CERT.LV

Informācijas tehnoloģiju
drošības incidentu
novēršanas institūcija

Atvērtie dati – iespējas un izaicinājumi

Kārlis Podiņš, CERT.LV

Saturs

CERT.LV
Informācijas tehnoloģiju drošības incidentu novēršanas institūcija

- **Atvērtie dati**
- **RD e-talona incidents**
- **Ieteikumi un secinājumi**

Atvērtie dati

CERT.LV
Informācijas tehnoloģiju drošības incidentu novēršanas institūcija

- Atvērta pieeja dažiem datiem brīvai tālākizmantošanai
- *Open data*
  - Turpinājums atvērtības ideoloģijai
    - open source, open hardware, open content and open access
- Labi
- Sarežģīti
  - *...idea that **some** data should be freely available...*

Anonimizācija

CERT.LV
Informācijas tehnoloģiju drošības incidentu novēršanas institūcija

- Visus datus nevar publicēt
- Personas dati?
  - Netiks apskatīti LV likumdošanas kontekstā
- Tieši/acīmredzami
  - Vārds uzvārds
  - Dažādi identificējoši numuri
    - p.k., pases nr, telefona nr,...
- Netieši
  - *Patterns*
  - Izmantojot citus publiskus datus

Deanonimizācija (1)

CERT.LV
Informācijas tehnoloģiju drošības incidentu novēršanas institūcija

- Sweeny, 1997
  - Slimnīcas pacientu dati
    - Pasta indekss + dz.datums + dzimums
  - Vēlētāju reģistrs
    - Vārds, adrese, **pasta indekss, dz.datums, dzimums**

Deanonimizācija (2)

CERT.LV
Informācijas tehnoloģiju drošības incidentu novēršanas institūcija

- Shmatikov & Narayanan, 2008
  - Netflix filmu noma
    - Datums, filma, klientaID
  - Lietotāja dati (piemēram sociālie tīkli)
    - Twitter 21.12.2012 12:43 “tiko netflix noskatījos Limuzīns jāņunakts krāsā”

Deanonimizācija (3)

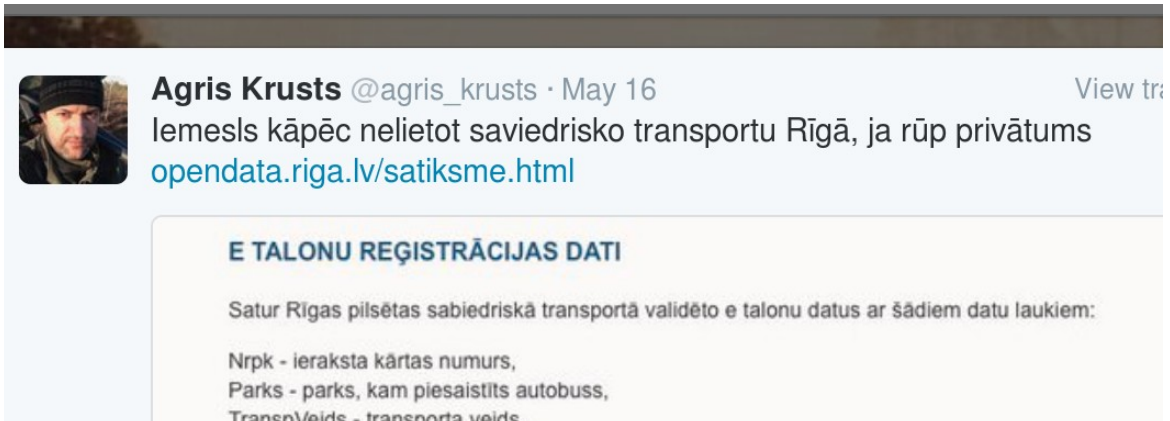
CERT.LV
Informācijas tehnoloģiju drošības incidentu novēršanas institūcija

