# BIEGŁY SĄDOWY

## SĄDU OKRĘGOWEGO W SZCZECINIE

**w zakresie informatyki**

**oraz badań nośników i dokumentów elektronicznych**

inż. Radosław Jedyński

tel. : 504990474, 695206999, e-mail: ekspertyzykomputerowe@interia.pl

70-110 Szczecin, Al. Powstańców Wlkp. 5A/8


Szczecin, dn. 15.11.2013 r.

Egz. nr 2

# OPINIA BIEGŁEGO SĄDOWEGO

# nr 11/2013/VIII GC 260/12

BIEGŁY SĄDOWY
SĄDU OKRĘGOWEGO W SZCZECINIE
*inż. Radosław Jedyński*

## Podstawa przeprowadzenia badań

Badania przeprowadzono na podstawie postanowienia z dnia 19.02.2013 roku wydanego przez Sąd Okręgowy w Szczecinie Wydział VIII Gospodarczy o dopuszczeniu dowodu z pisemnej opinii biegłego, w sprawie z powództwa IAI Spółki Akcyjnej z siedzibą w Szczecinie przeciwko Google Poland spółce z ograniczoną odpowiedzialnością w Warszawie o nakazanie i zapłatę

## Cel badania

1. Wyjaśnienie funkcjonowania systemu blokującego strony internetowe pod względem wyłudzania przez nie danych
2. Ustalenie, czy w przypadku stron internetowych powódki oraz obsługiwanych przez nią sklepów istniała uzasadniona potrzeba wprowadzania komunikatu o wyłudzaniu przez nią danych
3. Ustalenia działań podjętych przez pozwaną zmierzających do usunięcia komunikatu.
4. Ustalenia zakresu kompetencji pozwanej w odniesieniu do zarządzania katalogiem stron, na których znajduje się informacja o podmiotach wyłudzających dane.

## Definicja phishingu

Phishing jest jednym z głównych zagrożeń dla bezpieczeństwa użytkowników sieci i rozwoju usług komercyjnych w Internecie. W raporcie [1] zespołu **CERT Polska** z roku 2011, stwierdzono we wstępie: *„Wśród incydentów obsłużonych nieautomatycznie przez CERT Polska, ponad połowę stanowił Phishing umieszczony w polskich sieciach. Zanotowaliśmy* ***wzrost aż o 1/3 w stosunku do 2010 roku.***" i dalej w dziale dotyczącym phishingu: *„W 2011 roku otrzymaliśmy 266 300 informacji o tradycyjnym phishingu. Informacje te dotyczyły 222 214 różnych adresów URL w 139 770 domenach na 40 091 unikalnych adresach IP."*

Specjalizująca się w bezpieczeństwie elektronicznym firma RSA[1], opublikowała raport pt. *„The year in phishing"*[2] podsumowujący rok 2012, w którym stwierdzono, że całkowita ilość ataków phishingowych odnotowanych przez RSA Anti Fraud Comand Center **wzrosła w stosunku do roku 2011 o 59%.** W raporcie tym Polska znalazła się na 8 miejscu w klasyfikacji miejsc z których pochodziły ataki phishingowe. We wspomnianym raporcie, globalne straty finansowe spowodowane phishingiem w 2012 roku oszacowano u na 1,2 miliarda dolarów, co stanowi wzrost o 11% w stosunku do roku 2011.

W związku z tym, że zjawisko rozwija się dynamicznie, a metody wykorzystywane do wyłudzania danych ewoluują, trudno znaleźć jednolitą definicję phishingu. Biegły odnalazł następujące określenia:

**Polska strona Wikipedii** definiuje to zjawisko w następujący sposób: *„Phishing (spoofing) – w branży komputerowej, wyłudzanie poufnych informacji osobistych (np. haseł lub szczegółów karty kredytowej) przez podszywanie się pod godną zaufania osobę lub instytucję, której te informacje są pilnie potrzebne. Jest to rodzaj ataku opartego na inżynierii społecznej"*[3]

[1] RSA Security LLC – obecnie dział amerykańskiej firmy EMC Corporation

[2] http://www.emc.com/collateral/fraud-report/online-rsa-fraud-report-012013.pdf

[3] http://pl.wikipedia.org/wiki/Phishing

Organizacja **PhishTank**[4] definiuje phishing jako *„próbę oszustwa dokonywaną zwykle za pomocą poczty elektronicznej w celu kradzieży danych użytkownika .... Wiadomości email zawierające phishing zwykle wyglądają na pochodzące od dobrze znanych instytucji i zawierają prośbę o podanie osobistych informacji takich jak, numer karty kredytowej, numer ubezpieczenia, numer konta bankowego lub hasło. .... Aby skutecznie wyłudzić informacje osobiste użytkownika, internetowi przestępcy muszą go skłonić do przejścia z wiadomości email do strony internetowej. Wiadomości email zawierające phishing prawie zawsze sugerują użytkownikowi kliknięcie w link, który przeniesie go do strony internetowej na której wymagane jest wprowadzenie osobistych informacji."* [5]

Organizacja **Mozilla**, producent przeglądarki internetowej **Firefox** definiuje to zjawisko w następujący sposób: *„Fałszerstwo sieciowe (znane też jako "Phishing") jest formą kradzieży tożsamości które występuje kiedy złośliwa strona internetowa podszywa się i udaje legalną w celu skłonienia użytkownika do podania wrażliwych informacji takich jak hasło, szczegółów dotyczących konta, lub numerów kart kredytowych. Ataki phishingowe zwykle zaczynają się od wiadomości email, które usiłują skłonić odbiorcę do aktualizacji jego osobistych informacji na fałszywej lecz wyglądającej bardzo prawdziwie stronie internetowej."*[6]

Firma **Microsoft**, definiuje to zjawisko i instruuje swoich klientów w następujący sposób: *„Phishing (wymawiane "fiszing") to rodzaj kradzieży tożsamości w Internecie. Korzysta z poczty e-mail i fałszywych stron internetowych, które są stworzone by skraść Twoje osobiste dane, lub informacje tj. numer karty kredytowej, hasła, dane konta, i inne.*
*Oszuści mogą wysłać miliony fałszywych wiadomości e-mail zawierających linki do fałszywych stron internetowych, które wydają się pochodzić ze stron internetowych, którym ufasz, tj. Twój bank, firma ubezpieczeniowa, i wymagają by podać dane osobowe. Przestępcy mogą używać tych informacji dla wielu różnych rodzajów oszustw, tj. kradzież pieniędzy z Twojego konta, otwarcie nowych kont w Twoim imieniu, lub uzyskanie oficjalnych dokumentów wykorzystując Twoje dane"*[7]

---

[4] http://www.phishtank.com - strona, którą założyła firma OpenDNS, gromadzi i weryfikuje informacje o próbach phishingu i adresy stron phishingowych, zgłaszane przez społeczność użytkowników serwisu.

