

PV080 – Ochrana dat a
informačního soukromí

Vašek Matyáš

Konzultační hodiny: Po 15:00-55 &
Út 9:00-55

B415

Průběh kurzu

- Přednášky v D1 Út 10:00-11:xx
- Doplnkové čtení
 - Materiály (vč. slajdů) na IS
 - Vytisknuté komplety také v knihkupectví Pavla Marečka
 - Na IS a e-mailem diskuze, podněty, upozornění
- Aktivita v diskuzích, dobré odpovědi na otázky atd. průběžně hodnoceny (bonus až 10 %!!!)
- Polosemestrální písemná zkouška 35 %
- Závěrečná písemná zkouška 65 %

Hodnocení

A: 90 % (bodů) a více,

B: 80 % a více, ale méně než 90 %,

C: 70 % a více, ale méně než 80 %

D: 60 % a více, ale méně než 70 %

E: 50 % a více, ale méně než 60 %

F = neprospěl(a), za méně než 50 %.

- Kolokvium nebo zápočet alespoň 50 %.

Témata kurzu – I.

- Informační soukromí – úvod, pojmy atd.
- Ochrana osobních dat a legislativa.
- Etika, profesionalita a práce s informacemi.
- Úvod do informační bezpečnosti.
 - Ochrana dat ve vybraném oboru lidské činnosti.
- Od analýzy rizik k bezpečnostní politice.
- Úvod do kryptografie, digitální podpis.

Témata kurzu – II.

- Standardy bezpečnosti IT, kritéria hodnocení a standardizační procesy.
- Audit, řízení bezpečnosti, kontrola ochranných opatření. Ochrana dat a management.
- Internet a bezpečnost, ochrana soukromí.
 - Systémy podporující ochranu soukromí.
 - Anonymní komunikace.

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.

Soukromé informace

Soukromé informace jsou informace, které nechceme sdílet s jinými, nebo u kterých chceme osobně kontrolovat jejich pohyb (tzn. sdílíme je s někým, ale ne s „ostatními“).

[KC Laudon, Communications of ACM 9/96].

Úroveň ochrany osobních dat

- *Rozhodujícím ukazatelem úrovně ochrany je cena osobních dat „na ulici“ – na černém či šedém trhu. (Roger Needham, Cambridge U.)*
 - Zdravotní data „běžné“ osoby v UK lze získat za cca 150-200 liber
 - V kanadské provincii Quebec podle některých „inzerátů“ 20-60 liber.
 - Podle Needhama by měla cena být výrazně nad 500 liber.

Příklad aktuálních cen v UK (2006)

Požadovaná informace	Platba komplicům	Cena-klient
Adresa	neznámá	£17.50
Adresa podle tel. čísla	£40	£75
Adresa podle tel. čísla (mobil)	neznámá	£75
Seznam členů rodiny a přátel	£60 – £80	neznámá
Údaje o vozidle z registru	£70	£150-200
Trestní rejstřík	neznámá	£500
Tel. číslo – blokováné	£40	£65 – £75
Výpis z účtu mobilního telef.	neznámá	£750
Údaje z řidičského průkazu	neznámá	£250

Cenu osobních dat ovlivňují

1. Výše trestu těm, kdo data jiných řádně neohlídali a spolupodíleli se tak na jejich úniku.
2. Výše trestu těm, kdo s nimi neoprávněně manipulují.
3. Úroveň ochranných mechanismů.

Postoj občanů k zacházení s osobními daty (Anglie, 90. léta)

- Necelých 20 % občanů totálně lhostejných,
- Stejný počet velmi obezřetných až paranoidních
- Asi 60 % je ochotno část svých práv nechat omezit za “přiměřenou úhradu” - finanční, věcnou či nejčastěji v podobě výrazného zlepšení služeb.

Průzkum v Německu – I.

- *Privacy in e-commerce: stated preferences vs. actual behavior (Berendt a kol.), ACM Communications, April 2005*
- Soukromí si chránící – 30 %
- (Téměř) lhostejní – 24 %
- Citliví na profilování – 26 %
- Citliví na identitu – 20 %

Průzkum v Německu – II.

- Za určitých okolností je ovšem většina uživatelů online ochotna zapomenout na zábrany a sdělit osobní informace i bez skutečně závažných důvodů (takto učinit)
- I uživatelé, kteří podle vlastního názoru jsou citliví na ochranu osobních dat, tak při online interakci nekontrolují v tomto směru své chování

Experiment v Cambridge

- *How Much is Location Privacy Worth?*
(Danezis a kol.)
- Info studentům 1. ročníku o placeném výzkumu se sběrem informací o jejich pohybu (mobil – 28 dnů, 24 hodin denně)
 - Aukce!!!
- £10 medián, £27.4 průměr (max. £400, min. 0)
- Se zvažováním prodeje pro komerční účely pak £20 medián, £32.8 průměr (max. £300, min. 0)

Obdobný experiment ve větším měřítku...

- Následující slajdy jsou výjimečně v angličtině 😊
 - Prezentace připravená v souvislosti s rozbořením výsledků studie...

Starting Points

- Privacy – ensured by legal system or by technology
- Technologies to preserve privacy are really expensive
- Yet privacy intrusive technologies become more common
 - GSM system used for tracking down particular handsets (more precise than needed for the GSM system itself)
- What is the value of privacy?
 - How much are people willing to pay to protect their privacy (location privacy in this case)
 - What are black market prices and penalties
 - UK: £17.50 for address; up to £500 for criminal records check; £750 for mobile phone account details. (UK IPC, May 2006)
 - UK – penalties for privacy breaches in low £'000 per individual
 - US health data (HIPAA) – civil penalty \$100 per violation
- Design a study about how much we want to get for being tracked 24/7

New Study

- Organised within FIDIS project (www.fidis.net)
 - Spring 2006
 - Pseudonymity, with only email address provided
- Five countries involved
 - Belgium, Czech Republic, Germany, Greece, Slovakia
- Information advertised to
 - University students (IT) – all countries
 - University students (regardless on study) – CZ, DE, SK
 - Mobile phone community – CZ, DE

Organisation

- First form (webpage)
 - Language
 - Background (computers, law, other)
 - Gender
 - Network operator used (list of local operators)
 - Do you carry your mobile all the time?
 - How often are irregular movements (hourly, daily, weekly, monthly)?
 - Who do you talk to (friends, family, partner, business)?
- Second form
 - Commercial exploitation (decline, same bid, revised bid)
- Third form
 - Commercial use for one year (decline, write the bid)

