

Igor Margasiński  
Instytut Telekomunikacji  
Politechnika Warszawska  
E-mail: imargasi@tele.pw.edu.pl

# Zapewnianie anonimowości przy przeglądaniu stron WWW

Artykuł zawiera analizę zagrożeń związanych z utratą prywatności przy przeglądaniu stron WWW. Zaproponowano klasyfikację zagrożeń wraz z przykładami. Dalszą część pracy poświęcono przeglądowi stosowanych obecnie rozwiązań tego problemu. Systemy serwera pośredniczącego (*third-side proxies*), zasługujące na szczególną uwagę, opisano dokładniej. Przedstawiono opis własnej realizacji systemu zapewniania anonimowości przy przeglądaniu stron WWW opartego na serwerze pośredniczącym.

## 1. Wprowadzenie

Sieć Internet, której początki sięgają lat 60-tych (*ARPANET*), tworzona była początkowo spontanicznie w ośrodkach naukowych i akademickich. Przy formułowaniu tych fundamentów nie przywiązywano szczególnej wagi do bezpieczeństwa – nie przewidywano, że użytkownicy sieci mogą mieć względem siebie nieuczciwe intencje. Sieć miała pomagać w wymianie wyników badań czy spostrzeżeń naukowych.

Dzisiaj rozwój infrastruktury Internetu jest głównie domeną prywatnych firm. Również pole zastosowań sieci przybrało nową postać, w której udział przedsięwzięć komercyjnych zajmuje jedną z dominujących ról. System WWW (*World Wide Web*) stanowi w tych przemianach podstawowy czynnik.

Jesteśmy zatem świadkami zdecydowanego wzrostu liczby zastosowań WWW o charakterze komercyjnym. Handel elektroniczny, bankowość elektroniczną, a z drugiej strony komercyjna reklama coraz trwalej wpisują się w internetową rzeczywistość. Obserwujemy również szybki rozwój mechanizmów poszerzających możliwości i multimedialną atrakcyjność WWW. Pozostaje jednak luka w ochronie prawa do prywatności – obecnie tym bardziej widoczna, bo wykorzystywana bez pardonu właśnie w celach komercyjnych (np. jako środek reklamy).

## 2. Stan obecny

Bezpieczeństwo w kontekście WWW na ogół kojarzone jest z ochroną serwerów. Ataki na strony WWW są już przecież zagadnieniem klasycznym. Tymczasem, obecnie wyraźnie ujawnia się także zjawisko skierowane przeciwnie – ataki na użytkowników przeglądarek HTTP (*HyperText Transfer Protocol*). Działania te, to głównie pozyskiwanie i gromadzenie danych o osobach przeglądających strony WWW. Sprowadza się to do naruszania prywatności. Ingerencja w

prywatność jest w takich atakach celowa i przekładana na wymierny zysk. Dane osobowe użytkowników WWW stają się przedmiotem coraz większego zainteresowania.

## 2.1. Zagrożenia

Obecnie mamy do czynienia z tworzeniem się firm specjalizujących się w masowym gromadzeniu szczegółowych informacji o internautach. Istnieją już sieci zrzeszające strony współpracujące ze sobą w tworzeniu profili o użytkownikach. Składają się one z wielu serwisów internetowych pozyskujących dane, oraz strony gromadzącej, sortującej i interpretującej, do której dane te są przesyłane (*AdServer*). Informacje same w sobie stanowiące towar, są wykorzystywane np. do precyzowania treści reklam prezentowanych na stronach WWW i natarczywych ofert przysyłanych pocztą elektroniczną (*spam*). Zarówno proces zbierania i przekazywania informacji, jak i generacji odpowiednich działań wykorzystujących je, jest na ogół w całości zautomatyzowany. Umożliwia to bardzo szerokie i powszechne zastosowanie.

Protokoły Internetu nie dostarczają żadnych mechanizmów zapewniania anonimowości. Serwery WWW mają dostęp do informacji o maszynie klienta, takich jak adres IP, domena, adres poprzednio odwiedzanej strony, używana przeglądarka, informacje o konfiguracji itp.

Dodatkowo system WWW dostarcza mechanizmów wprowadzonych z myślą o polepszeniu komunikacji klient-serwer, uatrakcyjnienia prezentowanych stron, czy do podwyższenia komfortu nawigacji, które wykorzystywane są w sposób nieprzewidziany przez twórców i stanowią zagrożenie prywatności. Chodzi tu głównie o mechanizmy rozszerzające możliwości przeglądarek o funkcje wykraczające poza standard HTML (*HyperText Markup Language*), takie jak Cookies, ActiveX, Java, JavaScript, VBScript, Shockwave. „Życie własnym życiem” zaczął przede wszystkim mechanizm cookies\*. Za jego pomocą możliwe jest zdobycie szerokiej gamy informacji osobistych o użytkowniku (np. odwiedzane strony, zainteresowania, wyszukiwane hasła, adres e-mail, imię, nazwisko). Jest to dokonywane podczas typowych czynności internauty przy przeglądaniu stron WWW, w sposób niewidoczny dla niego.

