

FELHASZNÁLÓK NYOMKÖVETÉSE A WEBEN, ÉS KAPCSOLÓDÓ VÉDEKEZÉSI MÓDSZEREK

Gulyás Gábor György
Óraadó
BME Híradástechnikai Tanszék
gulyasg@hit.bme.hu

2011. május 3.,
Budapest

Privátszférát Erősítő Technológiák?

- PET – Privacy Enhancing Technologies
- Adatvédelem:
  - A személyes adatok gyűjtésének, feldolgozásának és felhasználásának korlátozását és az érintett személy védelmét biztosító alapelvek, szabályok, eljárások, adatkezelés eszközök és módszerek összessége.
- A PET technológiák célja az *adatvédelem biztosítása*.
  - *Felhasználó központúság*, vezérlés
  - Nem kötelez: *döntési lehetőség*
  - Idealizált cél, hogy szakértelem nélkül, a hozzá kapcsolódó *információkat* gond nélkül *menedzselje*
  - Üzleti igények
    - Kompromisszum szükséges, de nem zárják ki egymást!

PET-EK AZ INTERNETEN

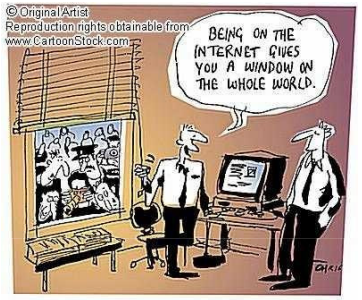
PET – csak egy elméleti igény?

- Bevásárlási szokások
- Mobiltelefonok helymeghatározása
- Kamera-túlterheltség
- Elektronikus jegy problémái
- RFID privacy
- E-kereskedelem
- Internet
  - levelezés
  - böngészés
  - különféle tartalmak megosztása (blog, hírek, fájlok)
  - vagy épp tartalmak elérése
- …

Miért érdekelne ez engem?

- Mindenkinek vannak érzékeny (személyes) adatai

  (amelyeket nem kötne csak úgy bárki orrára a hetes buszon).
  - Vallási beállítottság
  - Betegségek, egészségügyi állapot
  - Pénzügyi helyzet
  - Érdeklődési kör
  - Szexuális érdeklődés és kapcsolatok
  - Fiatalkori kicsapongások
  - Politikai meggyőződés
  - …
  - Data retention irányelv

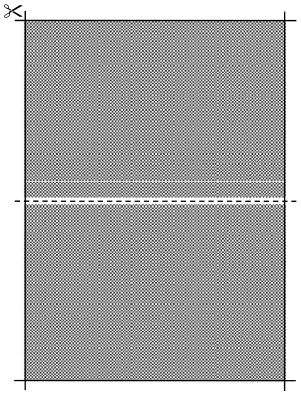


A PRIVÁTSZFÉRA A WEBEN

Mitől lesz egy technológia PET?

+

RFID árnyékolása az útlevélben



Adóbevallás 2011

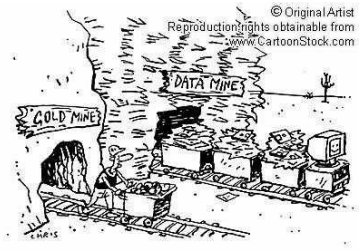
Miért fontos kérdés?

- Profilkészítés
  - Minden *személyre szabhatóvá* válik:
    - Reklámok
    - Web boltok árai
    - Persze a hagyományos tartalom és felület is…
  - Eszközök, módszerek:
    - *Követés*: érdeklődési kör, ízlésvilág feltérképezése
    - *Adatok gyűjtése*: loginok, e-mail címek, …
  - Pszeudonim vagy névvel ellátott profilok
- Identitás lopás, megszemélyesítés
- Érzékeny adatok ellopása (phishing, whaling)

A webes szféra szereplői, céljai

Felhasználók	Felhasználóközpontú vezérlés

Hirdetők	Profilalapú célzott hirdetések
Webes üzletek	Profilalapú dinamikus árazás
Adatgyűjtők	Profilkészítés, adat kereskedelem
Szolgáltatók	Naplózás, tiltás, együttműködés más szolgáltatókkal
Cenzúrázó szervek	Szabályozás, megfigyelés



NYOMKÖVETÉSES PROFILÍROZÁS

Potenciális veszélyforrások

I. **„Információs szuperhatalmak", nagy szolgáltatók szolgáltatásain belül**
(Pl. Google szolgáltatásai)

II. **Profilírozás publikus adatforrások, önkéntes adatszolgáltatás alapján**
(Pl. keresők, közösségi hálózatok adatai alapján, pl. Facebook adatvédelmi nyilatkozat 2009. nyarán)

