



# Introdução

**Segurança de Computadores:** proteção colocada num sistema automatizado de informação de maneira a atingir os objetivos aplicáveis de preservação da integridade, disponibilidade e confidencialidade dos recursos do sistema de info.

**Confidencialidade:** evitar divulgação não autorizada de informação — envolve a proteção de dados, fornecendo acesso àqueles que têm permissão para os ver enquanto proíbe outros de aprenderem qualquer coisa sobre o seu conteúdo

1. **Encriptação:** transformação da informação usando um regrado (chave de encriptação) para que a informação transformada só possa ser lida usando outro regrado (chave de desencriptação)

2. **Controlo de Acesso:** regras e políticas que limitam acesso a informação confidencial às pessoas e/ou sistemas que "têm de saber" para sua identidade ou papel

3. **Autenticação:** determinação da identidade ou papel que alguém tem, baseado em algo que tem, sabe ou é

4. **Anonimidade:** propriedade de certos registos ou transações não sejam atribuíveis  
↓  
a nenhum indivíduo

a. **Agregação:** combinação de dados de muitos indivíduos para que as somas ou médias reveladas não possam ser ligadas a nenhum indivíduo

b. **Mistura:** entrelacamento de transações, informação ou comunicações de forma a que não possam ser rastreadas para nenhum indivíduo

c. **Proxies:** agentes confiados que estão dispostos a participar em ações para um indivíduo de forma a que não possam ser rastreados para essa pessoa

d. **Pseudónimos:** identidades fictícias que podem ocupar o lugar de identidades reais em comunicações e transações, mas só conhecidas para uma entidade confiada

→ **Confidencialidade dos Dados:** informação confidencial não é divulgada indevida

→ **Privaçade:** indivíduos controlam as informações com eles relacionadas

**Integridade:** propriedade de a informação não ser alterada de forma não autorizada

1. Backups: arquivo periódico de dados

2. Checksums: computação de função que mapeia os conteúdos de um ficheiro num valor numérico

3. Códigos de Correção de Dados: pequenas alterações nos dados são facilmente detectadas e automaticamente corrigidas

→ **Integridade da Dados:** informação alterada só de forma autorizada

→ **Integridade de Sistemas:** sistema realiza a sua função pressuposta

**Disponibilidade:** propriedade de a informação ser acessível e modificável atempadamente por aqueles autorizados para o fazer — o serviço não é negado

1. Proteções Físicas: infraestrutura para manter a informação disponível

2. Redundância Computacional: computadores e dispositivos de armazenamento

**Autenticidade:** capacidade de determinar que afirmações, políticas e permissões impostas por pessoas ou sistemas são genuínas

1. Assinaturas Digitais: computações criptográficas que permitem a uma pessoa ou sistema comprovar a autenticidade dos seus documentos de forma única que garante não-repúcio — afirmações autênticas não podem ser negadas pelo seu autor

→ **Anonimato na Publicação de Dados:** não divulgar a identidade dos utilizadores

→ **Computação Segura:** garantiu a mínima fuga de informação possível

**Eavesdropping:** interceptação de informação enviada para ontem durante a transmissão

**Alteração:** modificação não autorizada de informação

**Denial-of-Service:** interrupção / degradação de um serviço de dados ou acesso à informação

**Masquerading:** fabrico de informação com autoria de alguém que não é o autor

**Correlação/Mashup:** integração de múltiplas fontes para determinar a origem de dados

# Publicação de Dados com Preservação da Privacidade

→ Ataque de Ligação: remover PII não é suficiente!

## Classificação de Atributos:

1. Atributo Chave: identificador único/explícito (PII)
2. Quase-Identificador: atributo (ou conjunto de atributos) que não identifica explicitamente um utilizador, mas pode ser combinado com outros dados de fontes públicas para desanonymizar o proprietário de um registo
3. Atributo Sensível: atributo privado específico do indivíduo que não deve ser publicamente divulgado - pode ser ligado para identificar indivíduos

O processo de desidentificação/anonymização é iterativo!

Distinção: grau para o qual as variáveis tornam os registos distintos

Separação: grau para o qual as combinações de variáveis separam os registos

→ Valores altos de distinção e separação indicam QID prováveis!