## 2.2. Źródła zagrożeń

Źródła zagrożeń, w zależności od umiejscowienia strony atakującej, można podzielić na trzy grupy:

- I. Wewnętrzne – umiejscowione w najbliższym otoczeniu użytkownika.
- II. Pochodzące z kanału komunikacyjnego przenoszącego transakcje HTTP.
- III. Pochodzące od serwera WWW, którego strony są przeglądane.

Utrata anonimowości w **grupie I** związana jest przede wszystkim z dostępnością adresu URL (*Uniform Resource Locator*) a w ogólności URI (*Uniform Resource Identifier*), wpisywanego przez osobę przeglądającą strony WWW, w pole adresu przeglądarki i wysyłanego tekstem otwartym. Zapis ten dostępny jest dla stron pośredniczących w transakcji HTTP bądź dla osób o uprawnieniach administratora sieci lokalnej danego komputera. Możliwość uzyskania wiedzy o przeglądanych przez użytkownika stronach (URL oraz inne dane przesyłane pomiędzy klientem a serwerem HTTP) dostępna jest między innymi dla:

- **Dostawcy usług internetowych** (*ISP – Internet Service Provider*)
- **Administratorsa sieci**
- **Pracodawcy**, np. właściciela sieci korporacyjnej

---

\* Mechanizm opracowany przez firmę Netscape Communications. Pierwsza przeglądarka wyposażona w tą technologię to Netscape Navigator 1.0. Nazwa wywodzi się od terminu znanego w informatyce jako określenie na nieprzejrzyste dane przechowywane przez stronę pośredniczącą.

- Również dla **innych osób korzystających z danego komputera** w przypadku kont typu „gość” (*guest*). Tu adresy URL odczytywane są z zapisu historii odwołań tworzonych przez przeglądarkę.

Trzeba zaznaczyć, że na podstawie adresów URL możliwe jest nie tylko określenie tego, jakie strony przeglądał użytkownik, ale także, jakie dane zastały wprowadzane przez niego do formularzy. Jest tak w przypadku, gdy dane z formularza przekazywane są do programów serwera HTTP za pomocą metody GET (w odróżnieniu od metody POST, gdzie dane te nie są zawierane w URL, ale w części zasadniczej wiadomości).

Niebezpieczeństwo w **grupie II** spowodowane jest łatwością podsłuchania (*sniffing*) przesyłanych danych w sieci Internet. Podstawowym protokołem w systemie WWW jest protokół HTTP. Bazuje on jedynie na niezawodności protokołów TCP/IP. Nie przewiduje możliwości ataków – specyfikacja HTTP zawiera jedynie elementarne metody uwierzytelnienia oraz sprawdzania integralności przesyłanych danych. Użytkownik przeglądarki musi więc liczyć się z faktem, iż wiedza o tym jakie strony są przez niego przeglądane oraz o tym jakie dane wprowadza do formularzy, łatwa jest do zdobycia dla innych użytkowników sieci Internet.

**W grupie III** podstawowym problemem jest dostępność adresu IP komputera użytkownika dla serwerów WWW. Serwer uzyskuje go już podczas nawiązywania połączenia. Dla serwera znane jest, więc pochodzenie poszczególnych żądań i skojarzenie ich z konkretnym hostem w sieci Internet. Serwer WWW tworzy plik z logami (*log file*) dostępny dla osób nim zarządzających. Poza adresami IP klientów zapisywane są tam również:

- Czas nadejścia usługi.
- Żądany adres URL.
- Czas przesłania.
- Nazwa użytkownika klienta – identyfikacja realizowana przez protokół HTTP.
- Informacje o błędach, jakie nastąpiły przy realizacji transakcji HTTP.
- Adres URL strony poprzednio odwiedzanej przez użytkownika przeglądarki (*refer link*) – w przypadku, gdy przejście do bieżącej strony nastąpiło przez odnośnik ze strony poprzedniej.
- Informacje o przeglądarce użytkownika.

Poważnym naruszeniem prywatności jest przekazywanie do serwera WWW adresu URL poprzednio odwiedzanej strony. Pomimo założenia, istniejącego w specyfikacji HTTP, mówiącego że dostępność informacji o adresie poprzednio odwiedzanej strony powinna być opcjonalna, dotychczas żadna przeglądarka internetowa nie wprowadziła możliwości wyłączenia tego mechanizmu. Znaczenie tego naruszenia potęguje fakt, iż w ciągu URL zawierane są często również inne dane poza adresem strony WWW. Tak jak to opisano w I grupie źródeł zagrożeń – w ciągu URL znajdują się często dane wpisywane przez użytkownika do formularzy, w szczególności hasła wyszukiwane za pomocą popularnych narzędzi wyszukiwujących (*web search engines*).