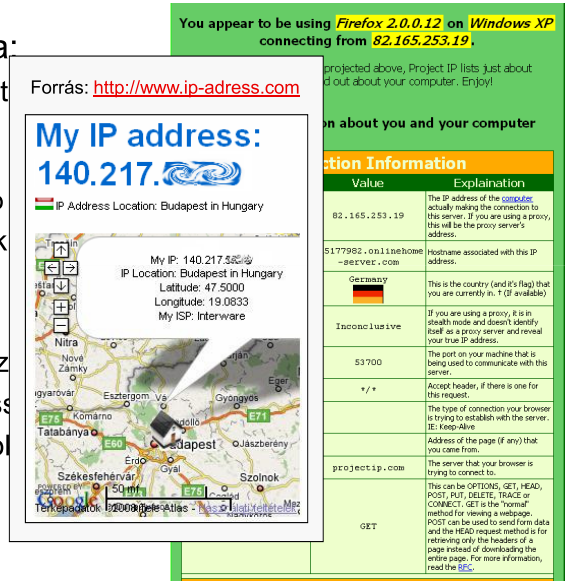
III. **Nyomkövetéses profilírozás**
(Pl. IP vagy süti alapú nyomkövetés)

Nyomkövetéses profilírozás

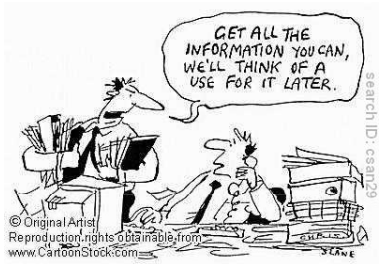
- Célja a profilépítés
  - Pszeudonim profil: szám jellegű azonosítóval ellátott
  - Azonosított profil: névvel ellátott, a felhasználó személyére utaló vagy azt pontosan jelölő profil
- Módszerei
  - A felhasználó tevékenységének követése online
  - Tevékenység rekonstrukciója a kiszolgálónál, vagy a felhasználónál keletkezett naplók, adatbázisok alapján

Árulkodó nyomok?

- A szolgáltató látja:
  - IP címünk (+ host
    - Port szám
    - Ország, város beazonosítható
  - Nyelvi beállítások
  - Böngésző
    - Pontos verzió!
  - Operációs rendsz
  - JavaScript képess
  - Telepített pluginol
  - …



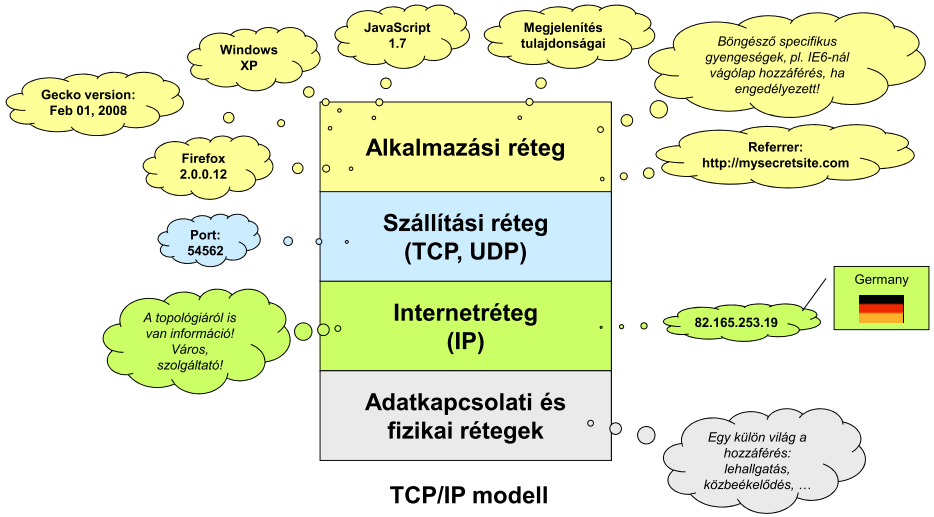
Forrás: http://projectip.com



- Nyomkövetéses profilírozás
  - **Alapvető technikák: IP és sütik**
  - Kiegészítők használata: Flash
  - Azonosításon alapuló módszerek
  - Evercookies
  - Offline adatbázisok
  - Védelmi módszerek