Demographics

- Number of participants per country
 - Belgium 37/3 (no of participants/females)
 - Czech Republic 744/131
 - Germany 251/33
 - Greece 30/6
 - Slovak Republic 152/46
- Students in all countries, mobile phone communities in Czech Republic and Germany
- Size of sample sets
 - Czech Republic, Germany, Slovak Republic – deep analyses
 - Belgium, Greece – too small, control sets

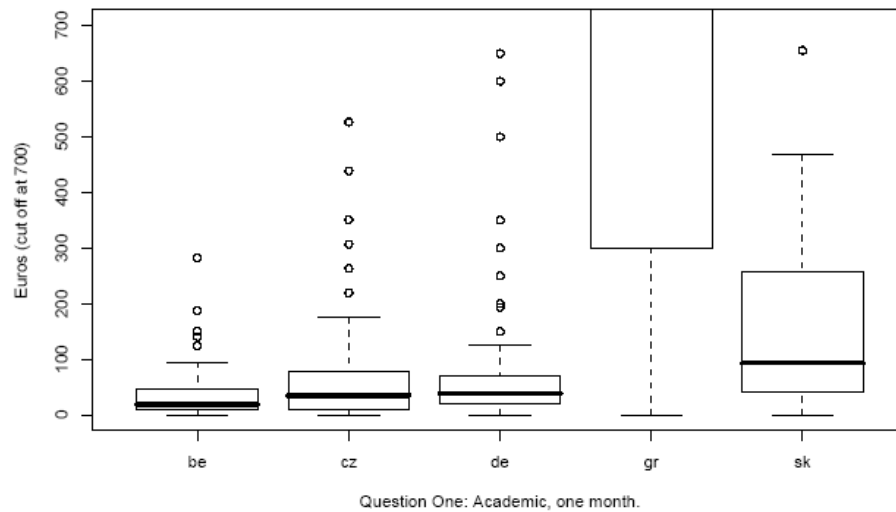
Cautiousness

- Drop-out rates
 - Early drop-outs (239 out of 2582)
 - BE 12 % CZ 6 % DE 12 % GR 25 % SK 12 %
 - Standard drop-outs
 - BE 56 % CZ 44 % DE 48 % GR 68 % SK 58 %
- Not interested
 - Greeks stand out, unfortunately the sample set too small
 - There is a remarkable number of really high bids
 - creating “linearity” from “not interested” to average bid

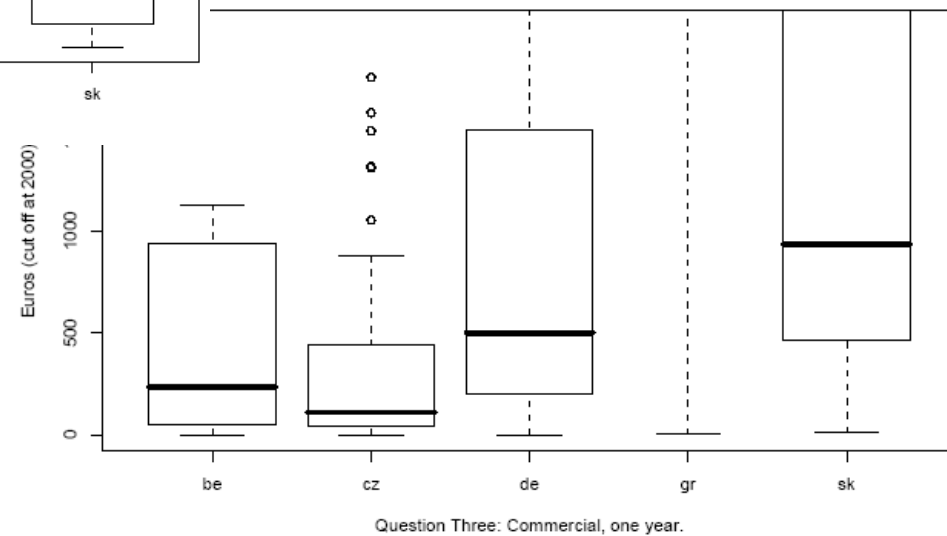
Differences among Countries

- 1st bids

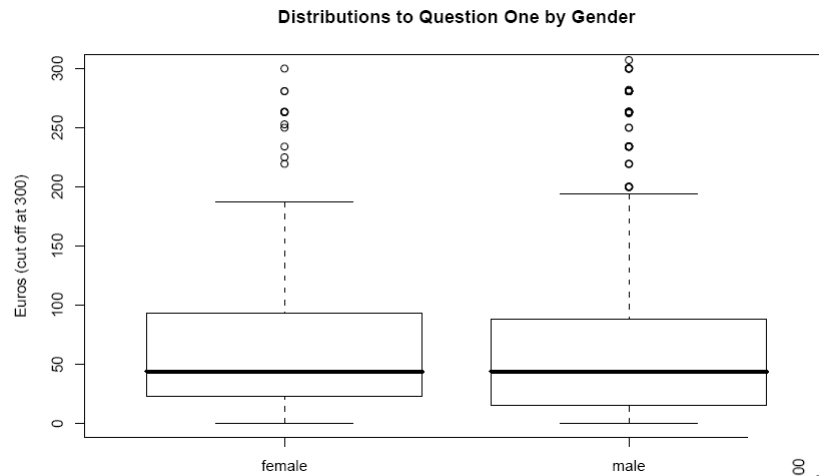
Distributions to Question One by Language



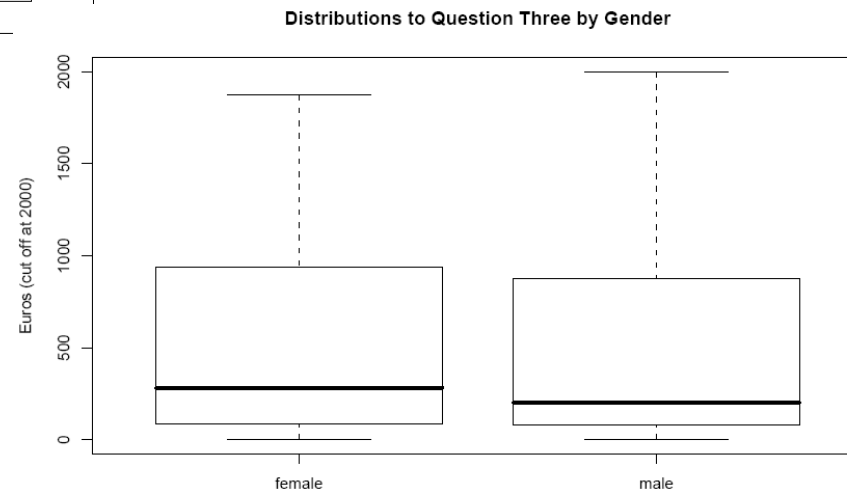
Distributions to Question Three by Language