[5] http://www.phishtank.com/what_is_phishing.php

[6] http://www.mozilla.org/en-US/firefox/phishing-protection/

[7] http://www.microsoft.com/pl-pl/security/online-privacy/phishing-faq.aspx

Firma **Google Inc.** przedstawiła własne podejście do zdefiniowania phishingu, podczas sympozjum *NDSS 2010* organizacji *Internet Society* w dokumencie [2]

Definicja, którą posługuje się firma Google Inc. w tym dokumencie, ma rozszerzony katalog działań uznawanych za noszące znamiona phishingu cyt.: *„Definiujemy jako stronę phishingową każdą stronę internetową, która bez zezwolenia twierdzi, że działa w imieniu strony trzeciej, w zamiarze wprowadzenia w błąd użytkownika i skłonienia go do podjęcia działań, które powinien podejmować wobec zaufanego przedstawiciela strony trzeciej.*

***Proszę zauważyć, że tak zdefiniowane działania phishingowe nie ograniczają się do zbierania danych osobowych na stronie internetowej.*** *W pewnym sensie, nasza definicja phishingu jest bliższa pojęcia „fałszerstwa sieciowego" - użytego w interfejsie użytkownika przeglądarki Firefox, niż tradycyjnej definicji phishingu.*

*Ta definicja obejmuje oczywiście klasyczny przypadek stron phishingowych, które wyświetlają elementy graficzne kojarzone z firmami finansowymi i żądają poświadczeń potrzebnych do zalogowania użytkownika. Ta definicja obejmuje także strony phishingowe, które wyświetlają logo zaufanej firmy i nakłaniają aby użytkownik pobrał i uruchomił nieznany program. Strony internetowe, które twierdzą, że są w stanie przeprowadzić działania za pośrednictwem strony trzeciej pod warunkiem uzyskania poświadczeń potrzebnych do zalogowania użytkownika również pasują do tak rozszerzonej definicji. Strony internetowe domagające się podania poświadczeń służących do logowania użytkownika, w celu odblokowania konta email lub chatu, także należą do tej ostatniej kategorii.*

*Proszę zauważyć, że jeśli istnienie jednej z tych stron internetowych jest usankcjonowane przez stronę trzecią, wówczas byłyby to działania odpowiednio upoważnione, a zatem nie jest to phishing."*

Jak widać z przytoczonych przykładów phishing jest na ogół definiowany jako działanie z pogranicza inżynierii społecznej i informatyki, polegające na wywieraniu wpływu na nieświadomych fałszerstwa użytkowników poprzez podszywanie się pod zaufany podmiot prowadzący działalność w Internecie, w celu wyłudzenia danych, umożliwiających korzystanie z usług tego podmiotu.

Najczęstszym celem działań phishingowych jest osiągnięcie korzyści majątkowych. Dlatego fałszowane są zwykle strony podmiotów prowadzących w Internecie działalność komercyjną: banków, sklepów internetowych, portali aukcyjnych, podmiotów oferujących wirtualne płatności, dostawców usług itp. Za pomocą odpowiednio spreparowanych stron internetowych wyłudzane są nazwy użytkowników, numery kont bankowych, hasła, numery kart płatniczych, adresy itp.

Znamiona działań phishingowych noszą także czyny, polegające na kradzieży tożsamości wyłącznie w celu przejęcia kontroli nad kontami mailowymi, kontami portali społecznościowych i komunikatorów internetowych. Takich wyłudzeń danych dokonuje się w celu wykorzystania ich do dalszych działań np. do nieuprawnionego wysyłania wiadomości w imieniu użytkownika, internetowego nękania itp.

W zakres nielegalnych działań klasyfikowanych jako phishingowe wchodzą takie zachowania jak:

- wysyłanie odpowiednio spreparowanych e-maili, mających budzić zaufanie odbiorcy i skłonić go do kliknięcia na zawarte w liście łącza prowadzące do sfałszowanych stron, lub podania nazw użytkownika, haseł, numerów kart płatniczych itp.
- tworzenie sfałszowanych stron internetowych, których adres internetowy oraz wygląd wskazuje na to, że ich właścicielem jest podmiot prowadzący legalną i znaną użytkownikom działalność w sieci,
- wyłudzanie danych od nieświadomych użytkowników, którzy odwiedzą tak spreparowane strony

# Zasada działania mechanizmu Google Safe Browsing

Usługa ***Google Safe Browsing***[8] ( nazywana w dalszej części opinii – **GSB** ) została udostępniona przez firmę Google Inc. w roku 2005. Powstała aby zapewnić zautomatyzowaną usługę dokonującą oceny „reputacji" stron internetowych pod kątem stwarzanego przez nie zagrożenia wyłudzania danych (*phishing*) lub rozprzestrzeniania złośliwego oprogramowania (*malware*). Istniejące wcześniej rozwiązania były oparte na czasochłonnej, „ręcznej" analizie stron, prowadzonej przez ekspertów skupionych wokół projektów służących bezpieczeństwu.

Usługa *Google Safe Browsing* składa się z dwóch, współpracujących ze sobą mechanizmów:

- **mechanizmu oceny stron podejrzewanych o wyłudzenia danych, lub rozprzestrzenianie złośliwego oprogramowania.** Ta część mechanizmu GSB sprawdza podejrzane adresy, automatycznie weryfikuje zawartość stron i ocenia czy stanowią zagrożenie dla użytkowników. Jeżeli strona zostanie uznana za stwarzającą zagrożenie jest zapisywana na „czarnej liście" w serwerowej bazie danych GSB. **Tylko ta część mechanizmu GSB dokonuje bezpośredniej oceny zawartości strony internetowej.**
- **usługi API[9] GSB, udostępniającej użytkownikom bazę informacji niebezpiecznych stron poprzez interfejs programistyczny.** Interfejs programistyczny GSB umożliwia utworzenie i okresową aktualizację lokalnej bazy danych GSB klienta oraz zadawanie bezpośrednich pytań *„on-line"* w przypadku wykrycia podejrzanego adresu w lokalnej bazie danych. **W trakcie korzystania z mechanizmów API GSB nie jest dokonywana żadna ocena aktualnej zawartości strony przez klienta.** Z tej części mechanizmu korzystają przeglądarki internetowe *Mozilla Firefox, Google Chrome* i *Apple Safari*. Ponieważ interfejs API usługi GSB jest publicznie dostępny, mogą z niego korzystać także inni programiści w swoich aplikacjach.