ALAPVETŐ TECHNIKÁK

Amit a szolgáltató lát – szakszerűbben



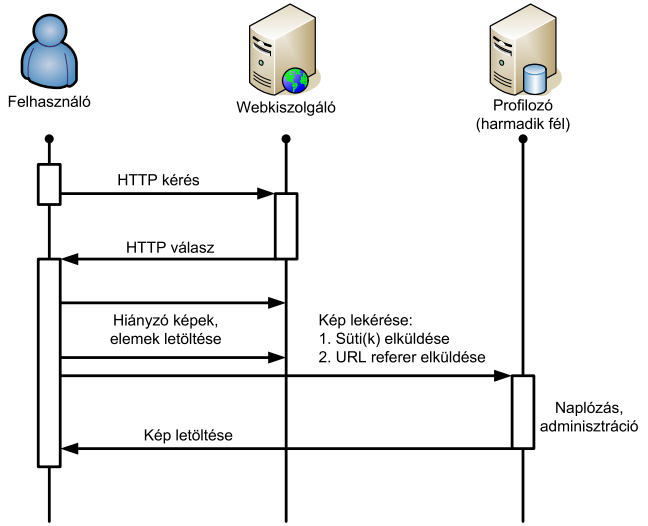
TCP/IP modell

Alapvető technikák

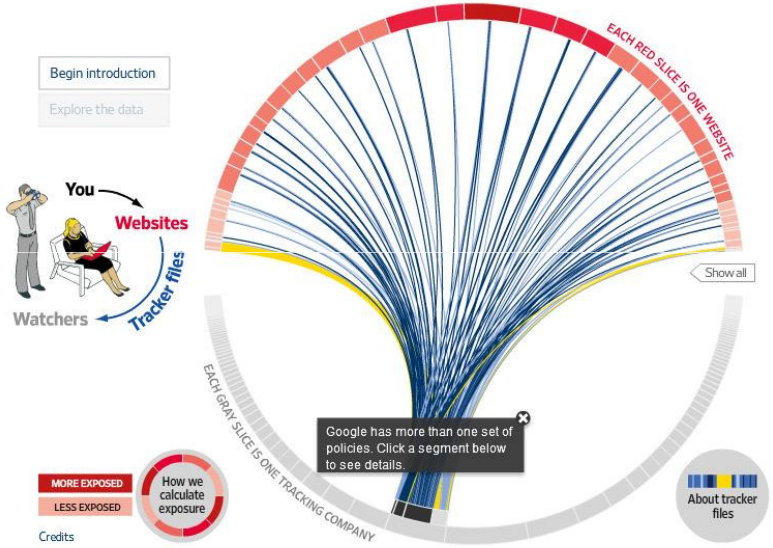
- IP alapú nyomkövetés
  - Kezdeti technika, egyedi IP címek esetén működik
  - A NAT, dinamikus IP miatt már nem hatékony
- Szolgáltatók együttműködése
  - Hirdetésekkel, szolgáltatói együttműködéssel
  - Egymás hirdetéseit megjelenítik
  - Pl. Yahoo Web Beacons
    - Nyilvánosan, adatvédelmi szabályzat szerint
    - Regisztrációhoz köthető azonosítás
    - Web poloskákhoz hasonló, sütis módszer (ld. később)

Böngésző sütik

- Angolul: cookie (tracking-cookie, third-party cookie)



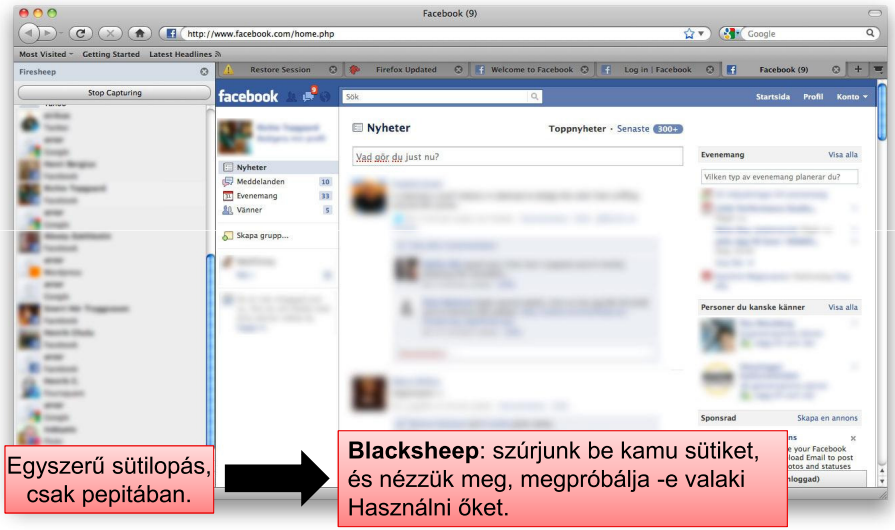
Wall Street Journal: What They Know



Böngésző sütik (2)

- Sütik ellopása lehallgatással és munkamenetek eltérítése
  - Pl. Wifi kapcsolat lehallgatása
- XSS – Cross Site Scripting
  - Sütik ellopása és átküldése a támadó oldalára
- Sütimérgezés
  - A kiszolgáló működésének befolyásolására
- Soha le nem járó sütik

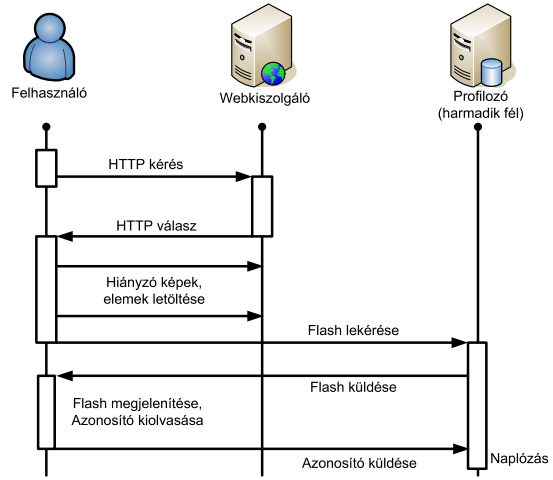
Firesheep (és Blacksheep)



- Nyomkövetéses profilírozás
  - Alapvető technikák: IP és sütik
  - **Kiegészítők és egyéb technikák**
  - Azonosításon alapuló módszerek
  - Evercookies
  - Offline adatbázisok
  - Védelmi módszerek

