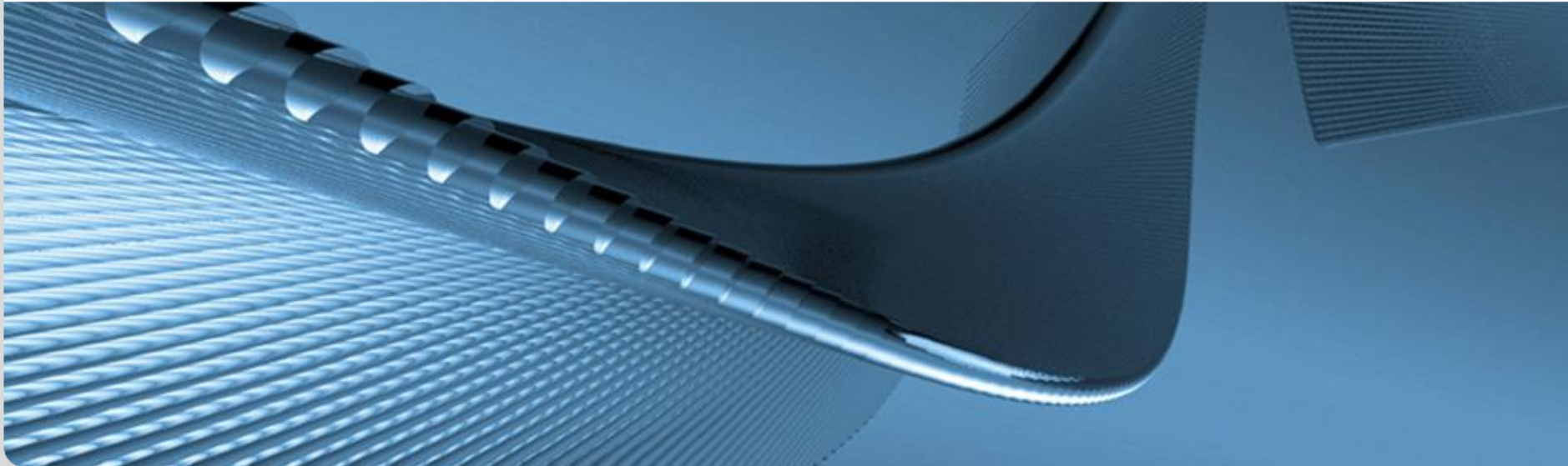


Datenschutz und Privatheit in vernetzten Informationssystemen

Kapitel 5: Datenschutz im Internet - Anwendungsebene

Erik Buchmann (buchmann@kit.edu)

IPD, Systeme der Informationsverwaltung, Nachwuchsgruppe „Privacy Awareness in Information Systems“



Inhalte und Lernziele dieses Kapitels

- Funktionsweise des WWW
 - Was erfährt ein Suchmaschinenbetreiber?
- Datenschutz auf Anwendungsebene
 - Web Bugs
 - Cookies
 - P3P
- Abschluss

- Lernziele
 - Sie können die Funktionsweise des Internets und des WWW beschreiben und aufzeigen, welche persönlichen Daten dabei wo und mit welchen Verfahren gesammelt werden.
 - Sie können P3P und dessen Stärken und Schwächen erklären.

Einführung: das WWW

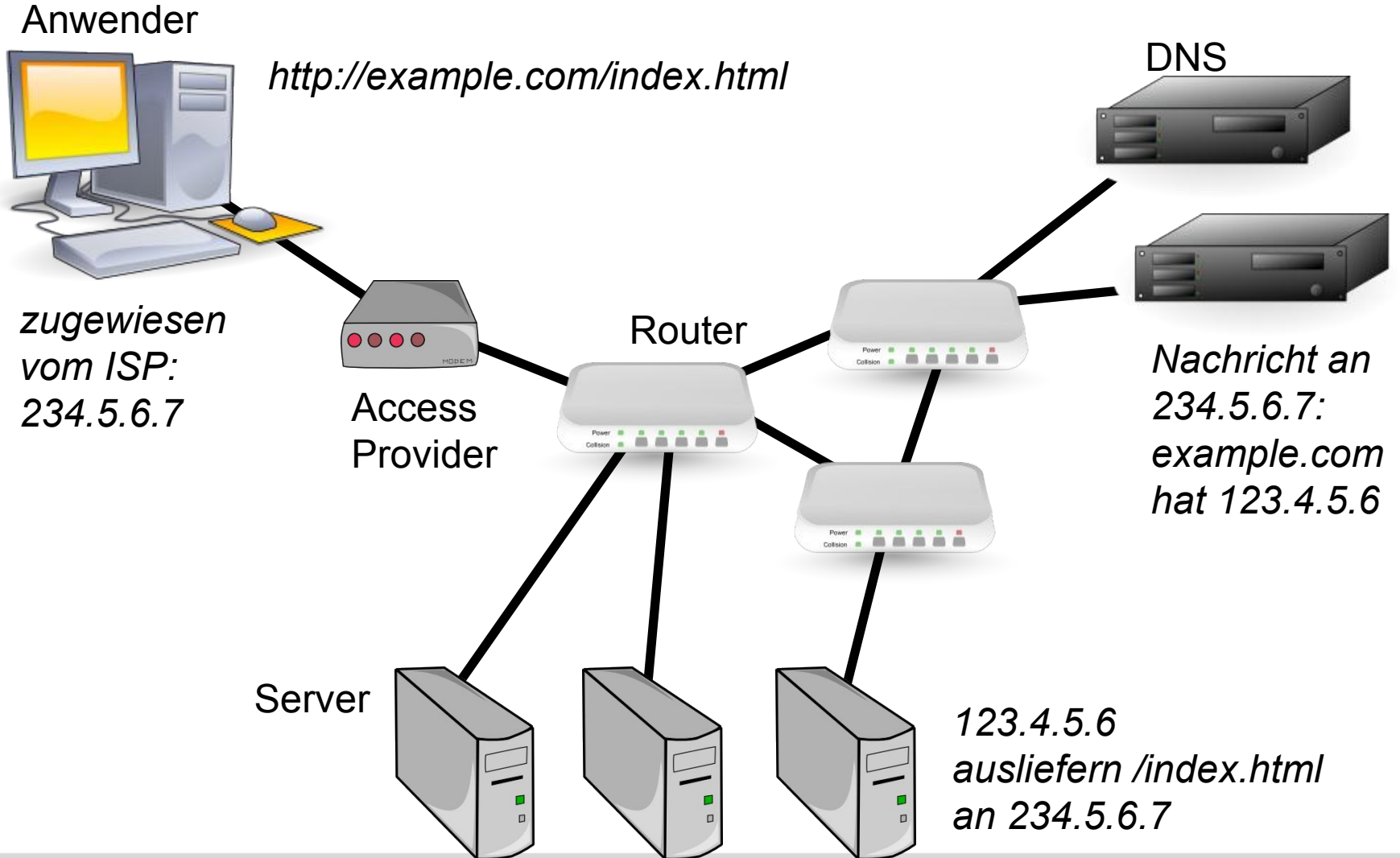
IPD, Systeme der Informationsverwaltung, Nachwuchsgruppe „Privacy Awareness in Information Systems“



Relevante Protokolle und Dienste

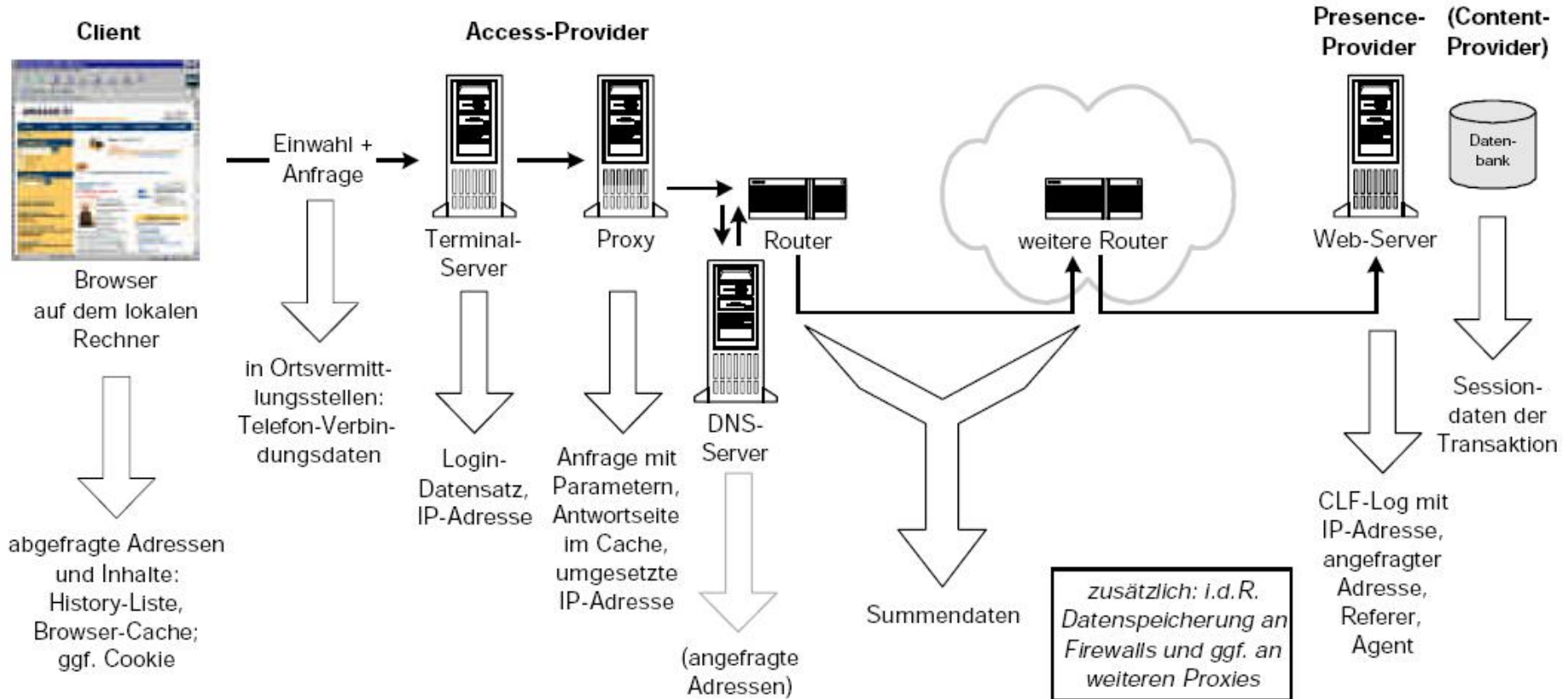
- DNS (Domain Name System)
 - Welche IP-Adresse hat Server für www.domain.de?
- TCP/IP
 - Verbindung zwischen lokalem Rechner und Server
- HTTP/HTTPS
 - Webseiten, Medien, Formulardaten etc. holen/senden
 - Cookie-Informationen austauschen
- HTML, CSS
 - WWW-Seiten selbst darstellen
 - Redirects, Daten von weiteren Servern anfordern
- Browser
 - HTML und Scripte interpretieren
 - Plugins von Dritten ausführen, z.B. Flash, Java Appletts
 - Externe Drittprogramme starten

