

Cr terios para Tecnologias de Melhoria da Privacidade

Paulo R. M. Cereda
Reginaldo A. Gotardo
S rgio D. Zorzo

Departamento de Computa  o
Universidade Federal de S o Carlos

T picos em Sistemas Distribu dos e Redes:
Privacidade e Personaliza  o

Agenda

- 1 Privacidade
- 2 Anonimato
- 3 Não-observância
- 4 Não-relacionamento
- 5 Pseudônimos
- 6 Uso da identidade

Aspectos de segurança

- São utilizados para a melhoria da privacidade
- Têm o objetivo de proteger:
 - Identidade de um usuário
 - Uso das identidades
 - Dados pessoais

- Anonimato
- Não-observância
- Não-relacionamento
- Pseudônimos

Anonimato

Definição

O anonimato de um indivíduo significa que este indivíduo não é identificado^a entre um conjunto de indivíduos, que é o conjunto de anonimato.

^aPor “não identificado” entende-se “não caracterizado de modo único”

Anonimato

Definição formal

Seja U um usuário, R um papel, E um evento, A um atacante, NC_A o conjunto de usuários que não estão cooperando com A , B uma observação, e R_U é o usuário U realizando o papel R durante um evento E . Então:

- U é anônimo, $U \in NC_A$, em um papel R para um evento E em relação a um atacante A se, para cada observação B ,
 $\forall U' \in NC_A : 0 < P(R_{U'}|B) < 1$
- U é totalmente anônimo se $\forall U' \in NC_A : P(R_{U'}) = P(R_{U'}|B)$

Anonimato

- Anonimato de um indivíduo
- Anonimato de um grupo
 - Anonimato global
- Anonimato do emissor
 - O usuário que envia uma mensagem é anônimo
- Anonimato do receptor
 - O usuário que recebe uma mensagem é anônimo

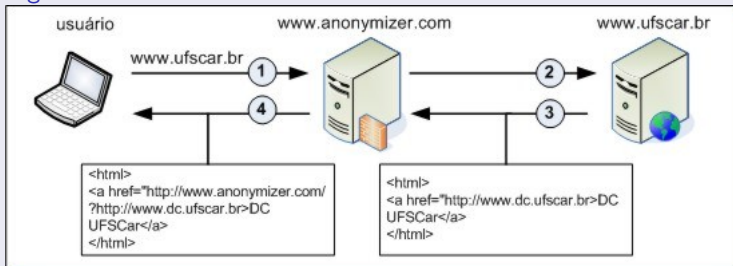
Privacidade
Anonimato
Não-observância
Não-relacionamento
Pseudônimos
Uso da identidade
Conclusões

Definição
Definição formal
Classificações
Exemplo

Anonimato

Exemplo

Figure:



Não-observância

Definição

É a situação onde um usuário utiliza um determinado recurso ou serviço sem que outros sejam capazes de observar essa ação.

Não-observância

Definição formal

Seja E um evento, A um atacante, e B uma observação. Então:

- Um evento E é não-observável para um atacante A se, para cada observação B que A pode fazer, $0 < P(E|B) < 1$
- Um evento E é totalmente não-observável se $P(E) = P(E|B)$

Não-observância

- Não-observância do emissor
 - Não é possível saber se um usuário enviou algo
- Não-observância do receptor
 - Não é possível saber se um usuário recebeu algo
- Não-observância do relacionamento
 - Não é possível saber se houve envio ou recebimento

Não-observância

Exemplo

Um usuário pode acessar um determinado serviço em uma página Web sem que seja notado.

Não-relacionamento

Definição

É a situação onde dois ou mais itens de interesse podem utilizar-se de um recurso ou serviço sem que outros sejam capazes de estabelecer uma relação entre esses usos

Não-relacionamento

Definição formal

Seja B uma observação, A um atacante, E e F são eventos, e $X_{E,F}$ significa que E e F têm uma característica correspondente. Então:

- Dois eventos E e F são não-relacionados em relação a uma característica para um atacante A se, para cada observação B que A pode fazer, $0 < P(X_{E,F}|B) < 1$
- E e F são totalmente não-relacionados se $P(X_{E,F}|B) = P(X_{E,F})$

Não-relacionamento

Exemplo

Existem duas pesquisas feitas em um determinado mecanismo de busca: a primeira pesquisa foi sobre “bola de futebol”, e a segunda foi sobre “jogo de panelas”. Essas pesquisas são relacionadas?

