Informační soukromí – služby, pojmy

Vašek Matyáš

PV080

Soukromí (angl. *Privacy*)

- *Je v obecném pojetí charakteristikou života jedince a jeho práva a možnosti kontroly informací o sobě a o své činnosti, spolu s ochranou proti nežádoucímu rušení.*

- Informační soukromí se vztahuje především na zmíněnou možnost kontroly informací osobních dat a jiných relevantních citlivých informací. Tento termín se váže na jiná práva jedince, a tak je přesná definice obtížná.

Informační soukromí

- Termín spíše pro neformální motivaci k zajištění ochrany osobních informací, pravidel pro jejich kontrolu a poskytování jiným subjektům atd.

- Příklady relevantních bezpečnostních funkcí:
  - anonymita,
  - pseudonymita,
  - nespojitelnost,
  - nepozorovatelnost.

Anonymita

Anonymita je vlastnost systému, který zajišťuje možnost použití zdrojů nebo služeb bez zjištění identity uživatele tohoto systému.

Pseudonymita

Vlastnost systému, který zajišťuje možnost použití zdrojů nebo služeb bez zjištění identity uživatele tohoto systému *tak, že uživatel je stále zodpovědný za toto použití*.

Určitá podobnost existuje s poštovními přihrádkami (PO Box).

Nespojitelnost (angl. *unlinkability*)

Vlastnost systému, který zajišťuje možnost *opakovaného* použití zdrojů nebo služeb s tím, že ostatní si tato použití nebudou schopni spojit.

- Spojení ve smyslu vzájemné souvislosti.
- Může se jednat o postupně i současně poskytované stejné i různé služby.
- Nezohledňuje identitu uživatele, ale rozsah služeb a zdrojů, které byly použity stejným uživatelem.

Nepozorovatelnost (angl. *unobservability*)

Vlastnost systému, který zajišťuje možnost použití zdrojů nebo služeb tak, že ostatní nemohou zpozorovat používání daného zdroje nebo služeb.

- Ochraňovanými hodnotami nejsou informace o uživatelích, ale o použití zdrojů nebo služeb.
- Příkladem aplikace může být ochrana proti tzv. analýze provozu (angl. *traffic analysis*).

K zamyšlení...

- Je e-mailová adresa ve tvaru `Jmeno.Prijmeni@NejakaFirma.cz` osobním údajem nebo nikoliv?

- Může zaměstnavatel sledovat e-mailovou komunikaci svého zaměstnance, který při nástupu na nové místo stvrdil písemně svůj souhlas s tím, že nebude používat e-mail pro soukromé účely?

Lze měřit/hodnotit informační soukromí?

