



Carsten Eilers | [www.ceilers-it.de](http://www.ceilers-it.de)

# Was weiß das Web über mich?

Verkettung digitaler Identitäten

# Vorstellung

- Berater für IT-Sicherheit
- Autor
  - About Security
  - Standpunkt Sicherheit
  - Buch „Ajax Security“
  - und anderes...
- Schwachstellen-Datenbank
  - „Security Aktuell“ auf [entwickler.de](http://entwickler.de)

# Agenda (1)

- Vorbemerkung
- (Digitale) Identität & Verkettung
- Anonym vs. Pseudonym
- Informationsquellen
- Personenbezug
- Übergang zur realen Welt
- Verkettung im Kontext
- Beispiel: Google

## Agenda (2)

- Was kann man wie verketteten?
- Beispiel: Communities
- Von Dritten geschaffene Identitäten
- Personensuchmaschinen
- Was weiß das Web über mich?
- Gegenmaßnahmen
- Fragen?

# Agenda

- **Vorbemerkung**
- (Digitale) Identität & Verkettung
- Anonym vs. Pseudonym
- Informationsquellen
- Personenbezug
- Übergang zur realen Welt
- Verkettung im Kontext
- Beispiel: Google
- ...

# Vorbemerkung

Wissen, was andere wissen können - wichtig wegen/für:

- Phishing / Whaling
- andere unschöne Sache
- Reputationsmanagement

# Agenda

- Vorbemerkung
- **(Digitale) Identität & Verkettung**
- Anonym vs. Pseudonym
- Informationsquellen
- Personenbezug
- Übergang zur realen Welt
- Verkettung im Kontext
- Beispiel: Google
- ...

# (Digitale) Identität

- Identität
  - die einen Menschen kennzeichnenden und von anderen Menschen unterscheidenden Eigenheiten (Attribute) seines Wesens
- Digitale Identität
  - technisch abgebildete Attribute eines Menschen

# Digitale Identität (1)

- Was für Daten?
  - Adressdaten
  - biometrische Daten
  - Vorlieben
  - ...
- Nicht alle Daten liegen von Anfang an digital vor, manche müssen digitalisiert werden

## Digitale Identität (2)

- Ausgehend von der Identität eines Menschen ist seine digitale Identität die Untermenge davon, die technisch abgebildet werden kann und auch abgebildet wird
- Werden nur Teile davon gespeichert, spricht man von Teilidentitäten

# Identität

technisch abbildbare Attribute

technisch abgebildete Attribute:  
Digitale Identität

digitale  
Teilidentitäten

eBay-  
Bewertungsprofil

Blog

Xing-Profil

Fingerabdruck  
im Reisepass

# Digitale Identität (4)

- zwei Ausprägungen:
  - selbst geschaffene, z.B. in Form von Pseudonymen, werden selbst fortentwickelt
  - von Dritten geschaffene und fortentwickelte, z.B. in Form von Kundenprofilen
- Daten können verfälscht sein (Hundefutter für die Katze vom Nachbarn)
- Verkettung je nach Standpunkt erwünscht oder unerwünscht

# Verkettung (1)

- Verkettung = In Beziehung setzen  
passiert ständig und ist Teil unseres Lebens
- Verkettung digitaler Daten  
Zusammenfassen und Verknüpfen
- Voraussetzung:  
Einheitliche Identifizierungsmerkmale

# Verkettung (2)

## Beispiel: Amazon

- „Was kaufen Kunden, die sich diesen Artikel angesehen haben?“
- „Inspiziert von Ihren Shopping-Trends“
- „Vervollständigen Sie Ihre Sammlung“

# Agenda

- Vorbemerkung
- (Digitale) Identität & Verkettung
- **Anonym vs. Pseudonym**
- Informationsquellen
- Personenbezug
- Übergang zur realen Welt
- Verkettung im Kontext
- Beispiel: Google
- ...