KIEGÉSZÍTŐK ÉS EGYÉB TECHNIKÁK

Flash sütik



- LSO = Local Shared Object, sütiszerű működés
- Böngészőfüggetlen tárolás

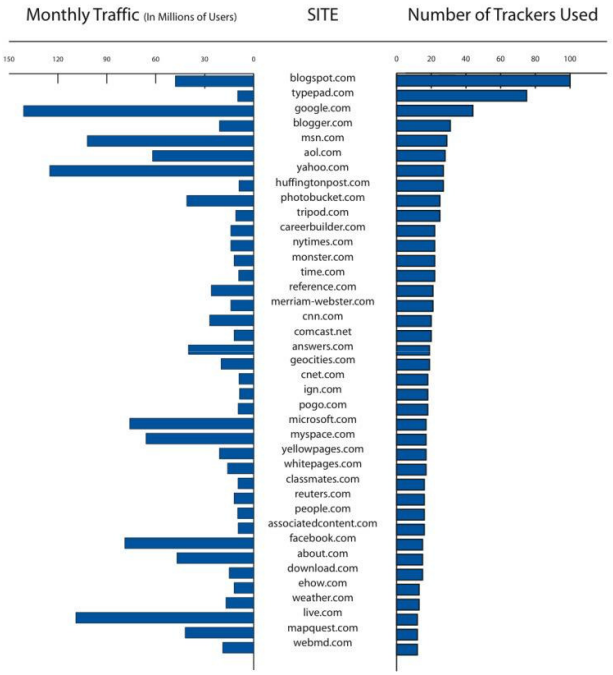
Flash (2)

- Felhasználás
  - Nyomkövetésre
    (PIE: Persistent Identification Element)
  - Böngésző sütik visszaállítására (backup cookie)
- A Flash egyéb veszélyei
  - AJAX?
  - Felhasználói tevékenység mérésére
- Microsoft Silverlight technológia

Egyéb módszerek

- Hirdetések, külső szolgáltatók képei
  - Az URL-ben azonosító is lehet!
  - URL referer felhasználása
  - Sütikre építhet
- Web poloska (web bug)
  - Láthatatlan: 1x1 pixeles, átlátszó GIF-ek
  - Sütivel, vagy IP és egyéb információk alapján

Egyéb módszerek (2): Web bugok



Egyéb módszerek (3)

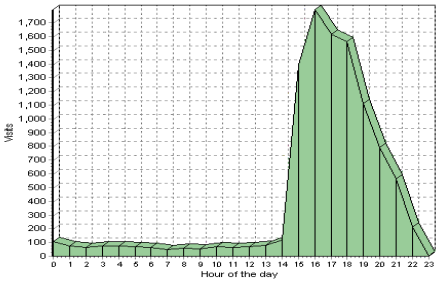
- URL-referer
  - Előző oldal, de környezet azonosítása is
  - Képek, Flash környezete
  - Kereső kifejezések is kiderülnek!
  - IFRAME környezete
- Trükkös e-mailben mellékletek
  - Mint képek, url-ek, web poloskák
  - Pl. SpyPig
- JavaScript gyorsítótár:

tracking.js:

```
var track_id = ”bZL0cjcLBYY”;
```


Egyéb módszerek (4): Meta-információk automatikus felfedése

- Mivel?
  - *RSS, Atom* hírcsatornák
  - *Kezdőlap* automatikus betöltése
- Azonosítás sütikkel, vagy mással



- Mi ezzel a probléma?
  - Felfedjük internetezési szokásainkat:
    - napi első letöltés
    - nap közbeni tevékenységek
  - IP címünk automatikus felfedése (hozzákapcsolható a sütiben tárolt azonosítóhoz)
- Csatolt lokalitás:
  - Munkahely
  - Otthon, szórakozás (kávézó)

- Nyomkövetéses profilírozás
  - Alapvető technikák: IP és sütik
  - Kiegészítők használata: Flash
  - **Azonosításon alapuló módszerek**
  - Evercookies
  - Offline adatbázisok
  - Védelmi módszerek

AZONOSÍTÁSON ALAPULÓ MÓDSZEREK

History Stealing

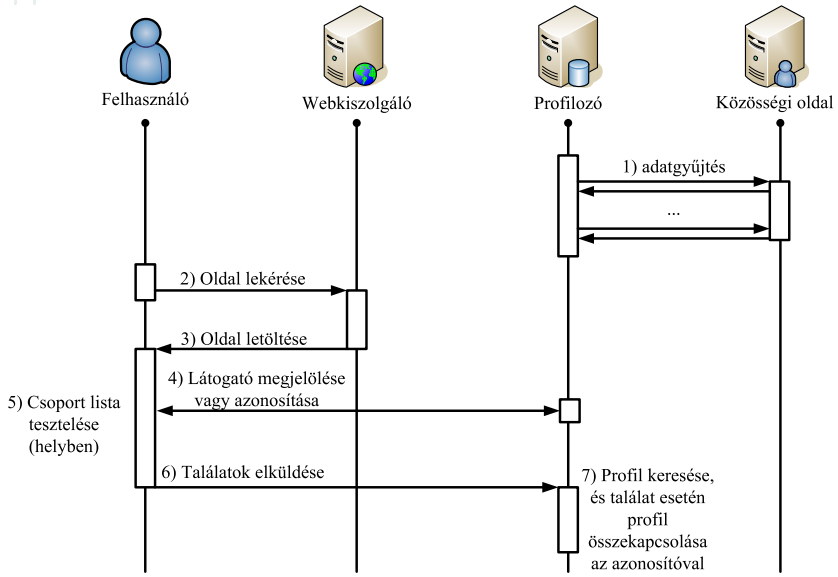
College for Enlisted Professional Military Educati
Columbus AFB MS
Fairchild AFB WA
Goodfellow AFB TX
Inter-American Air Forces Academy, Lackland AFI