Webseitenabruf im Internet



Datenspuren beim Webseitenabruf

World Wide Web: Abruf



Quelle: Marit Köhntopp, Kristian Köhntopp: *Datenspuren im Internet*; Computer und Recht 4/2000

Anfrage an den Webserver

```
/home/buchmann> telnet dbis.ipd.uni-karlsruhe.de 80  
Trying 129.13.182.137...  
Connected to dbis.ipd.uni-karlsruhe.de.  
Escape character is '^]'.  
  
GET / HTTP/1.0  
Referer: http://localhost/index.html  
User-Agent: Mozilla/5.0 (X11; U; Linux i686; en-US; rv:1.8.1.1) Firefox/2.0.0.1  
Host: localhost  
Accept: text/html, image/gif, image/jpeg, image/pjpeg, */*  
Accept-Language: en, de
```

Anfrage

Eingaben sind **fett**
2x Zeilenumbruch
ist wichtig!

(hier folgt jetzt die Antwort vom Webserver)

Antwort vom Webserver

```
/home/buchmann> telnet dbis.ipd.uni-karlsruhe.de 80
```

```
Trying 129.13.182.137...
```

```
Connected to dbis.ipd.uni-karlsruhe.de.
```

```
Escape character is '^]'.  
GET / HTTP/1.0
```

Eingaben sind **fett**
2x Zeilenumbruch
ist wichtig!

Antwort

```
HTTP/1.1 200 OK
```

```
Date: 11 April 2014 08:46:39 GMT
```

```
Server: Apache/2.2.8 (Unix) DAV/2 mod_ssl/2.2.8 OpenSSL/0.9.8h PHP/5.2.6
```

```
X-Powered-By: PHP/5.2.6
```

```
Set-Cookie: PHPSESSID=74e6a1911499e9578534f31814c357d4; path=/  
Expires: 19 Nov 2014 08:52:00 GMT
```

```
Cache-Control: no-store, no-cache, must-revalidate, post-check=0, pre-check=0
```

```
Pragma: no-cache
```

```
Connection: close
```

```
Content-Type: text/html
```

```
<!DOCTYPE HTML PUBLIC "-//W3C//DTD HTML 4.01 Transitional//EN" <...>
```

```
</html>Connection closed by foreign host.
```


Was loggt der Webserver?

- Direkt ablesbar:
 - Wo bin ich, wer bin ich? (IP-Adresse)
 - Was will ich? (URL)
 - Wo komme ich her? (Referrer, nicht im Beispiel enth.)
 - Wann habe ich die Seite abgerufen?
 - Welche Sprache spreche ich?
 - Welche Systemsoftware setze ich ein?

→ **Ungewöhnliche Kombinationen sind Quasi-Identifizier!**

```
/home/buchmann# tail -1 /var/log/apache2/access_log  
123.4.5.6 - - [11/May/2014:11:01:42 +0200]  
"GET / HTTP/1.0" 200 234  
"http://dbis.ipd.uni-karlsruhe.de/index.html"  
"Mozilla/5.0 (X11; U; Linux i686; en-US; rv:1.8.1.1) Gecko/20121208  
Firefox/2.0.0.1"
```

Was verrät mein Browser?

■ <https://panopticlick.eff.org>



A research project of the **Electronic Frontier Foundation**

Panopticlick

How Unique – and Trackable – Is Your Browser?

Is your browser configuration rare or unique? If so, web sites may be able to track you, *even if you limit or disable cookies.*

Panopticlick tests your browser to see how unique it is based on the **information** it will share with sites it visits. Click below and you will be given a uniqueness score, letting you see how easily identifiable you might be as you surf the web.

Only **anonymous data** will be collected by this site.



Test: Opera, SuSE Linux

browser characteristic	bits of identifying information	one in x browsers have this value	value
User Agent	13.41	10913.75	Opera/9.80 (X11; Linux x86_64) Presto/2.12.388 Version/12.14
HTTP_ACCEPT	8.58	382.27	text/html, */* gzip, deflate en-US,en;q=0.9
Browser Plugins	21.45+	2859402	Plugin 0: DivX Web Player; DivX Web Player version 1.4.0.233; libtotem-mully-plugin.so; (; video/divx; divx). Plugin 1: Gnome Shell Integration; [...]
Time Zone	2.85	7.22	-120
Screen Size and Color Depth	7.1	136.97	1600x1200x24
System Fonts	21.45+	2859402	LM Mono Slanted 10, GFS Baskerville, FreeMono, LM Mono 10, LM Mono 12, Droid Sans, [...]
Are Cookies Enabled?	0.42	1.34	Yes
Limited supercookie test	0.99	1.99	DOM localStorage: Yes, DOM sessionStorage: Yes, IE userData: No

Was verrät die IP-Adresse?

GEO IP TOOL

<http://www.geoipool.com/en/?IP=www.ira.uka.de>

language:    

[View my IP information](#)

[More info about IPs](#)


[Firefox Plugin](#)

[Now online](#)

[In your Website](#)

New tool for your Web!


[Karte](#) [Satellit](#) [Hybrid](#)



Host / IP: [View info](#)

Host Name: **irafs1.ira.uni-karlsruhe.de**

IP Address: **141.3.10.100**

Country: **Germany** 

Country code: **DE (DEU)**

Region: **Baden-Württemberg**

City: **Karlsruhe**

Postal code:

Calling code: **+49**

Longitude: **8.3858**

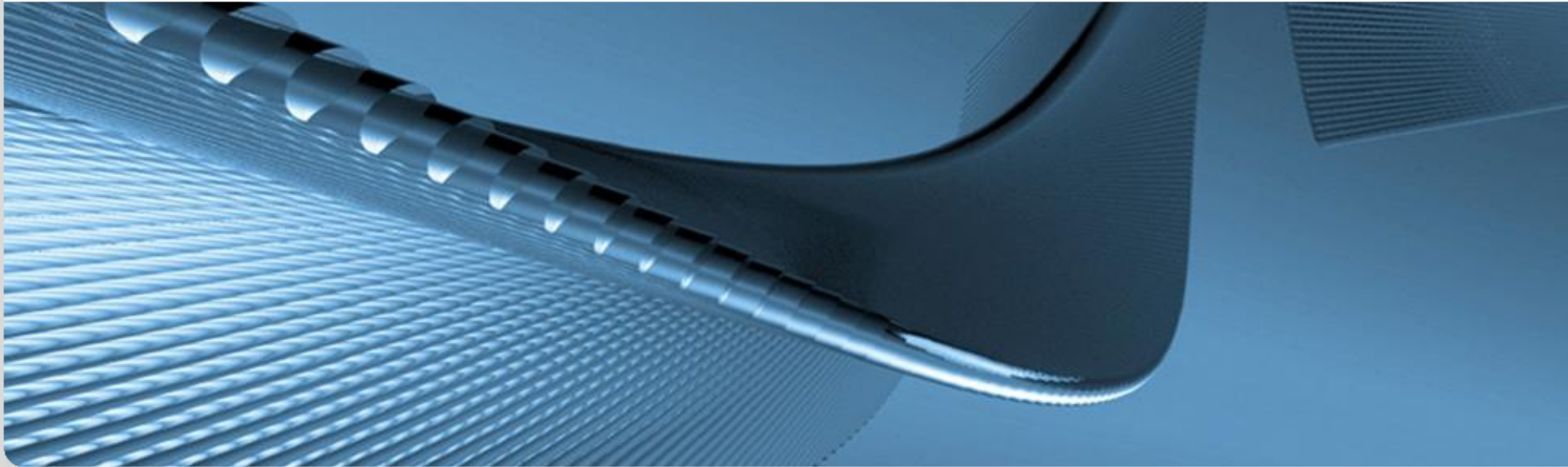
Latitude: **49.0047**

POWERED BY Google

Multimediale Kartendaten ©2007 TeleAtlas - Nutzungsbedingungen

Was erfährt ein Suchmaschinenbetreiber?