# Anonym vs. Pseudonym

- Anonym:  
Daten über eine Person können dieser Person von keinem Dritten zugeordnet werden
- Pseudonym:  
fiktiver Name als Alternative zum originalen Namen, kann bei Bedarf aufgedeckt werden

# Anonym (1)

Anonymität ist der Zustand, innerhalb einer Menge von Subjekten (Anonymitätsmenge) nicht identifizierbar zu sein

Anonymitätsgrad um so stärker, je größer die Anonymitätsmenge ist und je weniger sich das Verhalten der Mitglieder dieser Gruppe unterscheidet

# Anonym (2)

Rumpelstilzchen

möchte Anonym  
bleiben, posaunt  
aber seinen Namen  
in den Wald hinaus

nur einer holt morgen  
der Königin ihr Kind



# Pseudonym (1)

- Pseudonyme können anhand einer Zuordnungsregel den entsprechenden Personen zugeordnet werden
- Pseudonyme kommen u.a. dann zum Einsatz, wenn anonymes Agieren nicht möglich ist  
Bsp.: eBay - Pseudonym wird beim Kauf aufgedeckt

## Pseudonym (2)

„Der Komiker, der keinen Spaß versteht“  
wenn es um seinen realen Namen geht

Möchte nur unter seinem Pseudonym bekannt  
sein, registriert aber das Pseudonym beim  
DPMA auf seinen richtigen Namen

Was hat er davon? Wenn man nach  
Rumpelstilzchen-Bildern sucht

# Pseudonym (3)

Was hat er davon?

Wenn man bei Google nach Rumpelstilzchen-Bildern sucht, kommt ein Bild von ihm ohne seine Verkleidung

Treffer, versenkt!

# Agenda

- Vorbemerkung
- (Digitale) Identität & Verkettung
- Anonym vs. Pseudonym
- **Informationsquellen**
- Personenbezug
- Übergang zur realen Welt
- Verkettung im Kontext
- Beispiel: Google
- ...

# Informationsquellen (1)

- Informationsquellen sind vielfältig, z.B.
  - Äußerungen
  - Handlungen
  - Kontakte
- Informationen stammen sowohl vom Betroffenen als auch von Dritten

# Informationsquellen (2)

## Äußerungen

### Der Betroffene:

eigene Homepage, Blog, Profile,  
Forenbeiträge, ...

### Dritte:

Einschätzungen und Bewertungen des  
Betroffenen, ...

# Informationsquellen (3)

Handlungen

Der Betroffene:

Einkäufe, Fortbewegung, Teilnahme an  
Veranstaltungen,...

Dritte:

gekaufte Geschenke, Handlungen mit  
Vollmacht, ...

# Informationsquellen (4)

(Potenzielle) Kontakte

Der Betroffene:

Einträge Anderer im Adressbuch oder Social Network, besuchte Webseiten,...

Dritte:

durch den Betroffenen getätigte Kontakte in Social Networks, absichtlich gemiedene Kontakte,...

# Informationsquellen (5)

Bisher: Primäre Informationsquellen

Außerdem: Sekundäre Informationsquellen

- Äußerungen über die Gruppe, zu der der Betroffene gehört
- statistische Schlussfolgerungen
- indirekte Kontakte („Freund eines Freundes“)

# Informationsquellen (6)

## Außerdem: Kontextdaten

- Ort- und Zeitinformationen
- Datenspuren wie IP-Adressen, Cookies,...

# Agenda

- Vorbemerkung
- (Digitale) Identität & Verkettung
- Anonym vs. Pseudonym
- Informationsquellen
- **Personenbezug**
- Übergang zur realen Welt
- Verkettung im Kontext
- Beispiel: Google
- ...

# Personenbezogene Daten (1)

- Jede digitale Identität bezieht sich auf eine Person
- Aus Sicht eines Dritten nur, wenn die Person identifiziert werden kann

## Personenbezogene Daten (2)

- Ggf. ist ein Pseudonym bereits so aussagekräftig, das daraus auf die Person geschlossen werden kann
- Aus von AOL 2006 pseudonymisiert veröffentlichten Suchanfragen konnten Personen identifiziert werden

# Agenda

- Vorbemerkung
- (Digitale) Identität & Verkettung
- Anonym vs. Pseudonym
- Informationsquellen
- Personenbezug
- **Übergang zur realen Welt**
- Verkettung im Kontext
- Beispiel: Google
- ...