[8] **Google Safe Browsing** –ang. Bezpieczne Przeglądanie Google

[9] **API** (ang. Application Programming Interface – interfejs programowania aplikacji ) – ścisły opis reguł, który określa sposób, w jaki programy mogą ze sobą współpracować.

# Zasada działania mechanizmu oceny stron

Biegły zapoznał się z opublikowanym w 2010 roku opisem [2] sposobu działania mechanizmu wykorzystywanego przez firmę Google Inc. do oceny reputacji stron. Na podstawie tego dokumentu biegły przedstawił poniżej zasadę działania, oraz etapy i sposób pracy mechanizmu oceniającego strony internetowe.

### *Etap I -zbieranie informacji o podejrzanych adresach URL*[10]

Punktem wejścia danych mechanizmu klasyfikującego strony internetowe pod kątem stwarzanych przez nie zagrożeń jest aktualizowana na bieżąco kolekcja podejrzanych adresów, która jest zasilana informacjami z dwóch źródeł:

- **raportów przesłanych przez użytkowników** – są to zgłoszenia wyłudzeń przesyłane przez użytkowników końcowych mechanizmu GSB,
- **filtrów antyspamowych**[11] **usługi pocztowej Gmail**[12] **-** według opisu, dopiero odpowiednio duża ilość spamu zawierającego odnośniki URL przesłana na wiele unikalnych kont użytkowników usługi Gmail powoduje zaklasyfikowanie adresu URL do sprawdzenia przez mechanizm GSB.

### *Etap II- wyodrębnienie właściwości adresu URL*

Na tym etapie proces nazywany ***URL Feature Extractor***[13] ocenia konstrukcję adresu URL, sprawdzając, czy nie zawiera cech charakterystycznych dla adresów stron phishingowych. Każda sprawdzana właściwość adresu URL otrzymuje w wyniku tego procesu przypisaną wartość logiczną lub cyfrową i znajduje odzwierciedlenie w opisie badanej strony, który nazywany jest ***„zestawem własności strony"*** (*ang. page feature set*).

---

[10] **URL** – Uniform Resource Locator – ang. ujednolicony format adresowania zasobów, tu używany w znaczeniu adresów stron http. Standard URL opisany jest w dokumencie RFC 1738

[11] **filtr antyspamowy** – rozwiązanie pozwalające na rozpoznawanie, oznaczanie i ewentualne „odsiewanie" niechcianej, „śmieciowej" poczty elektronicznej

[12] **Gmail** – dostępna powszechnie usługa poczty elektronicznej firmy Google Inc.

[13] **URL Feature Extractor** - ekstraktor właściwości/funkcji URL

Na początku tego etapu, sprawdzana jest tzw. „biała lista", na której znajdują się dobrze znane i bezpieczne adresy zaufanych witryn internetowych. W roku 2010 lista zawierała 2778 „ręcznie" sprawdzonych stron internetowych (m.in. *citibank.com* i *cnn.com*). Na listę trafiają wyłącznie witryny, które spełniają jednocześnie dwa warunki: dużej ilości odwiedzin użytkowników i braku podstron z zawartością generowaną przez użytkowników. Jeżeli badany adres URL znajduje się na tak skonstruowanej „białej liście" nie jest przetwarzany dalej. Dzięki temu mechanizm oceniający odrzuca takie strony na samym początku procesu nie zajmując niepotrzebnie zasobów obliczeniowych.

Następnie proces ***URL Feature Extractor*** sprawdza następujące własności badanego adresu URL:

- **czy sprawdzany URL zawiera w nazwie adres numeryczny IP**, np. *http://88.113.105.22/.allegro.htm*[14] - często w atakach phishingowych stosuje się takie adresy URL, aby zamaskować przed użytkownikiem prawdziwego właściciela witryny.

- **czy URL jest zbudowany z wielu segmentów hosta** – przykładem tak spreparowanego adresu URL jest: *http://ssl.allegro.pl.abhinc.org.au*[15], który zawiera 3 segmenty nazwy hosta, udając przed użytkownikiem adres serwisu aukcyjnego Allegro, którym w rzeczywistości nie jest (domena zarejestrowana była w Australii).

**Przykład podejrzanego adresu URL zawierającego 3 segmenty nazwy hosta**


[14] źródło – www.phishtank.org – id 1943961 - potwierdzony przypadek strony wyłudzającej dane

[15] źródło – www.phishtank.org – id 1973953 - potwierdzony przypadek strony wyłudzającej dane

Ponieważ adresy URL stron phishingowych często zawierają charakterystyczne słowa kluczowe np.: *login, online banking, signin, customerservice, authentication, username,* oraz nazwy znanych marek handlowych, proces ***URL Feature Extractor*** wyodrębnia wszystkie obecne w adresie URL łańcuchy znakowe i zamienia każdy z nich w wartości logiczne np. „adres zawiera łańcuch *'login'*", dodając każdą taką własność do tworzonego zestawu własności strony.

Uzyskany w ten sposób zbiór łańcuchów znakowych jest wykorzystywany przez mechanizm GSB do badania częstości ich występowania oraz wyszukiwania wzorców takich fraz wyrazów, które są wspólne dla ataków phishingowych.

Ponadto proces ***URL Feature Extractor*** sprawdza za pomocą parametru ***PageRank***[16] popularność adresu URL oraz poziom *„reputacji domeny"*[17], wyliczony przez system antyspamowy usługi Gmail. Według firmy Google Inc. prawdopodobieństwo phishingu jest mniejsze dla adresów URL pochodzących z domen internetowych, które wysyłają mało spamu.

### *Etap-III - pobieranie zawartości strony*

Proces ***Content Fetcher***[18] pobiera treść strony dostępnej pod wskazanym adresem URL i gromadzi informacje na temat hostingu[19] badanej witryny:

- rozpoznaje nazwę i zapisuje adresy IP strony,
- sprawdza serwery DNS oraz adresy IP serwerów DNS
- sprawdza geograficzne położenie adresów IP oraz miasto, region i kraj.

Następnie badany adres URL jest przekazywany do puli bezobsługowych przeglądarek internetowych w celu wyświetlenia strony w symulowanym środowisku przeciętnego użytkownika. Kod źródłowy strony HTML, obiekty typu *iframe*, występujące na niej elementy graficzne oraz wbudowany kod javascript jest zapisywany w obiekcie opisu badanego adresu URL w bazie danych.