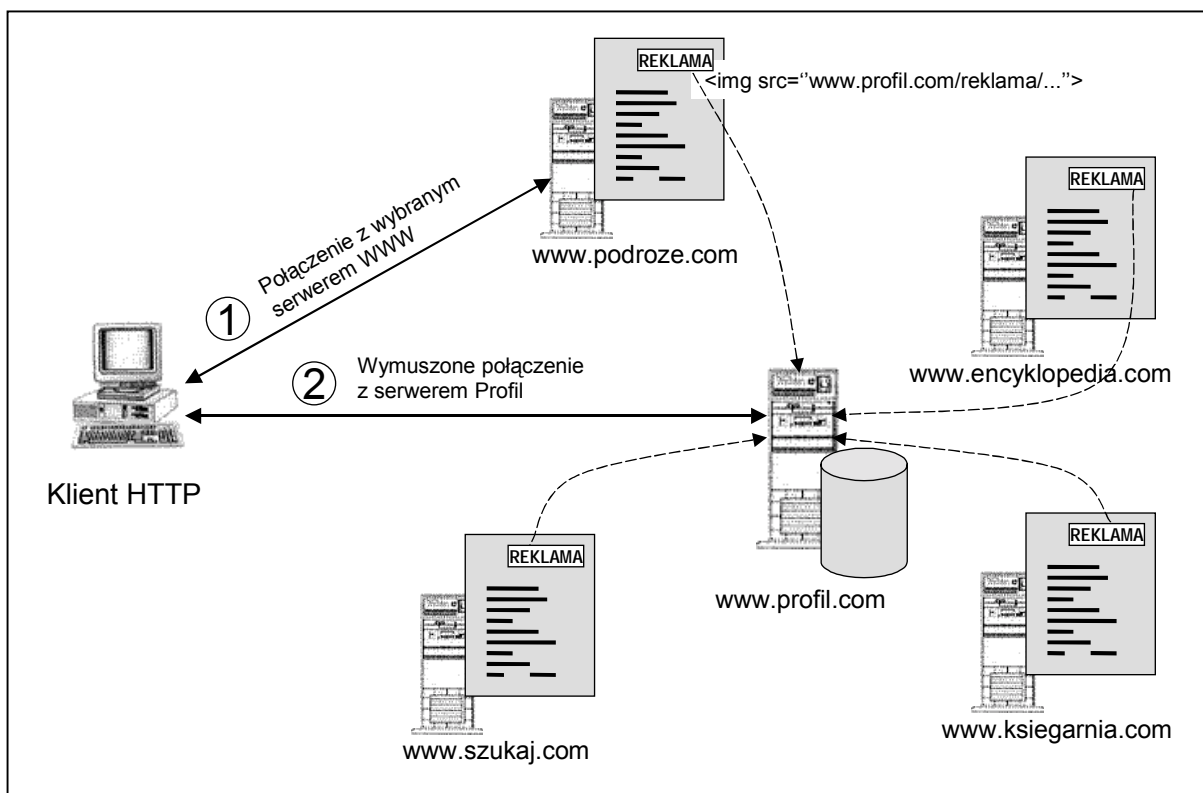
Kolejnym narzędziem, odbierającym użytkownikom przeglądarek anonimowość, jest wspomniany już mechanizm cookies – system zarządzania stanem.

Mechanizm ten został wprowadzony by w bezstanowym protokole, jakim jest protokół HTTP, umożliwić odróżnianie poszczególnych osób odwiedzających serwer. Cookies to informacje tworzone przez serwer WWW i przechowywane na komputerze użytkownika, gotowe do przyszłego odczytu. Mają postać ciągów par: zmienna, wartość. Cookies przesyłane są pomiędzy klientem i serwerem HTTP w zawartości HTML. Serwer WWW, chcąc umieścić zapis cookie, dołącza do nagłówka HTTP odpowiednie polecenie – „*Set-Cookie:*”, po którym następuje ciąg przekazywanych danych. W cookie zawsze znajduje się adres serwera, który je wysłał. Jest to kluczowe ze względu na identyfikację. Serwer może odczytać jedynie swój zapis cookies.

Cookies można sobie wyobrazić jako notatki dokonywane przez serwer HTTP na komputerze klienta, podczas przeglądania stron. Przy ponownym połączeniu się z danym serwisem internetowym, serwer z notatek tych przypomina sobie informacje w nich zawarte. Technologia ta pozwala więc na specjalizowanie po stronie użytkownika prezentowanych informacji. Np. jest używana do dostosowywania przeszukiwania zasobów (*Web search engines*), by umożliwić użytkownikom udział w konkursach, czy do zapamiętywania zawartości koszyka użytkownika dokonującego zakupów w sklepie internetowym.

Mechanizm cookies znalazł nowe zastosowania, nie przewidziane przez twórców. Obecnie mamy do czynienia z wykorzystaniem tego mechanizmu do tworzenia szczegółowych profili zainteresowań internautów.

