

Prywatność z protokołem P3P w transakcjach online

Igor Margasiński, Krzysztof Szczypiński
Instytut Telekomunikacji PW
E-mail: igor@margasinski.com, krzysztof@szczypiński.com

VIII Krajowa Konferencja Zastosowań Kryptografii Enigma 2004
Warszawa, 10-13 maja 2004

Plan prezentacji

- Furtki do prywatności
- Zarys technologii P3P
- Polityka P3P
 - Struktura
 - Najważniejsze elementy
 - Forma skrócona
- Perspektywy
- Podsumowanie

2

Margasiński, Szczypiński – Prywatność z protokołem P3P w transakcjach online

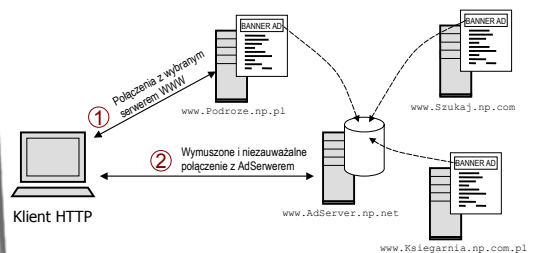
Furtki do prywatności

- Nawigacja WWW nie jest anonimowa
- Rejestrowane są zazwyczaj:
 - adres sieciowy
 - typ oprogramowania
 - czas
 - przeglądane zasoby
 - poprzednio odwiedzana strona – refer link
 - wiele innych – poprzez formularze HTML
- Zdarzenie wiązane są ze sobą poprzez Cookies

3

Margasiński, Szczypiński – Prywatność z protokołem P3P w transakcjach online

Schemat zastosowania Cookies w profilowaniu



4

Margasiński, Szczypiński – Prywatność z protokołem P3P w transakcjach online

Furtki do prywatności *cd*

- Zarejestrowane ścieżki nawigacji poddawane są automatycznym procesom odkrywania wiedzy
 - faza przetwarzania wstępnego
 - odkrywanie wzorców
 - statystyka matematyczna, wydobywanie wiedzy, uczenie się maszyn, rozpoznawanie obrazów
 - analiza wzorców
 - mechanizmy zapytań SQL, działania OLAP (*Online Analytical Processing*)
- Profile internautów:
 - dane osobowe
 - zainteresowania
 - zwyczaje
 - ...

5

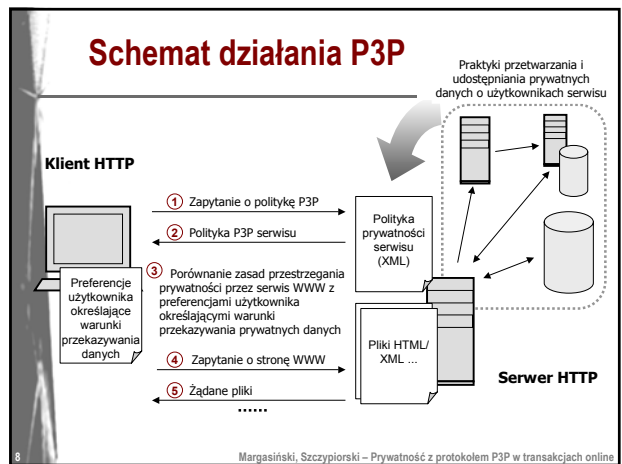
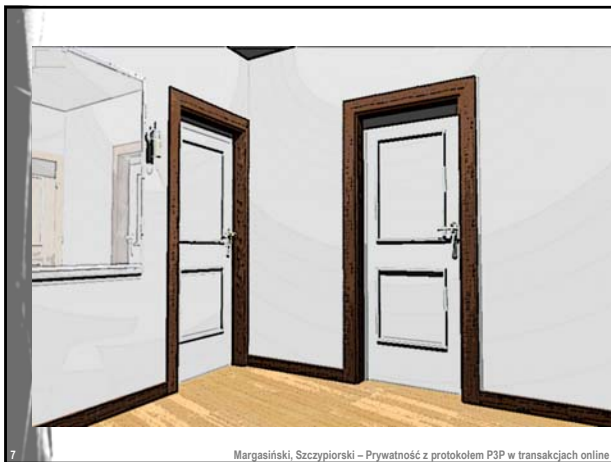
Margasiński, Szczypiński – Prywatność z protokołem P3P w transakcjach online