---

[16] **PageRank** – opatentowany w 1998 roku przez Uniwersytet Stanforda sposób oceny jakości i popularności stron internetowych. W metodzie tej im więcej innych stron o wysokim PageRank odwołuje się w swojej treści do badanej strony, tym lepsza jest z punktu widzenia użytkowników jakość treści tej strony i tym większa wartość PageRank jest jej przypisywana w systemie indeksowania i wyszukiwania Google Inc.

[17] **reputacja domeny** – (wg. Google Inc.) w przybliżeniu jest to procent emaili wysłanych z domeny, które nie są spamem.

[18] **Content Fetcher** - pobieracz treści

[19] **Hosting** – usługa polegająca na udostępnianiu zasobów serwerowni: łącza internetowego, serwerów.

### *Etap-IV - wyodrębnianie właściwości strony i jej hostingu*

Na podstawie uzyskanego w poprzednim etapie adresu IP, proces o nazwie ***Page Feature Extractor***[20] pogłębia informacje o hostingu badanej strony sprawdzając numer systemu autonomicznego **AS**[21]. Pozwala to na zweryfikowanie położenia geograficznego witryny, oraz sprawdzenie, czy inne strony podejrzewane o wyłudzanie danych są hostowane w podobny sposób (np. na tym samym serwerze, w tym samym **ASN**[22]). Przytoczone w dokumencie badania wskazują bowiem, że o ile gospodarzem największej ilości stron phishingowych są Stany Zjednoczone, o tyle procentowy udział stron wyłudzających dane jest największy w Meksyku (70,1%), Izraelu (48,6%) oraz krajach Europy Środkowej (Rumunia41,6%,Węgry 37,2%, Polska 32,7%).
Proces ***Page Feature Extractor*** analizuje odwołania do adresów IP wszystkich elementów składowych badanej strony i weryfikuje posiadane informacje o wykorzystywanych przez stronę numerach sieci ASN.

**Kolejną ze sprawdzanych własności strony jest stopień, w jakim występujące w kodzie odnośniki HTML (linki) oraz elementy graficzne, odwołują się do domen internetowych innych niż domena własna badanej strony**. W wielu przypadkach strony wyłudzające dane istnieją bardzo krótko, więc ich twórcom nie opłaca się kopiować wszystkich elementów graficznych z podrabianej strony. W związku z tym, często elementy graficzne na stronie sfałszowanej są pobierane podczas wyświetlania przez użytkownika bezpośrednio ze strony prawdziwej. **W kodzie HTML sfałszowanej strony można więc odnaleźć odwołania, które wskazują na domenę prawdziwej strony.**

Proces ***Page Feature Extractor***, bazując na uzyskanym w poprzednim etapie kodzie HTML strony, **bada kolejno częstość występowania wszystkich łańcuchów znakowych na podejrzanej stronie, w stosunku do częstości występowania tych samych wyrazów na innych stronach**. Stosowana jest metoda statystycznej oceny ważności słów

---

[20] **Page Feature Extractor** – ekstraktor właściwości strony.

[21] **AS** –System autonomiczny (ang. Autonomous System, AS) to zbiór prefiksów (adresów sieci IP) pod wspólną administracyjną kontrolą, w którym utrzymywany jest spójny schemat trasowania (ang. routing policy). Oryginalna definicja zawarta w RFC 1771 odnosi się do sieci lub grupy sieci opartych na protokole IP, lecz została później zmieniona w RFC 1930. – cyt. z Wikipedii

[22] **ASN** – identyfikator numeryczny określający system autonomiczny. Nadawany przez organizację IANA

kluczowych **TF-IDF**[23]. Wartość TF-IDF wyrazu na stronie jest wyrażana jako częstość występowania wyrazu na badanej stronie (*częstość wyrazu*), podzielona przez logarytm dziesiętny częstości występowania tego wyrazu na wszystkich stronach (*częstość dokumentu*). Przy określaniu częstości występowania wyrazu na innych stronach, brane są pod uwagę tylko strony występujące w mechanizmach wyszukiwania firmy Google Inc. w tym samym języku co strona podejrzana. W ten sposób, każdemu łańcuchowi znakowemu strony zostaje przypisana jego waga (wartość), która określa ważność (istotność) tego wyrazu dla strony.

Fałszerze tworzą witryny łudząco podobne do prawdziwych, więc korzystają z wyrazów i sformułowań występujących na stronach oryginalnych. Według firmy Google Inc., wyrazy z najwyższym parametrem TF-IDF na stronie fałszywej odzwierciedlają to podobieństwo do oryginałów, inne strony używają charakterystycznych zwrotów zbyt rzadko, aby osiągnęły one wysoki poziom TF-IDF.

**Ostatnim krokiem tego etapu, jest sprawdzenie, czy strona używa formularza z polem wprowadzania hasła**, ponieważ większość stron phishingowych stosuje taki wybieg w celu wyłudzenia danych użytkownika w tym kradzieży hasła. Przy tym, firma Google Inc. stwierdza w dokumencie opisującym mechanizm oceny reputacji stron, że strony posiadające taki formularz, które nie wyłudzają danych łatwo można odróżnić na postawie innych własności.

### *Etap-V - klasyfikacja strony*

Na podstawie danych zgromadzonych w poprzednich etapach, mechanizm GSB ocenia stronę w skali od 0.0 do 1.0, gdzie ocena 0.0 oznacza, że strona na pewno nie wyłudza danych, a 1.0 że definitywnie jest to strona wyłudzająca dane. Do obliczania tego prawdopodobieństwa stosuje się oddzielny proces uczący, który wylicza modele, którymi przeładowuje codziennie proces klasyfikatora.