Prześledźmy typowy **proces tworzenia profilu** (rys. 1):



Rys. 1. Poglądowy schemat tworzenia profilu zainteresowań

Wiele serwisów internetowych umieszcza na swoich stronach paski reklamowe (*banner ad*), na ogół w postaci animowanych obrazów w formacie GIF lub Shockwave, pochodzące od firm zajmujących się tworzeniem profili. Firma taka nazywana będzie dalej mianem Profil. Odwiedzając taką stronę, pobieramy w rzeczywistości pasek reklamowy z serwera Profil, choć jest to niewidoczne dla użytkownika. Nasz adres IP jest automatycznie wysyłany do „trzeciej strony” (Profil). Jeżeli jeszcze nie zostało na naszym komputerze umieszczony zapis cookie, to jest to dokonywane. Następnie Profil zapisuje informacje o nas takie jak data, czas, adres odwiedzanej strony. Na razie identyfikowani jesteśmy przez pewien numer id. Wiele stron wymaga jednak rejestracji przy udostępnianiu pewnych usług. Dokonując rejestracji podajemy dane osobowe, adres e-mail, itp. Strona, na której dokonujemy rejestracji, wysyła wprowadzone dane do firmy Profil, gdzie tworzona jest szczegółowa baza danych, aktualizowana przy naszych kolejnych odwiedzinach witryn należących do opisanej sieci (*banner ad network*). Informacje tak gromadzone pobierane są w przyszłości, przez strony należące tej sieci, w celu np. precyzowania treści reklam, tworzenia ofert wysyłanych pocztą elektroniczną, lub dalej idącej ingerencji w prywatność.

Przy standardowych ustawieniach przeglądarki, zapis cookies jest dokonywany bez wiedzy użytkownika. W większości przypadków nie tylko umieszczenie informacji o użytkowniku następuje niezauważalnie, ale również dostęp do nich. Jest on dokonywany automatycznie po połączeniu się z serwerem, który je umieścił.

Najnowsze przeglądarki wyposażane są w klienta protokołu P3P (opisany w rozdziale 3.2). Pozwala to na lepszą kontrolę mechanizmu cookies. Jest to jednak na razie faza eksperymentalna – bardzo mało serwisów WWW „obsługuje” technologię P3P. Istnieją także obawy, czy stosowanie jej nie obciąży zbytnio sieci Internet.

Warto też wspomnieć o niebezpieczeństwie, jakie stwarzają programy pochodzące z Internetu, uruchamiane na komputerze klienta. Mogą one stanowić tzw. konie trojańskie, które w sposób niewidoczny dla użytkownika przesyłają informacje o nim.

### 2.3. Potrzeby

Gwałtowny rozwój systemów pozyskujących dane o użytkownikach WWW powoduje, że poszukuje się rozwiązań chroniących prywatność. Wymaga się funkcji:

- **Ukrywanie adresu IP** przed serwerami WWW, których strony są przeglądane.
- **Ukrywanie adresu URL** przed „stroną trzecią” biorącą udział w transakcji HTTP (np. dostawca usług internetowych, administrator sieci, pracodawca).
- **Zarządzanie cookies** – mechanizm automatycznie selekcjonujący.
- **Odrzucanie pasków reklamowych** – opcjonalny filtr stanowiący duże utrudnienie profilowania oraz przyspieszenie pobierania zasobów.
- **Niedopuszczanie do tworzenia nowych okien**. Zdarzają się przypadki dynamicznego tworzenia przez skrypty bardzo małych, niewidocznych dla użytkownika okien (np. 1x1 piksel). Zawierać one mogą wspomniane wcześniej niebezpieczne elementy. Są także przeważnie źródłem reklam.
- **Blokowanie skryptów i programów**. Aplety Java zawierają ograniczenia w stosunku do programów, mające na celu ochronę przed nadużyciami. Jednak bezpieczeństwo innych technik, np. programów ActiveX, jest dużo bardziej problematyczne. W przypadku skryptów wykonywanych po stronie klienta (np. JavaScript) poziom niebezpieczeństwa jest mniejszy. Blokowanie ich powinno być co najwyżej opcjonalne.

### 3. Metody ochrony

W zależności od umiejscowienia rozwiązania, metody ochrony można podzielić na trzy kategorie:

- Oprogramowanie instalowane na komputerze użytkownika
- Wprowadzenie nowego protokołu
- Serwery pośredniczące

#### 3.1. Oprogramowanie instalowane na komputerze użytkownika

Rozwiązania te nie zapewniają pełnej anonimowości. Należy, więc widzieć w nich jedynie pomocniczą rolę. Programy takie pełnią najczęściej pojedyncze funkcje. Są to np.

- **Osobiste ściany ogniowe** (*personal firewalls*) – umożliwiające użytkownikom monitorowanie i kontrolę nad wszystkimi połączeniami od i do ich komputera.
- Programy **oczyszczające system** (*system cleaners*) – usuwające zawartość pliku cookies, zapis historii odwołań itp.
- Programy **zarządzające umieszczaniem cookies** – blokujące je lub odpowiednio filtrujące.

- Programy **blokujące reklamy** (*banner ads blockers*) – **wszystkie** lub pochodzące od „trzeciej strony”.
- Oprogramowanie **wykrywające konie trojańskie**.