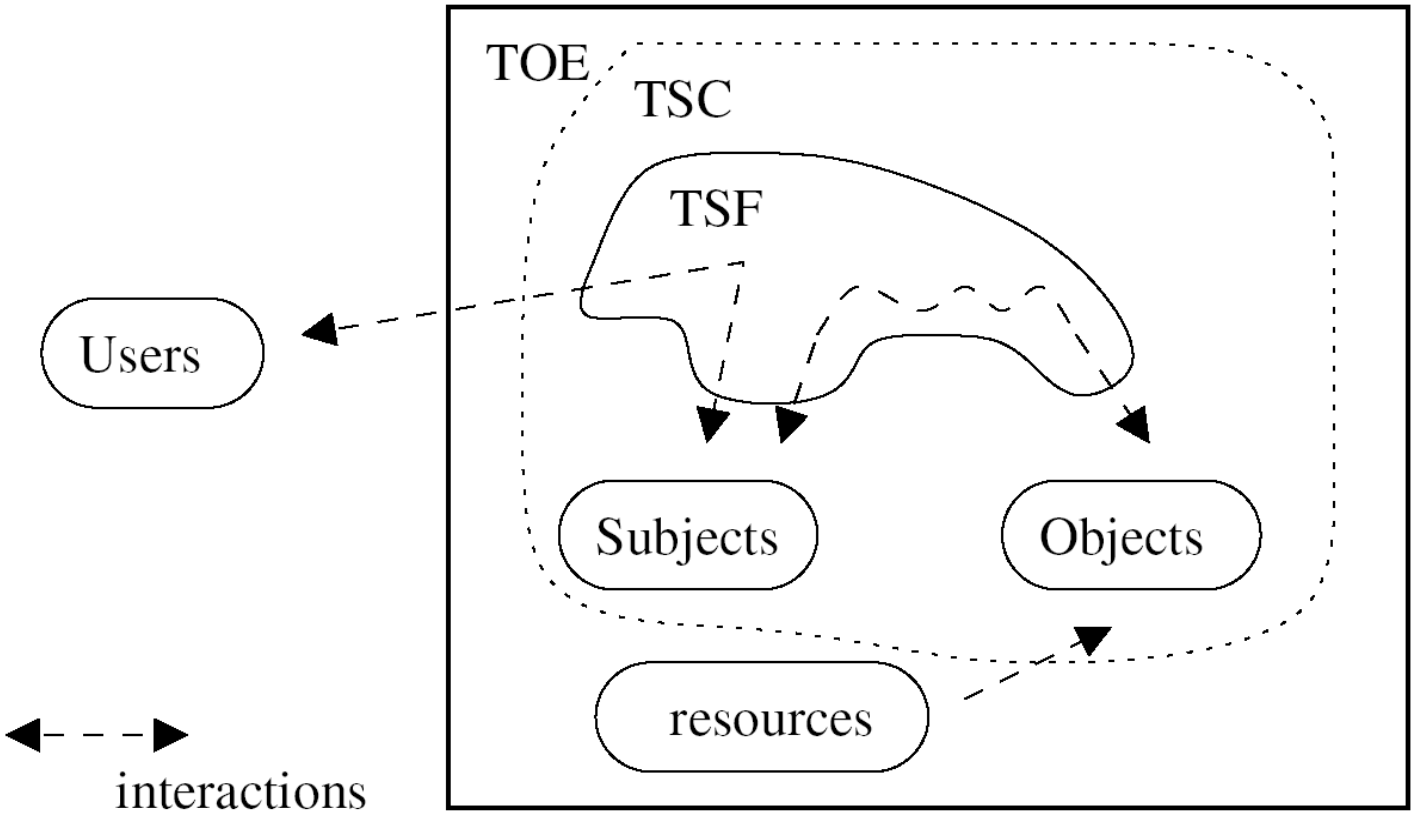
- Na jaké úrovni jsou data spojitelná s určitou osobou?
- Jakou míru jistoty máme při spojení různých datových položek?
- Na jaké úrovni je něco pozorovatelné?

Dva hlavní směry/pohledy

- *Mixy* – systémy pro posílání zpráv, obvykle s klamavými (kamuflovacími) zprávami, přeposílání mezi více účastníky (než je nejkratší/nejoptimálnější cesta)
  - Příklad systému později v semestru
- *Společná kritéria (systémy)* – standard (rozsáhlý) pro hodnocení bezpečnosti systémů, umožňuje lepší srovnávání systémů i specifikaci požadované funkčnosti
  - Více informací také později v semestru

Model Společných kritérií

TOE: Target of Evaluation – celý (hodnocený) systém
TSF: TOE Security Functions – HW, SW, FW který TOE využívá
TSC: TSF Scope of Control – interakce podléhající bezp. politice TOE security policy



Nepozorovatelnost (CC)

- Uživatel může použít zdroj nebo službu bez toho, aby ostatní byli schopni zjistit, že je daný zdroj nebo služba používán
- Úrovně:
  - specifikované entity nejsou schopny pozorovat specifikované operace prováděné specifikovanými entitami na specifikovaných objektech
    - a s podmínkami pro práci s relevantními informacemi
  - nebo specifikované subjekty poskytují specifikované služby bez vyžadování informací (TSF)
  - nebo specifikovaní uživatelé mohou pozorovat použití specifikovaných zdrojů nebo služeb

Anonymita (CC)

- Uživatel může využít zdroj nebo službu bez odhalení své identity
  - Jedná se o ochranu identity uživatelů, nikoliv ochranu identity subjektů v systému
- Úrovně:
  - specifikované entity nejsou schopny určit skutečné uživatelské jméno spojené se specifikovanými subjekty, operacemi, objekty
    - a specifikované subjekty získají specifikované služby bez vyžadování informací (TSF)

Pseudonymita (CC)