[23] **TF-IDF** (ang. TF – term frequency, IDF – inverse document frequency) - ważenie częstością termów - odwrotna częstość w dokumentach - jedna z metod obliczania wagi słów w oparciu o liczbę ich wystąpień, należąca do grupy algorytmów obliczających statystyczne wagi termów. Każdy dokument reprezentowany jest przez wektor, składający się z wag słów występujących w tym dokumencie. TFIDF informuje o częstości wystąpienia termów uwzględniając jednocześnie odpowiednie wyważenie znaczenia lokalnego termu i jego znaczenia w kontekście pełnej kolekcji dokumentów. – cyt. z Wikipedii

Klasyfikator przyporządkowuje wartości liczbowe poszczególnym cechom strony zebranym, sklasyfikowanym i zapisanym w poprzednich czterech etapach w obiekcie nazywanym **zestawem własności strony**.

| **Zestaw własności strony** | |
|---|---|
| **Własność strony** | **Typ danych** |
| **Własności adresu URL** | |
| Adres IP w nazwie hosta | boolean |
| Długa nazwa hosta | boolean |
| Łańcuchy wyrazów w URL | booleans |
| PageRank | float |
| reputacja w Gmail | float |
| **Własności hostingu** | |
| Numery ASN | booleans |
| Geolokalizacje | booleans |
| **Własności strony** | |
| Pole tekstowe do wprowadzania hasła | boolean |
| Top TF-IDF Terms | booleans |
| Częstość linkowania zewnętrznego | float |

W zależności od typu badanej własności, przyporządkowywane są następujące wartości:

- własności typu *boolean*[24]: prawda otrzymuje wartość 1.0; fałsz 0.0,
- własności typu *float*[25], otrzymują wartości ze skali od 0.0 do 1.0,
- własności typu określanego w dokumencie jako *booleans*, lub jako *rzadki zestaw własności binarnych*[26] , mają wartość „prawda" (1.0), jeżeli którykolwiek z ich elementów ma wartość „prawda" (1.0).
- jeśli cecha nie występuje w zestawie, przypisywana jest jej wartość 0.0

[24] **boolean** – dane typu logicznego, przyjmują wartość prawda lub fałsz.

[25] **float** – typ liczbowy, zmiennoprzecinkowy

[26] **sparse set of binary features** – sposób reprezentowania danych polegający na tym, że własności badanego obiektu przedstawia się w postaci długiego słowa binarnego, w którym każdy konkretna pozycja bitu w słowie odpowiada, za istnienie (1), lub nieistnienie (0) konkretnej cechy tego obiektu.

| Przykład zastosowania danych typu *booleans* do opisu łańcuchów znakowych występujących w adresie URL |
|---|
|  |



Proces klasyfikatora oblicza wynik w postaci logarytmu szans, podstawiając wartości cech z zestawu własności strony. Niestety, w dokumencie [2] nie podano szczegółów procesu obliczeniowego, ani przykładu takich obliczeń. Autorzy sugerują jedynie, że wykorzystywane są mechanizmy stosujące metodę regresji logistycznej. Model regresji logistycznej jest stosowany w statystyce, do badania wpływu wielu czynników na zmienną dychotomiczną (przyjmującą dwie wartości). Zastosowanie takiego podejścia pozwala na badanie wpływu wielu cech zebranych w ***zestawie własności strony*** na prawdopodobieństwo, że badana strona wyłudza dane. W dokumencie [2] podano wzór na ostateczny wynik:

$$wynik = \frac{e^{logodds}}{1 + e^{logodds}}$$

gdzie:

***wynik*** – oznacza prawdopodobieństwo, że strona wyłudza dane,

***e*** - jest podstawą logarytmu naturalnego[27],

***logodds***- jest logarytmem naturalnym ilorazu szans, wyliczanym z wzoru funkcji *logit(p)*, stosowanej w statystyce do przekształcania prawdopodobieństwa na iloraz szans:

$$logodds = logit(p) = \ln(odds) = ln\frac{p}{1-p}$$

gdzie:

***p*** – jest prawdopodobieństwem wystąpienia pewnego zdarzenia,
***odds*** – iloraz szans, jest stosunkiem prawdopodobieństwa wystąpienia zjawiska do prawdopodobieństwa, że zjawisko nie wystąpi

[27] **liczba Eulera** – podstawa logarytmu naturalnego, e≈2,7182818

Jeżeli wynik obliczeń jest większy niż 0,5, to prawdopodobieństwo, że strona wyłudza dane jest wyższe niż to, że jest stroną bezpieczną i klasyfikator oznacza adres URL jako podejrzany.

**Istotne dla niniejszej opinii jest to, że autorzy podkreślają w dokumencie** [2], **znaczenie parametru *Page Rank* w procesie klasyfikacji**. Zanim adres URL zostanie wpisany na *„czarną listę"*, klasyfikator sprawdza odpowiadający adresowi parametr ***Page Rank***. Jeżeli adres ma wysoką wartość ***Page Rank***, prawdopodobieństwo, że strona jest niebezpieczna jest niskie i może oznaczać błąd klasyfikacji. **Dla stron popularnych wymagana jest „ręczna" weryfikacja treści strony.** Według autorów dokumentu, taka dodatkowa procedura zabezpiecza mechanizm klasyfikacji stron przed wysokim poziomem błędów typu *false positive*[28]. Niestety, autorzy dokumentu nie określili znaczenia użytego w dokumencie sformułowania *„wysoka wartość Page Rank"*, więc nie wiadomo od jakiego poziomu tej wartości obowiązuje procedura dodatkowego sprawdzenia strony przez człowieka.

### *Etap-VI - agregacja i udostępnianie listy podejrzanych stron*

Proces nazywany ***Blacklist Agregator*** przygotowuje listy podejrzanych stron do publikacji. Czarna lista podejrzanych adresów URL jest przekształcana zgodnie z wymaganiami stawianymi przez ***Safe Browsing API v2***, w wyrażenia typu *host-suffix/ path–prefix*[29]. Następnie, w celu zmniejszenia rozmiaru czarnej listy i zmniejszenia występującej na niej liczby wzorców, lista podejrzanych adresów jest poddawana transformacji przez tzw. *algorytm poszerzający*. Jeśli pewna liczba adresów URL znajdujących się na czarnej liście, pasuje do szerszego wzorca, algorytm poszerzający umieszcza na czarnej liście ten wzorzec zamiast adresów URL. Jak wyjaśniono w dokumencie [2], ataki phishingowe często wykorzystują adresy, które różnią się tylko częścią hosta lub ścieżki w adresie URL. Stosowany *algorytm poszerzający* pozwala zmniejszyć ilość wpisów na czarnej liście do jednego adresu.

---

[28] **błąd I rodzaju** (ang. *false positive*) – polega na odrzuceniu hipotezy, która w rzeczywistości jest prawdziwa (błąd typu „fałszywy alarm")

[29] **host-suffix/path–prefix** – z ang. wyrażenia typu host-przyrostek/ ścieżka-przedrostek, np: "google.com/", "some.host.com/123/", "otherhost.net/some/url.html?q=123"

Takie podejście poprawia według autorów dokumentu tzw. *pokrycie* „czarnej listy" poprzez blokowanie adresów URL, które są częścią ataku typu phishing, ale zostały pominięte przez system klasyfikacji GSB. Aby uniknąć klasyfikacji zbyt agresywnej, *algorytm poszerzający* unika wzorców, które pasują do adresów witryn o wysokim **PageRank**. Takie rozwiązanie zabezpiecza mechanizm GSB przed dodaniem do czarnej listy adresów tzw. top-level URL popularnych witryn pomimo, że witryna może zawierać kilka stron phishingowych.