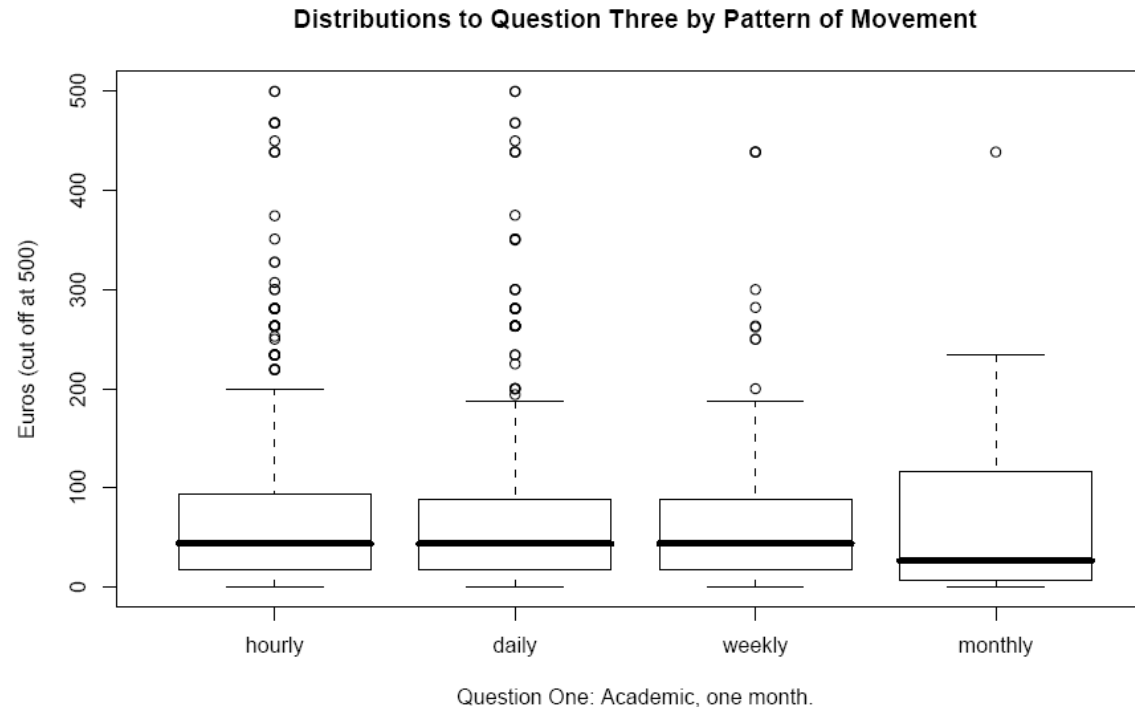
Men and Women



- Medians of the 2nd bids
 - 1.4 : 1
- Medians of the 3rd bids
 - 1.8:1

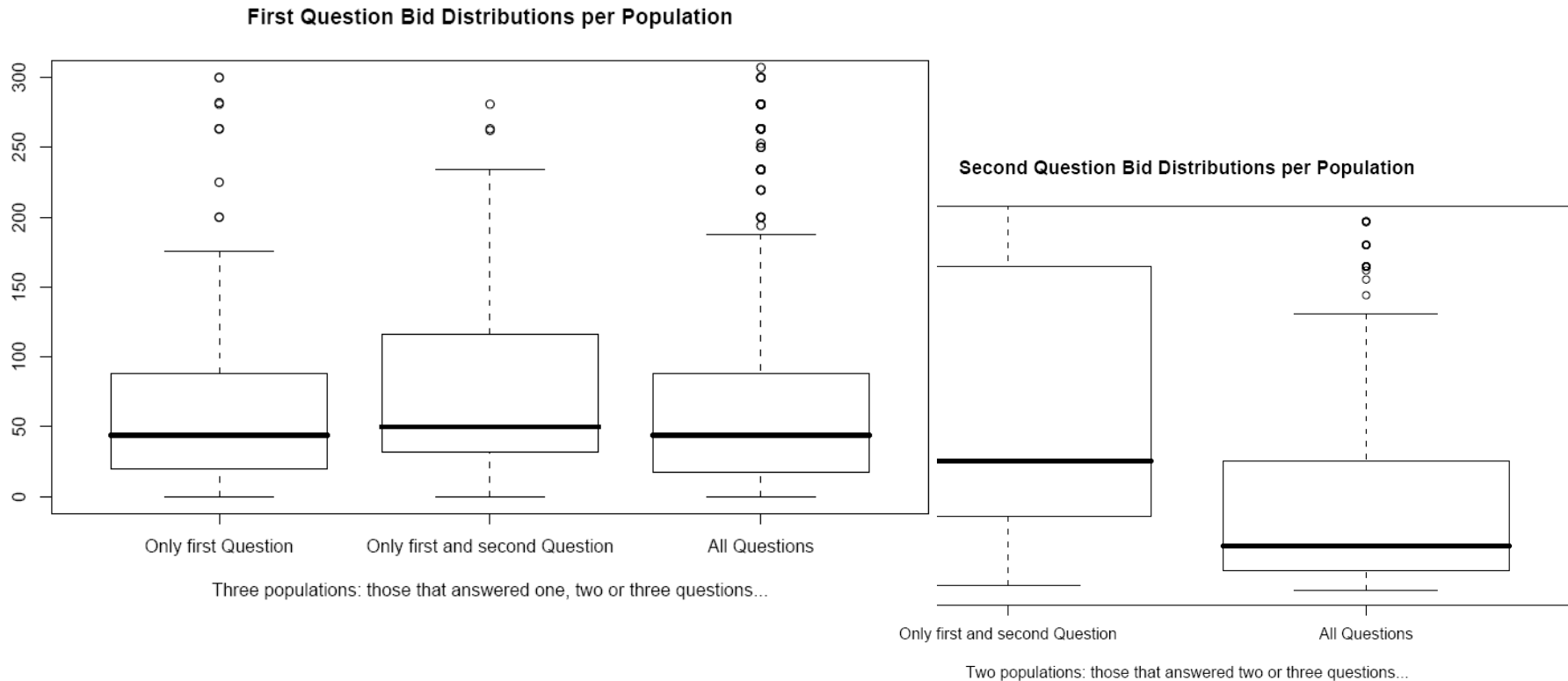


Mobility



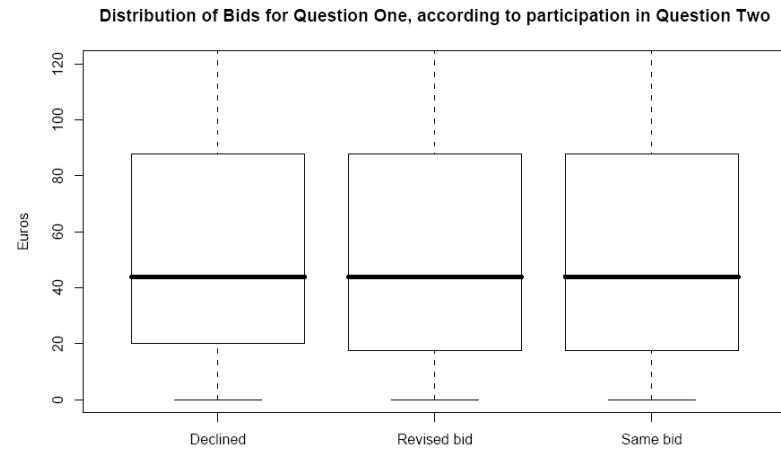
- Sizes of sample sets: daily 520, hourly 485, weekly 195, monthly 15
- Expectation was that there is correlation between value of irregular movements

Impact of Scenarios



