários que acessam tais conteúdos. Essas informações são processadas e utilizadas para fornecer serviços personalizados para os usuários. Ao mesmo tempo em que a coleta das informações pode oferecer personalização de serviços, pode também violar a privacidade dos usuários. A confiança do usuário em relação a um website é fator determinante para a assiduidade de seus usuários. Assim, nota-se que websites têm utilizado técnicas para melhoria da garantia de privacidade nos serviços disponibilizados. Entretanto, tais mecanismos não fornecem métricas diretas sobre a confiança de um determinado usuário em relação a um conteúdo de um website. Este artigo apresenta um conjunto de medidas utilizadas para calcular fatores de invasão de privacidade e graus de confiança e desconfiança de um determinado usuário. Essas medidas são calculadas através da relação estabelecida entre usuário e mecanismo de coleta, obtendo-se fatores de invasão de privacidade e graus de confiança e desconfiança do usuário para um determinado conjunto de informações a ser coletado. Um estudo de caso é apresentado para evidenciar a importância da quantificação da confiança e desconfiança do usuário durante a navegação por um website.

Palavras-chave: *privacidade, confiança, Web.*

Abstract

Data collection mechanisms have been used by websites in order to collect data about users that access such contents. This information is processed and used to provide personalized services to the users. At the same time the data collection may offer services personalization, it also may violate the users' privacy. The user's trustiness on a website is a decisive factor for attendance of users. Thus, it is observed that websites has been using privacy guarantee enhancing techniques in the available services. However, such mechanisms do not provide direct measures on the user's trustiness relating to a certain content of a website. This article presents a set of measures used to calculate privacy factors and trustiness and untrustiness degrees of a certain user. These measures are calculated through the established relationship between user and data collection mechanism, obtaining privacy invasion factors and trustiness and untrustiness degrees for a certain set of information to be collected. A case study is presented in order to show the importance of measurement of trustiness and untrustiness of the user while browsing in a website.

Keywords: *privacy, dependability, Web.*

Departamento de Computação

Universidade Federal de São Carlos

Rod. Washington Luís, Km 235
Caixa Postal 676, 13565-905
São Carlos — SP — Brasil

{zorzo,paulo_cereda}@dc.ufscar.br

1 Introdução

A abrangência e popularização da Internet contribuíram com o aumento dos serviços personalizados oferecidos por websites. Entretanto, a qualidade de tais serviços depende diretamente da quantidade disponível de informação sobre o usuário. Nos últimos anos, é possível observar a proliferação de websites de qualidade duvidosa de informação e websites que realizam a coleta das informações dos usuários de modo invasivo e indiscriminado [1]. A coleta invasiva caracteriza-se por violar a privacidade dos usuários, expondo e utilizando informações pessoais sem o consentimento de seus donos. Essa exposição e mau uso reduzem drasticamente a confiança desses usuários nos websites.

Existem alguns fatores que podem alterar de modo significativo a confiança dos usuários em um determinado website. Alguns desses fatores incluem design, facilidade de uso, reputação do website, familiaridadedo usuário comtecnologias,entre outros.Do ponto de vista de privacidade, a falta de compreensão ou desconhecimento das políticas de privacidade e utilização de mecanismos de coleta são fatores importantes que alteram a confiança dos usuários [2].

Este artigo apresenta um conjunto de medidas utilizadas para calcular fatores de invasão de privacidade e graus de confiança e desconfiança de um determinado usuário em relação a um conteúdo de um website. Os graus de confiança e desconfiança do usuário são fortes parâmetros para o tratamento da privacidade do usuário. Para obter essas medidas, estabelece-se uma relação entre usuário e mecanismo de coleta (questionários e formulários), e verificam-se as características dessa relação.

Na seção 2, são tratados os aspectos de privacidade e confiança do usuário. Na seção 3, trata-se da personalização de serviços, juntamente com os mecanismos usuais de coleta de dados. A seção 4 apresenta as medidas dos fatores de privacidade e confiança em websites. Um estudo de caso é apresentado na seção 5. A seção 6 contém os resultados obtidos no estudo de caso. As conclusões deste artigo são apresentadas na seção 7.

2 Privacidade e Confiança

A privacidade na web influencia diretamente a confiabilidade dos sites e vem sendo amplamente discutida entre pesquisadores e despertando grande interesse por parte dos usuários que diariamente no uso da rede fornecem dados sobre sua navegação, sobre seu comportamento,além de seus dados pessoais. Muitas vezes esses dados são coletados sem o conhecimento e consentimento, revelando seu perfil, seu comportamento, além de outras informações que podem ser utilizadas de forma inadequada, prejudicando o mesmo.

A questão da privacidade pode ser vista por diferentes ângulos entre os diferentes usuários inclusive entre os que convivem em uma mesma sociedade. Para Warren e Brandeis [3], "privacidade é o direito de estar sozinho" e neste mesmo artigo tem-se a seguinte regra: "O direito à privacidade termina com a divulgação de fatos pelo proprietário do fato ou com o seu consentimento". Assim, a privacidade é uma questão pessoal e uma vez sendo divulgado o fato pelo seu proprietário este não possui o direito de requerer novamente a privacidade sobre tal fato.

Segundo Wang e outros [4], "privacidade geralmente se refere a informações pessoais, e invasão de privacidade é geralmente interpretada como coleta, publicação ou outro uso não autorizado de informações pessoais, como um resultado direto de transações".