- E-talona lietojuma dati
  - Podiņš & Lavrenovs, 2016
- **Jebkura statistiska datubāze ietekmē privātumu (gan objektiem, kas iekļauti datubāzē, gan ārpus datubāzes atstātiem)**

E-talona atvērtie dati

CERT.LV
Informācijas tehnoloģiju drošības incidentu novēršanas institūcija

- 2009 Rīgas Satiksme ievieš elektronisku (RFID) biļešu sistēmu
- 25.03.2016 RD mājaslapā publicēti e-talona lietojumi par diviem mēnešiem
  - februāris 2015, janvāris 2016
- 16.05.2016 twitter disputs
- 28.01.2016 LATA balva
  - Atvērtākā iestāde



LATA
GADA BALVA
2015
ATVĒRTĀKĀ IESTĀDE

E-talona atvērtie dati

CERT.LV
Informācijas tehnoloģiju drošības incidentu novēršanas institūcija

- Papildus dati
- Pēdējo 6 braucienu vēsture glabājas e-talonā, brīvi pieejama
- Atvērtie dati
  - 2 mēneši
  - Katrai dienai savs csv fails ar ierakstiem par katru braucienu:
    - ParkaID
    - TransportaTips
    - TransportaID
    - Maršruts
    - Virziens
    - PubTalonaID (“šifrēts”)
    - Laiks (1s precizitāte)

Datu kvalitāte

CERT.LV
Informācijas tehnoloģiju drošības incidentu novēršanas institūcija

- 10 917 654 braucienu 2016. janvārī
- 5 505 785 braucienu 2015. februārī
- 2015. februāris:
  - Nav tramvaju un trolejbusu braucienu
  - 20.5% trūkst pubTalonaID
  - 2015. februāra dati nav uzticami un netiek izmantoti pētījumā
    - Vai tie reizē ir arī norēķinu dati ???
- 2016. janvāris:
  - 57 773 braucienu no iepriekšējā mēneša
  - Sinhronizācija?
    - 26 dienas???
  - 75 braucieni no nākamā mēneša

Labojamas kļūdas

CERT.LV
Informācijas tehnoloģiju drošības incidentu novēršanas institūcija

- Kritiski:
  - trūkst datu par reisu, t.i. cikos izbrauca no galapunkta
  - Trūkst datu par ģeogrāfisko vietu, kur veikta talona reģistrācija (piem. pēc pieturas “Dzirnavu iela”)
- Sekas: Dotie atvērtie dati ir nelietojami pasažieru plūsmas pētīšanai
- Pētījumā izveidotas statistiskas metodes reisa un iekāpšanas pieturas noteikšanai

Nelabojamas kļūdas

CERT.LV
Informācijas tehnoloģiju drošības incidentu novēršanas institūcija

- Nemainīgs/unikāls pubTalonaID
- Deanonimizācija
  - Datubāzes vaicājums, izmantojot pēdējo 6 braucienu vēsturi, kas pieejama talonā
    - Pieņemot, ka tiek publicēti visu braucienu dati
  - Transporta lietojuma *pattern*
    - 2 transporti, 15 minūšu laika intervāli rītam un vakaram
    - Veiksmīga deanonimizācija personai A
- Talona ID “šifrēšanas” reversēšana

PersonaA - deanonimizācija

CERT.LV
Informācijas tehnoloģiju drošības incidentu novēršanas institūcija

- Atrodam pazīstamu cilvēku ar fiksētu sabiedriskā transporta izmantošanas *pattern*
- Lejuplādējam datus un importējam datubāzē
  - `select distinct(validtalonaid) from braucieni where validtalonaid in (select validtalonaid from braucieni where tmarsruts='A 3' and hour(laiks)=6 and minute(laiks) between 40 and 50 and virziens='Forth') and validtalonaid in (select validtalonaid from braucieni where tmarsruts='Tr 15' and hour(laiks)=7 and minute(laiks) between 1 and 15 and virziens='Forth')`