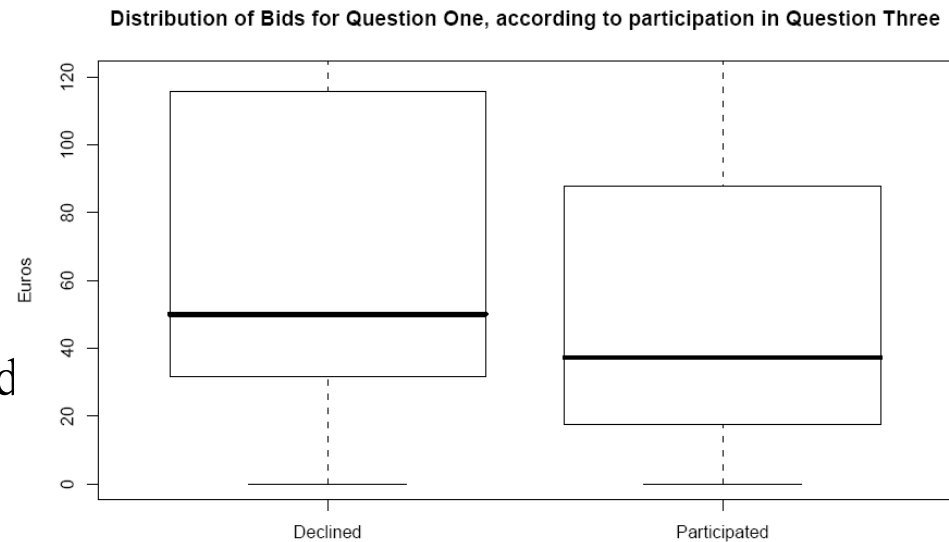
- Curiosity vs privacy cautiousness
 - Left – low bids: curiosity and falling off in the second round
 - Middle – higher bids, increased in the second round
 - Right – low first bid increased in each consecutive round

Impact of Scenarios II

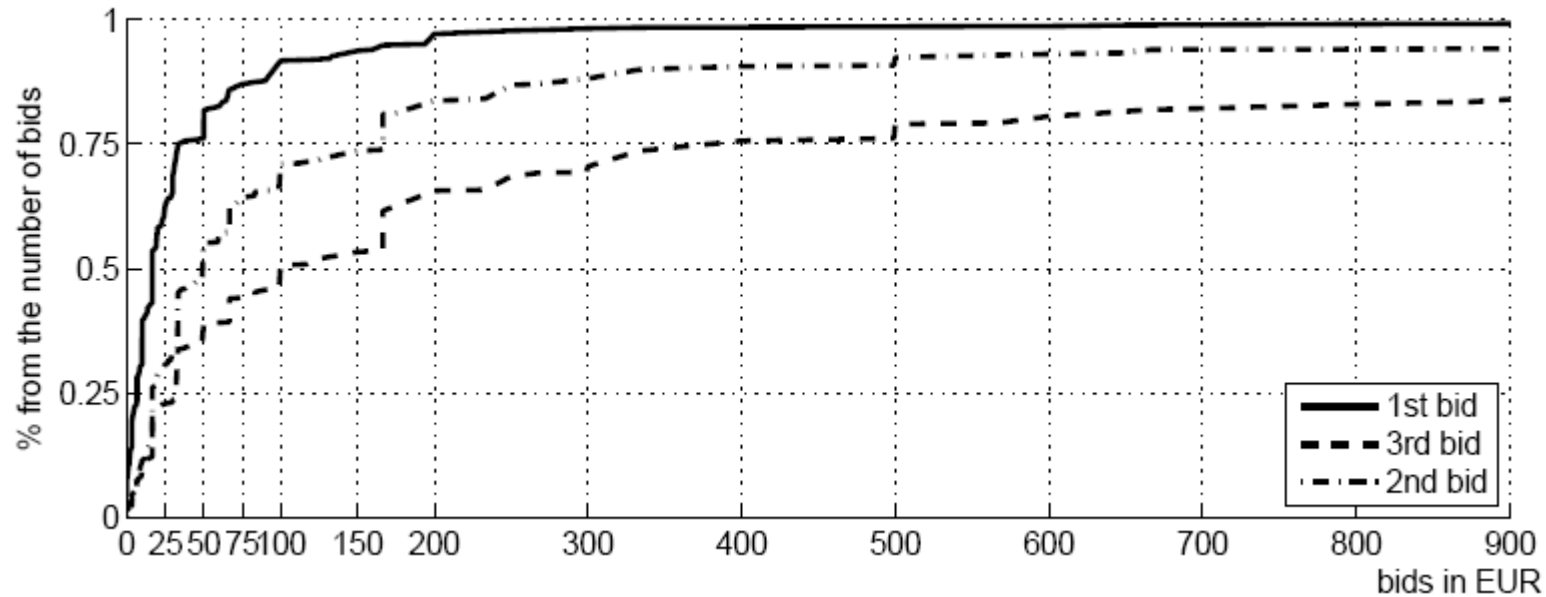


Value of bids according to answers in the second round (decline, same, revise)