- Uživatel může použít zdroj nebo službu bez odhalení své uživatelské identity, ale je stále zodpovědný za toto použití
- Úrovně:
  - [Anonymita 1.1] s přidělením aliasů pod kontrolou TSF a specifikovanou metrikou aliasů
    - a s právy zvrácení pro specifikované entity za specifikovaných podmínek
    - nebo se znovuvyužitím aliasu za specifikovaných podmínek

Nespojitelnost (CC)

- Uživatel může opakovaně využít zdroje nebo služby bez toho, aby ostatní byli schopni vzájemně spojit tato užití
- Další členění/úrovně nejsou

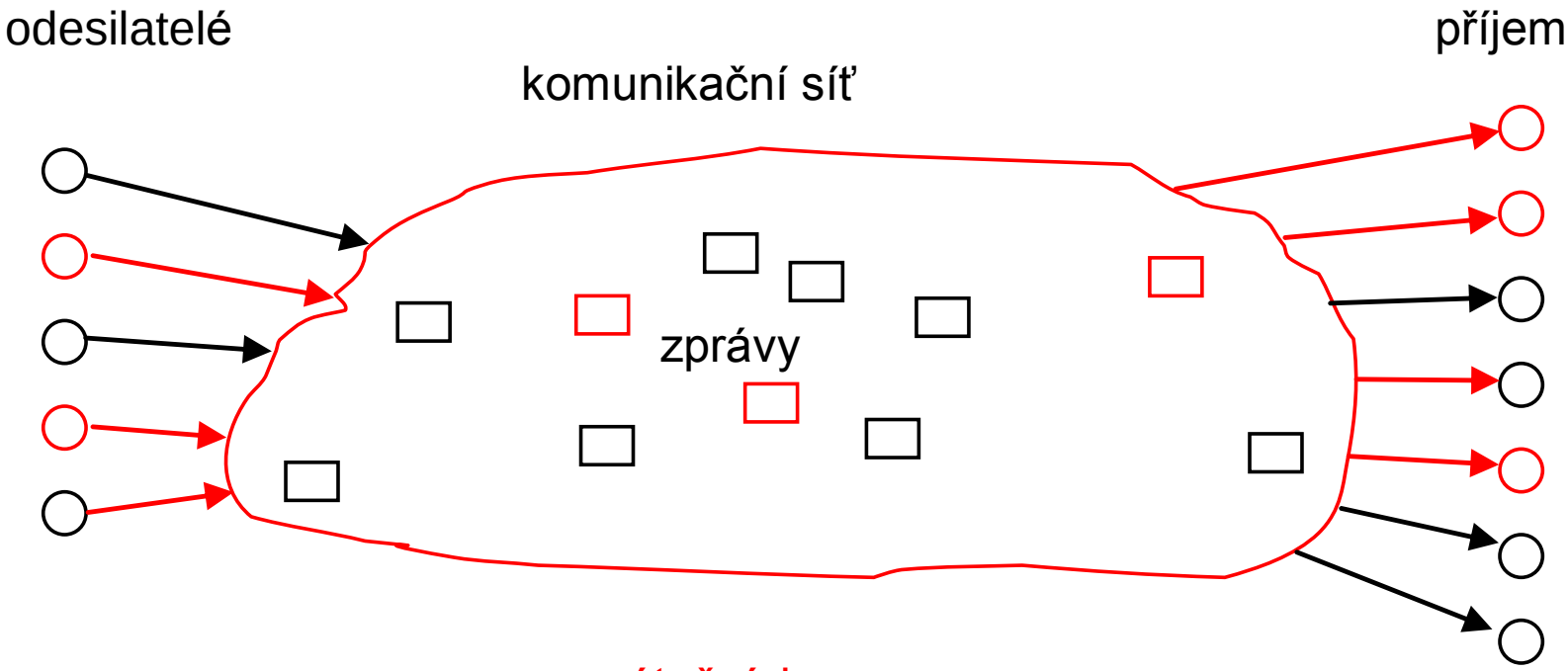
Pohled Společných kritérií

- Existenciální pohled – vlastnost buď je, nebo není
  - Kritéria neřeší (a ani to nemají za cíl) jak je vlastnosti dosaženo
  - Kritéria neumožňují jiné než diskrétní (Y/N) ohodnocení
    - Granularita jen podle stanovených úrovní

A. Pfitzmann a kol. - terminologie

- Anonymity, Unobservability, Pseudonymity, and Identity Management - A Proposal for Terminology
- Soustředí se pouze na prostředí, kde se posílají zprávy od odesilatelů k příjemcům
  - Specifickou (a nejvýznamnější) podmnožinou jsou tzv. mixy (sítě mixů zadefinoval David Chaum v roce 1981)

Takže obvyklé prostředí...