Chave/Identificador: subconjunto de atributos que identifica unicamente cada tuplo/registro numa tabela

QID  $\alpha$ -Distinto: subconjunto de atributos que se torna uma chave na tabela depois da remoção de no máximo  $1-\alpha$  (fracção) dos tuplos na tabela original

Um subconjunto de atributos separa um par de tuplos X e Y se X e Y têm valores diferentes em pelo menos um atributo no subconjunto

# OPERAÇÕES DE ANONIMIZAÇÃO

1. Generalização: substituição de um valor por outro mais geral
  - a. Full-Domain: todos os valores são generalizados para o mesmo nível
  - b. Local/Subtree: podem ser aplicados diferentes níveis de generalização
2. Supressão: remoção de alguns valores de atributos
  - a. Linha: remoção de registros/entregadas inteiros
  - b. Coluna: remoção de todos os valores de um atributo
  - c. Valores Selecionados
3. Anatomizações: desassociação de PIDs e atributos sensíveis:
  - ↪ dados não modificados nas tabelas PT e ST
  - ↪ número adicional de tabelas
4. Perturbação: substituição dos dados originais por valores sintéticos com informação estatística idêntica
  - a. Adição de Ruído: substituir o valor sensível original  $s$  com  $s+r$  onde  $r$  é um valor aleatório amostrado de alguma distribuição
  - b. Troca de Dados: trocar valores de atributos sensíveis entre registros individuais
  - c. Geração de Dados Sintéticos: gerar dados sintéticos que retêm informação estatística útil

Operações de Anonimização: blocos/componentes para desenvolver modelos de privacidade mais evoluídos

## ASSUNÇÕES FORTES

Modelo Sintático: especifica condições sintáticas para libertar dados

Modelo Semântico: usa informação nas características dos próprios dados para adicionar relativamente ruído ao output

## ASSUNÇÕES FRACAS

# Modelos Sintáticos de Privacidade

K-Anomimidade: os atributos identificáveis de qualquer registo da base de dados são indistinguíveis de pelo menos  $k-1$  outros registos, formando uma classe de equivalência — quaisquer QID presentes na base de dados devem aparecer em pelo menos  $k$  registos

Generalização: substitui quase-identificadores por valores menos específicos, mas semanticamente consistentes até obter  $k$  valores idênticos — "masking"

• K-Anomimidade não fornece privacidade se (1) os valores sensíveis numa classe de equivalência não tiverem diversidade e (2) o atacante tiver conhecimento de fundo

L-Diversidade: cada classe de equivalência tem pelo menos  $L$ -valores bem representados — os atributos sensíveis devem ser "diversos" nas classes de equivalência dos QID

1. L-Diversidade Distinta: há pelo menos  $L$  valores distintos em cada classe de equivalência — um valor pode ser mais frequente do que outros...

2. L-Diversidade por Entropia: a entropia da distribuição dos valores sensíveis em cada classe de equivalência é pelo menos  $\log(L)$  — pode ser restritivo

3. (C, L)-Diversidade Recursiva:  $r_i \leq c(r_1 + \dots + r_m)$ , sendo  $r_i$  a frequência do  $i$ -ésimo valor mais frequente — o valor mais frequente não aparece demasiado frequentemente

• L-diversidade não considera (1) a distribuição global dos valores sensíveis nem (2) a semântica dos valores sensíveis

1- Proximidade: a distribuição dos valores sensíveis em cada classe de equivalência deve estar próxima da distribuição correspondente na tabela original

↓ limitada superiormente pelo limite  $t$

$P$ : distribuição dos valores dos atributos sensíveis na Tabela original  
 $P$ : distribuição do mesmo atributo numa classe de equivalência

$$\text{Dist}(P, Q) \leq t$$

A medida de distância deve considerar o significado semântico, isto é, a distância entre valores

Earth Movers Distance (EMD): quantidade mínima de trabalho para transformar uma distribuição noutra

L-divisidade não considera as semânticas dos valores sensíveis

### Problemas Comuns:

1. Inssegurança contra atacantes com informação de fundo arbitrária
2. Não composição — anonimizações duas vezes revelam dados
3. Dificuldade em derivar
  - a. uma noção significativa de privacidade
  - b. uma noção significativa de utilidade
4. Como medir o risco de re-identificação?