IPD, Systeme der Informationsverwaltung, Nachwuchsgruppe „Privacy Awareness in Information Systems“



Suchmaschinen: Google, AOL, MSN und Co.

- Suche übermittelt an Suchmaschinenbetreiber:

Kategorie	Attribute
Suchterme (technisch unvermeidlich)	Aneinanderreihung von Schlüsselworten
Browser-Kommunikation (schwer vermeidbar)	Zeitstempel, IP, Browser, Betriebssystem, Spracheinstellungen, zuletzt besuchte Seite
Browser-Kommunikation (vermeidbar)	ausgewähltes Suchergebnis (Implementiert als Redirect)
Zusatzinformationen (vom Nutzer steuerbar)	Cookie-Informationen, Session-ID
	über JavaScript ermittelte Daten, Verweildauer auf der Seite, Bildschirmauflösung
	Nutzer-ID (z.B. Google-Login, Yahoo-ID)

Mögliche Verkettung dieser Attribute zur Profilbildung

Suchanfragen zu Profil verknüpfen

- Über die **Browser-Kommunikation**, insbes. IP-Adresse
 - dynamische IP-Adressen: überdauern mindestens eine Such-Session
 - Kombination aus Standort des IP-Adressbereichs, Betriebssystem, Sprache, Browser etc. können als Quasi-Identifizierer ausreichen
 - statische IP-Adressen, z.B. Uni-Netz: länger gültig
- Über **Cookies, Session-IDs, Nutzer-Login**
 - Identifiziert einen Browser (und damit oft dessen Benutzer) über lange Zeiträume eindeutig
 - auch bei wechselnder IP-Adresse
- Über die **Suchterme**
 - z.B. Suche nach eigenem Namen, seltene Hobbies
- **Kombinationen** aus allem

Zusatzdienste ins Profil integrieren

- Suchmaschinenanbieter oft Anbieter weiterer Dienste
 - Google: Youtube, Maps, Email, Verzeichnisdienste, Google Docs, Google Earth, News, Usenet
- Separate „digitale Identitäten“ werden verkettbar
 - über IP-Adresse Nutzerbewegungen über mehrere Dienste hinweg nachvollziehbar
 - oft übergreifendes Login für viele Dienste,
 - Microsoft Passport, Windows Live ID
 - Google Authentication for Web Applications (OAuth)
- Informationen über viele Lebensbereiche
 - Arbeit, Privatleben, Hobbies, Kommunikationspartner etc.

Haben Suchanfragen Personenbezug?

- kommt ganz drauf an...
 - zur Erinnerung: “...*Einzelangaben über persönliche oder sachliche Verhältnisse einer bestimmten oder bestimmbaren natürlichen Person.*”
 - Heute übliche Interpretation: kein Personenbezug, wenn “*für Einzelangaben zu einer Person die Wahrscheinlichkeit, dass diese der Person zugeordnet werden können, so gering ist, dass sie nach der Lebenserfahrung oder dem Stand der Wissenschaft praktisch ausscheidet.*”

Quelle: Roßnagel, A.; Scholz, P.: Datenschutz durch Anonymität und Pseudonymität, MMR 2000

Personenbezug von Suchbegriffen

- auch hier: kommt drauf an...
 - “Katzenfutter billig” → nein;
 - “Erik Buchmann Urlaub Italien” → ja
- Personenbezug ist *abhängig von den Benutzereingaben*; nicht ohne weiteres automatisch vom Betreiber entscheidbar!
 - je umfangreicher die Suchhistorie eines Nutzers, desto wahrscheinlicher kommt eine identifizierende Kombination von Identitätsattributen zusammen
(*Beispiel folgt*)

- Dynamische IP-Adresse
 - eher nicht personenbezogen
 - *Personenbezug erfordert Mithilfe des Kenners der Zuordnungsregel, d.h., Internet Service Provider*

- Statische IP-Adresse für festen Rechner
 - Suchmaschinenbetreiber kann Suchanfragen über lange Zeiträume einer Person zuordnen
 - mit zunehmender Zahl der Suchvorgänge steigt Wahrscheinlichkeit, dass sich der Suchende offenbart
(vgl. vorangegangene Folie)
 - daher: oftmals personenbezogen

Personenbezug von Zusatzinformationen

- Bildschirmauflösung, Sprache, Verweildauer auf der Seite
 - nicht als identifizierendes Merkmal geeignet
- Nutzer-Login, Session-ID, Cookie-Informationen
 - Betreiber hat Login, Session-ID bzw. Cookie selbst dem Nutzer zugeordnet
 - daher: oftmals personenbezogen, da *Suchmaschinenbetreiber gleichzeitig Kenner der Zuordnungsregel*

Aug. 2006: AOL-Datenleck (1/2)

- AOL Research veröffentlicht für Forschung 20 Mio. Suchanfragen von 650.000 Usern, gesammelt über 3 Monate
 - IDs pseudonymisiert, künstl. Schlüssel



User ID	Search Keywords	Date	Website
4417749	numb fingers	2006-03-06 18:35:02	http://irgendwas.de
...

- zwar ist keiner der Datensätze unmittelbar personenbezogen, aber schnell werden einzelne Identitäten und komplette Persönlichkeitsprofile offenbar
- 3 Tage später: Datenbank ist vom Netz, aber schon vielfach in Tauschbörsen kopiert, bis heute verfügbar

Aug. 2006: AOL-Datenleck (2/2)

- Nutzer 4417749: hunderte Suchanfragen in 3 Monaten
 - “dog urinates on everything”: *Hundebesitzer*
 - “60 single man”: *einsame ältere Frau*
 - “numb fingers”: *körperliche Gebrechen*
 - “homes sold in gwinnett county”: *Wohnung*
 - “xxx Arnold, yyy Arnold”: *suche nach Verwandten*
 - “school supplies for Iraq children”: *karitativ*
 - “best season to visit Italy”: *nächster Urlaub*

- identifiziert als Thelma Arnold:
“*My goodness, its my whole personal life!*”
 - Privat- und Alltagsleben, Ängste, Gebrechen
 - falsches Selbstbild durch Suchanfragen für Freunde

Quelle: <http://www.nytimes.com/2006/08/09/technology/09aol.html>

Wie handhaben es die Betreiber?

■ Google

- nutzt Cookies, verknüpft Daten der Suchmaschine und aller hauseigenen Dienste (Google Docs, Gmail, Google+, Google Analytics)
- Angabe 2011: nach 9 Monaten werden Suchlogs anonymisiert (was immer das heißt – im Moment ist die Angabe nicht zu finden)

■ Microsoft, AOL, Yahoo

- vergleichbare Praxis der Datensammlung
- Speicherdauer Yahoo, AOL: 6 Monate
- Speicherdauer Microsoft: *“After 18 months, we will completely anonymize all Search queries [...] by irreversibly removing all cross-session identifiers [...] including the full IP address and all cookie IDs.”*
→ schützt nicht vor persönlichen Suchtermen