### 3.2. Nowy protokół

W odpowiedzi na niepokojące praktyki pozyskiwania i gromadzenia danych o internautach, rozpoczęto prace nad protokołem mającym zapewnić użytkownikom przeglądarek większą kontrolę nad danymi o nich, obecnymi w Internecie.

Protokół ten to P3P – Platforma Preferencji Prywatności (*Platform for Privacy Preferences*).

P3P 1.0, rozwijany przez World Wide Web Consortium (W3C), jest promowany na standard przemysłowy mający umożliwić, w sposób prosty i zautomatyzowany dla użytkownika, kontrolę nad wykorzystywaniem danych osobowych przez strony WWW. Patrząc na P3P z najniższego poziomu jest to standaryzowany zestaw pytań wielokrotnego wyboru, zawierający wszystkie główne aspekty polityki prywatności stron webowych. Ma za zadanie przedstawiać przejrzysty wgląd w to, jak strony WWW wykorzystują dane osobowe swoich użytkowników.

Technologia ta umożliwia serwerom WWW tłumaczenie swoich praktyk związanych z prywatnością osób odwiedzających, do standaryzowanego formatu, odczytywanego przez maszynę (*Extensible Markup Language XML*).

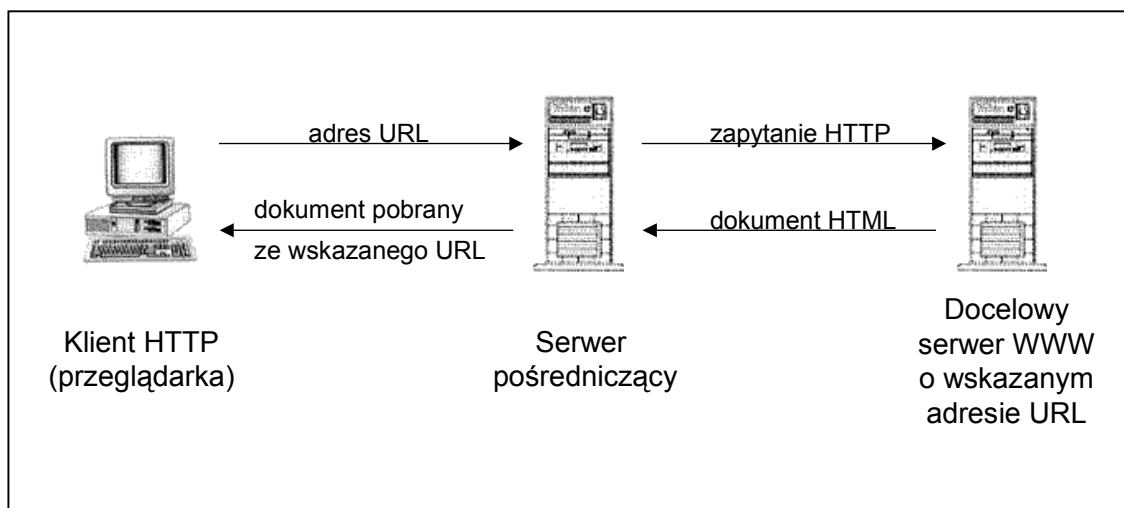
Zapis taki może być następnie automatycznie interpretowany przez przeglądarkę użytkownika. Z punktu widzenia użytkownika, klient P3P automatycznie pobiera i odczytuje zapis polityki bezpieczeństwa danego serwisu WWW. Przeglądarka użytkownika, wyposażona w technologię P3P, może sprawdzić politykę prywatności serwisu i poinformować użytkownika o jego praktykach informacyjnych. Przeglądarka może również porównać zapis ten z ustawieniami bezpieczeństwa wybranymi przez użytkownika. Na tej podstawie dokonywany jest wybór o przyjęciu, bądź blokowaniu zapisu cookies. Klient P3P może być częścią składową przeglądarki, stanowić *plug-in*, lub być zewnętrznym programem.

Obecnie w technologii P3P wyposażonych jest około 260 stron WWW. Czy P3P stanie się faktycznie standardem trudno jeszcze przesądzić. Wiadome jest natomiast, że nie rozwiąże to w pełni problemu utraty anonimowości użytkowników przeglądarek. Popularyzacja P3P może przyczynić się do ograniczenia praktyk tworzenia profili zainteresowań użytkowników, jednak nie pozwoli na ukrywanie informacji o maszynie użytkownika łączącego się z serwerem WWW.

### 3.3. Serwery pośredniczące

Serwery pośredniczące (*proxy*) mogą posiadać wszystkie wymienione wcześniej funkcje, a ponadto skutecznie ukrywać adres URL i IP klienta HTTP. Są także całkowicie niezależne zarówno od stosowanego systemu operacyjnego, oprogramowania (przeglądarka), sprzętu oraz sposobu dostępu do Internetu.