# Risco de Re-Identificação

Alvo: indivíduo a ser re-identificado - específico ou aleatório

D: base de dados original

U: base de dados publicada

$U \subseteq D$

## Modelos de Atacante:

1. Procurador: um indivíduo específico -  $U = D$
2. Jornalista: qualquer indivíduo -  $U \subset D$
3. Marketeer: tantos indivíduos quanto possível

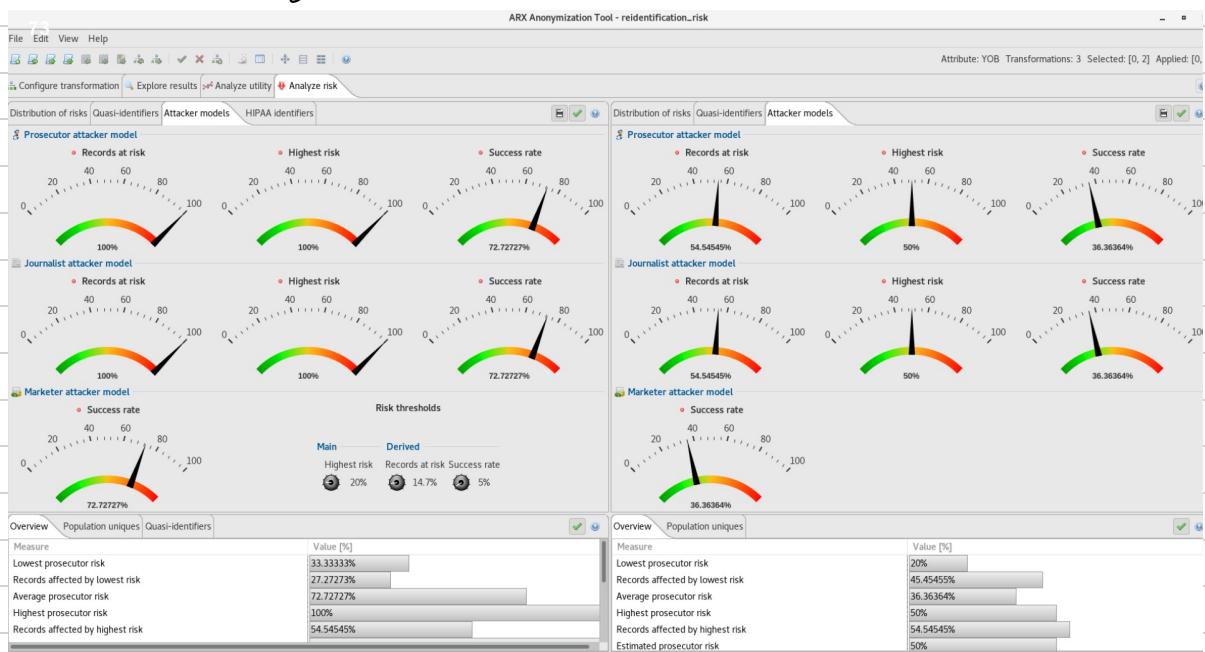
$F_j$ : tamanho da classe de equivalência

$J$ : conjunto de classes de equivalência

$$P_{\text{PROCURADORA}} = \frac{1}{F_j}$$

$$P_{\text{JORNALISTA}} = \frac{1}{\min(F_j)}$$

$$P_{\text{MARKETER}} = \frac{|J|}{n}$$



# Utilidade dos Dados

→ PRIVACIDADE vs. UTILIDADE

Objetivo: minimizar a perda de informação e o risco de re-identificação

Precisão: penalização para cada instância de valor de atributo suprimido ou generalizado — altura do nó na hierarquia sobre a altura hierárquica máxima

Perda de Informação: rácio de nós folha que não são generalizados

Discriminabilidade: penaliza registos porarem indistinguíveis de outros registos no conjunto de DID, usando o conceito de classes de equivalência

Tamanho Médio das Classes de Equivalência:  $C_{AVG} = \frac{\text{total de registos}}{\text{total de CEs}}$

Ganho de Informação, Perda de Privacidade:  $IGPL(\lambda) = \frac{IG(\lambda)}{PL(\lambda) + 1}$

• A generalização errata na localidade espacial **MAS** os dados são esparsos  
• A projeção para poucas dimensões perde informação → K-anonimidade inválida