Zarys technologii P3P

- World Wide Web Consortium
- P3P – Platform for Privacy Preferences – Platforma Preferencji Prywatności
- Standard zapisu i przekazywania polityk prywatności – w jednolity sposób
- Odczytywany przez maszyny
 - XML (*Extensible Markup Language*)

6

Margasiński, Szczypiński – Prywatność z protokołem P3P w transakcjach online



Polityka P3P

- Powinna odzwierciedlać rzeczywiste praktyki
 - dane identyfikacyjne organizacji
 - identyfikator pliku lub zasobu
 - opis zasobu
 - czy dane są gromadzone przy pobieraniu zasobu
 - jeśli tak – jakiego rodzaju (anonimowe, pseudoanonimowe, osobowe)
 - czy stosowany jest mechanizm Cookies
 - jeśli tak – jaki jest identyfikator oraz od kogo pochodzi zapis
 - gdzie dane są przechowywane
 - do jakich celów dane są wykorzystywane
 - jakim stronom dane są przekazywane
 - możliwość akceptowania praktyk (*opt-in*, *opt-out*)

9 Margasiński, Szczypiorski – Prywatność z protokołem P3P w transakcjach online

Polityka P3P cd

P3P o systemie Cookies

- Powinna odzwierciedlać rzeczywiste praktyki
 - nazwa (identyfikator) *Cookie*
 - opis
 - czy zapis jest usuwany po zakończeniu sesji
 - pochodzenie bezpośrednie, czy od trzeciej strony
 - jakie dane są gromadzone
 - anonimowe, pseudoanonimowe, osobowe
 - czy dane są kiedykolwiek wiązane z danymi osobowymi
 - od jakiej strony pochodzi zapis
 - do jakich celów uzyskane dane są wykorzystywane
 - jakim stronom dane są udostępniane
 - możliwość akceptowania praktyk (*opt-in*, *opt-out*)

10 Margasiński, Szczypiorski – Prywatność z protokołem P3P w transakcjach online

Struktura

- Plik odwołań (*Policy Reference File*)
 - dobrze znana lokalizacja
 - `/w3c/p3p.xml`
 - wskazanie w nagłówku HTTP
 - wskazanie w znaczniku HTML
- Pliki polityk

11 Margasiński, Szczypiorski – Prywatność z protokołem P3P w transakcjach online

Liczba polityk

	Mała	Duża
Za	Prostota tworzenia i modyfikacji Mniejsze ryzyko błędnego przyporządkowania	Dokładny opis praktyk Zachęcenie do korzystania chociaż z części serwisu
Przeciw	Bardziej inwazyjny charakter części polityk	Praktyczne trudności w implementacji i aktualizacji

12 Margasiński, Szczypiorski – Prywatność z protokołem P3P w transakcjach online

Przykład pliku polityki

```
<?xml version="1.0"?>
<POLICIES xmlns="http://www.w3.org/2002/01/P3Pv1">
  <!-- Expiry information for this policy -->
  <EXPIRY max-age="86400"/>
  <POLICY
    xml:lang="pl">
    <ENTITY>
      <DATA-GROUP>
        </DATA-GROUP>
      </ENTITY>
      <ACCESS><nonident></ACCESS>
      <STATEMENT>
        <EXTENSION optional="yes">
          <GROUP-INFO
            xmlns="http://www.software.ibm.com/P3P/editor/
            extension-1.0.html" name="Access log information"/>
          </EXTENSION>
        <CONSEQUENCE>
          Our Web server collects access logs containing this information
        </CONSEQUENCE>
        <PURPOSE>
          <admin/><current/><develop/>
        </PURPOSE>
        <RECIPIENT><ours></RECIPIENT>
        <RETENTION><indefinitely></RETENTION>
      <DATA-GROUP>
        <DATA ref="#dynamic.clickstream"/>
        <DATA ref="#dynamic.http"/>
      </DATA-GROUP>
    </STATEMENT>
  </POLICY>
</POLICIES>
```