Pseudônimos

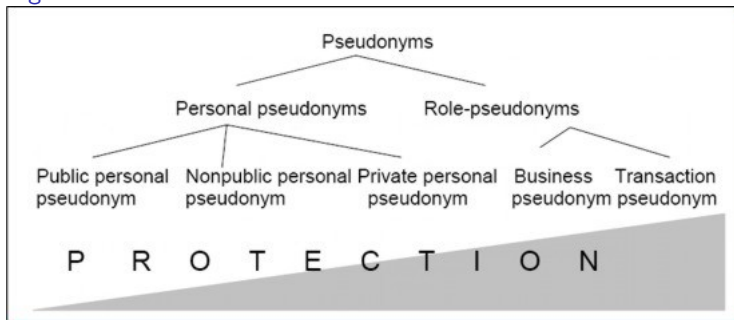
Definição

Um pseudônimo é uma identificação atribuída a um usuário que garante que este usuário poderá utilizar um recurso sem revelar sua identidade real.

Pseudônimos

- Classificação dos pseudônimos quanto ao nível de relacionamento

Figure:



Pseudônimos

- Pseudônimo pessoal
 - Substitui o nome do usuário
- Pseudônimo de papel/regra
 - Aplica uma determinada regra para atribuir pseudônimo
- Pseudônimo de relacionamento
 - Para cada comunicação, um pseudônimo diferente é utilizado
- Pseudônimo de regra-relacionamento
 - Para cada regra e comunicação, um pseudônimo diferente é utilizado
- Pseudônimo transacional
 - Para cada transação, um pseudônimo é utilizado

Privacidade
Anonimato
Não-observância
Não-relacionamento
Pseudônimos
Uso da identidade
Conclusões

Definição
Classificação
Exemplo

Pseudônimos

Exemplo

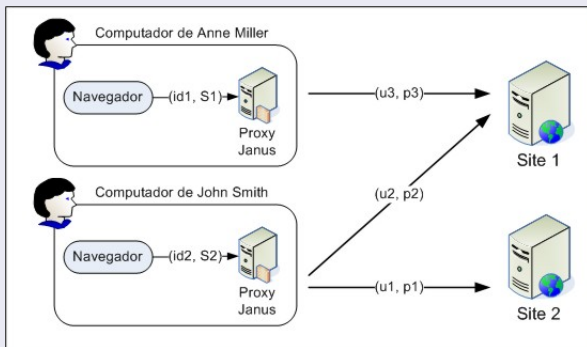


Figure:

Uso da identidade

- Despersonalização
- Reidentificação
- Entropia
- ANVC
- RR

Despersonalização

Despersonalização perfeita

Dados são renderizados anonimamente de tal modo que o usuário não é mais identificável.

Despersonalização prática

Há a modificação de dados pessoais de modo que determinadas informações não estejam atribuídas a um usuário identificado ou identificável.

Reidentificação

Registros coletados para fins estatísticos

- Dados de identidade
- Dados demográficos
- Dados de análise

Nível de anonimato de um dado estatístico

- Depende do tamanho do banco de dados
- Depende da entropia dos dados demográficos

Reidentificação

Entropia dos atributos de dados demográficos

- Depende do número de atributos
- Depende do número de valores possíveis para cada atributo
- Depende da distribuição de freqüência dos valores
- Depende das dependências entre atributos

Entropia

Definição

- Medida da variação ou desordem em um sistema
- Medida da desordem ou da imprevisibilidade da informação