(Zdroj: A. Pfitzmann)

Anonymita subjektu (A.P.)

- Stav bytí neidentifikovatelným v rámci dané množiny subjektů, tzv. anonymitní množině.
- Anonymitní množina je množinou všech možných subjektů (obvyklí podezřelí ☺)
  - s ohledem na odesilatele možných odesilatelů
  - s ohledem na příjemce možných příjemců atd.
- Anonymita subjektu je tedy vždy spojena s touto množinou!
  - Lze vnímat tak, že anonymita je silnější pro větší anonymitní množinu
    - Otázkou je někdy přínos tohoto pohledu – získáte více, když při stejné pravděpodobnosti víte, že pravděpodobnost spojení s nějakou identitou je pro daný subjekt různá pro různě velké množiny?

Nespojitelnost (A.P.)

- Nespojitelnost dvou nebo více prvků (např. subjektů, zpráv, událostí...) znamená, že v takovém systému nejsou prvky ani více, ani méně ve vzájemném vztahu s ohledem na předchozí znalost o systému
  - tzn. že pravděpodobnost spojení těchto prvků je stejná před a po (prů)běhu nějaké posloupnosti událostí v systému

Předmět zájmu

- Terminologie Pfitzmanna a kol. definuje *předmět zájmu* (item of interest) jako označení pro případ, že cílem zájmu není subjekt (jako např. u anonymity)
  - Pak lze definici anonymity subjektu rozšířit...

Nepozorovatelnost (A.P.)

- Stav (daných) předmětů zájmu, kdy nejsou odlišitelné od jiných předmětů zájmu.
  - U zpráv v mixech např. neodlišitelnost „skutečných“ zpráv od šumu
  - S ohledem na stejného útočníka pak lze říct: Nepozorovatelnost => anonymita
- Pro nepozorovatelnost a anonymitu u systémů pro posílání zpráv se používají mixy, příklad systému později v semestru na zvláštní přednášce

Pseudonym (A.P.)

- Z řeckého *pseudonumon* – falešně pojmenovaný
  - tzn. používající jiné než „skutečné jméno“
- Pozor – „skutečné jméno“ (např. dané oficiálními státními dokumenty) se během života mění
  - Mimo „obvyklých“ změn i otázky písma/abecedy
  - Jako pseudonym lze pak označit každé pojmenování (identifikátor)

Pseudonymita (A.P.)

- Bytí pseudonymním je stav používání pseudonymu jako identifikátoru (ID).

- Digitální pseudonym – řetězec bitů, který je
  - unikátní jako ID (s velmi velkou pravděpodobností) a
  - použitelný pro autentizaci jeho vlastníka a předmětů zájmu (např. odeslaných zpráv)

Poznámky k pseudonymitě

- Anonymita a prokazatelná zodpovědnost (accountability) jsou dva extrémy
- V praxi obvykle vhodná pseudonymita
  - Ovlivňuje spojitelnost mezi předměty zájmu a uživateli
- Opakované použití pseudonymu může uživateli umožnit ustavení reputace (důvěryhodnosti)
- Uživatelé používají větší počet pseudonymů
  - Odhalují spojitost mezi nimi jen v případě potřeby (zisku výhod, času, peněz...)