Modelo de Privacidade LkC: qualquer combinação de valores em  $DID' \subseteq DID$  com comprimento máximo  $L$ , partilhada por pelo menos  $K$  registos e com confiança de inferir quaisquer valores suspeitos menor do que  $C$

• Se o conhecimento do adversário não excede  $L$  valores, então as probabilidades de ligação de registos estão limitadas por  $1/K$  e de atributos por  $C$

# Privacidade Diferencial

Objetivo: a publicação de uma base de dados revela " pouco " sobre qualquer indivíduo, mesmo que um atacante saiba (quase) tudo sobre qualquer outro - probabilidade indistinguível - perturbação de dados

Privacidade Diferencial: a diferença entre (distribuições de) probabilidades de uma interrogação retornarem o mesmo resultado em dois conjuntos de dados distintos está limitada por  $\epsilon$

$D_1, D_2$ : bases de dados que diferem no máximo em 1 elemento (vizinhos)

$f$ : função de interrogação sobre a base de dados

$D$ : domínio dos dados

$M$ : função/algortimo aleatorizado que adiciona ruído às interrogações

$$M(D) = f(D) + \text{ruído}$$

$$\Pr[M(D_1) \in S] \leq e^{\epsilon} \Pr[M(D_2) \in S]$$

Para quaisquer bases de dados vizinhas que diferem em 1 registo, o adversário não é capaz de distinguir  $D_1$  e  $D_2$  do resultado  $M$

A inclusão/exclusão de um indivíduo é probabilisticamente indistinguível

A quantidade de ruído a adicionar depende da interrogação  $f$

Sensitividade:  $\Delta f = \max_{D_1, D_2} \|f(D_1) - f(D_2)\|$  - mede a diferença que 1 indivíduo faz no resultado da interrogação

→ Quanto maior a sensitividade, mais ruído deve ser adicionado!

→ Quanto menor  $\epsilon$ , mais ruído!

→ O resultado da composição é  $\epsilon_1 + \epsilon_2$  diferencialmente privado!

# Privacidade de Localização

Os dados de localização não são realmente sensíveis porque (1) os rastros de mobilidade são altamente únicos, (2) os pontos-de-interesse atuam como quase-identificadores, (3) podem revelar hábitos, religiões, condições de saúde... e (4) não são altamente previsíveis dado o histórico passado

Privacidade de Informação: a capacidade de indivíduos, grupos ou instituições determinarem por eles próprios quando, como e até que ponto a informação sobre eles é comunicada a outros

Privacidade de Localização: a capacidade de um indivíduo se mover no espaço público com a expectativa de que, sob circunstâncias normais, a sua localização não vai ser sistematicamente gravada para uso futuro

Agressão	1. Conhecimento de Fundo	a. dados disponíveis b. construção de conhecimento
	2. Ataque	a. objetivo - desanominalização/localização b. método - ligação de contexto/probabilidades

## Ataques de Identidade:

1. Ligação de Contexto: desanominalização através do cunhamento de dados públicos
2. Machine Learning + Ligação de Contexto: extração de demográficos por similaridade
3. Baseado em Probabilidades: desanominalização através de similaridade

## Ataques de Localização:

1. Baseado em Probabilidades: extração de locais suspeitos
2. Ligação de Contexto: correspondência com mapas

# MECANISMOS DE PRESERVAÇÃO DE PRIVACIDADE DE LOCALIZAÇÃO

## Mecanismos de Anonimização:

1. K-Anonymity: um sujeito é K-anônimo em relação à informação de localização se e só se a informação de localização apresentada é indistinguível da informação de localização de pelo menos  $K-1$  outros sujeitos
  - malhação da dimensionalidade
  - esparsidade dos dados
2. LKC-Privacy: qualquer combinação de valores de QID com comprimento máximo L (conhecimento de fundo assumido) é partilhado por pelo menos  $K$  registos com uma confiança de inferir qualquer valor numérico menor ou igual a C
3. Mix-Zones:
  - a. o espaço é dividido em zonas mistas e zonas aplicacionais
  - b. as interrogações só são feitas em zonas aplicacionais
  - c. os pseudónimos só são alterados em zonas mistas - mistura de identidades
  - d. os utilizadores podem requerer um mínimo de utilizadores numa zona mista
  - o tamanho e a forma das zonas mistas podem comprometer a anonimização