13

Najważniejsze elementy

- Polityka prywatności (**policy**)
 - podstawowe informacje o pojedynczej polityce prywatności
 - nazwa (**name**)
 - adres URI polityki w języku naturalnym (**discuri**)
 - adres gdzie podane są instrukcje w jaki sposób użytkownik może wyrazić zgodę lub niezgodę na praktyki zdefiniowane w polityce (**opturi**) w schemacie *opt-in* lub *opt-out*,
- Dane identyfikacyjne organizacji (**entity**)
 - adres oraz inne informacje potrzebne do nawiązania kontaktu z organizacją
- Zasady dostępu (**access**)
 - dostęp użytkowników do ich danych osobowych
 - np. umożliwia poprawienie danych adresowych

14

Margasiński, Szczypiorski – Prywatność z protokołem P3P w transakcjach online

Najważniejsze elementy *cd*

- Sprawy sporne (**disputes**)
 - informacje o nadrzędnych zasadach lub o innych prawach, którym podlegają praktyki serwisu,
- Sposoby rekompensaty (**remedies**)
 - zawarty w elemencie **disputes** i określa sposoby rekompensaty w przypadku naruszenia zasad zdefiniowanych w polityce
- Cele gromadzenia danych (**purpose**)
- Zbiór odbiorców (**recipient**)
- Warunki przechowywania danych (**retention**)

15

Margasiński, Szczypiorski – Prywatność z protokołem P3P w transakcjach online

Narzędzia

- Tworzenie polityk

IBM P3P Policy Editor →

16

Margasiński, Szczypiorski – Prywatność z protokołem P3P w transakcjach online

- Testowanie poprawności
- ← W3C's P3P Policy Validator

Skrócona forma zapisu

- Polityka kompaktowa
- Obejmuje tylko praktyki *Cookies*
- Przekazywana w nagłówku HTTP
- Ciąg trzyliterowych słów oddzielonych spacjami, każde słowo:
 - przynależy do określonej kategorii, takiej jak np.:
 - cele przetwarzania danych
 - zbiór odbiorców
 - reprezentuje określone praktyki
- Przeglądarki WWW obecnie bazują na politykach kompaktowych

17

Margasiński, Szczypiorski – Prywatność z protokołem P3P w transakcjach online

Przykład przekazania polityki kompaktowej

Klient

```
GET /bardzo_ciekawe.html HTTP/1.1
Host: np_portal_informacyjny.pl
Accept: */* Accept-Language: en, pl
User-Agent: WonderBrowser/1.0
```

Serwer

```
HTTP/1.1 200 OK
P3P:
policyref="http://np_portal_informacyjny.pl/P3P/
PolicyReferences.xml",
CP="NON DSP ADM DEV PSD OUR IND STP PHY PRE"
Content-Type: text/html
Content-Length: 1234
Server: WonderServer/1.0
```

...dane...

18

Margasiński, Szczypiorski – Prywatność z protokołem P3P w transakcjach online

Interpretacja przykładu

Dostęp	NON Brak		
Sprawy sporne	DSP sprawy sporne są określane w pełnej polityce P3P		
Rekompensata			
Brak identyfikacji			
Cele	ADM do administracji systemem i serwerem WWW	DEV do badań i rozwoju	PSD do celów marketingowych w oparciu o pseudo-anonimową identyfikację
Odbiorcy	OUR odbiorcą jest wyłącznie firma		
Warunki przechowywania	IND wraz z informacjami identyfikującymi	STP dla celów stanowych	
Kategorie	PHY informacje o potrzebnym do nawiązania fizycznej komunikacji	FRE informacje o preferencjach	