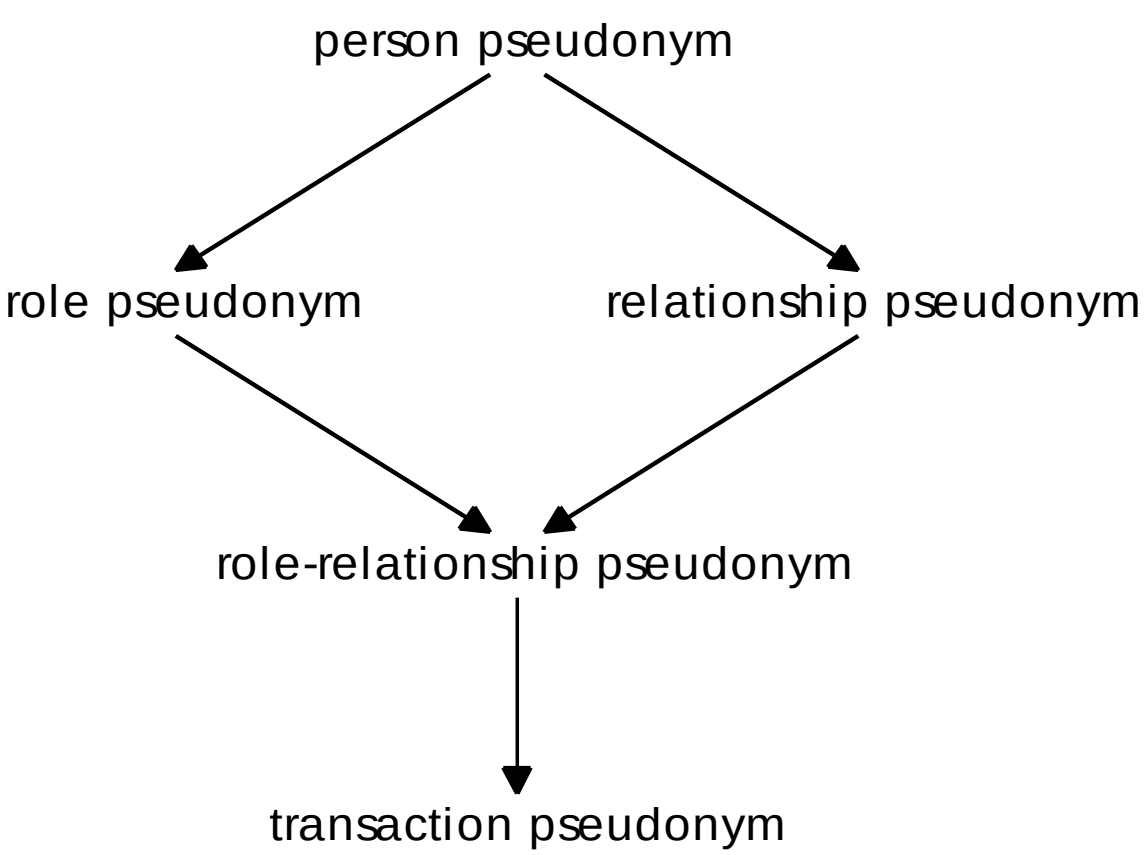
Vztah mezi pseudonymem a vlastníkem (A.P.)

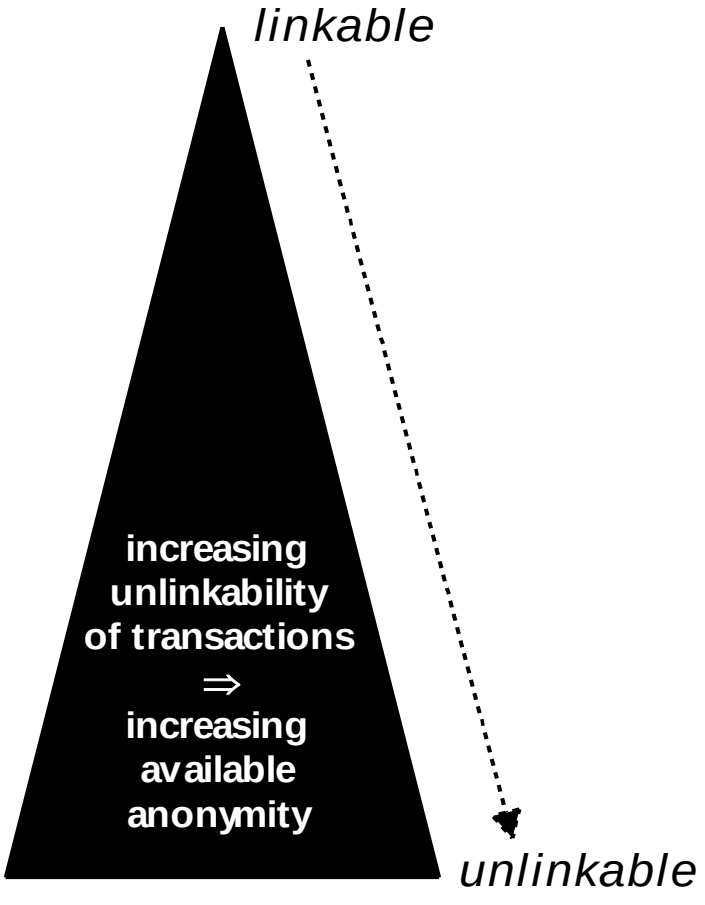
- *Veřejný pseudonym* – veřejně znám od počátku, např. v seznamu osob
- *Původně neveřejný pseudonym* – není od počátku znám veřejně, např. číslo účtu, pseudonymní certifikát veřejného klíče (podpisový certifikát)
- *Původně nespojený pseudonym* – není od počátku znám nikomu mimo jeho vlastníka, např. ID v chatu