- Alapkoncepció
  - A böngészési előzmények nem explicit lekérdezhetőek, csak igen/nem jellegű
  - Pl. nem látható linkek állapotának lekérdezése
- URL lista látogatottságának ellenőrzése
  - Hibatűrés: előzmények törlése miatt szükséges!
- Mire használható?
  - Célzott hirdetések
  - Dinamikus árazás
  - Profilírozás
- Védelem
  - Előzmények automatikus törlése, privát böngészési mód, vagy például SafeHistory Firefox kiegészítő
  - Vagy Firefox 4.0 ☺

History Stealing (2)

- Azonosítási módok
  - Pszeudonim -> egyedi URL minta szerint
  - Személyes: közösségi hálózatban is!
- Azonosítás közösségi hálózatban
  - Profilinformációk gyűjtése (crawling), adatbázis építés
  - Csoport tagságok tárolása, amelyeknek egyedi URL-jük van
  - Ellenőrzés: milyen csoportokat látogatott meg
- Valódi személyt is azonosít a módszer!

History Stealing (3)



Panopticlick

Panopticlick

How Unique — and Trackable — Is Your Browser?

Your browser fingerprint **appears to be unique** among the 742,632 tested so far.

Currently, we estimate that your browser has a fingerprint that conveys **at least 19.5 bits of identifying information.**

The measurements we used to obtain this result are listed below. You can read more about the methodology here, and about some defenses against fingerprinting here.

Help us increase our sample size:

Browser Characteristic	bits of identifying information	one in *x* browsers have this value	value
User Agent	14.14	18112.98	Mozilla/5.0 (Windows; U; Windows NT 6.1; hu; rv:1.9.2) Gecko/20100115 Firefox/3.6 (.NET CLR 3.5.30729)
HTTP_ACCEPT Headers	9.69	824.23	text/html, */* ISO-8859-2,utf-8;q=0.7,*;q=0.7 gzip,deflate hu-hu,hu;q=0.8,en-us;q=0.5,en;q=0.3

Panopticlick (2)

- Előnyök
  - Kis műveletigénye van (és csak) kliensoldalon
  - Nem igényel adattárolást
  - Pontos azonosításra képes
- Hátrányok
  - Böngészőfüggő
  - Plugin függő (Java, Flash)
  - Géptermekben nem működik!
- Süti újragenerálását is tudja!

Fingerprinting

Az Ön számítógépének ujjlenyomata:

Ujjlenyomat:	8bf2bcc936e098497691f3782c3c6fa62705f51e
A böngésző adatait tartalmazó karakterlánc:	Mozilla/5.0 (Windows; U; Windows NT 6.1; hu; rv:1.9.2.9) Gecko/20100824 Firefox/3.6.9 (.NET CLR 3.5.30729)
Operációs rendszer:	Windows
Képernyő felbontás:	1680x1050
Időzóna:	-120
Telepített betűkészletek:	

Aharoni, ALGERIAN, Andalus, AngsanaNew, AngsanaUPC, Aparajita, Arabic Transparent, Arabic Typesetting, Arial, Arial Baltic, Arial Black, Arial CE, Arial CYR, Arial Greek, Arial Narrow, Arial TUR, Arial Unicode MS, Baskerville Old Face, Batang, BatangChe, Bauhaus 93, Bell MT, Berlin Sans FB, Berlin Sans FB Demi, Bernard MT Condensed, Bodoni MT Poster Compressed, Book Antiqua, Bookman Old Style, Bookshelf Symbol 7, britannic bold, Broadway, BrowalliaNew, BrowalliaUPC, Brush Script MT, Calibri, Californian FB, Cambria, Cambria Math, Candara, Centaur, Century, Century Gothic, Chiller, Colonna MT, Comic Sans MS, Consolas, Constantia, Cooper Black, Corbel, CordiaNew, CordiaUPC, Courier, Courier New, Courier New Baltic, Courier New CE, Courier New CYR, Courier New Greek, Courier New TUR, cursive, DaunPenh, David, DejaVu Sans, DejaVu Sans Condensed, DejaVu Sans Light, DejaVu Sans Mono, DejaVu Serif, DejaVu Serif Condensed, DFKai-SB, DilleniaUPC, DokChampa, Dotum, DotumChe, Ebrima, Estrangelo Edessa, EucrosiaUPC, Euphemia, Ezra SIL, Ezra SIL SR, FangSong, fantasy, footlight mt light, Franklin Gothic Medium, FrankRuehl, FreesiaUPC, French Script, Gabriola, Garamond, Gautami, Gentium Basic, Gentium Book Basic, Georgia, Gisha,

Fingerprinting (2)

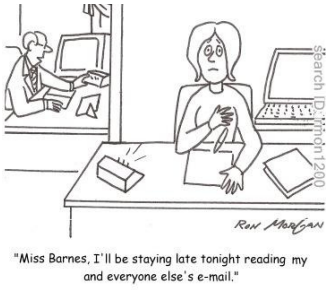
- Alapkoncepció
  - Rendszerjellemzők
  - Font lista: nagy halmazban is többé-kevésbé egyedi, korellál a rendszer egyediségével
  - IP cím
- Célok
  - Böngészőfüggetlen
  - Plugin független
  - Cross-domain követés
  - Süti újragenerálás (adattárolást nem igényel)
  - Csak kliensoldali műveletvégzés
- Veszélye: a privát módot is kijátssza!
- Hátrány: géptermekben ugyanúgy nem működik!