Em certas ocasiões, os usuários necessitam fornecer seus dados para concretizar uma transação na web. Um exemplo seria a compra de um livro em determinado site. Para que isto ocorra devemos informar no mínimo o local de entrega e a forma de pagamento. Outras informações, muitas vezes de forma obrigatória, são solicitadas pelo site a fim de mapear o ato da compra, por exemplo, nome, data de nascimento, sexo, estado civil entre outras. Uma vez fornecidas tais informações, o usuário pode ser surpreendido com o recebimento de propagandas sem sua autorização ou ter problemas com o uso de seus dados para a realização de outras transações de forma fraudulenta.

Em determinadas situações, as navegações na web são monitoradas sem o consentimento de seus usuários.Aqui se enquadram sites que armazenam o caminho da navegação e que guardam informações no computador do usuário para posteriormente identificá-lo. Com isto pode-se identificar a frequência de visita, rastrear o caminho percorrido, o que o usuário está procurando e também descobrir outros dados que revelam o seu perfil e comportamento. Os usuários preocupados com sua privacidade procuram identificar locais realmente seguros e sérios que não fazem mau uso dos seus dados e que coletam somente os dados permitidos por eles. Segundo uma pesquisa realizada por Teltzrow e Kobsa [5], a preocupação com relação à privacidade dos dados dos usuários ficou evidente quando 64% dos usuários relataram que deixaram de acessar algum site ou mesmo de fazer compras online por não terem conhecimento de como suas informações seriam utilizadas.

A fim de estabelecer uma relação de confiança onde o usuário se sinta tranquilo para realizar suas pesquisas e transações, deve-se fornecer aos usuários clareza nas políticas de privacidade, garantir a segurança no armazenamento dos dados e o sigilo dos mesmos juntamente com serviços de personalização.

A confiança do usuário em um determinado website está diretamente associada a quatro fatores: histórico, *feedback*, resposta imediata e conformidade.

O histórico está associado à origem da fonte da informação. Ao navegar por um determinado website, o usuário pode influenciar-se por informações de histórico sobre o mesmo, obtidas de várias formas. Por exemplo, o usuário pode acessar um determinado website de comércio eletrônico porque este foi considerado idôneo por uma revista de informática ou por colegas.

O *feedback* é um dos fatores mais utilizados pelos usuários na Internet na busca da confiança e na credibilidade das informações disponibilizadas. O *feedback* é uma das formas mais comuns para troca de experiências. O usuário pode decidir realizar um acesso a um website de acordo com a opinião de outros usuários. Geralmente, o *feedback* leva em consideração as opiniões dos usuários no tempo (experiências passadas) e no espaço (quantidade de informações).

A resposta imediata é uma característica que considera a forma dinâmica da Web e requer que os websites apresentem respostas imediatas para as dúvidas do usuário. O contato entre usuário e website proporciona uma relação direta e aumenta a confiança. A conformidade está relacionada com a constatação de realização do website frente aos seus objetivos elencados.

A conformidade está muito associada aos aspectos de privacidade, pois requer que os websites demonstrem que de fato seguem suas políticas. Por exemplo, um website deve assegurar ao usuário que não utilizará as informações deste para fins alheios aos que foram previamente acordados.

Os aspectos apresentados devem ser considerados para que o usuário se sinta confiante em relação a um determinado website. É importante destacar que a confiança é algo subjetivo, extremamente pessoal e que é construída em um processo gradual, mas não necessariamente linear.

A confiança influencia na privacidade, e viceversa. Um usuário desconfiado de um determinado website pode sentir-se dessa forma devido à preocupação com a coleta dos dados e a influência dessa coleta em sua privacidade.Assim,é importante avaliar a confiança e suas implicações também do ponto de vista da privacidade.

3 Personalização de Serviços

A personalização pode ser descrita como tornar algo pessoal, individual, dependente das características e dos interesses humanos. Ao personalizar um objeto de acordo com um usuário, cria-se uma relação de afinidade. De acordo com Grande [6], "um produto ou serviço pode atender as necessidades fundamentais de uma pessoa por suas funcionalidades e características primárias."

Para utilizar a personalização, algumas informações relevantes necessitam ser obtidas para saber algo a respeito da preferência de um usuário [7]. As vantagens da personalização incluem sistema "mais próximo" do usuário, perfis diferentes de usuários, tendências de comportamento, melhorias na navegabilidade, entre outros.

A personalização pode oferecer a um site algumas funcionalidades, como reconhecimento de visitas recorrentes, interface de usuário e personalização de conteúdo, vendas colaterais, vendas por impulso, filtragem colaborativa ativa, eventos de calendário, eventos de estilo de vida e localização [8]. Há um grande benefício na utilização da personalização por sites, principalmente os de e-commerce. Com isso, diversas técnicas e ferramentas são criadas ou melhoradas para facilitar a tarefa de prover personalização.

Para a coleta de informações, geralmente são utilizadas análise de dados em formulários e análise de navegação do usuário. Com os dados obtidos, são aplicadas técnicas de mineração de dados para determinar preferências, tendências, entre outros. Os mecanismos mais comuns de coleta são os *cookies*, *clickstream*, *web bugs* e formulários, que são descritos a seguir.