Tak przygotowane czarne listy podejrzanych adresów są przekształcane przez ***Blacklist Agregator*** do formatu zdefiniowanego przez ***Safe Browsing API v2*** i udostępniane programom klientów API.

# API GSB - udostępnianie listy podejrzanych stron

W tej części opinii, biegły opisze zasadę działania systemu udostępniania informacji o podejrzanych stronach mechanizmu ***Google Safe Browsing***. Ze względu na to, że ta część mechanizmu GSB współpracuje z aplikacjami użytkowników końcowych, jest dobrze udokumentowana przez firmę Google Inc. [3] [4] [5]. Z tej części usługi mogą korzystać programy przeglądarek internetowych, programy antywirusowe itp.

Usługa ***Google Safe Browsing*** umożliwia aplikacjom użytkowników końcowych, sprawdzania „reputacji" adresów URL za pomocą dwóch rodzajów interfejsów programowania aplikacji (API):

1. **Safe Browsing API v2** – interfejs API, który umożliwia klientom usługi pobranie z serwerów firmy Google Inc. zaszyfrowanej[30] tabeli w celu lokalnego (*off-line*) sprawdzania adresów URL. Metoda ta cechuje się następującymi właściwościami:
   - zapewnia lepszą ochronę prywatności – klient, zamiast podejrzanych adresów URL wymienia z serwerami GSB wyłącznie skróty SHA256 sprawdzanych adresów.
   - lokalna baza podejrzanych adresów URL umożliwia szybsze wyszukiwanie, które nie wymaga każdorazowego komunikowania się z serwerami GSB.
   - wymaga od aplikacji klienta API znajomości struktury wewnętrznej bazy danych GSB, która przechowuje skróty SHA256 podejrzanych adresów URL.
   - wymaga od aplikacji klienta API zaimplementowania funkcji skrótu SHA256.
   - aplikacja klienta API musi w wymaganym czasie kontaktować się z serwerami GSB, sprawdzać aktualizacje podejrzanych adresów URL oraz pobierać te aktualizacje.
   - aplikacja klienta API musi implementować funkcję, która przekształca badany adres URL do postaci kanonicznej opisanej przez Google Inc. w [4] [5].

[30] tabela przechowuje pierwsze 32 bity (4 bajty) skrótów z adresów URL otrzymanych za pomocą jednokierunkowej (nieodwracalnej) funkcji skrótu SHA-256, która każdy łańcuch znakowy przekształca w 256 bitowy (32 bajtowy), unikalny ciąg sumy kontrolnej

2. **Safe Browsing Lookup API** – interfejs API, który umożliwia wyszukiwanie adresów URL oraz sprawdzanie ich stanu bezpośrednio (*on-line*) na serwerach usługi GSB. *Safe Browsing Lookup API* charakteryzuje się następującymi cechami:
   - jest prostszy w implementacji – aplikacja klienta API musi tylko odpowiednio „opakować" sprawdzany URL i przesłać go za pomocą metody HTTP GET lub POST do serwerów GSB.
   - gorzej chroni prywatność użytkowników – sprawdzany adres URL jest przesyłany do serwerów GSB.
   - nie daje żadnej gwarancji w zakresie czasu odpowiedzi serwerów GSB na żądanie sprawdzenia URL.

## *Opis metody Safe Browsing API v2*

Klient API[31] usługi GSB bazuje, na publikowanych przez Google Inc. listach znanych stron internetowych, wyłudzających dane (*phishing*), lub rozprzestrzeniających złośliwe oprogramowanie (*malware*). Listy zawierają 32 bajtowe skróty SHA256 wyrażeń utworzonych z adresów URL określonych w [3] jako *host-suffix/ path–prefix*.[32]

Aby zapewnić odpowiednią wydajność, aplikacja korzystająca z tego rodzaju API, pobiera z serwerów GSB dane w postaci tzw. *chunks* (jednostek/kawałków), zawierających 4 początkowe bajty – prefiksy skrótów SHA256 podejrzanych adresów. Safe Browsing API przewiduje 2 rodzaje takich danych:

- ***add chunks*** – zawierają 4 bajtowe prefiksy skrótów nowych adresów URL, które klient API powinien dodać (ang. *add*) do lokalnej bazy danych
- ***sub chunks*** - zawierają 4 bajtowe prefiksy skrótów, które klient powinien usunąć ze swojej bazy danych (np. fałszywe alarmy, po weryfikacji)

Usługa GSB prowadzi dwie listy adresów, podzielone na niezależnie numerowane porcje danych (*chunks*):

- ***goog-phish-shavar*** – lista adresów URL wyłudzających dane (*phishing*)
- ***goog-malware-shavar*** – lista adresów URL rozpowszechniających złośliwe oprogramowanie (*malware*)

---

[31] **klient API** - program korzystający z opisywanego API, np. przeglądarka internetowa, program antywirusowy itp.

[32] **host-suffix/path–prefix** – z ang. wyrażenia typu host-przyrostek/ ścieżka-przedrostek, np: "google.com/", "some.host.com/123/", "otherhost.net/some/url.html?q=123"

Klient API musi rozpocząć proces aktualizacji lokalnej bazy przed upływem 5 minut od swojego uruchomienia. Drugie żądanie aktualizacji powinno nastąpić po czasie określonym przez serwer GSB lub w okresie od 15 do 45 minut po pierwszym. Kiedy program klienta API chce zaktualizować swoją lokalną bazę danych, korzysta z metody API: ***HTTP Request for Data*** opisanej w [4] i [5]. Program kontaktuje się za pomocą metody POST protokołu HTTP z serwerami GSB pod adresem:

```
http://safebrowsing.clients.google.com/safebrowsing/downloads?clien
t=myapplication&appver=1.5.2&pver=2.2
```