# Übergang zur realen Welt (1)

Auswirkungen der Nutzung der gesammelten Informationen:

- Biometrie liefert evtl. medizinische Informationen
- Nach Identitätsdiebstahl ist die Wiederherstellung des Ausgangszustands schwierig

# Übergang zur realen Welt (2)

Auswirkungen der Nutzung der gesammelten Informationen:

- Abhängig vom Surf- und Klickverhalten angepasste Preise im Webshop
- Scoring digitaler Daten entscheidet über Kreditwürdigkeit

# Übergang zur realen Welt (3)

Auswirkungen der Nutzung der gesammelten Informationen:

- Phishing / Whaling
- Reputation
  - gezielte Verunglimpfung
  - zufällige Namens-/Pseudonymgleichheit
- „Spionage“

# Agenda

- Vorbemerkung
- (Digitale) Identität & Verkettung
- Anonym vs. Pseudonym
- Informationsquellen
- Personenbezug
- Übergang zur realen Welt
- **Verkettung im Kontext**
- Beispiel: Google
- ...

# Verkettung im Kontext (1)

- Prinzipiell ist alles mit allem in Beziehung zu setzen
- Verkettung von Daten i.d.R. Zielorientiert, Ziel bestimmt anzuwendende Methoden

## Verkettung im Kontext (2)

- Verkettung über eindeutige Kennungen (Name, Pseudonyme, Adressen, ...)
- oder z.B. Raum- und Zeitkoordinaten (Bewegungsprofil erstellen, Bewegungsmuster erkennen)

# Agenda

- Vorbemerkung
- (Digitale) Identität & Verkettung
- Anonym vs. Pseudonym
- Informationsquellen
- Personenbezug
- Übergang zur realen Welt
- Verkettung im Kontext
- **Beispiel: Google**
- ...

# Beispiel: Google (1)

## Dienste und gesammelte Daten

- Suche - jedes eingegebene Suchwort
- Desktop - Index der Dateien des Benutzers einschließlich E-Mails, Multimedia-Dateien, Chat- und Browserhistory
- Groups - alle veröffentlichten Postings

# Beispiel: Google (2)

## Dienste und gesammelte Daten

- Talk - Chats zwischen den Benutzern
- Maps - alle Anfragen, bei Routenplanung ist meist der aktuelle Standort dabei
- Mail - alle E-Mails (bei entsprechender Einstellung)

## Beispiel: Google (3)

Dienste und gesammelte Daten

- Calendar - eingegebene Termine
- Orkut - Name, Adresse, Beziehungen
- Reader - gelesene ATOM-/RSS-Feeds
- Video und YouTube - angesehene Videos

# Beispiel: Google (4)

## Dienste und gesammelte Daten

- Checkout - Kreditkarten- und Zahlungsinformationen
- Health - Krankheitsgeschichte, sofern eingegeben
- Analytics - Besuche auf beteiligten Websites

## Beispiel: Google (5)

Google gehört DoubleClick:

- Cookies
- Welche Webseite wurde von welcher IP-Adresse aufgerufen

Verkettung liefert Surf- und Interessenprofil

Kann durch Verknüpfung mit Google-Daten verdichtet werden

## **Beispiel: Google (6)**

Beispiel: Nutzung als Suchmaschine,  
Mailprovider und IM-Server

Nutzung hinterlässt z.B. die folgenden  
Informationen:

## Beispiel: Google (7)

- 08:00 gmail.com  
Anmeldung mit dem Benutzernamen 'Joe'
- 08:01 gmail.com  
Abruf neuer E-Mails
- 08:24 google.de  
Eingabe des Suchbegriffs 'irgendwas'
- 08:25 google.de  
Eingabe des Suchbegriffs 'wasanderes'

## Beispiel: Google (8)

- 08:26 google.de  
Anklicken des ersten Suchergebnisses
- danach Surfen auf der angeklickten Seite  
wenn die DoubleClick nutzt, landen  
Informationen über den Besuch bei Google

## Beispiel: Google (9)

- 08:55 gmail.com  
Anmeldung mit dem Benutzernamen 'Joe'
- 08:56 gmail.com  
Schreiben einer E-Mail
- 09:01 gmail.com  
Senden einer E-Mail an 'Jim' mit Betreff 'wasweissich' und Inhalt 'Hallo Jim, ...'