Bids according to answers In the third round (declined participated)



Overall Distribution of Bids

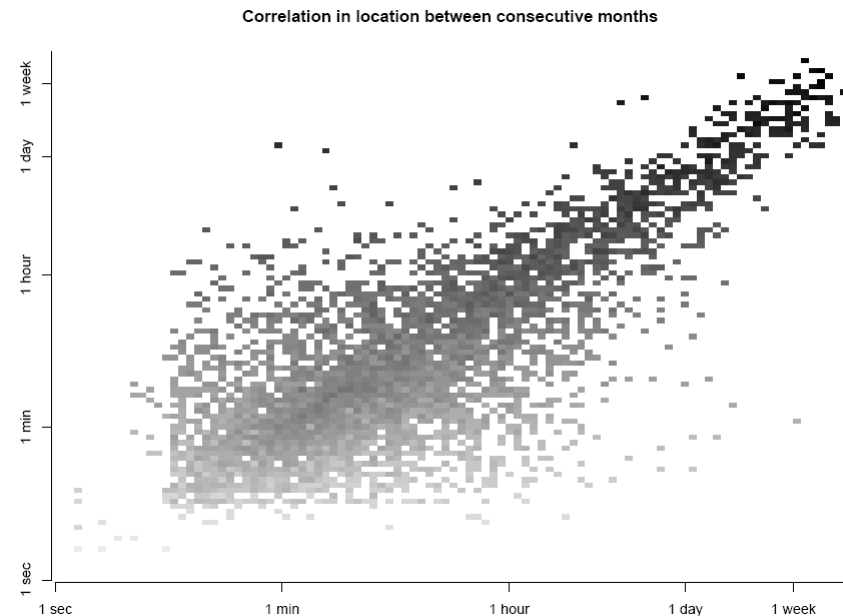


- Second bids (2.5x first bids)
- Third bids (2x second bids)

Non-linearity in Time

- 12-fold increase in the experiment length
 - 2x increase of the bids
- Hypothesis
 - Data after the 1st month are of less value
 - Little information in consecutive data

Correlation between consecutive months (MIT Reality Mining project)



Why Participating in The Study

- Questioned after the experiment
 - 300 responses (25 % of the participants)
- Why did you take part in the experiment
 - Money (38 %), results (32 %), fun (30 %)
- Correlated bid values (medians)
 - 1st auction: 12, 8, 9 (roughly)
 - 2nd auction: 9, 5, 6 (roughly)
 - No substantial difference between bids

Conclusions

- 10 % of participants bidding < 1 EUR
 - Curiosity and enthusiasm for cover story
- Greek sensitivity to privacy breaches
 - Eavesdropping scandal a couple of months before
- Non-linearity in regard of the study length
- No correlation between bids and movements
- Medians of Cambridge study correspond to our results (43 EUR to 28 GBP)

Introduction – second study

- Usage of online communication tools
- Email or instant messaging used every day
- Network administrators can track their users
- Risk of profiling or another analyzes of data
- People can sense the value of such information

Organisation of the study

- How much money for being tracked for two weeks
 - email
 - instant messaging
 - all tracking data
- **First form** (webpage) – do you want to take part?
 - **Academic research**
 - Yes, with a PC only
- **Second form** – partially supporting our cover story
 - Age?, Gender?
 - Own or shared hardware?
 - Level of IT-knowledge?
- **Second bid – commercial exploitation** (decline, revised bid)
- **Third bid – use by national governments** to improve terrorist activity detection and tracking tools

Structure of responders

- Intent to participate in the first step (academic research usage of data) of the study – 498 subjects (of 1080 loads)
 - BE(3.4%), CZ(40.2%), DE(8%),
 - SK(32.1%), EN(16.3%)
- 284 then actually bid (first scenario)
- Those who saw the introtext and answered
 - will participate – 46.1 %, (26.3 % – first scenario)

Academic usage (quartiles)

First bids			First bids – males			First bids - females		
email	messaging	all	email	messaging	all	email	messaging	all
10	10	12	10	9.5	12	10	10	15
30	30	50	32.5	25	50	30	35	50
100	100	200	100	100	200	275	150	300

- Quartiles instead of min, max, average values
- 23 participants (almost 10%) explicitly opt out for the next scenario, but 27% left

Commercial usage

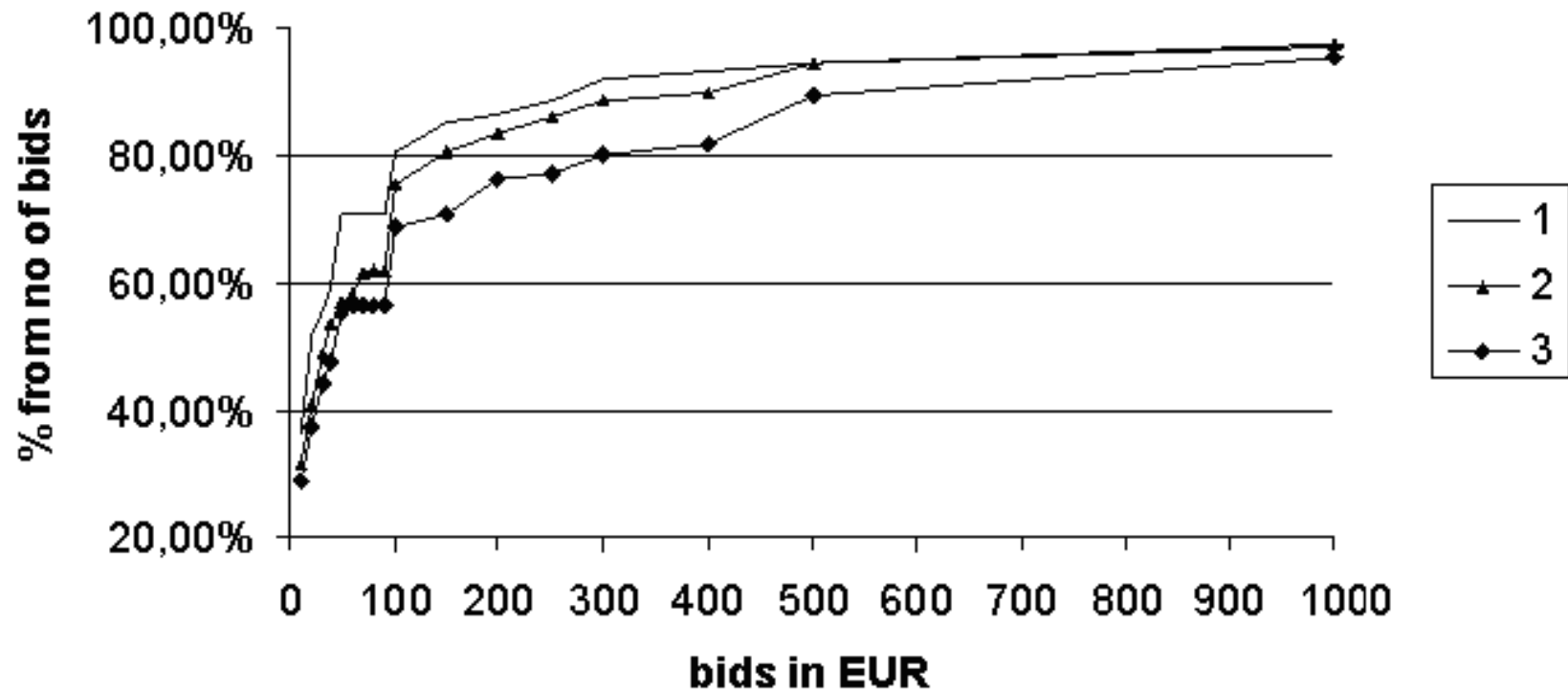
Academic			Commercial			% increase
email	messaging	all	email	messaging	all	
10	8.3	10.4	10	10	15	22%
20	22.5	40	40	40	50	57%
100	80	150	100	100	200	21%