Entropia

Definição formal

Dados m atributos $X_1 \dots X_m$, com valores $x_{i1} \dots x_{in_i}$ para X_i , e $\sum_{j=1}^{n_i} p(x_{ij}) = 1$, então a entropia $H(X_i)$ do atributo X_i é dada por:

$$0 \leq H(X_i) = \sum_{j=1}^{n_i} p(x_{ij}) * \log_2(1/p(x_{ij})) \leq \log_2(n_i)$$

Entropia

Definição

A entropia de um atributo X_i :

- aumenta à medida em que o número n_i de possíveis valores aumenta.
- diminui à medida que a distribuição dos valores dos atributos se torna mais e mais distorcida. $H(X) = 0$ quando $p(x_{ij}) = 1$ para qualquer valor x_{ij}
- é máxima se todos os valores são semelhantes.

Entropia

Definição formal

A entropia de uma combinação de atributos $X_1 \dots X_m$ é definida por:

$$\begin{aligned} 0 &\leq H(X_1 \dots X_m) \\ &= \sum_{j_1=1}^{n_1} \sum_{j_m=1}^{n_m} p(x_{1j_1} \dots x_{mj_m}) * \log_2(1/p(x_{1j_1} \dots x_{mj_m})) \leq \\ &H(X_1) + \dots + H(X_m) \end{aligned}$$

Entropia

Exemplo

Seja o atributo *sexo* com $p(\text{male}) = 0.469$ e $p(\text{female}) = 0.531$.

Então $H(\text{sexo}) =$

$$p(\text{male}) * \log_2(1/p(\text{male})) + p(\text{female}) * \log_2(1/p(\text{female})) =$$

$$0.469 * \log_2(1/0.469) + 0.531 * \log_2(1/0.531) = 0.997$$

ANVC

Definição

ANVC e o número médio de combinações de valores.

ANVC

Medida para o número médio de combinações de valores para os atributos $X_1 \dots X_m$ que podem realmente ser utilizadas para reidentificação.

ANVC

Definição formal

A função ANVC é definida por $ANVC(X_1 \dots X_m) = 2^{H(X_1 \dots X_m)}$.

Privacidade
Anonimato
Não-observância
Não-relacionamento
Pseudônimos
Uso da identidade
Conclusões

Despersonalização
Reidentificação
Entropia
ANVC
RR

ANVC

Exemplo

$$ANVC(\textit{sexo}) = 2^{0.997} \approx 1.996$$

ANVC

Definição formal

Se não existem dependências entre os atributos $X_1 \dots X_m$, então

$$ANVC(X_1 \dots X_m) = \prod 2^{H(X_i)}$$

Se para cada atributo X_i todos os valores $x_{i1} \dots x_{in_i}$ são bem semelhantes, então $ANVC(X_1 \dots X_m) = \prod n_i$

RR

Definição

RR é uma função utilizada para estimar o risco médio de reidentificação. Para o seu cálculo, utiliza a função ANVC.

Definição formal

A função RR é definida por:

$$RR(X_1 \dots X_m) =$$

- $ANVC(X_1 \dots X_m)/N$, se $ANVC(X_1 \dots X_m) \leq N$
- 1, se $ANVC(X_1 \dots X_m) > N$

Exemplo

Dados:

- banco de dados com $N = 100$ registros para N indivíduos diferentes
- atributos demográficos *sexo* e *estado civil*, onde $H(\text{sexo}, \text{estadocivil}) = 2.61$

Então:

$$ANVC(\text{sexo}, \text{estadocivil}) = 2^{H(\text{sexo}, \text{estadocivil})} = 2^{2.6} = 6.105$$

Assim, o risco de reidentificação pode ser calculado por:

$$RR(\text{sexo}, \text{estadocivil}) = 6.105/100 = 0.0615$$

A estimativa de porcentagem de indivíduos reidentificáveis é:

$$RR(X_1 \dots X_m) * 100\%$$

Conclusões

- Necessidade de proteger informações pessoais!
- Existem mecanismos que buscam fornecer privacidade ao usuário, através dos aspectos citados nesta apresentação

A seguir...

Proteção em Nível de Comunicação

- Jantar dos Criptógrafos - Cláudio
- Mix Net - Filipe