# Beispiel: Google (10)

- 09:45 google.de  
Eingabe des Suchbegriffs 'wasweissich'
- 09:46 google.de  
Anklicken des dritten Suchergebnisses
- danach Surfen auf der angeklickten Seite  
wenn die DoubleClick nutzt, landen  
Informationen über den Besuch bei Google

# Beispiel: Google (11)

- 09:58 gmail.com  
Chat mit Benutzer 'Jack' mit dem Inhalt 'Hallo Jack, ...'
- 10:14 google-Calendar  
Eintragen eines Termins mit Person 'Bill' am 11.12. um 13:00 Uhr in A-Stadt
- 10:17 maps.google.de  
Routenplanung von B-Stadt nach A-Stadt

# Beispiel: Google (12)

Google weiß:

- wer der Benutzer ist ('Joe')
- das er sich für 'irgendwas', 'wasanderes' und 'wasweissich' interessiert
- an wen er welche E-Mail geschickt hat
- mit wem er gechattet hat und worum es dabei ging
- wann er sich wo mit wem trifft
- wo er vermutlich jetzt ist (B-Stadt)

# Beispiel: Google (13)

Alles nicht so wild, schließlich ist Google ja einer der Guten?

1. Wer immer wieder betont, das er ein Guter ist, hat das wohl nötig, oder?
2. Erinnerst sich noch jemand an AOLs Veröffentlichung von Suchanfragen im August 2006?

# Beispiel: AOL (1)

AOL:

- über 20 Millionen Suchanfragen
- von über 500.000 Benutzer
- aus 3 Monaten (März - Mai 2006)
  
- 440 MByte

## Beispiel: AOL (2)

- Benutzernamen durch Ziffernkombinationen ersetzt
- Trotzdem Identifikation einzelner Benutzer aufgrund der Suchanfragen möglich

## Beispiel: AOL (3)

- 17556639 how to kill your wife
- 17556639 how to kill your wife
- 17556639 wife killer
- 17556639 how to kill a wife
- 17556639 poop
- 17556639 dead people
- 17556639 pictures of dead people
- 17556639 killed people

## Beispiel: AOL (4)

- 17556639 dead pictures
- 17556639 dead pictures
- 17556639 dead pictures
- 17556639 murder photo
- 17556639 steak and cheese
- 17556639 photo of death
- 17556639 photo of death
- 17556639 death

## Beispiel: AOL (5)

- 17556639 dead people photos
- 17556639 photo of dead people
- 17556639 [www.murderdpeople.com](http://www.murderdpeople.com)
- 17556639 decapitated photos
- 17556639 decapitated photos
- 17556639 car crashes3
- 17556639 car crashes3
- 17556639 car crash photo

# Beispiel: AOL (6)

Was wird das?

- Mordplan gegen die Ehefrau?
- Material für einen neuen Roman?
- Profiler bei der Suche nach einem Mörder?
- Ganz was anderes?

# Agenda

- ...
- **Was kann man wie verketteten?**
- Beispiel: Communities
- Von Dritten geschaffene Identitäten
- Personensuchmaschinen
- Was weiß das Web über mich?
- Gegenmaßnahmen
- Fragen?