- Medians increased significantly
- 41 participants (18%) explicitly opt out in the next scenario, but 28% actually left

Usage by governments

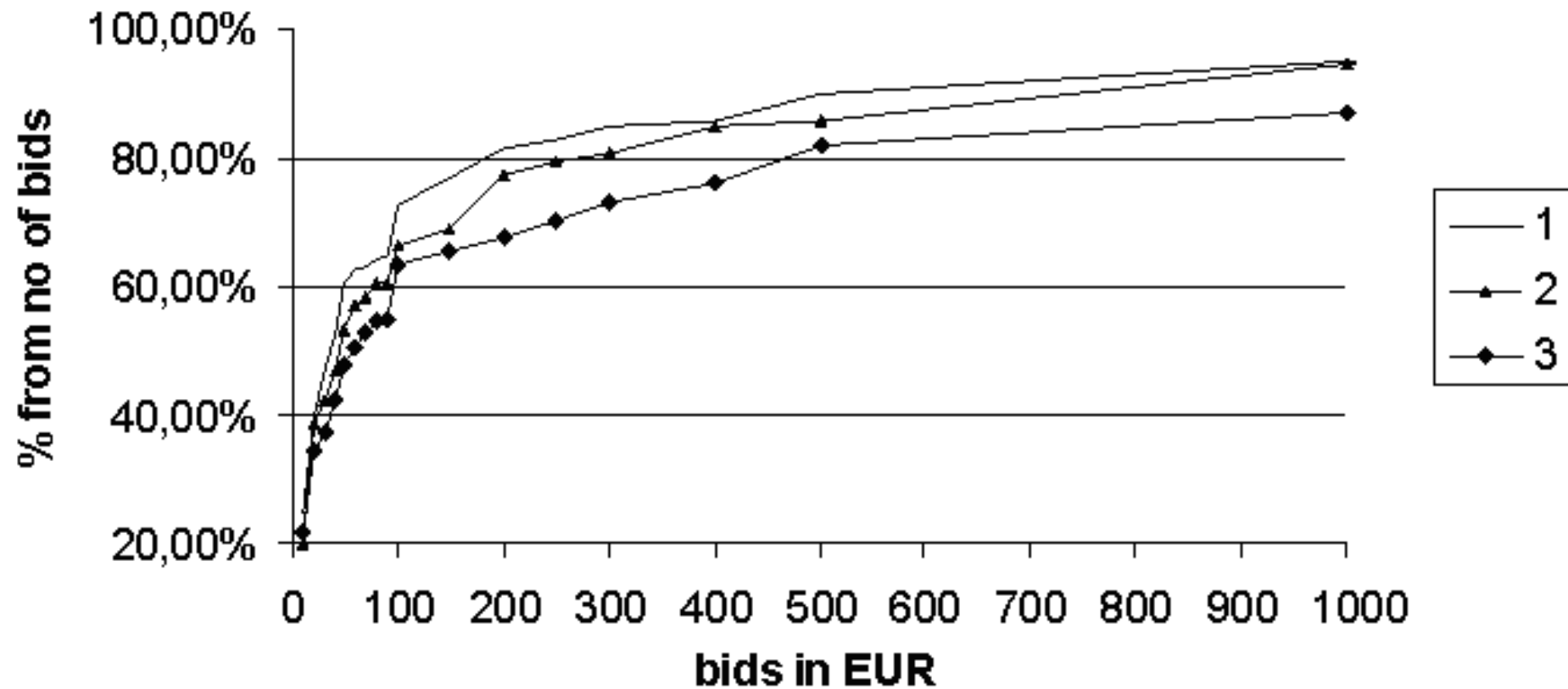
Second bids			Third bids		
email	messaging	all	email	messaging	all
10	10	15	10	10	15
40	40	50	50	50	60
100	100	200	200	200	400

Histogram – 1st bid, all scenarios



- Higher differences expected

Histogram – 3rd bid, all scenarios



Highlights of the second study

- 284 responses for at least the first scenario
 - responses from more than four countries
- EUR 30 for being tracked (email or instant messaging) for academic purposes
 - EUR 50 for all tracking data
 - No considerable differences between males and females
- Increasing tendency to opt out with changing purpose of tracking
 - 1/10 academic -> commercial usage (real dropout 27%)
 - 1/5 commercial -> governmental usage (real dropout 28%)
- Governmental usage
 - After dropouts, ie valuation of all-consenting subjects
 - €50 for one type of data (cf. €40 commercial, €20/25 acad.)
- No significant difference between value of email and other messaging traffic data

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 uživateli, 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