Alternativen existieren

- Es gibt Suchmaschinen, die keine IP-Adressen oder Suchangaben protokollieren



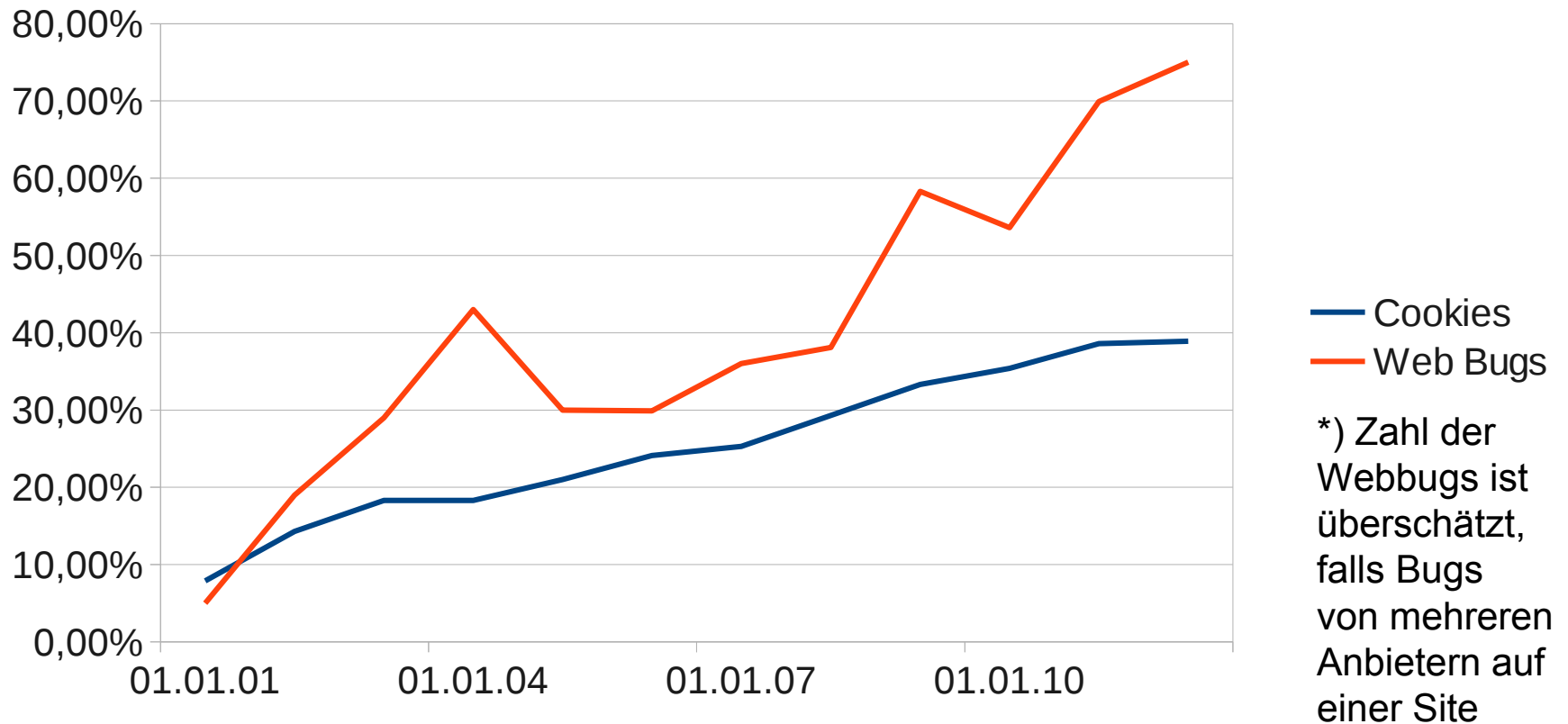
Mehr über den Nutzer lernen: Cookies und Web-Bugs

IPD, Systeme der Informationsverwaltung, Nachwuchsgruppe „Privacy Awareness in Information Systems“



- Wenn nur das technisch Notwendige durchgeführt wird
 - Nutzer ist quasi-anonym
 - dynamische IP-Adressen:
Änderung nach jedem neuen Einwählen beim ISP
 - statische IP-Adressen: können zu Proxies, Hubs, Firewalls oder Gateways gehören; nicht zwingend einem Einzelnen zugeordnet
 - Datenspuren sind verteilt
 - Jeder Webseitenzugriff hinterlässt Datenspuren nur auf dem jeweiligen Webserver
 - keine Verfolgung des Nutzers über mehrere Sites
- Aufbau von umfangreichen Nutzerprofilen kaum möglich

Anteil der Webseiten mit Cookies/Web Bugs

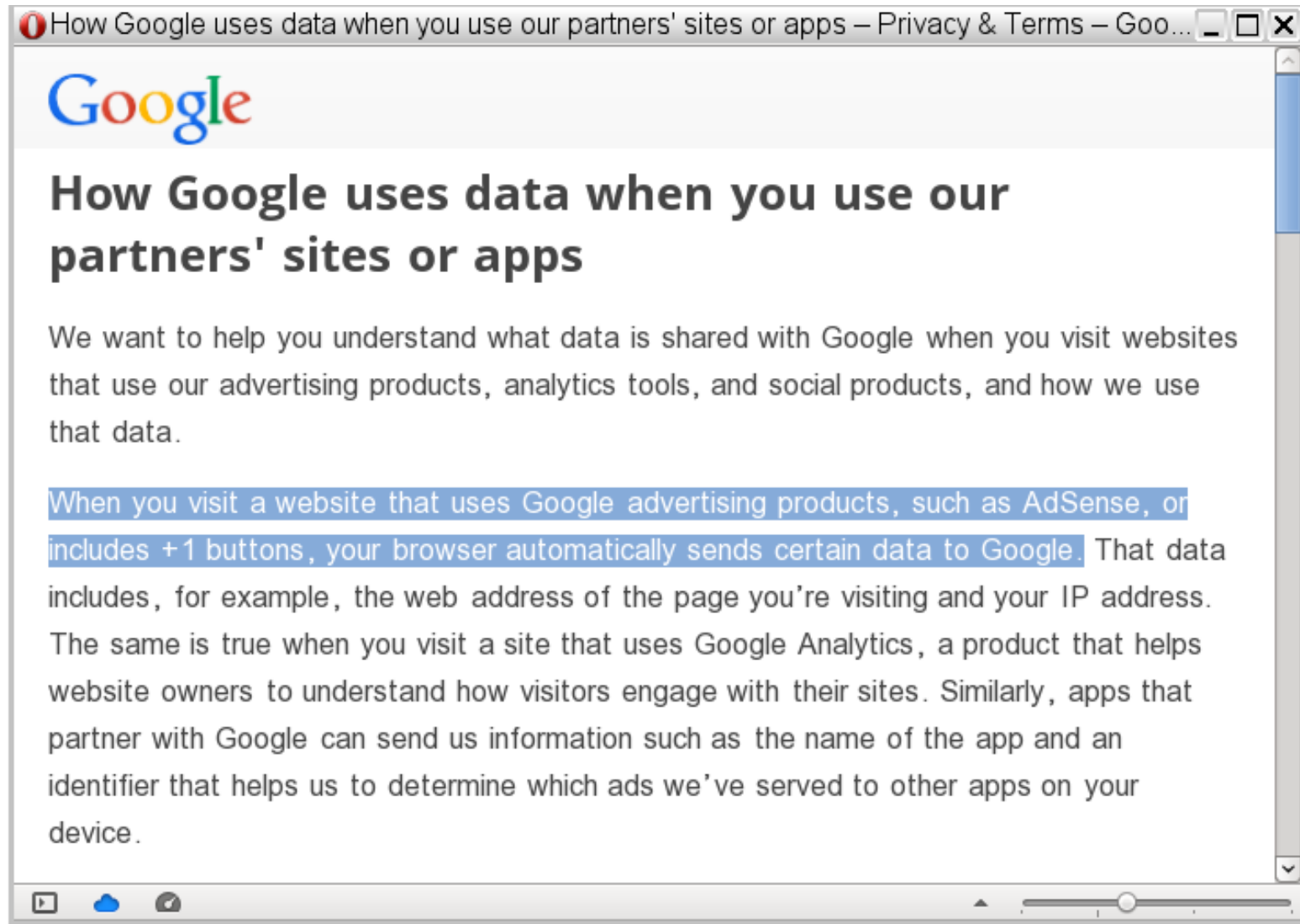


<http://www.securityspace.com>

- Ziel: Überwachung des Nutzers, Nachvollziehen seiner Bewegungen (Clickstream-Analyse)
 - auf einer einzelnen Webseite
 - über mehrere Webseiten hinweg
- Idee:
 - Browser ruft **präparierte Datenobjekte** auf verschiedenen Servern ab, und
 - hinterlässt dort **Spuren im Log**
- Methode:
 - Verweise auf Datenobjekte werden in HTML, EMails, PDFs etc. so eingebunden, dass sie der Betrachter automatisch nachlädt
 - dynamisch generierte Namen, damit Web-Bugs nicht aus dem Browsercache geladen werden

- Viele Webseitenbetreiber nehmen bei einem Analysedienst teil
 - jeder Webseitenbetreiber bindet den Web-Bug des Dienstes ein, z.B. ein unsichtbares Iframe oder einen Werbebanner
- Webseitenzugriff durch den Benutzer
 - Browser lädt Webseite
→ *Eintrag im Log des Betreibers*
 - Browser lädt iframe vom zentralen Server des Analysedienstes → *Eintrag im Log des Dienstes*
- Log-Analyse beim Dienst
 - Daten über Zugriffe auf unterschiedliche Webseiten *von unterschiedlichen Anbietern zentral in einem Log*

Beispiel: Google



■ Bilder

```

```

■ Frames, IFrames

```
<frame src="http://spy.com/verifyme.cgi?id=X">
```

■ Scripts

```
<script src="http://spy.com/verifyme.cgi?id=X" type="text/javascript"></script>
```

■ Styles

```
<link rel="stylesheet" media="screen" href="http://spy.com/verifyme.cgi?id=X">
```

■ Layer

```
<layer top="80" left="40" src="http://spy.com/verifyme.cgi?id=X">
```

■ und noch einige mehr...

Daten zu Web Bugs

■ Top-10 der Web Bug-Verwender:

Domain	Sites
google-analytics.com	176668 (6.7%)
googlesyndication.com	166945 (6.4%)
google.com	112760 (4.3%)
youtube.com	71912 (2.7%)
fc2.com	35320 (1.3%)
yadro.ru	32341 (1.2%)
addthis.com	30717 (1.2%)
statcounter.com	29845 (1.1%)
rambler.ru	21905 (0.8%)
cnzz.com	20824 (0.8%)

Stand: 1.5.2010,
<http://www.securityspace.com>
untersucht: 35.275.490 Seiten

Domain	Sites
googlesyndication.com	179940 (6.6%)
google.com	148664 (5.4%)
google-analytics.com	139681 (5.1%)
youtube.com	106370 (3.9%)
facebook.com	79746 (2.9%)
googleapis.com	63848 (2.3%)
addthis.com	48514 (1.8%)
fc2.com	38954 (1.4%)
cnzz.com	34090 (1.2%)
yadro.ru	33345 (1.2%)

Stand: 1.5.2012,
<http://www.securityspace.com>
untersucht: 41.618.980 Seiten

Für Web-Bugs nutzbare Software

- Alles, was **automatisch** Inhalte aus dem Web nachlädt
 - Web-Browser, Browser-Plugins
 - Microsoft Office, OpenOffice
 - Mail-Clients (weniger anfällig; hier haben viele Hersteller reagiert)
 - PDF-Dateien, Windows-Hilfedateien
 - sämtliche Produkte mit automatischem Update
 - und viele mehr...