Serwer pośredniczący jest trzecią stroną – „lustrem” odbijającym docelowe zasoby. Pośredniczy w pobieraniu zasobów z WWW. Z punktu widzenia przeglądarki jest to serwer HTTP. Użytkownik, odwiedzając stronę takiego serwera, wypełnia formularz adresem, jaki chce odwiedzić (URL). W odpowiedzi otrzymuje dane, które otrzymałby, wpisując URL w pole adresu przeglądarki. Z punktu widzenia serwera docelowego jest to klient HTTP. Generuje, na podstawie zgłoszonego w formularzu adresu, odpowiednie zapytanie i wysyła je do serwera docelowego. Następnie otrzymuje żądane dane.



Rys. 2. Schemat serwera pośredniczącego WWW w sieci Internet

Dzięki zastosowaniu serwera-lustra, ukrywane są wszelkie informacje o użytkowniku komputera klienta. Dla serwera, z którym chcemy się połączyć, dostępne są jedynie dane o serwerze pośredniczącym.

Serwer pośredniczący daje również szerokie możliwości kontroli i filtrowania przesyłanych zasobów. Możliwe jest, więc zarządzanie mechanizmem cookies oraz blokowanie niepożądanych dodatków (okna reklamowe itp.), jak również usuwanie skryptów, czy programów. Obecnie istnieje już wiele takich systemów, różniących się między sobą głównie poziomem złożoności filtrowania oraz obecnością mechanizmów kryptograficznych. Są to w chwili obecnej m.in.:

- Anonymize.net <http://Anonymize.net>
- Anonymizer <http://www.anonymizer.com>
- Iprive <http://www.iprive.com>
- Magusnet Proxy <http://www.magusnet.com/proxy.html>
- Rewebber <http://www.rewebber.de>
- Surfola <http://www.surfola.com>
- SafeWeb <http://www.safeweb.com>

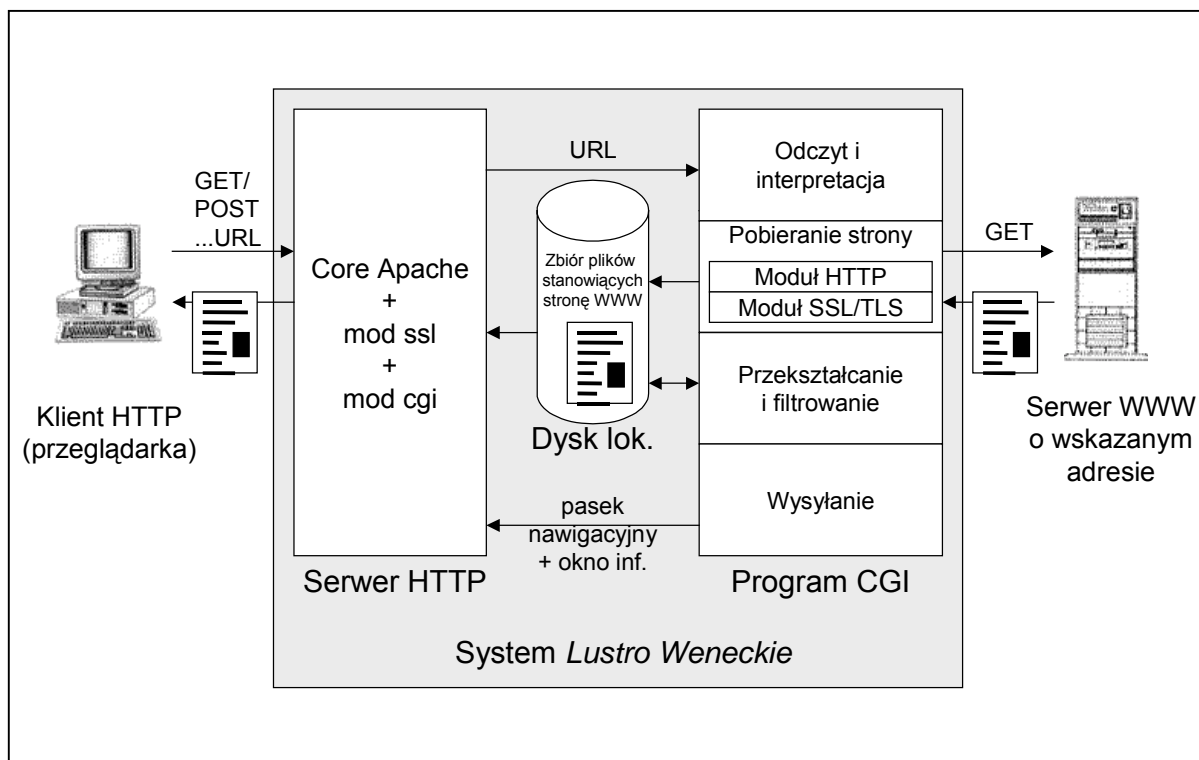
### 3.3.1 Serwery pośredniczące – realizacja

W ramach pracy inżynierskiej autor stworzył system – *Lustro Weneckie* – zapewniający anonimowość przy przeglądaniu stron WWW. Koncepcja działania systemu oparta jest na idei serwera pośredniczącego, stanowiącego dodatkowy węzeł przy pobieraniu stron WWW.