# Was kann man wie verketten? (1)

- Verketten ist immer dann möglich, wenn es einheitliche Identifizierungsmerkmale gibt:
  - (Benutzer-)Name, Pseudonym
  - Benutzerkennung
  - IP-Adresse
- Betreiber kann den Benutzer identifizieren
- ISP kann über IP-Adresse den Anschluss und evtl. den Benutzer identifizieren

# Was kann man wie verketteten? (2)

Beispiel: Usenet

Enthält der Header eine eindeutige Benutzer-ID, können alle Postings dem entsprechenden Benutzer zugeordnet werden

Die ggf. als From: eingetragenen Pseudonyme werden damit verkettet

# Was kann man wie verketteten? (3)

Beispiel: Yahoo! Groups

Der Header enthält die Ursprungs-IP-Adresse

Verwendet ein Benutzer mehrere IDs und sendet mehrere Nachrichten nacheinander, können sie über die IP-Adresse verkettet werden (Proxies etc. außen vor gelassen)

# Was kann man wie verketten? (4)

Beispiel: Kommentare unter Blogs etc.

Immer, wenn ein einheitlicher Name oder z.B. die eigene Homepage angegeben wird, sind die Einträge untereinander und/oder mit einer Person verkettbar

# Was kann man wie verketteten? (5)

Beispiel: (Social) Bookmarks

Enthalten u.U. Namen, IDs, ... über die der Benutzer identifiziert und/oder verkettet werden kann

# Agenda

- ...
- Was kann man wie verketteten?
- **Beispiel: Communities**
- Von Dritten geschaffene Identitäten
- Personensuchmaschinen
- Was weiß das Web über mich?
- Gegenmaßnahmen
- Fragen?

# Beispiel: Communities (1)

Communities basieren auf der Verknüpfung ihrer Benutzer untereinander, diese ist erwünscht und wird durch die Betreiber gefördert

Identifikation der realen Personen teilweise erwünscht

## **Beispiel: Communities (2)**

evtl. Preisgabe der Identität eines Benutzers  
durch andere Benutzer, auch gegen dessen  
Willen

## Beispiel: Communities (3)

- Flickr:  
z.B. Nachverfolgen von Reisen,  
Ermitteln des aktuellen Aufenthaltsorts
- Social Networks:  
Beziehungen zwischen den einzelnen  
Benutzern

# Beispiel: Communities (4)

## Verkettung **innerhalb** der Communities

- Erweiterung der Angebote um immer mehr eigentlich extern bereits vorliegende Dienste liefert immer mehr Daten

Beispiel: Statt E-Mails interne Nachrichten zwischen den Mitgliedern

# Beispiel: Communities (4)

## Verkettung über mehrere Communities

- Verwendung gleicher Pseudonyme in mehreren Communities erleichtert das Wiedererkennen, aber auch die Verkettung
- Querverweise zwischen den Diensten verknüpfen auch die jeweiligen Digitalen Identitäten

# Beispiel: Communities (5)

## Verkettung **über mehrere** Communities

- Absichtliche oder unabsichtliche Verwendung des gleichen Pseudonyms führt zu fehlerhaften Ergebnissen bei der Verkettung dieser Pseudonyme

# Agenda

- ...
- Was kann man wie verketten?
- Beispiel: Communities
- **Von Dritten geschaffene Identitäten**
- Personensuchmaschinen
- Was weiß das Web über mich?
- Gegenmaßnahmen
- Fragen?

# Von Dritten geschaffene Identitäten (1)

Selbst erzeugte digitale Identitäten hat man unter Kontrolle, nicht aber ihre Verkettung mit anderen digitalen Identitäten

Das größere Problem: Von Dritten geschaffene digitale Identitäten, z.B. in Form von Blog-Einträgen oder Forenbeiträgen, die man nicht kontrollieren kann

# Von Dritten geschaffene Identitäten (2)

## Beispiele:

- negative Blogeinträge, Forenbeiträge, ...
- Fakes:
  - Steve Jobs blogt:  
<http://fakesteve.blogspot.com/>
  - Wolfgang Schäuble twittert:  
<http://twitter.com/schaeuble>
  - Sogar Gott hat eine Zeitlang gebloggt:  
<http://gottsblog.de/blog/>

# Von Dritten geschaffene Identitäten (3)

Folgen:

- Schaden am eigenen Ruf
- Verfälschung durch Verkettung erstellter Ergebnisse/Profile
- falsche Schlussfolgerungen führen zu falschen Entscheidungen

# Agenda

- ...
- Was kann man wie verketten?
- Beispiel: Communities
- Von Dritten geschaffene Identitäten
- **Personensuchmaschinen**
- Was weiß das Web über mich?
- Gegenmaßnahmen
- Fragen?