O *cookie* é um grupo de dados trocados entre o usuário e o servidor,armazenados em um arquivo de texto criado no computador do usuário [9]. As informações armazenadas nos cookies podem refletir algo sobre o perfil do usuário. Os *cookies* são amplamente utilizados, principalmente para identificação de usuários. Pelo fato deste mecanismo estar incluído em todos os navegadores (e, na maioria dos casos, ativado de forma padrão), ele persistirá durante muito tempo como a ferramenta primária para propiciar personalização [8].

O *clickstream*, também conhecido como *clickpath* ou sequência de cliques, representa o caminho que o usuário percorre enquanto está visitando um determinado site. A sequência de cliques obtida, quando aplicada de forma correta, pode proporcionar informações muito importantes para as empresas.Todos os servidores Web têm a capacidade de registrar as requisições em arquivos de log ou banco de dados. Os dados de log de um servidor Web são a fonte primária de dados do *clickstream*, pois toda vez que o servidor Web responde uma requisição HTTP, uma entrada é anotada no arquivo de log. O *clickstream* é muito importante, do ponto de vista comercial, pois é possível identificar as preferências e os padrões de comportamento do usuário; isso inclui qual área lhe interessa, a frequência com que a procura e quais as informações úteis para criar estratégias de marketing mais direcionadas ao usuário e, consequente, maior chance de sucesso [10].

Web Bugs são mecanismos que tentam obter algum tipo de identificação de um usuário. Geral-

mente, estão inseridos em mensagens de e-mail ou páginas da Web. Os *Web Bugs* são muito utilizados em mensagens de e-mail, por motivos diversos. Alguns deles incluem verificar quantas pessoas leram um determinado anúncio de campanha ou determinar se uma pessoa visualizou um e-mail de *spam* (pessoas que não visualizaram são removidas das listas de e-mail) [11]. Os *Web Bugs* podem trabalhar em conjunto com os *cookies*, monitorando quais sites o usuário visita. Dessa forma, os *banners* de anúncios são específicos para cada usuário [12].

A linguagem HTML apresenta controles que permitem enviar informações para um determinado processamento no servidor. Tais informações são preenchidas através de formulários eletrônicos; através dos métodos POST ou GET fornecidos pelo protocolo HTTP, o servidor recebe a requisição e realiza o tratamento das informações submetidas. Através dos formulários, podem ser realizados dois tipos de coleta: a coleta explícita, onde o envio das informações pelo usuário envolve seu consentimento, e a coleta implícita, onde o usuário pode enviar informações adicionais, sem tomar conhecimento da existência delas. Em ambos os casos, as informações são armazenadas no lado do servidor, por exemplo, em um banco de dados.

É importante ressaltar que os próprios navegadores enviam informações interessantes, como sistema operacional do usuário, idioma, tipo de navegador, a página de referência, e outros, podendo essas informações ser utilizadas para serviços de personalização.

4 Fatores de Privacidade e Confiança em Websites

A quantidade e a qualidade de informações fornecidas pelos usuários refletem no quanto tais usuários apresentam confiabilidade do site [13]. Assim, através da análise das respostas fornecidas pelo usuário, é possível medir o quanto esse usuário confia em um determinado site.

Esta seção apresenta fatores para calcular a privacidade e a confiança dos usuários em um determinado website, através da relação estabelecida entre usuário e mecanismo de coleta explícita de dados.

Definição 4.1 (Função Nota para Itens de Entrada de Dados Optativos). Seja $V = \{v_1, v_2, \ldots, v_n\}$ o conjunto de voluntários, e $O = \{o_1, o_2, \ldots, o_m\}$ o conjunto de itens de entrada de dados optativos. Define-se a função *Nota para Itens de Entrada de Dados Optativos* M como $M : V \times O \rightarrow X$, onde $X = \{x | x \in R, 0 \geq x \geq 10\}$.

A função *Nota para Itens de Entrada de Dados Optativos* $M(v_j, o_i)$ denota o grau de invasão de privacidade do item de entrada de dados optativo o_i atribuído pelo voluntário $v_j, o_i \in O$ e $v_j \in V$. O grau de invasão de privacidade varia no intervalo real de zero (não oferece invasão alguma) até 10 (o voluntário jamais responderia tal item). Os itens de entrada de dados obrigatórios não recebem notas, pois requerem a inserção de valores pelo usuário.

Definição 4.2 (Fator de Invasão de Privacidade dos Itens de Entrada de Dados Optativos). Seja $V = \{v_1, v_2, \ldots, v_n\}$ o conjunto de voluntários, $O = \{o_1, o_2, \ldots, o_m\}$ o conjunto de itens de entrada de dados optativos, e M a função *Nota para itens de entrada de dados optativos*. O *Fator de Invasão de Privacidade do item de entrada de dados optativo* o_i é definido como:

$$\theta_i = \frac{\sum_{j=1}^{n} M(v_j, o_i)}{n}$$

onde $j, i, n \in \mathbb{N}$.

O *Fator de Invasão de Privacidade do item de entrada de dados optativo* o_i denota a média das notas atribuídas pelos voluntários ao item de entrada de dados optativo o_i. O valor de θi estará sempre no intervalo real de zero (não oferece invasão alguma) até 10 (indicando que os voluntários jamais responderiam tal item).