- ## Mecanismos de Ofuscamento:
- reduzem a precisão/granularidade da informação
1. "Dummies": adicionar vários falsos positivos juntamente com a localização real
    - criação de "dummies" realistas SOLUÇÃO
    - consistência de movimento; rastreabilidade;
    - custos de comunicação
    - área anónima adaptável
  2. Privacidade Diferencial → Geo-indistinbibilidade: seja  $K(\cdot)$  um mecanismo de ofuscamento que, dada uma posição  $x$ , atribui uma função de densidade de probabilidade de registar  $x' \sim K(\cdot)$  satisfaça  $E\text{-geo-indistinbibilidade}$   $\| = E_x$

Certeza: confiança/ambiguidade de um adversário

Correção: probabilidade de um adversário estar correto, - erro estimado esperado

Ganho/Perda de Informação: ganho de informação do adversário

Geo-Indistinbibilidade:  $E = \| / n \Rightarrow \| = E n$

Tempo: tempo de que um adversário precisa para comprometer a privacidade de um

# Secure Multiparty Computation

- A partilha de dados é necessária para utilização total  
**MAS**
- As organizações não podem partilhar diretamente os seus dados  
**PORQUE**
- A grande quantidade de dados disponíveis significa que é possível aprender muita informação sobre indivíduos a partir de dados públicos

## Mineração de Dados com Preservação da Privacidade:

1. Publicar dados estatísticos, previamente modificados de modo que não comprometem a privacidade de ninguém e ainda permitem obter resultados significativos
2. Dividir os dados por várias partes diferentes, executando o algoritmo de mineração de dados na união das bases de dados nem permitir que qualquer parte veja outras bases de dados

**Privacidade & Autonomia:** informação considerada pelos próprios como pessoal, confidencial ou privada não deve ser distribuída ou publicamente conhecida

**Privacidade & Controle:** informação pessoal/privada não deve ser mal utilizada

**Computação Segura:** mostra como computar uma função que todas as partes querem computar da forma mais segura, garantindo a mínima fuga de informação, **MAS**  
**NÃO** considera se o próprio output da função revela informação suspeita ou se as partes devem concordar em computar a função

**"Secure Multiparty Computation":** dado um conjunto de partes com inputs privados, que querem computar conjuntamente uma função dos seus inputs de modo que certas propriedades de segurança sejam preservadas, mesmo que alguma das partes tente atacar deliberadamente o protocolo — privacidade e corretude

**Modelo Ideal:** as partes enviam inputs para uma parte confiável que computa a função  
**Modelo Real:** as partes executam um protocolo real nem ajuda confiável

Um protocolo é seguro se qualquer ataque num protocolo real puder ser levado a cabo no modelo ideal — como nenhum ataques podem ser conduzidos no modelo ideal, a segurança está implicada

**Distinguibilidade Computacional:** qualquer observador probabilístico em tempo-polinomial que receba a distribuição de input/output das partes honestas e do adversário faz output de 1 ao receber a distribuição gerada em IDEAL com probabilidade negligivelmente perto de quando é gerada em REAL

O protocolo computa de forma segura se a probabilidade de distingibilidade é pequena!

**Privacidade:** o adversário não consegue aprender mais sobre o input da parte honesta do que o que foi revelado pelo output da função

**Correção:** a função é computada corretamente

- Independência de Inputs
- Entrega Garantida de Outputs
- Justiça

**Parte:** envia input para a parte confiável

**Adversário:** escolhe os inputs das partes corrompidas

# PRIMITIVAS DE SECURE MULTIPARTY COMPUTATION

## Oblivious Transfer (OT):



1. Receptor escolhe  $(pk, sk)$  e  $pk'$  (nem  $sk'$ )
2. Receptor define  $pk_0 = pk$  e  $pk_{1-\sigma} = pk'$   $\sigma \in \{0, 1\}$
3. Receptor envia  $pk_0$  e  $pk_1$
4. Emissor computa  $c_0 = E(pk_0, m_0)$  e  $c_1 = E(pk_1, m_1)$
5. Receptor decifra  $c_\sigma$  usando  $sk$  e obtém  $m_\sigma$

O protocolo assume comportamento semi-honesto do receptor!