- Nyomkövetéses profilírozás
  - Alapvető technikák: IP és sütik
  - Kiegészítők használata: Flash
  - Azonosításon alapuló módszerek
  - **Evercookies**
  - Offline adatbázisok
  - Védelmi módszerek

MINDENT A LEVESBE: EVERCOOKIES

Evercookies alapok

- Koncepció
  - Egy módszer önmagában kevés → redundancia
  - „Gyakorlatilag törölhetetlen sütik”
  - Az ujjlenyomat módszer a vállalati, iskolai, stb. környezetekben becsődölnek
- Több módszer együttes alkalmazása
  - Hagyományos sütik, flash, silverlight, history stealing, stb.
  - Cache: PNG pixelekben (RGB értékekkel)
  - HTML5 adatbázisok
  - További gyorsítótárakban



"Miss Barnes, I'll be staying late tonight reading my and everyone else's e-mail."

- Nyomkövetéses profilírozás
  - Alapvető technikák: IP és sütik
  - Kiegészítők használata: Flash
  - Azonosításon alapuló módszerek
  - Evercookies
  - **Offline adatbázisok**
  - Védelmi módszerek

OFFLINE ADATBÁZISOK

Offline is elérhető adatbázisok

Kliens oldal

- Spyware-k ideális célpontjai
  - Offline hozzáférés
  - „Korlátlanul” rendelkezésre álló idő és erőforrások
  - A cél: profilkészítése
- Előzmények adatbázisa
  - Teljes böngészési profil!
- Sütik teljes adatbázisa
  - Böngésző és flash sütik
  - Adat- és szövegbányászati módszerek
- Gyorsítótár (cache)
  - Tartalmak szemantikus elemzése

Szerver oldal

- Web-logok összemosása
  - Felhasználói tevékenység rekonstrukciója több kiszolgáló logjai szerint
- Leginkább az előző két támadási csoportra jellemző (I.-II. csoportok)
  - Regisztrációhoz köthető tevékenységek
  - Publikált adatok elemzése, összekapcsolása

Spyware védelem

- Tevékenységek
  - Felhasználói tevékenység követése
  - Érzékeny adatok lehallgatása
  - Offline adatbázisokhoz hozzáférés
    - Sütik, Flash sütik, gyorsítótár, előzmények, űrlapok, weblap beállítások, jelszó adatbázis, HTML 5 adatbázisok, stb.
    - Komplex adatbányászat, elemzés és rekonstrukció
- El kell távolítani, aktív védelem szükséges (Anti-Spyware).
- Offline működésről van szó, így az élő kapcsolat védelméről külön kell gondoskodni!

Ad-Aware

Spyware Doctor

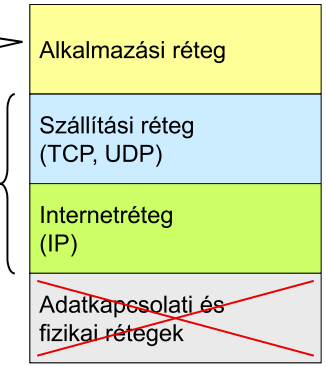
"Visible fences make invisible neighbors."

- Nyomkövetéses profilírozás
  - Alapvető technikák: IP és sütik
  - Kiegészítők használata: Flash
  - Azonosításon alapuló módszerek
  - Evercookies
  - Offline adatbázisok
  - **Védelmi módszerek**

VÉDEKEZÉSI MÓDSZEREK

Védekezés általában

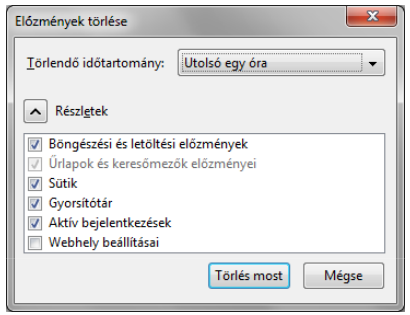
- **Részleges megoldások**
  - Böngésző megfelelő beállítása
  - Weboldal tartalmának szűrése
  - Spyware-ek elleni védekezés
  - Anonim proxyk, anonimizáló hálózatok
- **Komplex megoldás**
  - Anonim böngészők
    - Teljes anonimitás
    - Identitásmenedzsment

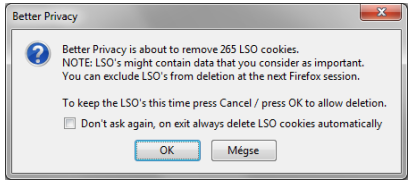


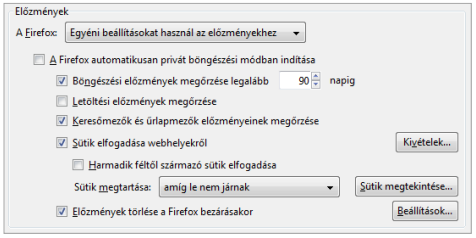
TCP/IP modell

A böngésző megfelelő beállítása…

- Személyes adatok törlése kilépéskor
- Előzmények, gyorsítótár, mezők, űrlapok
  - Privát mód, ha szükséges
- Sütik csak az eredeti kiszolgálótól
- Képek is (néhol hátrányos lehet)
- (Régebbi böngészőben SSL 2.0 letiltása)