Definição 4.3 (Grau de Invasão de Privacidade). Seja $\Theta = \{\theta_1, \theta_2, \ldots, \theta_m\}$ o conjunto dos fatores de invasão de privacidade dos itens de entrada de dados optativos. O *grau de invasão de privacidade* α é definido como:

$$\alpha = \frac{\sum_{i=1}^{m} \theta_i}{10m}$$

onde $i, m \in \mathbb{N}$.

O *grau de invasão de privacidade* é um valor real no intervalo fechado entre zero e um e denota a quantificação amostral por voluntários da possibilidade de violação da privacidade de seus usuários.

Definição 4.4 (Grau de Privacidade). Seja α o grau de invasão de privacidade. O grau de privacidade β é definido como:

$$\beta = 1 - \alpha$$

O *grau de privacidade* β denota o fator da privacidade geral, onde β é um valor real no intervalo fechado entre zero e um.

Observe que o grau de privacidade foi obtido por respostas dadas pelos usuários – de forma amostral – considerando todos os aspectos pessoais e subjetivos do conceito em questão.

Definição 4.5 (Função Resposta para Itens de Entrada de Dados Optativos). Seja U o conjunto de usuários, $U = \{u_1, u_2, \ldots, u_l\}$, O o conjunto de itens de entrada de dados optativos, $O = \{o_1, o_2, \ldots, o_m\}$ e $\Theta = \{\theta_2, \ldots, \theta_m\}$ o conjunto dos fatores de invasão de privacidade dos itens de entrada de dados optativos. Define-se a *Função Resposta para Itens de Entrada de Dados Optativos* S como $S : U \times O \rightarrow Y$, onde $Y = \{y | y \in \{0\} U \Theta\}$. Se o usuário u_i preencheu o item de entrada de dados optativo o_k , então $S(u_i, o_k) \rightarrow 0$, caso contrário $S(u_i, o_k) \rightarrow \theta_k$.

A função *Resposta para Itens de Entrada de Dados Optativos* $S(u_j, o_i)$ denota o grau de invasão de privacidade do item de entrada de dados o_i associado ao usuário u_j, $o_i \in O$ e $u_j \in U$. Se o usuário preencheu o item de entrada de dados optativo, o grau de invasão é igual a zero. Caso contrário, o grau de invasão recebe o valor do fator de invasão de privacidade do item de entrada de dados optativo.

Definição 4.6 (Função Indicador de Preenchimento dos Itens de Entrada de Dados Obrigatórios). Seja U o conjunto de usuários, $U = \{u_1, u_2, \ldots, u_l\}$, e B o conjunto de itens de entrada de dados obrigatórios, $B = \{b_1, b_2, \ldots, b_m\}$. Define-se a *Função Indicador de Preenchimento dos Itens de Entrada de Dados Obrigatórios* I como $I : U \times B \rightarrow W$, onde $W = \{w | w \in \{0, 1\}\}$. Se o usuário u_i preencheu o item de entrada de dados obrigatório b_k, então $I(u_i, b_k) \rightarrow 1$, caso contrário $I(u_i, b_k) \rightarrow 0$.

A função *Indicador de Preenchimento dos Itens de Entrada de Dados Obrigatórios* denota o preenchimento do item de entrada de dados obrigatório b_k associado ao usuário $u_i, b_k \in B$ e $u_i \in U$. Se o usuário preencheu o item de entrada de dados obrigatório, a função retorna o valor 1. Caso contrário,a função retorna o valor zero.

Definição 4.7 (Fator de Relevância dos Itens de Entrada de Dados Obrigatórios). Seja B o conjunto de itens de entrada de dados obrigatórios, $B = \{b_1, b_2, \ldots, b_n\}$, e O o conjunto de itens de entrada de dados optativos, $O = \{o_1, o_2, \ldots, o_m\}$. O *Fator de Relevância dos Itens de Entrada de Dados Obrigatórios* é definido como:

$$\delta_B = \frac{m}{m+n}$$

onde $0 \geq \delta_B \geq 1$, $\delta_B \in \mathbb{R}$.

Definição 4.8 (Fator de Relevância dos Itens de Entrada de Dados Optativos). Seja B o conjunto de itens de entrada de dados obrigatórios, $B = \{b_1, b_2, \ldots, b_n\}$ e O o conjunto de itens de entrada de dados optativos, $O = \{o_1, o_2, \ldots, o_m\}$. O *Fator de Relevância dos Itens de Entrada de Dados Optativos* é definido como:

$$\delta_O = \frac{m}{m+n}$$

onde $0 \geq \delta_O \geq 1$, $\delta_O \in \mathbb{R}$.

Os *fatores de relevância dos itens de entrada de dados obrigatórios e optativos* denotam a importância de tais itens em um determinado mecanismo de coleta explícita.

Definição 4.9 (Grau de Desconfiança do Usuário). Seja U o conjunto de usuários, $U = \{u_1, u_2, \ldots, u_l\}$, B o conjunto de itens de entrada de dados obrigatórios, $B = \{b_1, b_2, \ldots, b_n\}$, O o conjunto de itens de entrada de dados optativos, $O = \{o_1, o_2, \ldots, o_m\}$, I a função *Indicador de Preenchimento dos Itens de Entrada de Dados Obrigatórios*, e δ_o o *fator de relevância dos itens de entrada de dados optativos*. O *Grau de Desconfiança do Usuário* u_i é definido como:

$$\Psi_i = \frac{\sum_{k=1}^{m} S(u_i, o_k)}{(\sum_{j=1}^{m} \theta_j) + 1} \delta_O (\prod_{h=1}^{n} I(u_i, b_h)) + 1 \tag{4.9a}$$

onde $0 \geq \psi_i \geq 1$, $\psi_i \in \mathbb{R}$.