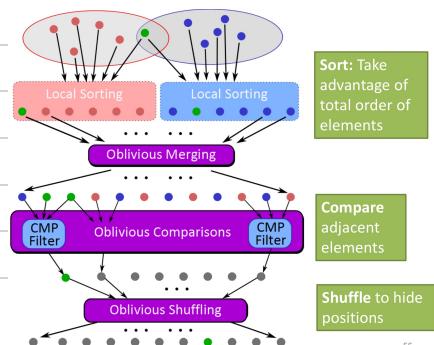
## Bit Commitment:

- COMMIT**
1. "Committer" tem um bit  $\sigma$
  2. Receptor obtém uma string de compromisso  $c = (pk, c)$
- REVEAL**
3. "Committer" envia uma mensagem de "decommit" para o receptor
  4. Receptor usa a mensagem de "decommit" e  $c$  para obter o

Vínculo: para qualquer  $c$ , não existe um  $\sigma$  para o qual o "decommit" é aceite

Dúltigão: o receptor não consegue distinguir strings de compromisso

Zero Knowledge: um provador quer provar uma afirmação para o verificador de modo que (1) o verificador não vai aprender nada para além do facto de que a afirmação está correta, e (2) o provador não vai ser capaz de convencer o verificador de uma afirmação errada



## → Intersecção de Conjuntos Privados

# Aplicações de SMC

Computação Segura: computar um algoritmo de mineração de dados nos dados de modo que nada para além do output seja aprendido — não lida com o processo de computar a função

Privacidade: decidir se a função deve ser computada

Segurança: aplicar técnicas de computação segura para computar a função

Jornal Personalizado: a computação segura fornece privacidade dos dados, mas não AUTONOMIA — as regras de organização devem ser públicas ou definidas pelos próprios utilizadores

Catálogo Personalizado: a computação segura fornece privacidade dos dados, mas não protege contra DISCRIMINAÇÃO de preços — PROFILING / CLUSTERING

Agrupamento com Preservação de Privacidade: agrupam um conjunto de entidades, sem revelar qualquer valor no qual o agrupamento é baseado

→ Dados Verticalmente Particionados

K-Means:

1. usar pontos iniciais para particionamento
2. computar novos centroides da partição atual
3. renovar pertença com base novos centroides

K-Means com Preservação de Privacidade:

1. desfigurar as componentes de distância com valores aleatórios que se cancelam
2. comparar distâncias para que só K saiba a comparação
3. permutar a ordem dos grupos para que o real significado seja desconhecido

$$A: \vec{V} = (v_1, \dots, v_n), \quad B: \vec{X} = (x_1, \dots, x_n) \quad \pi(\vec{X} + \vec{V})$$

# Privacidade na Internet

Cliente: navegador que pede, recebe e mostra objetos Web

Servidor: serviço Web que envia objetos em resposta a pedidos

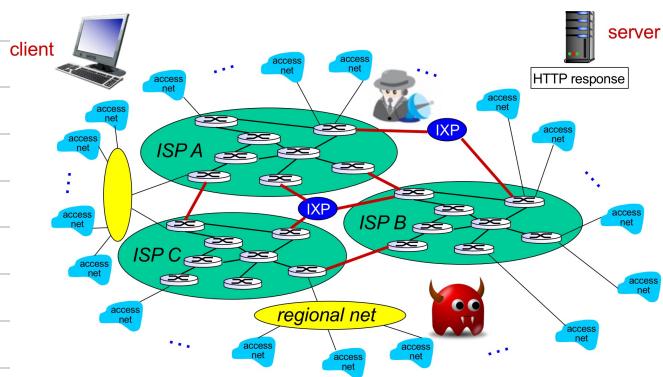
} HTTP

Cookie: arquivo enviado nos pedidos e respostas HTTP, mantido no armazenamento do utilizador e gerido pelo navegador do utilizador — estocado nas mensagens

Internet: infraestrutura que fornece serviços a aplicações — rede de redes

Dados milhões de acessos a ISPs, como conectá-los em conjunto?

Se um ISP global é negócio viável, então vai haver competidores... que têm de estar interligados... e podem trunghi redes regionais para ligar redes de acesso a ISPs.



A Internet está projetada como uma rede pública, com informação de rotas pública — a encriptação não esconde identidades, só conteúdos!

