

Datenschutz im Web 2.0

Hauptseminar Technischer Datenschutz SS 2007
Tobias Lange - tl@einbecker.net

Agenda

- ▶ *Web 2.0*: Technik, Mensch, Geschäftsidee
- ▶ Datenschutzbegriff im *Web 2.0*
- ▶ Die Onlineidentität
- ▶ Angriffsarten
- ▶ Soziale Netzwerke
- ▶ Zusammenführung zu Profilen
- ▶ Fazit

Web 2.0 eine ungenaue
Wolke von „sozialen“
Anwendungen auf einer
ebenso ungenauen
technischen Basis.

Technik

- ▶ offene APIs, die zusammenspielen
 - ▶ RSS, XML, SOAP...
 - ▶ AJAX (JavaScript + XML + Request, asynchron)
 - ▶ Web Applications
- ▶ seit Jahren im Einsatz
 - ▶ bewährt und verbreitet
- ▶ Neu: hochverknüpft

Soziale Konzepte

- ▶ Nutzer verbringen immer mehr Zeit online
- ▶ digitale Freundschaft
- ▶ Zusammenarbeit / Teilen
- ▶ „Verschlagwortung“ (Tagging - Nutzer erzeugen Semantik)
- ▶ vom Nutzer generierter Inhalt

Beispiele

- ▶ Wikipedia
- ▶ Weblogs
- ▶ Video-/Fotoportale (Flickr, YouTube)
- ▶ soziale Netzwerke (MySpace, StudiVZ)
- ▶ Google (APIs!)

Geschäftlich

- ▶ MySpace: 580 M\$ (News Corp.)
- ▶ YouTube: 1.65 G\$ (Google)
- ▶ StudiVZ: ~100 M€ (Holzbrink)

- ▶ MySpace: 10 \$ / Mitglied
- ▶ YouTube: 256 \$ / Mitglied
- ▶ StudiVZ: 100 € / Mitglied

- ▶ Ziel bei fast allen Unternehmen: Marketingmaßnahmen

Datenschutz?

Datenschutzkriterien

- ▶ ähnlich schwer definierbar wie *Web 2.0*
- ▶ nur notwendige Daten
- ▶ Datenweitergabe erkenntlich
- ▶ Der Nutzer hat die Kontrolle über die Daten
- ▶ Pseudonymität
- ▶ sichere Protokolle

Datenschutzkriterien: Wie gut sind die Dienste?



nur notwendige Daten



Datenweitergabe erkenntlich



Der Nutzer hat die Kontrolle über die Daten



Pseudonymität



sichere Protokolle

Identität?

Identität Summe der unter
einem Pseudonym bereit-
gestellten Informationen

einbecker

Personenpseudonym

\$ whois einbecker.net

Admin Contact

Tobias Lange (TL148-ABC)

Tobias Lange

admin@einbecker.net

Foerstereistrasse 45

Dresden, De, 01099, Germany

phone: +49 3517969645

einbecker.net/top5
einbecker.net/eingebeckert
hommage.at

Weblogs

myspace.com/einbecker
last.fm/user/einbecker
flickr.com/photos/einbecker

Dienste

Identität im Web 1.0



Identität im Web 2.0



Welcome to my Blog!

Flickr-Stream of my
last tour

My music from
last.fm

My visitors on a
map (using
google-analytics)

Verkettung Der Nutzer ermöglicht die Verknüpfung der Inhalte. Sie ist erwünscht.

(Identitätsmanagement)

- ▶ OpenID (single sign-on)
- ▶ SAML (enterprise applications)
- ▶ Microsoft CardSpace

Angriffe

- ▶ Angreifermodell noch schwammiger als *Web 2.0* oder der Datenschutzbegriff hierin
- ▶ Unterscheidung generell in:
 - ▶ allgemeine Sammlung von Daten
 - ▶ gezielte Sammlung zu einer Person
- ▶ „Angriff“ auf Daten, die der Nutzer selbst zur Verfügung stellt.

Angriff: Orte & Zeit

- ▶ Flickr: GeoTagging (Koordinaten des Photos)
- ▶ Plazes (Loginpunkte)
- ▶ Qype (Bewertung von Kneipen, Läden, ...)