- Web-Bugs vs. nützlichen Anwendungen?
Beispiele für Anwendungen der Technik sind:
 - kleinere E-Mails, Grafiken bei Bedarf vom Server
 - Daten in Spreadsheets automatisch aktualisieren
 - Rechtemanagement für geschützte Inhalte über Authentifizierungsmechanismen des Webservers
- kein wirkungsvoller Schutz möglich, aber Teillösungen:
 - Nachladen von allen Inhalten aus dem Internet unterdrücken (Mailclients)
 - zusätzliche Inhalte nur aus der Domain des Ursprungsdokuments nachladen (Web-Browser)

- Daten, die ein Webserver auf dem Rechner des Anwenders speichern und jederzeit wieder abrufen darf
 - beliebige kurze Zeichenketten
 - rudimentäre Sicherheitsfeatures
 - Verfallsdatum

- Webbrowser darf Cookies jederzeit löschen

- Cookie-Arten im Webbrowser
 - Cookie im Webbrowser
 - Supercookie (DOM local/session Storage)
 - Internet Explorer User Data
 - Flash Cookie

Typische Anwendungen für Cookies

- Viele sinnvolle Anwendungen
 - Single-Signon auf Webseiten
 - Speichern von Benutzereinstellungen im Browser des Anwenders → *Werkzeug gegen Profiling!*
 - Vormerken von Artikeln in Web-Shops

- Aber auch: viele problematische Anwendungen
 - Wiedererkennen eines Nutzers
 - Nachverfolgen der Nutzerbewegungen im Web

Datenschutzprobleme von Cookies

- Ziel: Eindeutige Identifikation des Nutzers, Nachvollziehen seiner Bewegungen
 - auf einer einzelnen Webseite
 - über mehrere Webseiten hinweg
- Idee:
 - Browser speichert **eindeutige Kennung** auf dem Rechner des Benutzers
- Methode:
 - Cookies, 3rd Party-Cookies
 - Flash Local Stored Objects (“Flash Cookies”)

Speicherdauer von Cookies

- Je größer die Speicherdauer,
 - desto größer der Zeitraum, über eine Person wiedererkannt werden kann, und
 - desto einfacher ist die Profilbildung

Stand: 1.5.2012,
<http://www.securityspace.com>

Speicherdauer	2010	2012
Set a cookie for longer than a day	25.2%	29.4%
Set a cookie for longer than a year	7.2%	10.6%
Set a cookie for longer than a decade	2.1%	3.3%

*) der Rest sind zumeist kurzlebige Session-Cookies

- **einmal** gesetzte Cookies werden vom Browser **automatisch** bei **jedem** Klick auf einen Link an Server übermittelt
- In PHP (auf Server-Seite) genügt eine Zeile zum Setzen von Cookies und zwei zum Abfragen:

```
<?php
    if (isset($_COOKIE['PHPSESSID'])) {
        $id = $_COOKIE['PHPSESSID'];
    } else {
        $id = generateId();
        registerInDatabase($id);
        setcookie('PHPSESSID', $id, time() + 241920000);
    }
?>
```

- Cookie-Fähigkeit: Webbrowser, Java-Script, Flash
→ Browser-Einstellungen gelten nicht für alle Cookies!

Cookies vom Webserver

```
/home/buchmann> telnet dbis.ipd.uni-karlsruhe.de 80
```

```
Trying 129.13.182.137...
```

```
Connected to dbis.ipd.uni-karlsruhe.de.
```

```
Escape character is '^]'.  
GET / HTTP/1.0
```

```
HTTP/1.1 200 OK
```

```
Date: 11 May 2014 08:46:39 GMT
```

```
Server: Apache/2.2.8 (Unix) DAV/2 mod_ssl/2.2.8 OpenSSL/0.9.8h PHP/5.2.6
```

```
X-Powered-By: PHP/5.2.6
```

```
Set-Cookie: PHPSESSID=74e6a1911499e9578534f31814c357d4; path=/  
Expires: Thu, 19 Nov 2014 08:52:00 GMT
```

```
Cache-Control: no-store, no-cache, must-revalidate, post-check=0, pre-check=0
```

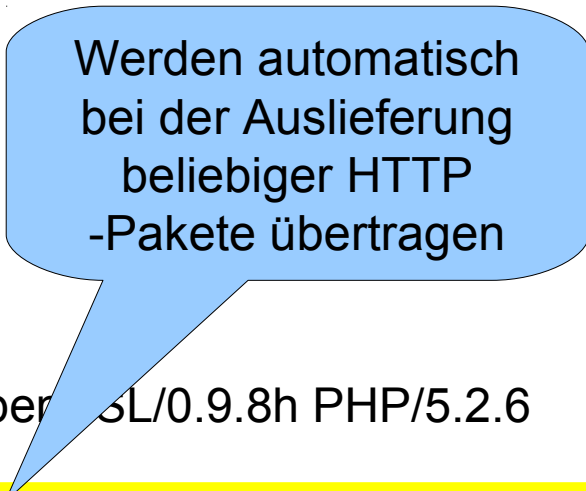
```
Pragma: no-cache
```

```
Connection: close
```

```
Content-Type: text/html
```

```
<!DOCTYPE HTML PUBLIC "-//W3C//DTD HTML 4.01 Transitional//EN" ...
```

```
</html>Connection closed by foreign host.
```



Werden automatisch bei der Auslieferung beliebiger HTTP-Pakete übertragen

Was speichern Cookies?

- Domain, die den Cookie gesetzt hat und lesen kann
 - Sicherheitsmechanismus; Cookie kann nicht von Dritten gelesen werden
- Ob alle Rechner der Domain Zugriff auf Cookie haben
 - wichtig z.B. bei Serverfarmen, Lastverteilung
- Pfad der Domain, in der der Cookie gültig ist
- Ob Cookie-Zugriff nur SSL-verschlüsselt möglich ist
- Lebensdauer des Cookies
- Name des Cookies
- Wert des Cookies
 - beliebiger Text, oft Identifikationsnummer

Cookies auf der Festplatte

■ Beispiel:

```
/home/buchmann> cat ~/.mozilla/firefox/5p4dyjr8.default/cookies.txt
.advertising.com    TRUE / FALSE    1380216294    BASE    cxqDMdeP80sVzIE!
.advertising.com    TRUE / FALSE    1373322096    ACID    c1000121564209600
.advertising.com    TRUE / FALSE    1383850713    C2      1+IFJ14DFMQtFe+h
.affilinet.parship.de  FALSE/ FALSE    1272740302    VID     par-sv-53%3ASRSP
.advertising.com    TRUE / FALSE    1380216294    F1      BYGbejkAAAAAAnXo
.dilbert.com        TRUE / FALSE    1609459211    RMAM    01cen8_1006.4Y9ZK
.dilbert.com        TRUE / FALSE    1609459275    OAX     VKP4KEi0ZVYACirH
```

■ Spalten

- Domain
- Zugriff von der ganzen Domain / Zugriff nur bei HTTPS
- Verfallsdatum (Millisekunden ab 01.01.1970)
- Name
- Wert

■ Standard zu Cookies:

<http://www.w3.org/Protocols/rfc2109/rfc2109>

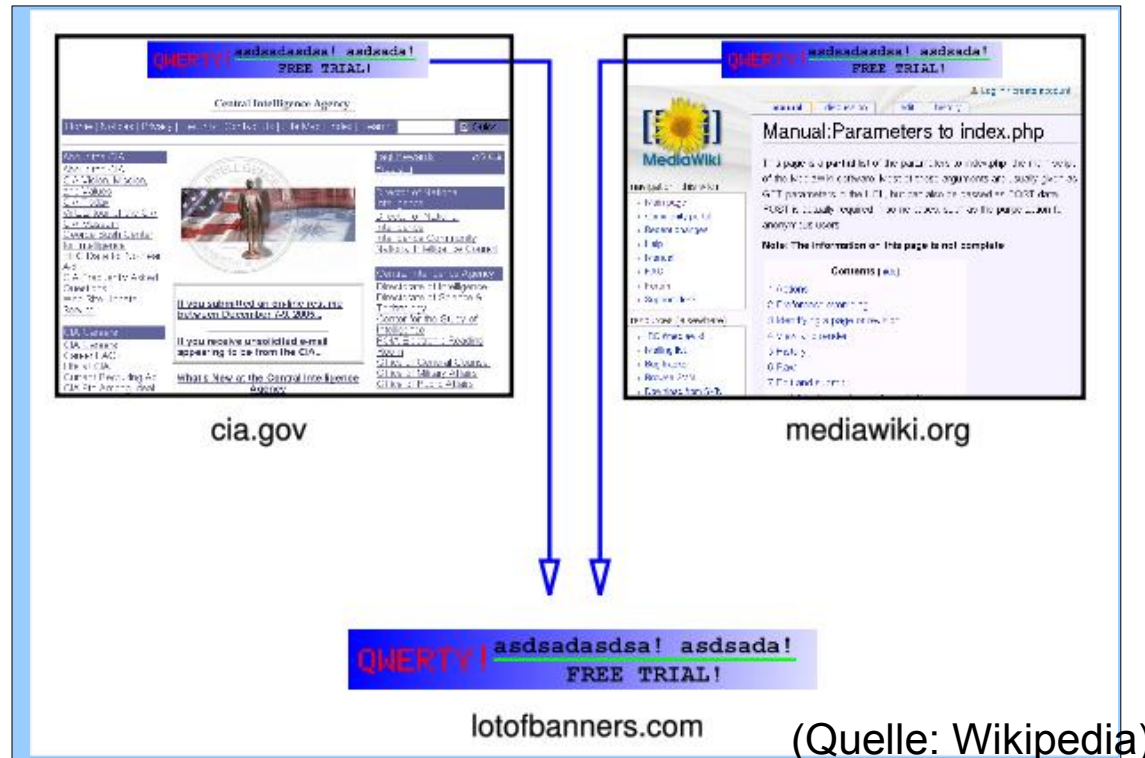
- "...user agents' cookie support should have no fixed limits ... should strive to store as many frequently-used cookies as possible..."
- "...minimum capabilities:
 - at least 300 cookies
 - at least 4096 bytes per cookie
 - at least 20 cookies per unique host or domain..."