Anonimato: estado de não ser identificável num conjunto de sujeitos

1. Emissor: MENSAGEM ~~X~~ EMISSOR

2. Receptor: MENSAGEM ~~X~~ RECEPTOR

3. Relação: EMISSOR ~~X~~ RECEPTOR

Não-Ligações: as ações não têm ligação às identidades

Não-Observabilidade: os itens de interesse não distinguíveis entre si

### Ataques:

1. Análise Parcial de Tráfego: inferir do tráfego de rede os intervenientes da comunicação
2. Análise Ativa de Tráfego: injetar pacotes OU colocar uma assimetria temporal
3. Compromisso de Nós de Rede

Chains Mic: as mensagens não são enviadas através de uma sequência de misturas (que podem formar uma rede) em que basta uma boa mistura para garantir anonimato, que fazem "padding" e "buffering" do tráfego para dificultar ataques de correlação - o adversário sabe todos os emissores e receptores, mas não consegue ligar uma mensagem enviada com uma mensagem recebida

Onion Routing: o emissor escolhe uma sequência aleatória de roteadores - a informação de rotas para cada ligação é encriptada com a chave pública do roteador, pelo que cada roteador só aprende a identidade do roteador seguinte (custo computacional, latência)

Tor: "onion routing" que anonimiza a origem do tráfego - as mensagens são repetidamente encriptadas enviadas por vários nós de rede, em que cada nó remove uma camada de encriptação para descobrir informações de roteamento, mas não conhece origem, destino e conteúdo da mensagem

O cliente estabelece chaves de sessão simétricas com os "onion routers"!

IPSec

VPN: tecnologia que permite estender de forma segura sobre longas distâncias físicas ao fazer uso de uma rede pública como meio de transporte

Acesso Remoto: permite que clientes remotos acedam a uma rede privada (intranet)

Sit-to-Site: ponto seguro entre duas ou mais redes fisicamente distantes

# Privacidade Sem Fios

Ligação a Longo-Prazo: fácil identificam e relacionam dispositivos no tempo  
Ligação a Curto-Prazo: fácil isolam fluxos distintos de pacotes

Problema: muitos bits expostos não (ou podem ser usados como) identificadores

Objetivo: fazer com que todos os bits pareçam aleatórios

Desafio: filtrar atm identificadores

Quando A gera uma mensagem para B, envia

$$PM = f(A, B, M)$$

Confidencialidade: só A e B conseguem determinar M

Autenticidade: B consegue verificar que A criou PM

Integridade: B consegue verificar que M não foi modificado

Denunciabilidade: só A e B conseguem ligar PM ao emissor e receptor

Eficiência: B consegue processar PM tão rápido como receber PM

Pseudônimos MAC: mudar o endereço MAC periodicamente (por razão ou em reposo)  
i os outros campos permanecem

Encryptar Tudo: usar chaves de inicialização para encryptar tudo  
Chave Pública? lenta!  
Chave Simétrica? escala!

Sly Fox: identificar a chave de forma inviolável

1. A e B concordam em todos  $T_1^{AB}, T_2^{AB}, \dots$

2. A anexa  $T_i^{AB}$  ao pacote encryptado para B

i Terceiros não conseguem ligar  $T_i^{AB}$  e  $T_j^{AB}$  se  $i \neq j$

i A e B conseguem computar  $T_i^{AB}$  independentemente

$$T_i^{AB} = AES(K_{AB}, i)$$

	Confidentiality	Authenticity	Integrity	Unlinkability	Efficiency
	Only Data Payload				
802.11 WPA	Only Data Payload	Only Data Payload	Only Data Payload	Only Data Payload	✓
MAC Pseudonyms	Only Data Payload	Only Data Payload	Only Data Payload	Long Term	✓
Public Key Symmetric Key	✓	✓	✓	✓	Only Data Payload
SlyFi	✓	✓	✓	✓	✓