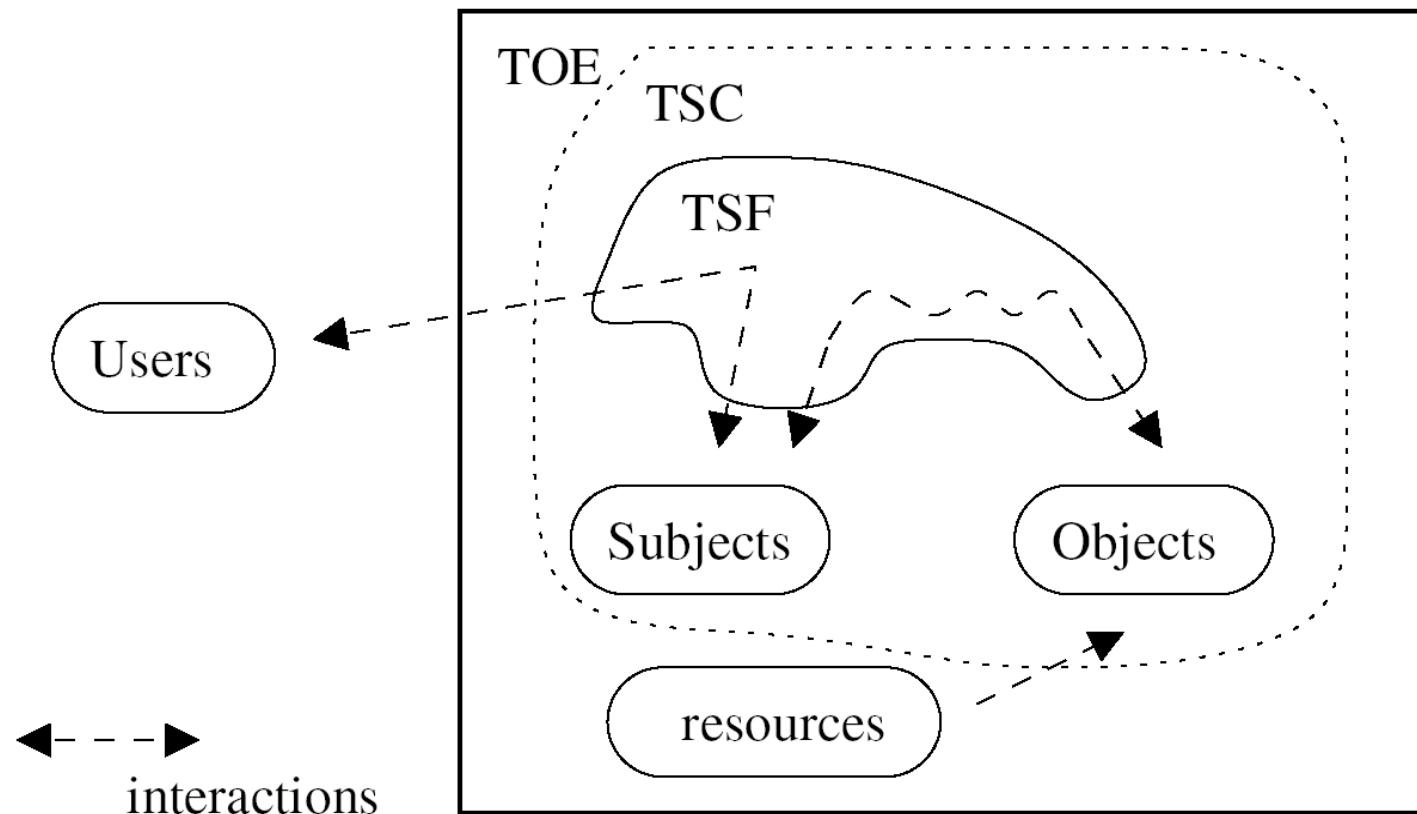
- *Mixy* – systémy pro posílání zpráv, obvykle s klamavými (kamuflovacími) zprávami, preposílání mezi více účastníky (než je nejkratší/nejoptimálnější cesta)
 - Příklad systému později v semestru
- *Společná kritéria (systémy)* – standard (rozsáhlý) pro hodnocení bezpečnosti systémů, umožňuje lepší srovnávání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

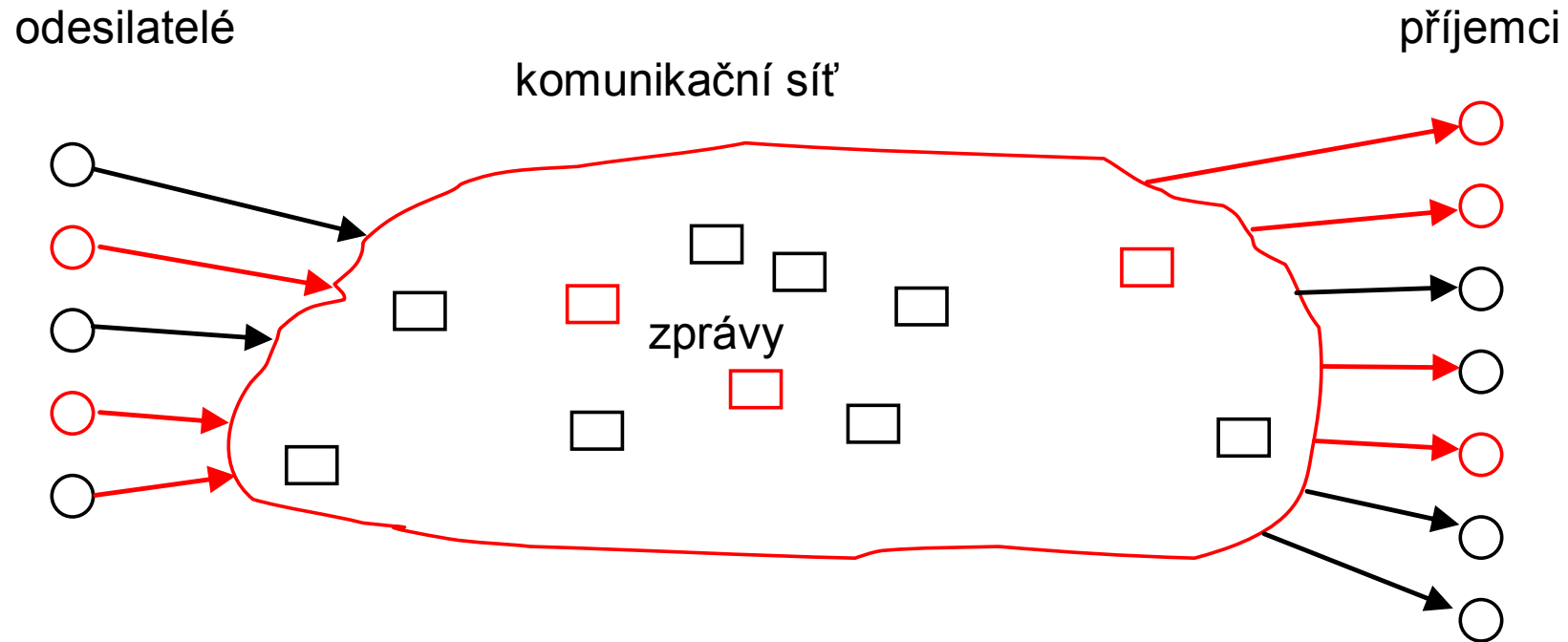
Pohled Společných kritérií

- 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 odesílatelů k příjemcům
 - Specifickou (a nejvýznamnější) podmnožinou jsou tzv. mixy (sítě mixů zdefinoval David Chaum v roce 1981)