Podstawowe funkcje systemu to:

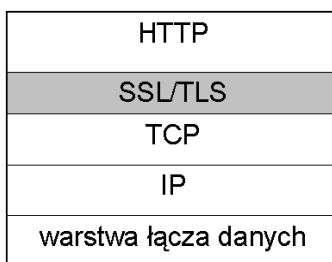
- Ukrywanie wszelkich danych o kliencie HTTP przed serwerem docelowym – w szczególności adres IP.
- Szyfrowanie wszelkich danych przekazywanych od i do klienta HTTP – w szczególności adresu URL zasobów przeglądanych przez użytkownika.
- Opcjonalne szyfrowanie przekazu pomiędzy systemem a docelowym serwerem WWW.
- Blokowanie zapisu cookies pochodzących od serwera docelowego. Blokowanie skryptów i programów pochodzących od serwera docelowego.

System oparty jest na serwerze Apache (*The Apache Software Foundation*), nasłuchującym w systemie operacyjnym Unix. Sercem systemu jest program CGI, odbierający od serwera Apache dane (przede wszystkim URL żądanej strony), pobierający zasoby, filtrujący, oraz przekazujący je do serwera HTTP.



Rys. 3. Schemat własnej realizacji serwera pośredniczącego

Transakcje HTTP zabezpieczane są przez protokół SSL/TLS (*Secure Socket Layer / Transport Layer Security*). Realizuje on tzw. warstwę usług bezpieczeństwa, stanowiącą połączenie między protokołami aplikacyjnymi (m.in. HTTP), a grupą protokołów TCP/IP zarządzającą tworzeniem połączeń. SSL/TLS dostarcza usług poufności, uwierzytelnienia oraz integralności dla poddanych enkapsulacji wiadomości protokołów aplikacyjnych. Używany więc jest identyfikator dostępu URL "https" charakterystyczny przy zastosowaniach do ochrony protokołu HTTP.



Rys. 4. Umieszczenie warstwy pośredniczącej SSL/TLS w modelu warstwowym sieci TCP/IP. Podział zawężony do zastosowań WWW.

Komunikacja między komputerem użytkownika a serwerem pośredniczącym jest zawsze objęta tym zabezpieczeniem. Serwer Apache wzbogacony jest o moduł *mod\_ssl*, autorstwa Ralfa S. Engelschall'a. Natomiast między serwerem pośredniczącym a serwerem docelowym połączenie SSL/TLS nawiązywane jest wtedy, gdy serwer docelowy pozwala na to (obecnie nie wszystkie serwery WWW zawierają obsługę tych protokołów). Do implementacji użyta została biblioteka OpenSSL.

Za pomocą interfejsu CGI (*Common Gateway Interface*) serwer Apache przekazuje do programu (język C++ ze względu na wysokie wymagania co do prędkości) URL żądanej strony wraz z parametrem poziomu bezpieczeństwa ustalonym przez użytkownika. Program ten realizuje dalszą część operacji. W początkowej fazie występuje w roli klienta HTTP. Następuje połączenie z docelowym serwerem WWW. Pierwsza próba połączenia dokonywana jest na porcie 443, jeśli nie uzyskana zostanie odpowiedzi to na porcie 80. Nawiązane zostaje połączenie SSL/TLS lub połączenie niezabezpieczone. Następnie generowane jest odpowiednie zapytanie o dokument. Odbierane dane z serwera docelowego poddawane są filtracji. Konieczne jest przekształcenie odnośników (*links*) na postać kierowaną na serwer pośredniczący. Program pobiera następnie poszczególne pliki (elementy, do których istnieją odwołania na pobieranej stronie) – przede wszystkim pliki graficzne. Następnie dokument może być poddawany dalszej obróbce. Np. może być dokonywane usuwanie skryptów, programów, reklam, itp. Ostatecznie przetworzone dane wysyłane są do klienta. Program generuje również pasek nawigacyjny oraz dodatkowe okno z informacjami o przebiegu połączenia. Uzyskiwane jest to za pomocą JavaScript.

### 3.3.2 Rozwój

Przedstawiony system daje szerokie możliwości rozwoju. Mogą one dotyczyć doskonalenia algorytmów filtrujących pliki HTML. Filtry takie powinny być aktualizowane w stosunku do pojawiających się nowych technologii w Internecie. Również możliwe jest wykorzystanie mechanizmu cookies do zapamiętywania ustawień parametrów bezpieczeństwa poszczególnych użytkowników. Dodatkowo warto wprowadzić algorytm zarządzania cookies tak, by wyeliminować zagrożenia, jakie ten mechanizm niesie oraz zapewnić wygodę korzystania z niego. Rozwiązanie tego problemu można zrealizować w sposób niezależny bądź opierając się na nowym protokole P3P.