Spojitelnost s ohledem na použití pseudonymu v různých kontextech (A.P.)

- Pseudonym *osoby* – vnímán jako reprezentace dané osoby
- Pseudonym *role* – osoba používá různé pro různé role (může někdy i stejné)
- Pseudonym *vztahu* – pro každého partnera je použito jiné jméno
  - může být stejné pro komunikaci se stejným partnerem v různých rolích
- Pseudonym *role-vztahu* – unikátní pro roli a vztah (partnera)
- Pseudonym *transakce* – unikátní pro transakci

Úroveň anonymity/nespojitelnosti transakcí podle druhu pseud. (A.P.)





Identita (A.P.)

- Libovolná podmnožina atributů určitého jedince, která tohoto jedince jednoznačně určuje v jakékoliv množině jedinců.
  - Tzn. není jedna identita, ale několik.
  - Částečná identita se pak vztahuje k určitému kontextu či roli, tzn. i k omezené množině jedinců.
    - Pak může být i pseudonym za určitých okolností identifikátorem pro částečnou identitu.

Systémy řízení identity

- Angl. *Identity Management System – IMS*
- Využívají technologie pro návrh a správu atributů (popisů) identity
- V jednodušší podobě známy dříve jako, resp. stavějí často na využití
  - Single sign-on (systémy jednoduchého přihlašování)
  - Public-key infrastructures (infrastruktury veřejných klíčů – nejčastěji pro spolehlivé spojení klíče a informací o osobě)

Rozsáhlé databáze osobních informací

Vašek Matyáš

PV080

Agregace dat

- Seskupování (osobních) dat do rozsáhlých databází. Agregace (z angl. *aggregation*).
- Tímto kombinováním dat o určité citlivosti lze získat informace daleko citlivější, které jinak spadají do kategorie s vyššími požadavky na ochranu.

Zákon o ochraně osobních údajů (101/2000 Sb.) – Povinnosti správce

Mj. zákon říká:

- nesdružovat osobní údaje, které byly získány k rozdílným účelům, pokud zvláštní zákon nestanoví jinak

Žadatel o investici

- Chodil roky ke stejnému obvodnímu lékaři.
- Uzavřel před měsícem vysokou živ. pojistku.
- V minulém čtvrtletí byl u specialisty.
- Přede dvěma měsíci změnil obvodního lékaře.

Odvození (Inference, i angl.)

- Odvození informací o vyšší citlivosti zpracováním a analýzou skupiny informací o nižší citlivosti.

nebo

- Nepřímý přístup k informacím bez přímého přístupu k datům, která tyto informace reprezentují.

Příklad politiky klinických IS, British Medical Association

- Musí být zavedena účinná opatření proti agregaci osobních zdravotních informací.
- Pacienti, k jejichž seznamu řízení přístupu má být přidána další osoba, musí být zvlášť upozorněni, pokud již tato osoba má přístup ke zdravotním informacím velkého množství lidí.

Co když máte informace o finanční situaci a zdrav. stavu

1. Přítele/kyně, resp. manžela/ky.
2. Spolupracovníka, nadřízeného...
3. Všech studentů/zaměstnanců FI.
4. Všech obyvatel místa, kde žijete.
5. Všech klientů určité firmy (banky, zdravotní pojišťovny...).
6. Všech (většiny) občanů.