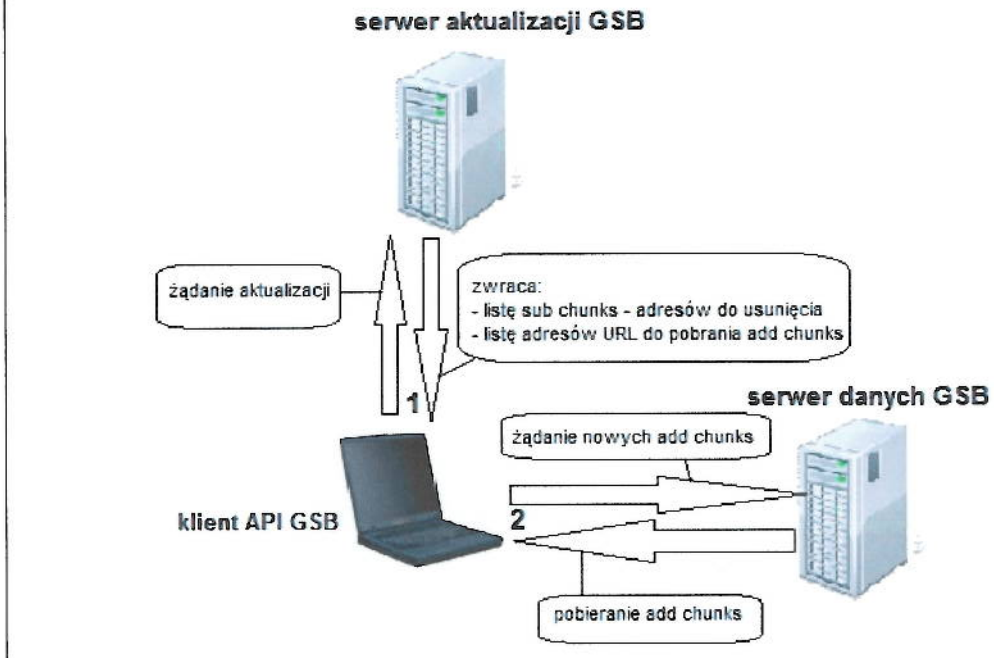
Następnie wysyła listę wszystkich posiadanych jednostek danych *add chunks* i *sub chunks* w następujący sposób:

```
goog-phish-shavar;a:1-3,5,8:s:4-5
```

co oznacza, że klient posiada wszystkie jednostki danych *add chunks* od 1 do 3, 5 i 8 oraz *sub chunks* od 4 do 5.

W odpowiedzi serwer zwraca serię przekierowań do adresów URL, które zawierają jednostki danych *add chunks* i *sub chunks* do pobrania dla klienta API oraz czas, przez który klient API musi odczekać przed ponownym kontaktem z serwerem. Klient pobiera jednostki danych typu *add chunks* i dodaje do lokalnej bazy danych. Pobiera również dane typu *sub chunks*.

**Przebieg procesu aktualizacji danych klienta API Safe Browsing v2**



Zanim program klienta API pobierze z adresu URL zawartość strony i wyświetli ją użytkownikowi, musi sprawdzić, czy adres URL znajduje się na „czarnej liście". W celu sprawdzenia, program musi stworzyć zbiór wszystkich wyrażeń typu *host-suffix/ path–prefix*, możliwych do utworzenia dla danego adresu URL. Jak wspomniano w [3], dla przykładowego adresu:

```
"http://www.nazwahosta.com/path/page.html?args",
```

będą to wyrażenia:

```
www.nazwahosta.com/path/page.html?args
www.nazwahosta.com/path/page.html
www.nazwahosta.com/path/
www.nazwahosta.com/
nazwahosta.com/path/page.html?args
nazwahosta.com/path/page.html
nazwahosta.com/path/
nazwahosta.com/
```

Program oblicza skróty SHA256 dla każdego z powstałych wyrażeń. Następnie klient API przeszukuje lokalną bazę danych i porównuje pierwsze 4 bajty każdego obliczonego skrótu, z posiadanymi w lokalnej bazie danych prefiksami skrótów podejrzanych stron.

**Znalezienie w bazie dopasowania nie oznacza, że sprawdzany adres URL jest podejrzany i nie jest to wystarczający powód do zablokowania adresu URL przez program**. W takim wypadku program musi zażądać od serwera GSB listy wszystkich 32 bajtowych skrótów SHA256 adresów URL, których pierwsze 4 bajty pasują do odnalezionych przez klienta. W celu uzyskania listy pełnych skrótów, program korzysta z metody API: ***HTTP Request for Full-Length Hashes*** i kontaktuje się za pomocą metody POST protokołu HTTP z serwerem GSB pod adresem:

```
http://safebrowsing.clients.google.com/safebrowsing/gethash?client=
myapplication&appver=1.5.2&pver=2.2
```

Serwer odpowiada listą pełnych, 32 bajtowych skrótów, z których każdy zawiera nazwę listy (*goog-phish-shavar* lub *goog-malware-shavar*) i numer jednostki danych (*chunk*). Każdy pełny skrót SHA256 adresu URL zwrócony przez serwer GSB, musi być zapamiętany przez klienta w lokalnej bazie danych i przypisany odpowiadającemu mu 4 bajtowemu prefiksowi, tak długo, jak długo prefiks skrótu jest aktualny.
Program klienta API sprawdza listę pełnych skrótów. **Jeżeli znajdzie dopasowanie do skrótu pełnego adresu URL, klient API zatrzymuje pobieranie danych i wyświetla komunikat z ostrzeżeniem.**

Firma Google Inc. zastrzega jednak [4] [5], że program klienta nie może używać danych starszych niż 45 minut, a wyświetlenie komunikatu ostrzegającego użytkownika może nastąpić wyłącznie, gdy występuje jeden z 3 scenariuszy:

1. Sprawdzany adres URL pasuje do pełnego 32 bajtowego skrótu SHA256, który został zwrócony przez serwer GSB jako jednostka danych *add chunk*, w odpowiedzi na żądanie API: ***HTTP Request for Data***, pod warunkiem, że taki skrót nie został usunięty z listy (np. poprzez jednostkę danych typu *sub chunk*), oraz pod warunkiem, że lista została pomyślnie zaktualizowana za pomocą żądania API: ***HTTP Request for Data*** (gdy cała aktualizacja zakończyła się powodzeniem) w ciągu ostatnich 45 minut liczonych od chwili ,w której komunikat ostrzegawczy ma być wyświetlony, lub
2. Sprawdzany adres URL pasuje do pełnego 32 bajtowego skrótu SHA256, który został zwrócony przez serwer GSB w odpowiedzi na żądanie API: ***HTTP Request for Full-Length Hashes***, pod warunkiem, że odpowiadający mu 4 bajtowy prefiks nie został usunięty z listy (np. poprzez jednostkę danych typu *sub chunk*), oraz pod warunkiem, że lista została pomyślnie zaktualizowana za pomocą żądania API: ***HTTP Request for Data*** (gdy cała aktualizacja zakończyła się powodzeniem) w ciągu ostatnich 45 minut liczonych od chwili ,w której komunikat ostrzegawczy ma być wyświetlony, lub
3. Sprawdzany adres URL pasuje do pełnego 32 bajtowego skrótu SHA256, który został zwrócony przez serwer GSB w odpowiedzi na żądanie API: ***HTTP Request for Full-Length Hashes***, w ciągu ostatnich 45 minut liczonych od chwili ,w której komunikat ostrzegawczy ma być wyświetlony, pod warunkiem, że taki skrót nie został następnie usunięty z listy (np. poprzez jednostkę danych typu *sub chunk*)