Tartalomszűrés, kiegészítők

- Tartalmi elemek szűrése
  - Flash (AdBlock, AdBlock Plus)
  - JavaScript (NoScript)
- Hasznos kiegészítők
  - BetterPrivacy (sütimenedzsment)
  - Ghostery
  - RefControl
- http://pet-portal.eu/links/

Firefoxhoz

- Komplex
  - Web poloskák
  - Hirdetések

BUGNOSIS

- Nem nyújt teljes értékű megoldást
  - Csak alkalmazási réteg
  - Spyware-t egyszer elég elkapni

Anonim proxy-k (esetleg webes)

- Egyszerű anonim proxy
  - Csak IP címet rejt
  - Hagyományos proxyként funkcionál
  - Titkosított kliens-proxy kapcsolat
  - Legfeljebb a HTTP fejléceket módosítja minimális mértékben, például sütik kiszűrése
- Webes felületű anonim proxy:
  - Előzmények és gyorsítótárazás mellőzése
  - Ma már egyáltalán nem jellemző
  - Helyettük privát mód vagy anonim böngészők

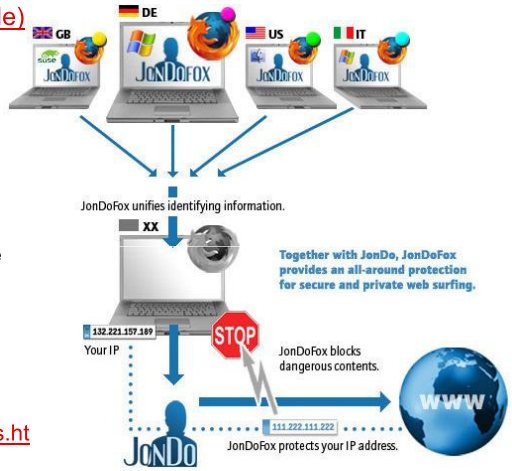
Anonim böngészők (1)

- Speciális webes proxy-k: beavatkoznak a belső protokollba
- Webes felület, vagy lokális proxy alkalmazás
- Funkciók lehetnek:
  - MIX használata
  - Alternatív csatlakozási pontok, dedikált proxy szervek
  - Titkosított kliens-proxy kapcsolat
  - Sütik tiltása, szűrése
  - Java, Flash, ActiveX objektumok szűrése
  - HTTPS átjászása
  - JavaScript szűrése, tiltása
  - Hibás JavaScript kódok szűrése
  - Reklámok szűrése
  - Pop-up tiltása
  - URL-referrer szűrés
  - Böngésző, rendszerinformációk szűrése
  - Előzmények, gyorsítótár mellőzése

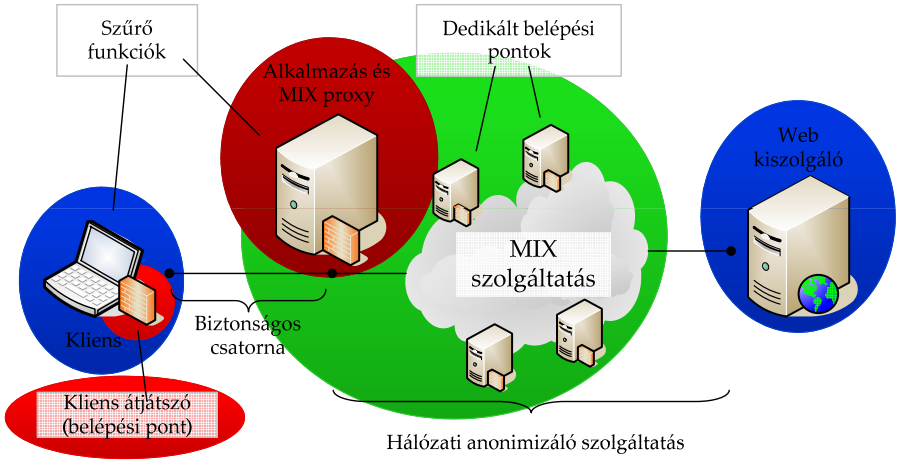
Anonim böngészők (2)

Anonim böngészők (3)

- JonDoFox (https://www.jondos.de)
  - JonDo hálózat (többrétegű titkosítás, mix-kaszkádon keresztül)
  - Futtatási lehetőségek:
    - Hagyományos telepítés
    - Pendrive-ra másolható, onnan futtatható (proxy is)
  - Szűrés kliens oldalon (Firefox kiterjesztések)
    - Adblock Plus, NoScript, CS Lite
    - JonDoFox: referrer, proxy
    - (Private Browsing)
  - Alternatív csatlakozási pontok (cenzúra)
  - Open Source
- Színvonalas gyűjtemények:
  - http://www.epic.org/privacy/tools.html#surf
  - http://pet-portal.eu/links/#link_15