# Personensuchmaschinen (1)

- Verketteten Fundstellen zu einer Person zu einem Profil
- Erlauben teilweise, dieses zu bearbeiten
- Auswertung von normalen Suchmaschinen-Ergebnissen sowie Social Networks, Wikipedia, Nachrichtenartikeln, ...

# Personensuchmaschinen (2)

- Spock.com  
<http://www.spock.com/>
- 123people.com  
<http://www.123people.com/>
- Yasni  
<http://www.yasni.de/>

# Agenda

- ...
- Was kann man wie verketten?
- Beispiel: Communities
- Von Dritten geschaffene Identitäten
- Personensuchmaschinen
- **Was weiß das Web über mich?**
- Gegenmaßnahmen
- Fragen?

# Was weiß das Web über mich? (1)

- Usenet seit 1992
  - mit 'X-No-archive: yes'
  - alte Artikel weitgehend gelöscht
- diverse Mailinglisten
- Homepage seit 1994/2001/2004
- About Security seit 2005
- Standpunkt Sicherheit seit 2006

# Was weiß das Web über mich? (2)

Außer den von mir veröffentlichten Informationen findet man mindestens:

- mein Geburtsdatum
- einige Bilder
- meine eBay-Bewertungen

# Was weiß das Web über mich? (3)

Wäre ich im Web 2.0 aktiv, wäre es deutlich mehr!

# Agenda

- ...
- Was kann man wie verketten?
- Beispiel: Communities
- Von Dritten geschaffene Identitäten
- Personensuchmaschinen
- Was weiß das Web über mich?
- **Gegenmaßnahmen**
- Fragen?

# Gegenmaßnahmen allgemein

- nicht auf die Profile der Personensuchmaschinen vertrauen - wer Negatives finden will, nutzt die gar nicht erst, da sie ja geschönt sind
- Reputationsmanagement wird m.E. überbewertet - das Internet vergisst nichts

# Gegenmaßnahmen allgemein

Verkettung ist nicht möglich, wenn es kein eindeutiges Identifizierungsmerkmal gibt

Also: Getrennte digitale Teilidentitäten dürfen keine Gemeinsamkeiten haben

## **Gegenmaßnahmen: Foren etc.**

- Pseudonyme geschickt wählen
- nicht die eigene Domain angeben
- keine Mailadressen nach dem Muster  
vorname-nachname-ort@provider

# Gegenmaßnahmen: Sicher Suchen

- IP-Adresse über Anonymisierungsdienst verstecken
  - Proxy
  - AN.ON
  - Tor
- Nicht alle Cookies annehmen  
Schützt nicht vor Webbugs
- „Sichere“ Suchmaschine wählen

## Gegenmaßnahmen (4)

- Man kann leicht beweisen, das man einen Text selbst verfasst hat (indem man ihn signiert), aber nur schwer, das man ihn nicht verfasst hat

# Reputationsmanagement (1)

Die meisten Tips laufen darauf hinaus, selbst aktiv zu werden:

- Bloggen,
- Social Networks nutzen,
- Beziehungen durch Kommentare und Forenbeiträge aufbauen

Kurz: Gezielt eine Reputation aufbauen

# Reputationsmanagement (2)

Ist „aktiv werden“ eine gute Lösung?

Zumindest eine zweischneidige:

- „Tue Gutes und rede darüber“ funktioniert
- liefert aber auch Material für Verkettungen
  
- Was einmal im Netz ist, bleibt im Netz!

# Fragen?

**Vielen Dank...**

... für Ihre Aufmerksamkeit

Material und Links auf  
[www.ceilers-it.de/konferenzen/](http://www.ceilers-it.de/konferenzen/)