A fórmula anterior leva em consideração os itens de entrada de dados optativos e obrigatórios. Caso o mecanismo de coleta explícita possua apenas um dos tipos de itens de entradas de dados, o *Grau de Desconfiança do Usuário* u_i é definido a seguir, levando-se em conta:

i) somente itens de entrada de dados optativos

$$\Psi_i = \frac{\sum_{k=1}^{m} S(u_i, o_k)}{(\sum_{j=1}^{m} \theta_j)} \tag{4.9b}$$

ii) somente itens de entrada de dados obrigatórios

$$\Psi_i = \prod_{h=1}^{n} I(u_i, b_h) \tag{4.9c}$$

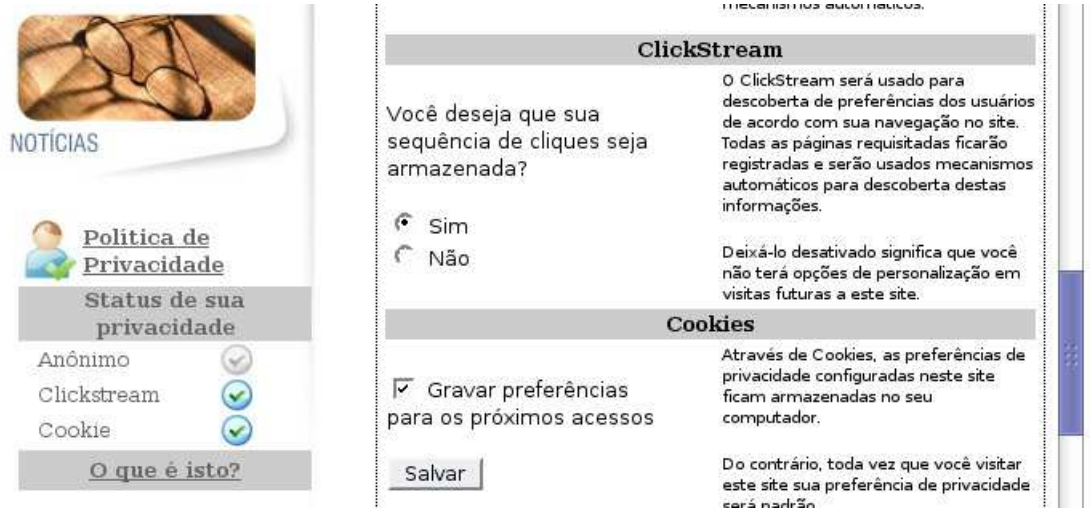

Figura 1: *Módulo de Política de Privacidade*

O *grau de desconfiança* do usuário denota a porcentagem de desconfiança do usuário em relação a um determinado mecanismo de coleta explícita contendo itens de entrada de dados obrigatórios e optativos. O valor de ψ_i estará sempre no intervalo real de zero até 1.

Definição 4.10 (Grau de Confiança do Usuário). Seja U o conjunto de usuários, $U = \{u_1, u_2, \ldots, u_l\}$, Ψ o conjunto dos graus de desconfiança dos usuários, $\Psi = \{\psi_1, \psi_2, \ldots, \psi_l\}$, e $\psi_i \in \Psi$ o grau de desconfiança do usuário $u_i \in U$. O *Grau de Confiança do Usuário* u_i é definido como:

$$\eta_i = 1 - \psi_i$$

onde $0 \geq \eta_i \geq 1$, $\eta_i \in \mathbb{R}$.

O *grau de confiança do usuário* denota a porcentagem de confiança do usuário em relação a um determinado mecanismo de coleta explícita contendo itens de entrada de dados obrigatórios e optativos. O valor de η_i estará sempre no intervalo real de zero até 1.

Definição 4.11 (Grau de Desconfiança Geral). Seja U o conjunto dos usuários, $U = \{u_1, u_2, \ldots, u_l\}$, Ψ o conjunto dos graus de desconfiança dos usuários, $\Psi = \{\psi_1, \psi_2, \ldots, \psi_l\}$ e $\psi_i \in \Psi$ o grau de desconfiança do usuário $u_i \in U$. O *Grau de Desconfiança Geral* é definido como:

$$\omega = \frac{\sum_{k=1}^{l} \psi_i}{l}$$

onde $0 \geq \omega \geq 1$, $\omega \in \mathbb{R}$.

O *grau de desconfiança geral* denota a porcentagem geral de desconfiança de todos os usuários em relação a um determinado mecanismo de coleta explícita contendo itens de entrada de dados obrigatórios e opcionais. O valor de ω estará sempre no intervalo real de zero até 1.

Definição 4.12 (Grau de Confiança Geral). Seja ω o grau de desconfiança geral. O *Grau de Confiança Geral* é definido como:

$$\alpha = 1 - \omega$$

onde $0 \geq \alpha \geq 1$, $\alpha \in \mathbb{R}$.