■ <http://webdesign.about.com/od/cookies/f/cookies-per-domain-limit.htm>

- Chrome 9 allowed 180 cookies per domain
- Firefox 3.6.3 allowed 50 cookies per domain
- Internet Explorer 8 allowed 50 cookies per domain
- Opera 9 and 10 allowed 30 cookies per domain
- Safari 5, Opera 11 – unlimited

3rd Party Cookies

- Beispiel: lotofbanners.com platziert Banner auf cia.gov und mediawiki.org, erzeugt Cookie mit ID



- Besucher beider Webseiten können zugeordnet werden

- typisches Fallbeispiel: <http://www.msn.com>

- Name: MUID
Content: 19721833AA754D52AD0145F1F06BE895
Domain: *.msn.com
Send For: Any type of connection
Expires: 01.01.2024

- 3rd Party Cookies auf der selben Seite:

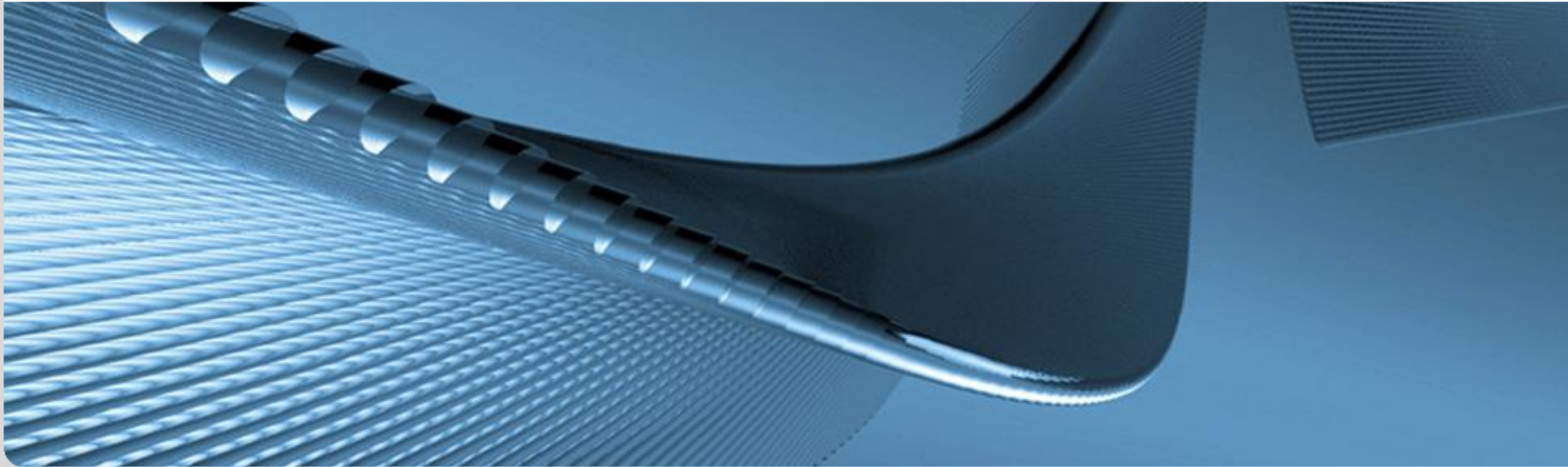
- live.com (Microsoft dependance)
- 2o7.net (“Measure customer behavior in real-time...”)
- atdmt.com (“...analyze their online advertising, rich media, search marketing and website behavior.”)

Schutz vor Cookies?

- viele sinnvolle Anwendungen für Cookies
- keine automatische Erkennung von 'guten' und 'bösen' Cookies
 - die gesammelten Informationen sind verborgen im Server des Cookie-Setzers gespeichert
- einzige Abhilfe:
 - Filter im Webbrowser installieren, Regelwerk aufsetzen, Filter im Flash-Plugin konfigurieren
→ erfordert Zeit und umfangreiche Kenntnisse

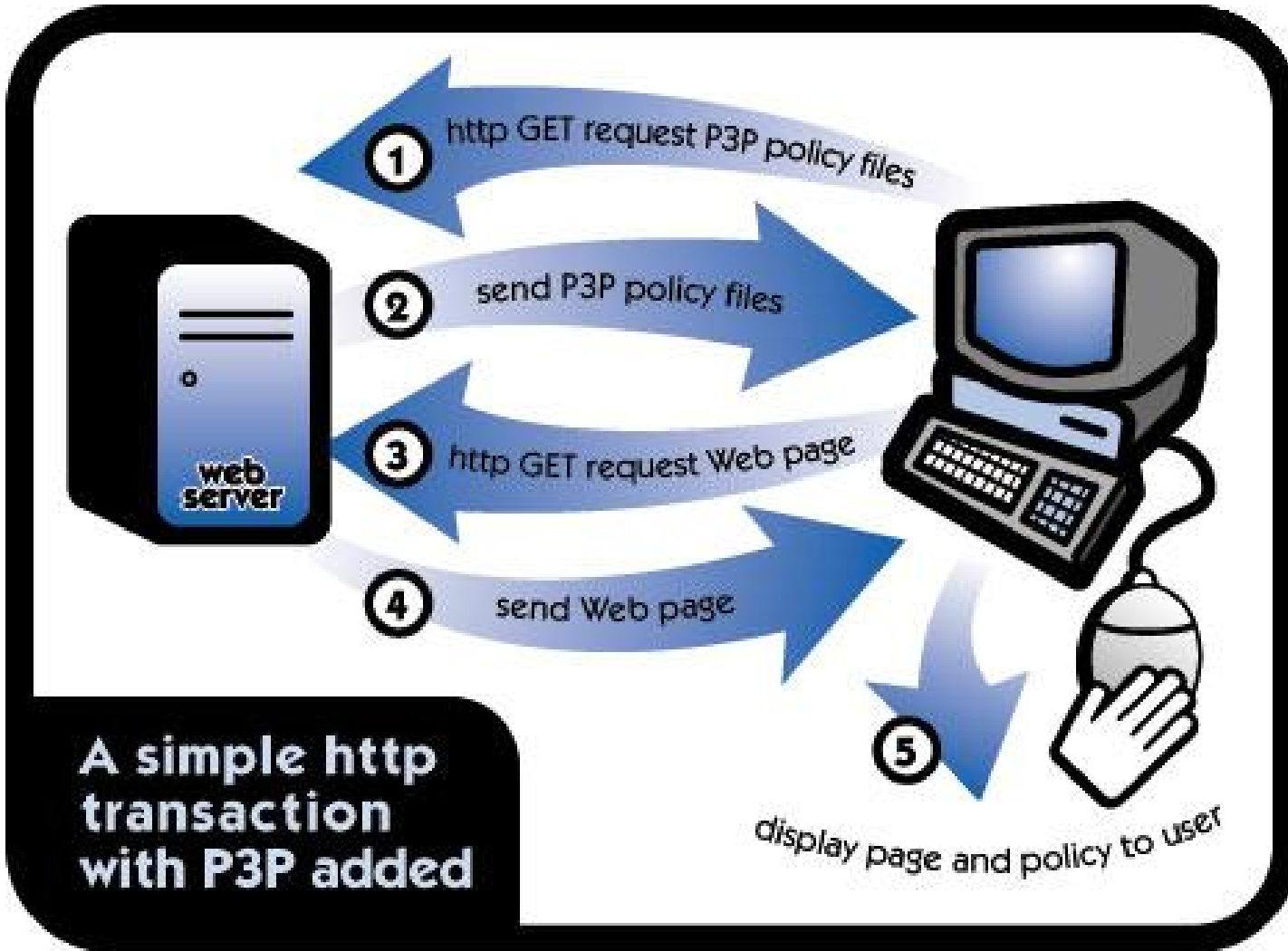
Platform for Privacy Preferences

IPD, Systeme der Informationsverwaltung, Nachwuchsgruppe „Privacy Awareness in Information Systems“