Pravděpodobnost neoprávněného použití

- Počet osob, které mají k informacím přístup (operátoři, uživatelé systému ap.).
- Hodnota informací.
  - Výše trestu těm, kdo data jiných řádně neohlídali a spolupodíleli se tak na jejich úniku.
  - Výše trestu těm, kdo s nimi neoprávněně manipulují.
  - Úroveň ochranných mechanismů.

Řešení?

- U menších souborů osobních dat provádět agregaci jen v nutných případech.
- U větších souborů neprovádět agregaci.

- Statistické databáze!

Statistické databáze

- Obsahují citlivé údaje o jednotlivcích.
- Jejich využití má být **jen** pro statistické dotazy k vytvoření obrazu o celkových potřebách obyvatelstva a formulování (vládní) politiky.
  - podpora církví, regionů/měst atd.
- Výsledky dotazů v takovýchto databázích nesmějí poskytnout údaje o jednotlivcích.

Studium statistických databází

- USA, 70. léta, databáze ze sčítání lidu.
- Dorothy Denning
  - Studium používaných způsobů pro formulaci dotazů a získávání odpovědí.
  - Ty povolovaly (netriviální!) dotazy, které umožnily získat údajně tajné informace o jednotlivci.
  - ☺ Údajně nedůvěra ve zjistění Denningové – dokud nezjistila plat svého šéfa sérií legitimních dotazů.

Příklad kritického dotazu

Kolik je měst s 15-16 000 obyvatel

& s muži, evangelíky, slovenské nár., 36-40 let

& jejich ženy, 28-30 let žijí mimo toto město

& 2 děti do 10 let žijí s těmito ženami

& 1 dítě nad 18 žije s těmito muži

& muž žije ve vlastním domě, plocha nad 200m^2
a domácnost má/používá aspoň 2 automobily.

Kompromitace databáze

- Výsledkem série dotazů je jeden záznam
  - Databáze byla pozitivně kompromitována
- Následný pokus o získání dalších informací
  - Výsledkem je buď 1 nebo 0 záznamů
  - Pozitivní/částečná kompromitace databáze
- Částečná kompromitace
  - Informace o entitě i když neznáme konkrétní hodnotu

Protiopatření ve statistických databázích I.

- Omezení dotazu
  - Např. i sledování předchozích dotazů
- Úmyslná změna zdrojových dat
  - Např. orig. hodnoty nahrazeny novým vzorkem se stejným rozložením pravděpodobnosti hodnot
- Úmyslná změna výsledku dotazu
  - Např. zaokrouhlování
- Cílem je zabránit situacím, kdy je možné získat informace o jedné entitě.

Protiopatření ve statistických databázích II. – *Náhodný výběr*

Každý dotaz je zodpovězen na základě vyhodnocení náhodně vybraných záznamů ze všech existujících záznamů.

- Kontrola překrytí množiny záznamů u vícenásobných dotazů na tutéž informaci.
  - Má zabránit situaci, kdy několik uživatelů databáze začne spolupracovat.
- Technika nyní používaná v americké databázi údajů ze sčítání lidu.

Protiopatření ve statistických databázích III. – *Minimální rozsah dotazu*

- Minimum celkového počtu záznamů použitých pro tvorbu odpovědí.

nebo

- Minimum počtu záznamů použitých pro tvorbu odpovědí na každou část dotazu.

Protiopatření ve statistických databázích III. – *Perturbační (zmatečné) techniky*

Přidání pseudonáhodného „šumu“:

- Odpovědi konzistentní, ale získání spolehlivé odpovědi na sérii podobných dotazů není možné.

1. K záznamům zahrnutým pro vyhodnocení dotazů se přidají další náhodně vybrané podobné záznamy
2. Vypočtená hodnota nebo mezihodnoty jsou zaokrouhlovány nebo mírně pozměněny.

- Podle některých definic zahrnují *náhodný výběr*.

De-anonymizace uživatelů

- Narayanan a Shmatikov (2008)
  - Huge de-anonymization of large sparse datasets (ACM)
- Databáze hodnocení filmů
  - Databáze zpřístupněna „anonymizovaně“
- Uživatel hodnotí filmy (filmů jsou stovky) na škále 1-10
- Uživatele se podařilo de-anonymizovat – spojit se skutečnou identitou pokud:
  - Víme jeho hodnocení pro 5-8 filmů