Általános architektúra

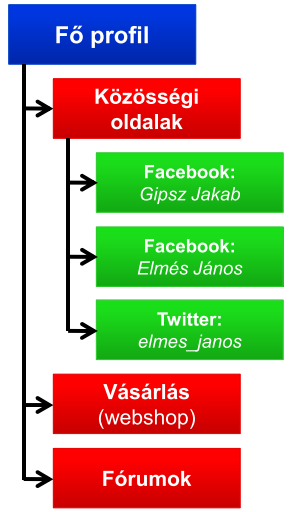


Identitásmenedzsment

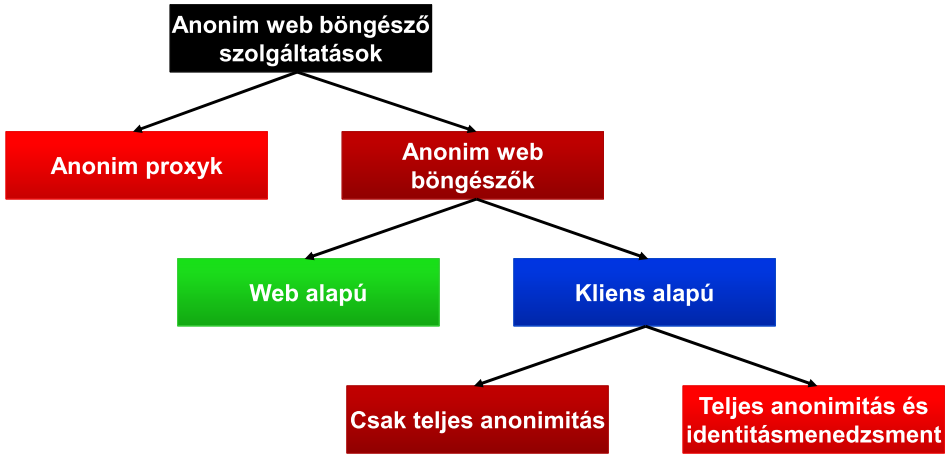
- Amit felfedünk, általában nem vonható vissza
- Egyirányúság (Nymity Slider koncepció)
  - A cél: adatközlés minimalizálása
  - Alapvetően semmit nem osztunk meg (teljes anonimitás)
  - Ha szükséges, felfedünk bizonyos információkat
  - Így garantálható a pszeudonim jelenlét, és az összeköthetetlenség elve

Identitásmenedzsment (2): anonim böngészőkben

- Szerep alapú identitásmenedzsment
  - Más helyzetekben más információ számít
  - Köthetőek: partnerhez, szerepkörhöz
  - Időtartam: hosszú távú, tranzakcióhoz kötött
- A szerepek az áttekinthetőség, kezelhetőség miatt hierarchiába vannak szervezve



Anonim szolgáltatás típusok

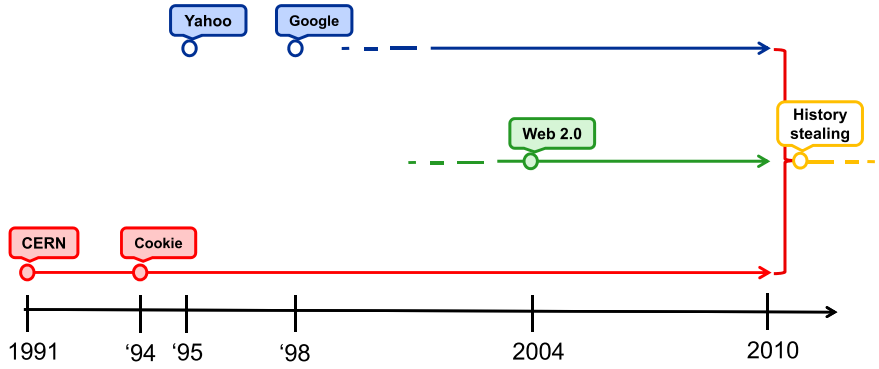


Nyomkövetés: összefoglalás

- Léteznek ígéretes próbálkozások, de
  - Komplex, minden tekintetben jó megoldás nincs
  - A privacy nem finomhangolható
  - Az ígéretes szolgáltatások lassúak, nem megbízhatóak (hibás működés szerver és kliensoldalon egyaránt), vagy pénzbe kerülnek
  - Identitásmenedzsment nincs
- Ko-evolúció jelensége
  - A támadó azért mindig egy-két lépéssel előrébb jár…

Összefoglalás

I. Információs szuperhatalmak
II. Profilírozás publikus forrásból
III. Nyomkövetéses profilírozás



- PET-ek az interneten
- A privátszféra a weben
  - Információs szuperhatalmak
  - Profilírozás publikus források alapján
  - Nyomkövetéses profilírozás
- **Következtetés**

KÖVETKEZTETÉS

Összefoglalás (2)

- Szolgáltató: mi a helyes szemléletmód?
  - Kontextus szerint célzott hirdetések vagy személyes profilok?
- Felhasználói tudatosság növelése
  - „Tudatos fogyasztás”
  - Jogok érvényesítése
  - Felhasználói nyomásgyakorlás
- Védelmi módszerek: nincs mindig rá szükség, de néha van elég jó is…

KÖSZÖNÖM A FIGYELMET!

Gulyás Gábor György
gulyasg@hit.bme.hu

Híradástechnikai Tanszék

http://www.hit.bme.hu

international PET portal and blog

http://pet-portal.eu