O *grau de confiança geral* denota a porcentagem geral de confiança de todos os usuários em relação a um determinado mecanismo de coleta explícita contendo itens de entrada de dados obrigatórios e optativos. O valor de α estará sempre no intervalo real de zero até 1.

5 Estudo de Caso

Para avaliar os fatores de privacidade e confiança apresentados na seção anterior, foi realizado um estudo de caso no website do curso de pós-graduação Lato Sensu oferecido pelo Departamento de Computação da Universidade Federal de São Carlos. Esse curso possui como público-alvo profissionais já formados na área de computação.A troca de informações entre usuários e website é realizada no momento da inscrição do aluno no curso. Nessa etapa, os aspectos de privacidade e confiança devem ser observados,pois os usuários devem fornecer seus dados pessoais para efetivação da inscrição no curso.

O website possui dois modos de exibição de sua política de privacidade: o primeiro apresenta as informações referentes à coleta dos dados de forma centralizada, na política de privacidade; o segundo apresenta tais informações de forma contextual, à medida que as coletas serão realizadas (política de privacidade distribuída). Para a análise, foi inserido um controle para alternar exibições; para cada usuário distinto, o website é apresentado através de um dos modos, alternando-se logo em seguida para o próximo usuário.

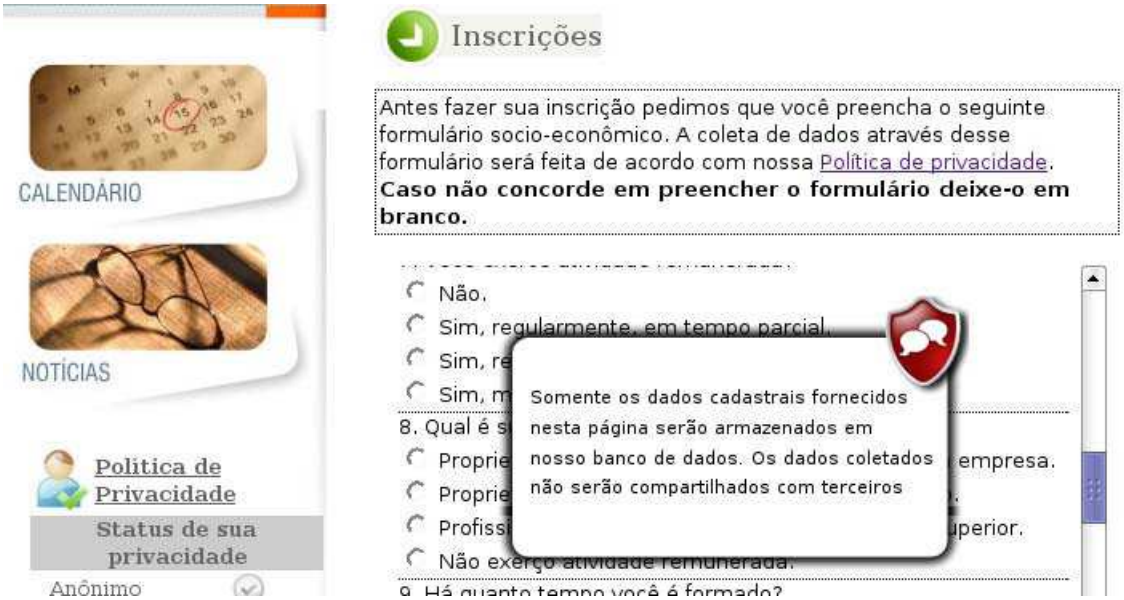

Figura 2: *Balões de contexto com um trecho da política do site*

Em ambos os modos de exibição, o módulo de privacidade é exibido de forma visível, e clicando em "Política de Privacidade" o usuário pode alterar seu status de privacidade, como ilustrado na Figura 1.

O status de privacidade do usuário indica se ele permite que seja coletado seu endereço IP, caminho percorrido no site através do *clickstream* e se permite que sejam gravados *cookies* em sua máquina. A Figura 2 ilustra um dos balões de contexto com um trecho da política de privacidade do site. Conforme o usuário respondia o questionário ou navegava pelo site, estes balões eram exibidos fazendo com que o usuário naturalmente ficasse ciente da política de privacidade.

Para realizar a coleta das informações deste experimento, foi utilizado um questionário sócioeconômico composto por 17 perguntas. Ao efetuar a inscrição, o questionário era apresentado ao usuário, podendo este preenchê-lo ou não. Após a apresentação do questionário, o usuário completaria sua inscrição fornecendo os dados necessários. O questionário utilizou apenas perguntas com opções de resposta pré-definidas e não coletou nenhuma informação sobre a identidade do usuário.

O questionário sócio-econômico aplicado poderia conter perguntas que inibissem o usuário, seja por questões pessoais, privacidade ou outra razão que o impedisse em fornecer tal informação ao website. Como não era obrigatório o preenchimento do formulário, cada pergunta foi considerada como um item de entrada de dados optativo, e calculou-se o fator de invasão de privacidade, apresentado na seção 4 (definição 4.2) para cada pergunta.

Para calcular o fator de invasão de privacidade, o experimento contou com 97 alunos dos cursos de Bacharelado em Ciência da Computação, Engenharia de Computação e Mestrado em Ciência da Computação da Universidade Federal de São Carlos, que se voluntariaram para atribuir notas às perguntas do questionário.