Angriff: Beziehungen

- ▶ Beziehungsstatus: single / vergeben / offen für alles
- ▶ Freundeslisten (friend-of-a-friend)
- ▶ Blogrolls (Auflistung gelesener Weblogs)

Angriff: Interessen

- ▶ Wunschliste bei Amazon (intern auch: gespeicherte Einkäufe / betrachtete Artikel)
- ▶ Musikstile bei last.fm / Pandora / MySpace
- ▶ Produktbewertungen (Amazon / Qype / Dealjäger)
- ▶ Bookmarks (del.icio.us)

Angriff: Reputation

- ▶ Bewertungen bei Ebay
- ▶ Kommentare in einem Blog
- ▶ Eingehende Links zum Blog
- ▶ Freunde bei MySpace, Flickr, StudiVZ

soziale Netzwerke

- ▶ berufliches Umfeld: Xing (OpenBC)
 - ▶ Studenten: Facebook / StudiVZ
 - ▶ Generell: Friendster, orkut, MySpace...
 - ▶ Musik: Last.fm, MySpace
-
- ▶ riesiges Nutzerpotenzial: mindestens 41 Dienste mit mehr als 1 Mio. Nutzer, MySpace: 174 Mio.

Paralleluniversum soziale Netzwerke

- ▶ Traffic nicht mehr zwischen den Nutzern bzw. deren (selbst ausgesuchten Service Providern), sondern nur innerhalb der jeweiligen Plattform
- ▶ Verschlüsselung, MIX-Einsatz etc. nicht mehr möglich / sinnvoll
- ▶ Daten liegen in homogener Umgebung vor, teilweise sogar semantisch aufbereitet ⇒ Spidern
- ▶ Login erfolgt größtenteils im Klartext

Web 2.0 ⊖ soziale Netzwerke als Lösung?

- ▶ Vorteile:
 - ▶ Daten liegen beim Nutzer bzw. dem selbstgewählten Dienst
- ▶ Nachteile:
 - ▶ Verkettung immer noch möglich
 - ▶ Interoperabilität teilweise nicht gegeben
 - ▶ Nutzer konzentrieren sich auf populäre Dienste
 - ▶ ⇒ vorgenannten Probleme bestehen weiter

Mehrdimensionale Verkettung

- ▶ Zeit, Ort, Beziehungen, Interessen, Reputation
- ▶ unglücklicherweise genau das, was die Nutzer wollen
- ▶ Personen- bzw. Gruppenprofile möglich
 - ▶ personalisierte Werbung
 - ▶ Ansprechen von Peer-Groups (early adopters)
 - ▶ Bewegungsprofile (Einbruch, Verfolgung...)

Identitätsfetzen? - Profile!

- ▶ Automatisches Spidern durch Verknüpfung aller APIs möglich
- ▶ Suche nach Schlagwörtern: Bier, liberal, Datenschutz...
- ▶ Erinnerung: Durch Tagging und XML-Derivate erzeugen Nutzer selbst Semantik

Horrorszenario oder schon existent?

- ▶ friendster, AOL u. a.: Kündigungen wegen privaten Blogeinträgen
- ▶ StudiVZ: Mindestens drei unabhängige Komplettextraktionen der Daten, Stalking-Gruppen (~700 Mitglieder)
- ▶ MySpace: Mindestens 7 Mädchen (höchstens 11 Jahre) von Stalkern im realen Leben belästigt
- ▶ Subtile Werbung: Blogeinträge, Produkttests, Konzertberichte, Alternate Reality Games

Headhunter!

- ▶ Personalabteilungen nutzen Web-2.0-Daten
 - ▶ selbst
 - ▶ über Headhunting-Agenturen
- ▶ In den USA: 45% Google, 23% Soziale Netzwerke, 14% Blogs
- ▶ Rezeption von Partybildern, Gruppenmitgliedschaften, Pinnwandeinträgen?

Fazit

- ▶ Dienste haben selbst nur wenige Daten
- ▶ Vernetzung als Problem
- ▶ Kaum reglementierbar
- ▶ „Verantwortungsvoller“ Nutzer in der Pflicht

Fragen?

Gerne.