# Protocolos de Autenticação

Autenticação: determinar se um utilizador deve poder aceder a um sistema

Autorização: que ações um utilizador pode realizar no sistema

Protocolo: regras seguidas para comunicação — idealmente eficiente e robusto

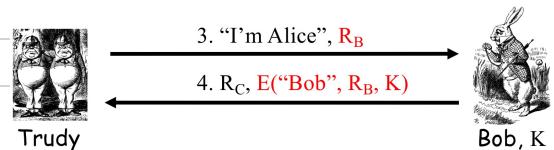
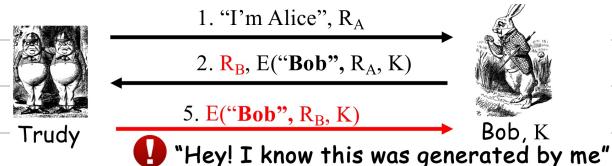
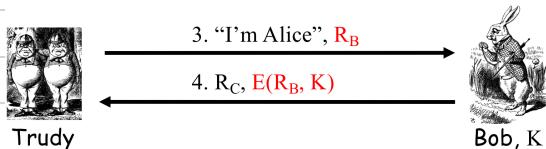
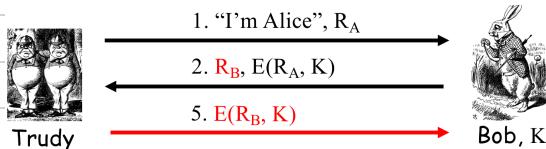
O derivação-resposta previne ataques de repetição ao garantir frescura  
NONCE - HASH

Autenticação Simples:  $E(n, k)$  chave simétrica partilhada

Autenticação Mútua: o protocolo de autenticação simples não é seguro

→ Ataque de Reflexão

• Deve usar-se um par de chaves para encriptação/assinatura e outro para autenticação!



# Autenticação Anónima

Autenticação Anónima: o servidor de autenticação sabe que o utilizador pertence a um dado conjunto de utilizadores autorizados **MAS** o servidor de autenticação não sabe que membro do conjunto se autenticou

Autenticação Segura: nenhum utilizador não autorizado deve ser capaz de enganar o servidor para lhe conceder acesso

Anonimato Completo Perfeito: o servidor consegue adivinhar o utilizador autenticado com probabilidade de, no máximo,  $1/N$

Ideia: o servidor computa  $E(pk_i, w)$  para todos os  $i$  e envia para o utilizador que se pretende autenticar - o utilizador computa  $D(sk_i, c_i)$  e envia

Problema: o servidor pode encriptar um  $w_i$  diferente para cada utilizador - o utilizador não sabe, mas o servidor fica a saber quem se autenticou

Solução: o servidor tem de provar que todos os criptogramas encriptam o mesmo  
Como?

Prova da Igualdade da Encriptação: o servidor envia o padding aleatório usado para encriptar cada criptograma - o utilizador pode reencriptar com cada chave e verificar que todos os criptogramas encriptam o mesmo valor

O anonimato é verificável!

Eficiência: o cliente e o servidor só têm de computar  $N$  encriptações

Escalabilidade: o utilizador pode escolher um subconjunto aleatório de utilizadores

Anonimato Revogável: impossível com autenticação anónima

## **1. Privacy-preserving Data Publishing**

- 1.1. One can identify 4 basic anonymization operations: generalization, suppression, anatomization and perturbation. Anatomization consists on de-associating QIDs and sensitive attributes. Explain what is the advantage and disadvantage of anatomization?

A vantagem da anatomização é que os dados não são modificados nas tabelas QID e ST.  
A desvantagem da anatomização é o número adicional de tabelas necessárias para publicação.

- 1.2. Differential privacy MÍNIMIZA the risk of an individual/register joining or leaving the database.

## **2. Secure Multiparty Computation (SMC) and Privacy**

- 2.1. In SMC two or more parties wish to jointly compute a function of their inputs while preserving certain security properties, such as privacy, correctness and independence of inputs. Considering the auction example, where users bid for a product, explain what privacy means in this context.

No contexto de leilões, privacidade significa que os bens leiloados por cada utilizador não são conhecidos por outros utilizadores. Nem que uno compreto o funcionamento dos leilões. Por exemplo, os utilizadores A e B devem poder licitar o bem C de forma anónima, nem o conhecimento um do outro, mas apenas da melhor oferta, e o bem deve ser entregue ao utilizador com a maior licitação sem revelar a sua identidade.

## **3. Authentication Protocols and Anonymous Authentication**

- 3.1. Encryption and signing/authentication should be done with C.
- A) the same key pair for both
  - B) a different set of symmetric-keys for each
  - C) two different public-key cryptography key pairs for each