**Rozwiązanie niezależne** należy rozumieć przez takie, które stanowi samodzielny mechanizm i nie wymaga zmiany działania węzłów współpracujących z systemem (przeglądarka, docelowy serwer WWW). Rozwiązanie takie może opierać się na trzech filarach:

- „Czarna lista” – baza danych systemu – zawierająca listę adresów serwerów WWW, których zapis cookie jest ignorowany (firmy profilujące).
- Automatyczne odrzucanie cookies pochodzących od „trzeciej strony”.
- System wystosowuje zapytanie do użytkownika o akceptację zapisu cookie w przypadku zmiany adresu IP pobieranych zasobów (nawigacja użytkownik przenosi się do innego serwisu WWW). Jeżeli użytkownik potwierdzi swoje zaufanie do danego serwisu internetowego, system będzie przyjmował cookies, aż do kolejnej zmiany adresu IP.

Innym sposobem zapewnienia bezpiecznego i wygodnego korzystania z mechanizmu cookies jest wzbogacenie systemu o **klienta protokołu P3P**. Dzięki temu system mógłby odczytywać i odpowiednio interpretować politykę prywatności poszczególnych serwisów internetowych, i na tej podstawie podejmować decyzje o przyjęciu bądź odrzuceniu cookie. W chwili obecnej nie może to być rozwiązanie praktyczne, a jedynie eksperymentalne, ponieważ protokół P3P jest w fazie rozwoju i nie stanowi jeszcze faktycznego standardu.

Pasek nawigacyjny, poza polem adresu oraz strzałkami historii, zawierać może menu pozwalające użytkownikowi na konfigurację parametrów bezpieczeństwa. Zrealizowane to może być z wykorzystaniem JavaScript, oraz reguł CSS (*Cascading Style Sheets*).

Kolejną funkcją może być również składowanie (*cache*) na serwerze pośredniczącym stron WWW, do których odnosi się duża liczba odwołań. Takie rozwiązanie stanowić będzie mechanizm podnoszący szybkość pobierania zasobów.

## 4. Przyszłość

Przyszłość tego typu systemów związana jest ściśle z ich wiarygodnością. Prace nad doskonaleniem serwerów pośredniczących powinny być zatem związane z wprowadzeniem mechanizmów potwierdzających brak gromadzenia informacji o użytkownikach usługi.

Do kontrowersyjnych należy możliwość integracji tej usługi z portalami internetowymi, wciąż rozszerzającymi zakres swoich działań także z dziedziny bezpieczeństwa. Mogłoby to przyczynić się do popularyzacji opisanego rozwiązania. Jednak oznaczałoby udostępnienie portalowi szczegółowych danych o nas i o naszych zainteresowaniach. Portale internetowe zabiegają o takie wiadomości, wykorzystują je często w sposób sporny i do końca nie określony, np. udostępniając na zasadach komercyjnych innym firmom.

## 5. Podsumowanie

Serwery pośredniczące cechują się wysoką skutecznością w ukrywaniu danych o kliencie HTTP. Ich dużą zaletą jest także ogólnodostępność oraz szerokie możliwości rozwoju. Izolacja danych, wysyłanych przez serwis WWW i ich analiza jeszcze poza hostem użytkownika, to kolejny ważny czynnik takiego rozwiązania. Do zalet należy również zaliczyć możliwość szyfrowania danych protokołu HTTP wychodzących i przychodzących do komputera użytkownika a przesyłanych do węzła, gdzie wszelkie informacje prywatne użytkownika są usuwane.

Rozwiązanie takie ma jednak również poważne wady. Przede wszystkim jest nią udział trzeciej strony w transakcji HTTP – wszelkie informacje, które ukryte są przed serwerem docelowym, dostępne są dla serwera pośredniczącego. Wadą jest również obniżenie prędkości pobierania stron.

Przedstawiony system, obecnie jako jedyny, pozwala skutecznie chronić prywatność przy przeglądaniu WWW i jednocześnie cechuje się powszechną dostępnością.

## Literatura

1. S.Garfinkel, G. Spaffords, *Web Security & Commerce* O'Reilly and Associates 1999
2. S. Garfinkel, G. Spaffords, *Practical Unix and Internet Security, 2nd Edition* O'Reilly and Associates 1991.
3. R. Kristol, L. Montulli, *HTTP State Management Mechanism*. RFC 2965, October 2000
4. T. Berners-Lee, R. Fielding, H. Frystyk, *Hypertext Transfer Protocol – HTTP/1.0*. RFC 1945, May 1996.
5. R. Fielding, J. Gettys, J. Mogul, H. Frystyk, L. Masinter, P. Leach, T. Berners-Lee, *Hypertext Transfer Protocol – HTTP/1.1*. RFC 2616, June 1999.
6. T. Dierks, C. Allen, *The TLS-Protocol Version 1.0*. RFC 2246, January 1999.
7. L. Cranor, M. Langheinrich, M. Marchiori, M. Presler-Marshall, J. Reagle, *The Platform for Privacy Preferences 1.0 (P3P1.0) Specification* W3C Working Draft, 28 September 2001.
8. I. Margasiński, *Zapewnianie anonimowości przy przeglądaniu stron WWW*, praca dyplomowa, Instytut Telekomunikacji Politechniki Warszawskiej 2002.