Margasiński, Szczypiorski – Prywatność z protokołem P3P w transakcjach online

Konfrontacja z potrzebami

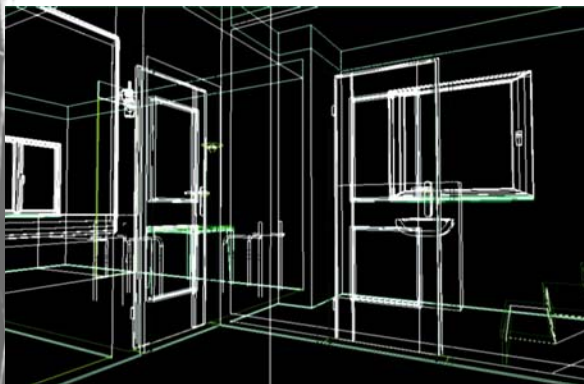
- Wysokie oczekiwania
- P3P w istocie
 - umożliwia serwerom WWW zaprezentowanie deklaracji swoich praktyk z zakresu prywatności w jednolity sposób
 - umożliwia użytkownikom odwiedzającym serwisy WWW poznanie deklaracji:
 - jakie dane o nich będą gromadzone
 - jak będą użyte
 - w jaki sposób użytkownik może wyrazić zgodę lub niezgodę na poszczególne działania
 - nie należy do technologii rozszerzających prywatność (PET – Privacy Enhancing Technologies)

Margasiński, Szczypiorski – Prywatność z protokołem P3P w transakcjach online

Postulaty P3P a prywatność

- Platforma (prezentacji) preferencji prywatności
- Nie podwyższa bezpieczeństwa
- Pełni jedynie rolę informacyjną
- Nie gwarantuje zgodności opisu z rzeczywistością
- Opieranie ochrony prywatności tylko na P3P jest zgubne
- Element pomocniczy w budowaniu technik zapewniających prywatność

Margasiński, Szczypiorski – Prywatność z protokołem P3P w transakcjach online



Margasiński, Szczypiorski – Prywatność z protokołem P3P w transakcjach online

Perspektywy

- Komponent nowych technik zapewniania prywatności
- Rozwój semantyki protokołu
- Mechanizmy kontroli korelacji rzeczywistych praktyk z deklaracjami P3P
 - systemy zarządzania prywatnością (Privacy Manager)
 - Uwierzytelnienie i niezaprzeczalność w oparciu o podpisy cyfrowe i certyfikaty X.509

Margasiński, Szczypiorski – Prywatność z protokołem P3P w transakcjach online

Podsumowanie

- Niepełna odpowiedź na obawy internautów z zakresu prywatności WWW
- Technologia nie ogranicza bezpośrednio nadużyć
- Pozwala na prezentację deklaracji praktyk
- Ujednolicenie zapisu polityki prywatności to dobre narzędzie regulujące sposób informowania konsumentów o zasadach przetwarzania ich danych
- Środek ustalenia kompromisu pomiędzy ekonomicznymi potrzebami firm, a prawem użytkowników do prywatności
- Gdy strony gromadzące dane nie są wiarygodne, rozwiązanie okazuje się mało przydatne, a zaufanie do protokołu może być szkodliwe

Margasiński, Szczypiorski – Prywatność z protokołem P3P w transakcjach online

Koniec

Czy mają Państwo pytania?

Igor Margasiński, Krzysztof Szczypiorski
Instytut Telekomunikacji PW
E-mail: igor@margasinski.com, krzysztof@szczypiorski.com

*(...) I know what is wrong,
And I know what is right.
And I'd die for the truth
In My Secret Life.*