Os voluntários receberam o questionário juntamente com uma breve descrição dos riscos de se fornecer dados pessoais a websites. Cada voluntário atribuiu uma nota entre zero e 10 a cada pergunta do questionário. A Tabela 1 apresenta o fator de invasão de privacidade ϕ_i para cada pergunta p_i do questionário.

ϕ_1	1,89	ϕ_6	3,24	ϕ_{11}	4,39	ϕ_{16}	7,76
ϕ_2	2,51	ϕ_7	4,11	ϕ_{12}	5,35	ϕ_{17}	6,49
ϕ_3	3,32	ϕ_8	3,11	ϕ_{13}	4,00		
ϕ_4	3,44	ϕ_9	3,43	ϕ_{14}	4,04		
ϕ_5	3,19	ϕ_{10}	5,29	ϕ_{15}	5,24		

Tabela 1: *Fatores de invasão de privacidade das 17 perguntas do questionário*

Devido à não-obrigatoriedade do preenchimento de algumas perguntas do questionário, esperavasse que alguns usuários, inconscientemente ou não, deixassem algumas ou todas as perguntas sem preenchimento. Para tratar essa situação, a função *Resposta para Itens de Entrada de Dados Optativos*, apresentada na seção 4 (definição 4.5), verifica se o usuário em questão respondeu ou não uma determinada pergunta e retorna o valor correspondente.

Neste estudo de caso, o grau de desconfiança do usuário (definição 4.9) foi obtido para cada usuário que respondeu o questionário. Como o questionário é constituído apenas de itens optativos, utilizou-se a fórmula 4.9b para calcular o grau de desconfiança do usuário. Os questionários respondidos de forma mecânica (que tivessem a mesma alternativa assinalada para todas as perguntas) foram considerados como não res-

Grupo	Descrição	ω_{PPC}	ω_{PPD}
1	Usuários que enviaram o formulário sócio-econômico (vazio ou preenchido)	0,2294	0,2642
2	Usuários que não completaram a inscrição	0,2848	0,2707
3	Usuários que completaram a inscrição	0,1684	0,4960
4	Usuários que não responderam o questionário e não completaram a inscrição	1	0,9333
5	Usuários que responderam o questionário e não completaram a inscrição	0,0344	0
6	Usuário que responderam o questionário e completaram a inscrição	0,0046	0
7	Usuários que não responderam o questionário e completaram a inscrição	0,9870	0,9920

Tabela 2: *Graus de desconfiança geral obtidos nos dois modos de exibição do website do estudo de caso*

pondidos. Dessa forma, os graus de desconfiança dos usuários que assim procederam serão totais, $\psi_i = 1$. Os questionários que apresentaram mais de 50% das questões não respondidas foram agrupados com os questionários não preenchidos.

6 Resultados

O experimento foi realizado durante o período de inscrições do curso de pós-graduação Lato Sensu, de 12/11/2007 a 08/12/2007. Neste período, 66 usuários completaram o questionário sócio-econômico solicitado. Devido ao controle de alternância de exibições, 33 usuários acessaram o website com a política de privacidade centralizada, e 33 usuários acessaram com a política de privacidade distribuída.

Os graus de invasão de privacidade (definição 4.3) e de privacidade (definição 4.4) foram calculados para o website do estudo de caso, resultando nos valores $\alpha = 0,4164$ (grau de invasão de privacidade) e $\beta = 0,5836$ (grau de privacidade).

Para comparar os dois modos de exibição do website, o grau de desconfiança geral (definição 4.11) foi calculado para cada modo de exibição, além da utilização de algumas divisões para distinguir os possíveis grupos comportamentais na amostra. A Tabela 2 apresenta os graus de desconfiança geral obtidos nos modos exibição com as políticas de privacidade centralizada (PPC) e distribuída (PPD).

Os usuários foram divididos em sete grupos comportamentais. De acordo com a Tabela 2, somente o grupo 3 apresentou uma diferença considerável entre ω_{PPC} e ω_{PPD}. É possível verificar que os usuários que acessaram o website no modo de política de privacidade distribuída apresentaram um grau de desconfiança consideravelmente maior do que os usuários que acessaram no modo de política de privacidade centralizada, $\omega_{PPD} >> \omega_{PPC}$ (onde $>>$ representa "muito maior que"). Os valores obtidos através das medidas estão relacionados à privacidade e a confiança dos usuários em acessar determinados conteúdos de websites e podem ser combinados para avaliar de forma mais consistente os impactos dos aspectos de privacidade e personalização disponíveis ao usuário.

É importante destacar que os graus de privacidade e confiança gerais podem ser mais precisos com a inclusão da medição da privacidade implícita, obtida através de técnicas e métodos descritos em outros artigos [14], [15], [16].

7 Conclusões

Através do estudo de caso, é possível inferir que a partir do momento em que o usuário toma ciência sobre as questões de privacidade, ocorre um aumento da preocupação com seus dados pessoais e das informações fornecidas, aumentando, conseqeentemente, a sua desconfiança em relação ao website.

Este artigo apresentou medidas utilizadas para calcular fatores de invasão de privacidade e graus de confiança e desconfiança de um determinado usuário em um website. Tais medidas podem ser obtidas através da análise de questionários sócio-econômicos, formulários ou mesmo uma combinação de ambos.