- P3P (Platform for Privacy Preferences)
 - Entwickelt vom World-Wide-Web-Consortium in Zusammenarbeit mit zahlreichen Unternehmen
- Bestandteil des HTTP-Protokolls, mit dem Dienste dem Nutzer mitteilen können
 - welche Daten gesammelt werden
 - wie diese verarbeitet werden
 - wem die Daten zugänglich gemacht werden
- Auszeichnung von
 - per http zugänglichen Medien (Webseiten, Scripten...)
 - Cookies
- Browser warnt Nutzer bei Datenschutzkonflikten

Prinzipielle Funktionsweise



Quelle: <http://p3ptoolbox.org>

Automatischer Abgleich

- Nutzer legt im Browser seine Privacy-Einstellungen fest
- Browser lädt bei jedem Medienzugriff/Cookie die P3P-Policy vom Server
- Automatischer Abgleich von Nutzerpräferenz und maschinenlesbarer P3P-Policy
 - keine Übereinstimmung
 - Kommunikation wird geblockt und/oder Warnung ausgegeben
 - Übereinstimmung
 - Nutzer bekommt vom Abgleich nichts mit; P3P arbeitet verborgen im Hintergrund
- **Achtung: P3P ist dient nicht der Durchsetzung!**
 - **Anbieter kann durchaus Falschangaben spezifizieren**

- **Langformat:** XML-Datenformat, Tags für
 - Wer sammelt die Daten?
 - Welche Daten?
 - Für welche Zwecke?
 - Gibts Opt-in oder Opt-out Möglichkeiten?
 - Wer bekommt die Daten zu sehen?
 - Welche Informationen kann der Betroffene abrufen?
 - Wann werden die Daten gelöscht?
 - Wie werden Streitfragen zur Policy gelöst?
 - Wo ist die Klartext-Policy gespeichert?

Kurzformat: Kürzel in Klartext

→ **Aber: deutsches Datenschutzgesetz nicht vollständig abbildbar!**
(z.B. keine Unterscheidung nach Ländern mit äquiv. DS-Standard)

```
<POLICY name="forBrowsers"
discuri="http://www.catalogshop.example.com/PrivacyPracticeBrowsing.html">
<ENTITY><DATA-GROUP>
<DATA ref="#business.name">CatalogShop</DATA>
<DATA ref="#business.contact-info.postal.street">4000 Lincoln Ave.</DATA>
<DATA ref="#business.contact-info.postal.city">Birmingham</DATA>
<DATA ref="#business.contact-info.postal.stateprov">MI</DATA>
<DATA ref="#business.contact-info.postal.postalcode">48009</DATA>
<DATA ref="#business.contact-info.postal.country">USA</DATA>
</DATA-GROUP></ENTITY>
<DISPUTES-GROUP>
<DISPUTES resolution-type="independent"
service="http://www.PrivacySeal.example.org">
</DISPUTES>
</DISPUTES-GROUP>
<STATEMENT><br /> <PURPOSE><admin/><develop/></PURPOSE>
<RECIPIENT><ours/></RECIPIENT>
<RETENTION><stated-purpose/></RETENTION>
<DATA-GROUP>
<DATA ref="#dynamic.clickstream"/>
<DATA ref="#dynamic.http"/>
</DATA-GROUP>
</STATEMENT>
</POLICY>
```

Beispiel: Tags für Purpose (Zweck)

<code><current/></code>	Daten für Dienstleistung erforderlich
<code><admin/></code>	Web Site Administration
<code><develop/></code>	Forschung und Entwicklung
<code><tailoring/></code>	maßschneidern der Webseite, z.B. behindertengerecht anzeigen
<code><pseudo-analysis/></code>	Analyse von pseudonymen Profilen
<code><pseudo-decision/></code>	Entscheidung mit pseud. Profilen
<code><individual-analysis/></code>	Individuelle Analysen
<code><individual-decision/></code>	Individuelle Entscheidungen
<code><contact/></code>	Kontaktierung, z.B. für Werbung
<code><historical/></code>	Archivierung
<code><telemarketing/></code>	Marketing
<code><other-purpose> string </other-purpose></code>	

Beispiel für das P3P-Kurzformat

“NON DSP ADM DEV IVD_o OTP_i OUR IND STP PHY PRE UNI”

Access Policy?

NON = None

Disputes Policy?

DSP = There is a DISPUTES-GROUP section in the full P3P policy

Purposes of Data Collection?

ADM = Used for web site and system administration

DEV = Used for research and development

IVD_o = Used for Individual decision making, users can "opt-out"

OTP_i = Used for other purposes if users "opt-in" to such purposes.

Recipients of the data?

OUR = The Web site organization itself receives the data.

IND = indefinitely

STP = Retained for stated purpose.

Categories of Data Collected?

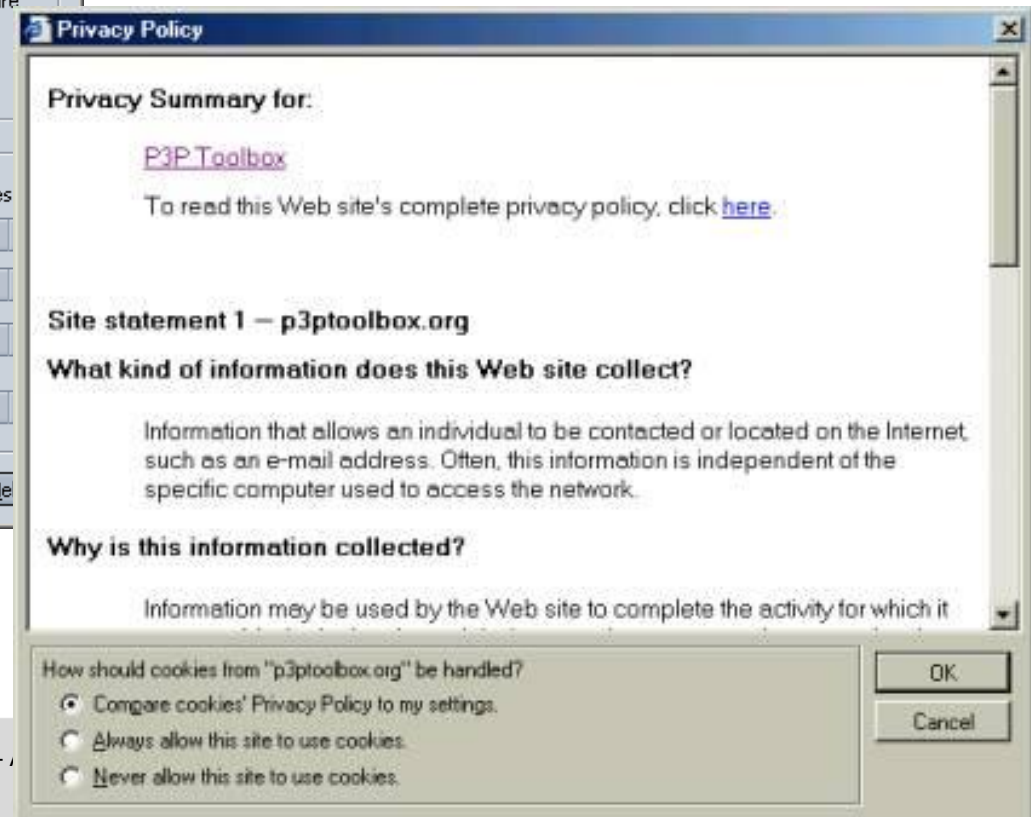
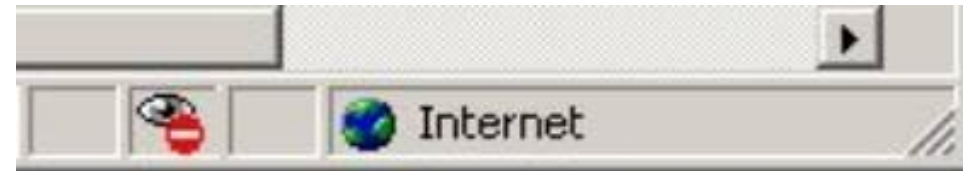
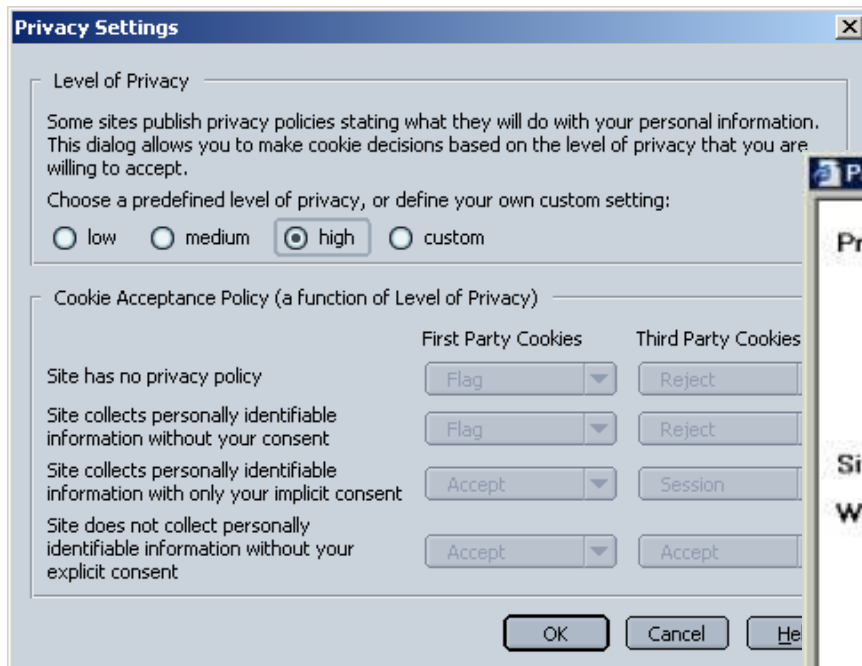
PHY = Physical contact information

PRE = Preference information.