Leonard Cohen - In My Secret Life

Literatura

- [1] Agrawal, R., Imielinski, T., Swami A. Mining Association Rules Between Sets of Items in Large Databases. Materialy: ACM SIGMOD Conference on Management of Data, Washington DC, USA, May 1993.
- [2] Agrawal, R., Srikant, R. Fast algorithms for mining association rules. Materialy: 20th VLDB Conference, str. 487-499, Santiago, Chile, 1994.
- [3] Agrawal, R., Srikant, R. Mining Sequential Patterns. Materialy: 11th Int'l Conference on Data Engineering (ICDE), Taipei, Taiwan, March 1995.
- [4] Anonymizer (<http://anonymizer.com>).
- [5] Berners-Lee, T., Fielding, R., Fraytek, H. Hypertext Transfer Protocol – HTTP/1.0. RFC 1945, 1996.
- [6] Catalogue L.D., Pitkow J.E. Characterizing Browsing Strategies in the World Wide Web. Materialy: 3rd Int'l World Wide Web Conference, 1995.
- [7] Chavira, D. Untraceable Electronic Mail, Return Addresses, and Digital Pseudonyms. Communications of the ACM Vol. 21 no. 2, str. 84 – 88, 1981.
- [8] Cooley, R., Mobasher, B., Srivastava, J. Grouping Web Page References into Transactions for Mining World Wide Web Browsing Patterns. Materialy: 1997 IEEE Knowledge and Data Engineering Exchange Workshop (KDEX), Newport Beach, California, November 1997.
- [9] Fayyad, U., Piatesky-Shapiro, G., Smyth P. From data mining to knowledge discovery. An overview. Materialy: ACM KDD, 1994.
- [10] Feltenpough, C. Online Security and Privacy Concerns on the Increase in Carats, Inca-Red, 2001.
- [11] Fielding, R., Gathys, J., Mogul, J., Frystyk, H., Masinter, L., Leach, P., Berners-Lee T. HyperText Transfer Protocol – HTTP/1.1. RFC 2616, 1999.
- [12] Goldberg, I., Shostack, A. Freedom Network: 1.0 Architecture and Protocols, Zero-Knowledge Systems, White Paper 1999.
- [13] Goldschlag, D. M., Reed, M. C., Sykeson, P. F. Onion Routing for Anonymous and Private Internet Connections. Communications of the ACM Vol. 42 no. 2, str. 39-41, 1999.
- [14] Hartigan, J. Clustering Algorithms. John Wiley, 1975.
- [15] Jamaguchi, L. P3P Implementation Guide. The Internet Education Foundation, 2003.
- [16] Kranz, D., Light, L., Grawith D. Privacy On and Off the Internet: What Consumers Want. Harris Interactive (2002).
- [17] Kissel, D., Montulli L. HTTP State Management Mechanism. RFC 2965, October 2000.
- [18] Cranor, L., Langheinrich, M., Marchioni, M., Preiser-Marchal, M. The Platform for Preferences 1.0 (P3P 1.0) Specification, W3C Recommendation, April 2002.
- [19] Cranor, L. Web Privacy with P3P. O'Reilly & Associates, September 2002.
- [20] Margasiński, I. Zapewnianie anonimowości przy przeglądaniu stron WWW. Materialy: 18. Krajowe Sympozjum Telekomunikacji – KST 2002, Bydgoszcz, 2002.
- [21] Margasiński, I., Szczypiorski, K. Web Privacy: an Essential Part of Electronic Commerce. Materialy: 3rd International Interdisciplinary Conference on Electronic Commerce "ECOM-03", Gdańsk, 2003, str. 65-72.
- [22] Margasiński, I., Szczypiorski, K. VAST: Versatile Anonymous System for Web Users. Materialy: The Tenth International Multi-Conference on Advanced Computer Systems ACS2003, Mińskopole, 2003.
- [23] Preiser-Marchal, M. The Platform for Privacy Preferences 1.0 Deployment Guide, W3C Note, May 2001.
- [24] Reiter, M.K., Rubin, A.D. Crowds: Anonymity for Web Transactions. ACM Transactions on Information and System Security, str. 66-92, 1998.
- [25] Schneier, J. What They (Don't) Know About You. Wired, May 2001.
- [26] SixFour System (<http://www.hackwisdom.com/projects>).
- [27] Srikant, R., Agrawal, R. Mining Sequential Patterns: Generalizations and Performance Improvements. Materialy: 5th Int'l Conference on Extending Database Technology (EDBT), Avignon, France, March 1996.
- [28] Sykeson, P. F., Goldschlag, D. M., Reed, M. C. Anonymous Connections and Onion Routing. IEEE Symposium on Security and Privacy, 1998.
- [29] Weiss, S.M., Kulkowski, C.A. Computer Systems that Learn: Classification and Prediction Methods from Statistics, Neural Nets, Machine Learning, and Expert Systems. Morgan Kaufmann, San Mateo, CA, 1991.
- [30] Yezou, F., Bry, T., Paek, J., Sperberg-McQueen, C. M., Maler, E. Extensible Markup Language (XML) 1.0 (Third Edition), W3C Recommendation February 2004.