Os fatores e graus obtidos através das medidas apresentadas podem ser utilizados como parâmetros para a definição de políticas de privacidade mais direcionadas ao usuário e ao próprio conteúdo do websites, estabelecendo regras sobre a utilização dos mecanismos de coleta existentes.

Agradecimentos

Os autores agradecem aos alunos Erlon R. Cruz, Danilo A. Moschetto, David O. Lorente e Wenceslau E. Marcomino pela realização do estudo de caso, apresentado na seção 5, para a disciplina "Tópicos em Sistemas Distribuídos: Privacidade e Personalização" do Programa de Pós-Graduação em Ciência da Computação da Universidade Federal de São Carlos, e que foi utilizado neste artigo.

Referências

[1] B. J. Fogg, J. Marshall, O. Laraki, A. Osipovich, C. Varma, N. Fang, J. Paul, A. Rangnekar, J. Shon, P. Swani, and M. Treinen, "What makes web sites credible?: a report on a large quantitative study," in *CHI '01: Proceedings of the SIGCHI conference on Human factors in computing systems*, (New York, NY, USA), pp. 61–68, ACM Press, 2001.

[2] M. J. Culnan and P. K. Armstrong, "Information privacy concerns, procedural fairness, and impersonal trust: An empirical investigation," *Organization Science*, vol. 10, no. 2, pp. 104–115, 1999.

[3] S. D. Warren and L. D. Brandeis, "The right to privacy," *Harward Law Review*, vol. 5, no. 5, pp. 193–220, 1890.

[4] H. Wang, M. K. O. Lee, and C. Wang, "Consumer privacy concerns about internet marketing.," *Communications of the ACM*, vol. 41, no. 3, pp. 63–70, 1998.

[5] M. Teltzrow and A. Kobsa, "Communication of privacy and personalization in e-business," in *Proceedings of the Workshop "WHOLES: A Multiple View of Individual Privacy in a Networked World"*, (Stockholm, Sweden), 2004.

[6] R. E. D. Grande, "Sistema de integração de técnicas de proteção de privacidade que permitem personalização," dissertacao eletrônica, Biblioteca Digital de Teses e Dissertações da Universidade Federal de São Carlos, São Carlos, SP, Brasil, outubro 2006.

[7] M. Koch and K. Möslein, "User representation in e-commerce and collaboration applications," in *Proc. 16th Bled eCommerce Conference*, (Bled, Slowenien), pp. 649–661, Jun. 2003.

[8] R. Kimball and R. Merz, *Data Webhouse: construindo o data warehouse para a Web*. Rio de Janeiro: Campus, 2000.

[9] D. Kristol and L. Montulli, "HTTP State Management Mechanism." RFC 2965 (Proposed Standard), October 2000.

[10] A. L. B. Nogueira and L. R. Oliveira Jr., *Uma análise da aplicabilidade de Data Warehouse em ambientes empresariais*. Salvador, BA, Brasil: Faculdade Ruy Barbosa, 2004.

[11] R. M. Smith, *Web Bugs: Frequently Asked Questions (FAQ)*. Public Comments: Online Profiling. FTC., 1999.

[12] W. Aiello and P. McDaniel, *Lecture 1, Intro: Privacy – Stern School of Business*. New York University, 2004.

[13] M. J. Culnan and G. R. Milne, "The culnan-milne survey on consumers & online privacy notices: summary of response," in *Interagency Public Workshop: Get noticed: Effective Financial Privacy Notices*, (Washington, D.C., USA), 2001.

[14] S. D. Zorzo, R. A. Gotardo, P. R. M. Cereda, B. Y. L. Kimura, R. A. Rios, and R. E. Grande, "An approach to treat the user's preferences about personal data," in *European Computing Conference*, vol. 27, (Vouliagmeni, Athens, Greece), pp. 1279–1286, Lecture Notes in Electrical Engineering, 2009.

[15] L. L. Lobato and S. D. Zorzo, "Avaliação dos mecanismos de privacidade e personalização na web," in *Proceedings da 32ª Conferência Latino-Americana de Informática*, (Santiago, Chile), 2006.

[16] L. L. Lobato, T. J. Bittar, and S. D. Zorzo, "Abordagem para definição de taxonomia de privacidade e personalização para design de interação e gestão do conhecimento em comunidades cscl para licenciatura em computação," in *XVII Simpósio Brasileiro de Informática na Educação*, (Brasília, DF, Brasil), 2006.

Mini-currículo do autor: Sérgio Donizetti Zorzo possui graduação em Bacharelado em Ciência da Computação pela Universidade Federal de São Carlos (1978), mestrado em Ciências da Computação pela Universidade de São Paulo (1985) e doutorado em Engenharia Elétrica pela Universidade de São Paulo (1996). Pró Reitor Adjunto de Extensão e Professor Associado da Universidade Federal de São Carlos. Pesquisador na área de Ciência da Computação, com ênfase em Teleinformática e atuando no tema de privacidade e personalização, com aplicações em serviços na web, em tv digital e em ambientes de computação sem fio.

Mini-currículo do autor: Paulo Roberto Massa Cereda possui graduação em Computação: Sistemas de Informação pelo Centro Universitário Central Paulista (2005) e mestrado em Ciência da Computação pela Universidade Federal de São Carlos (2008). Tem experiência na área de Ciência da Computação, com ênfase em Linguagem Formais e Autômatos.