UNI = A unique ID is associated with the cookie

Nutzung von P3P heute

- praktisch ausschließlich Cookies,
Beispiele: Netscape 7, Internet Explorer 6



AT&T Privacy Bird

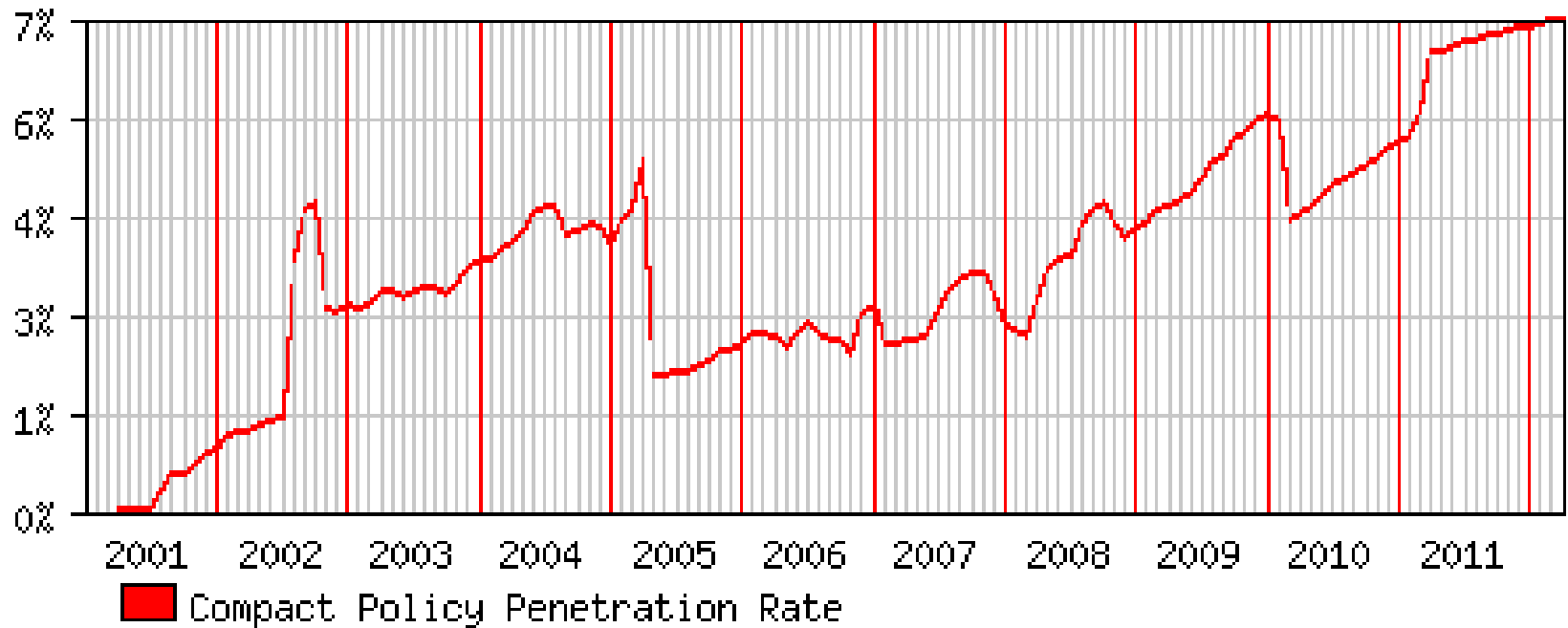
- Eines der komfortabelsten Interfaces für P3P-Policies
 - detaillierte Einstellung von Präferenzen (übliche Browser bieten nur grobe Einstellungen)
 - Intuitives Icon informiert über Einschätzung der Webseite gemäß Nutzerpräferenzen



- <http://www.privacybird.com>

- P3P hat geringe Marktdurchdringung:

P3P Compact Policy Penetration Rates



- Stand: 01.04.2012,
<http://www.securityspace.com>,
3,289,487 untersuchte Websites

Top-5 Verwendungszwecke in P3P-Angaben

- Schlüssel **PSA(a,i,o)**: **70%**
 - create a pseudonymous record,
 - determine the habits, interests, ..., of individuals
- Schlüssel **ADM**: **62%**
 - technical Support of the Web site and its computer
- Schlüssel **PSD(a,i,o)**: **21%**, wie wie **PSA**, aber
 - to make a decision that directly affects that individual
- Schlüssel **CON(a,i,o)**: **15%**
 - used to contact the individual
 - for the promotion of a product or service
- Schlüssel **CUR**: **14%**
 - complete the activity the information was provided for

Stand: 01.05.2010, <http://www.securityspace.com>, 12.3781 untersuchte P3P-Angaben

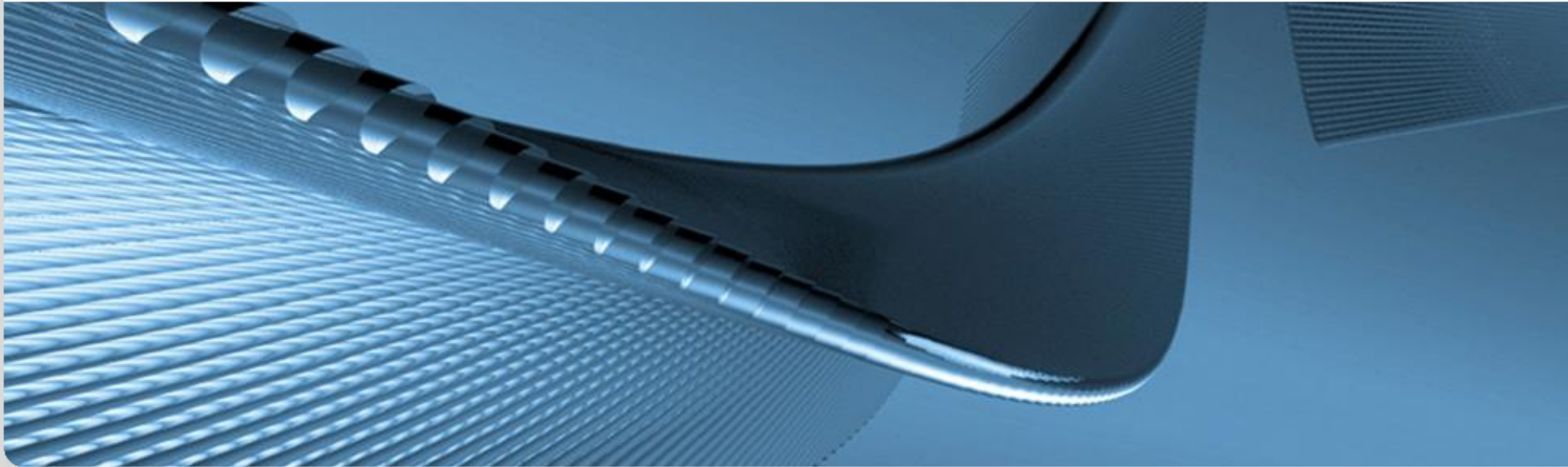
- Entlastet den Nutzer
 - automatischer Abgleich von Präferenzen und Policy
→ vermeidet unnötige “geistige Arbeit”
- mehr Transparenz
 - P3P-Clients können Datenschutzerklärung bei jeder Änderung prüfen
 - Detaillierte Datenschutzerklärungen für alle Objekte einer Webseite, incl.
 - Cookies
 - Partnerprogramme
 - Werbebanner von Drittservern
 - Web-Bugs, Scripten von Analysediensten

- Falsche Herangehensweise
 - P3P: “Wieviel Privatsphäre bist du aufzugeben bereit, um diesen Dienst in Anspruch zu nehmen?”
 - sinnvoller: “Welche persönlichen Daten braucht der Dienst mindestens zum funktionieren?”
- Nutzer kauft die Katze im Sack
 - P3P verhindert keinen Datenmißbrauch
- Umständliche Handhabung
 - P3P-Regeln sind komplex, erfordern Expertenwissen
- Ausschluss 'guter' Seiten ohne P3P

“P3P is a protocol that requires Internet users to reveal their privacy preferences before they are allowed to access information on the Internet.” (www.epic.org)

Abschluss

IPD, Systeme der Informationsverwaltung, Nachwuchsgruppe „Privacy Awareness in Information Systems“



- **Datenspuren im Internet**
 - auf Rechner des Nutzers, beim Internet-Anbieter, beim Weiterleiter, Logs in Web-, DNS-, Mail-Server, Analysedienste von Dritten
 - basieren auf Log-Informationen des HTTP-Protokolls
Cookies, Web-Bugs
- **P3P ist ein interessanter Lösungsansatz mit Schwächen**
 - beim Konzept (Katze-im-Sack-Problem, Benachteiligung des Nutzers)
 - bei der Nutzerakzeptanz