Talona ID "šifrēšanas" reversēšana

CERT.LV
Informācijas tehnoloģiju drošības incidentu novēršanas institūcija

- PubTalonaID
  - 9 dec.cipari – 130673
  - 10 dec.cipari – 185681
  - 13 dec.cipari – 574246
  - 14 dec.cipari – 60351
- Secinājums – netiek lietots vienots *hash*
  - Zelta likums – *paštaisīta kriptogrāfija = problēmas*
- Datu analīze
  - 13,14 cipari – dzeltenie taloni
  - 9,10 - pārējie

Talona ID "šifrēšanas" reversēšana

CERT.LV
Informācijas tehnoloģiju drošības incidentu novēršanas institūcija

- Apskatam starpību starp secīgiem pubTalonsID
  - 3,6,9,12,15,18 – veido 76,8% no visiem
- Atbilde ir ~~42~~ 3
- y=ax+b ???
  - a=3
- Zināms viens pāris (x,y) no deanonimizācijas
- b=-3072913

Talona ID "šifrēšanas" reversēšana

CERT.LV
Informācijas tehnoloģiju drošības incidentu novēršanas institūcija

- Vispār mums bija zināmi divi pāri (x1,y1) un (x2,y2)
- Lineāru vienādojumu sistēma
  - y1=ax1+b
  - y2=ax2+b
- VISC, matemātikas standarts 9.klasei:
  - Vienādojums ar diviem nezināmiem (gan 1., gan 2. pakāpes).
  - Vienādojumu grafiks.
  - Vienādojumu sistēmas jēdziens.

Nu un tad?

CERT.LV
Informācijas tehnoloģiju drošības incidentu novēršanas institūcija

- Kā iegūt talonaID?
  - Tieši
    - Nolasīt RFID
    - Uzdrukāts
  - Netieši
    - No sociālo tīklu ierakstiem
    - No attēliem

Nu un tad?

CERT.LV
Informācijas tehnoloģiju drošības incidentu novēršanas institūcija

- No talonaID varētu iegūt
  - Mājas un darba/studiju vietas
  - Piederība kādai konkrētai grupai
    - Reliģija, partija, u.c.

Feature creep

CERT.LV
Informācijas tehnoloģiju drošības incidentu novēršanas institūcija

- E-talona un bankas kartes kombinācija
- Zinot talonaID iespējams noskaidrot bankas kartes derīguma termiņu
  - Nav reālistiski
  - PCI compliance

Ko darīt? E-talona gadījums.

CERT.LV
Informācijas tehnoloģiju drošības incidentu novēršanas institūcija

- Lietotājiem
  - RFID-droši konteineri
  - Vienreizējie taloni
  - Mazaizsargātām sociālām grupām nav iespējas aizsargāt savu privātumu
- Sistēmas turētājam
  - Lietot kriptogrāfiski drošas jaucējfunkcijas
  - Lietot unikālu salt vērtību katrai dienai

Ko darīt?

CERT.LV
Informācijas tehnoloģiju drošības incidentu novēršanas institūcija

- Domā. Dari
  - LMT – CERT.LV konferences sponsors
  - Ievērot secību!!!
- Sistēmas arhitektiem/programmētājiem nepieciešama pamatskolas izglītība
  - Vēlams augstākā
- Sabiedrībai – kļūt aktīvākai
  - pieprasīt savu tiesību ievērošanu
  - sodīt pārkāpējus

Ko darīt? 2

CERT.LV
Informācijas tehnoloģiju drošības incidentu novēršanas institūcija

- Privātajā sektorā – mazāk aktuāli
  - Dati = biznesa pamats
- Publiskajā sektorā – aktuāli
  - Prognoze – augoša
  - Politiskā/vadības vēlme izskatīties moderniem
  - Riski un sekas netiek apzinātas
    - Vislabāk palīdz sliktie piemēri – skat. iepriekš


Paldies!