Takže obvyklé prostředí...



útočníci
(příklad jejich domény červenou barvou)

(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 odesílatelů
 - 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 Pfizmanna 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...)

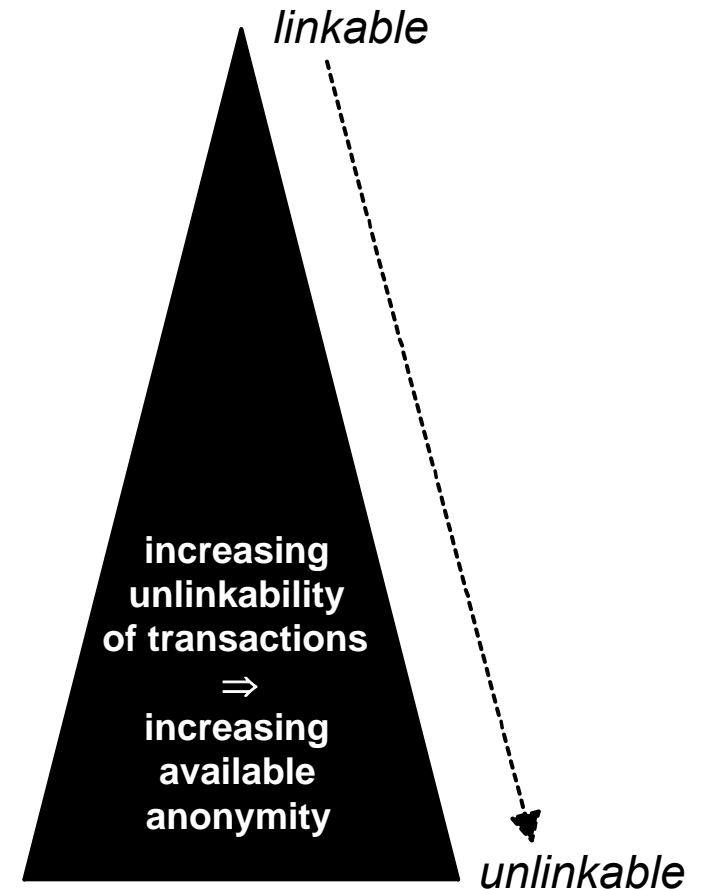
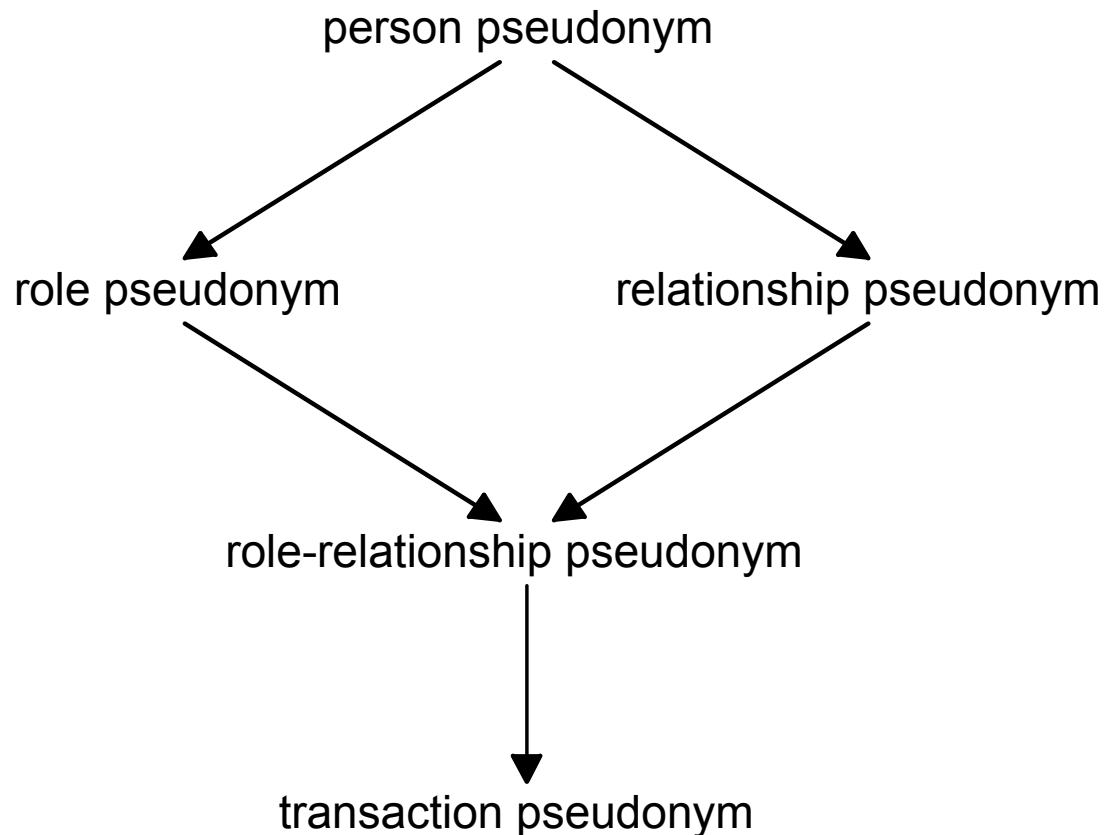
Vztah mezi pseudonymem a vlastníkem (A.P.)

- *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.

Systemy 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 (systemy jednoduchého přihlašování)
 - Public-key infrastructures (infrastruktury veřejných klíčů – nejčastěji pro spolehlivé spojení klíče a informací o osobě)