Google Inc. w opisie API[33] oraz w *Warunkach korzystania z usługi* [5] **wymaga, aby przestrzegać poniższych zasad w zakresie ostrzegania użytkowników końcowych przed podejrzanymi stronami i używanych w komunikacie ostrzegawczym sformułowań:**

1. Jeżeli program lub usługa korzystająca z API GSB informuje użytkowników końcowych, że zapewnia ochronę przed wyłudzaniem danych lub złośliwym oprogramowaniem, **musi także informować, że mechanizmy ochrony nie są doskonałe.**

[33] https://developers.google.com/safe-browsing/developers_guide_v2?csw=1#UserWarnings

2. **informacja musi być widoczna dla użytkowników zanim włączą ochronę** i musi dawać im do zrozumienia, że korzystanie z mechanizmów ochrony nie jest wolne od występowania zarówno ***błędów I rodzaju***[34] (bezpieczne strony oznaczone jako niebezpieczne), jak i ***błędów II rodzaju***[35] (ryzykowne strony, które nie zostały oznaczone jako niebezpieczne). Sugerowany, język komunikatu powinien wyglądać następująco: „*Google pracuje nad tym, aby zapewnić najbardziej dokładne i aktualne informacje o stronach wyłudzających dane i rozprzestrzeniających złośliwe oprogramowanie. Jednak nie może zagwarantować, że jego (Google) informacje są pełne i bezbłędne: niektórych ryzykownych adresów nie można zidentyfikować, a niektóre bezpieczne witryny mogą być zidentyfikowane błędzie jako ryzykowne.*"
3. Widoczne dla użytkownika końcowego **ostrzeżenie przed odwiedzeniem podejrzanego adresu URL nie może prowadzić do przekonania, że strona której dotyczy ostrzeżenie, jest bez wątpienia stroną wyłudzającą dane lub rozprzestrzeniającą złośliwe oprogramowanie.** W komunikacie odnoszącym się do strony, która została zidentyfikowana jako potencjalnie ryzykowna dla użytkownika, jako ostrzeżeń **należy używać terminów takich jak: podejrzenie, potencjalnie możliwe, prawdopodobne, może być.**
4. **Ostrzeżenie musi umożliwiać użytkownikowi pogłębienie wiedzy na temat zagrożenia** poprzez umożliwienie mu przejścia do stron
   - *http://www.antiphishing.org/* (dla ostrzeżeń o phishingu)
   - *http://www.stopbadware.org/* (dla ostrzeżeń o malware).
5. Ostrzeżenie musi zawierać sformułowanie *„Porada podana przez Google"*, z linkiem do: *http://code.google.com/apis/safebrowsing/safebrowsing_faq.html#whyAdvisory*. Jeśli klient API GSB pokazuje również ostrzeżenia na podstawie innych źródeł, to ostrzeżenia pochodzące z tych źródeł nie mogą wskazywać w komunikacie na Google jako źródło tych ostrzeżeń.
6. **Klient API GSB nie może wyświetlać komunikatu ostrzegającego użytkownika** o podejrzanym adresie URL ani blokować jego wyświetlania **jeśli w ciągu ostatnich 30 minut nie otrzymał z serwerów GSB aktualizacji danych** dotyczących podejrzanych adresów

---

[34] **błąd I rodzaju** (ang. *false positive*) – polega na odrzuceniu hipotezy, która w rzeczywistości jest prawdziwa (błąd typu „fałszywy alarm")

[35] **błąd II rodzaju** (ang. *false negative*) – błąd przyjęcia, polega na nieodrzuceniu hipotezy, która w rzeczywistości jest fałszywa (błąd „przeoczenia")

W opisie interfejsu **Safe Browsing API** [5], w rozdziale *„Sugerowany język ostrzeżeń"* przedstawiono rekomendowane przez Google Inc. formy komunikatów, które zgodnie z intencją twórcy można skopiować i stosować w komunikatach ostrzeżeń klienta API :

- **Ostrzeżenie: Strona podejrzana o wyłudzanie danych (phishing).** Ta strona może być falsyfikatem lub imitacją innej strony, powstałą w celu wyłudzenia danych osobowych lub finansowych. Wprowadzanie jakichkolwiek informacji osobistych na tej stronie może skutkować kradzieżą tożsamości lub innym nadużyciem. Możesz dowiedzieć się więcej na temat phishingu pod adresem *www.antiphishing.org.*
- **Ostrzeżenie: przejście do tej witryny może spowodować uszkodzenie komputera**. Ta strona wydaje się zawierać złośliwy kod, który może być pobrany do komputera bez Twojej zgody. Możesz dowiedzieć się więcej na temat szkodliwych treści internetowych, w tym wirusów i innego złośliwego kodu, oraz jak chronić komputer na *StopBadware.org.*

## *Opis metody Safe Lookup API*

Interfejs **Safe Lookup API** umożliwia sprawdzanie „reputacji" adresów URL bezpośrednio (*on-line*) na serwerach usługi GSB. Od poprzedniego interfejsu odróżnia się tym, że nie daje żadnej gwarancji szybkiego dostarczenia odpowiedzi przez serwery usługi GSB. Klient API ma do wyboru dwa rodzaje komunikowania się z usługa GSB za pomocą tego API. Obie odmiany (HTTP GET i HTTP POST) korzystają z połączenia szyfrowanego protokołem HTTPS.

### *Metoda Safe Lookup API - HTTP GET*

Za pomocą tej metody klient API może sprawdzić tylko jeden URL w jednym żądaniu do serwera GSB. Sprawdzany adres URL musi być odpowiednio zakodowany i przesłany jako jeden z parametrów w sekcji zapytania URL. Klient API stosując tą metodę wysyła żądanie na adres serwera GSB w następujący sposób:

```
https://sb-ssl.google.com/safebrowsing/api/lookup?client=CLIENT
&apikey=APIKEY &appver=APPVER&pver=PVER&url=URL
```

gdzie URL oznacza zakodowany zgodnie z RFC 3986 adres internetowy, który ma być sprawdzony w usłudze Google Safe Browsing.

Jeżeli żądanie wysłane przez klienta API, było poprawne składniowo, serwer odpowiada:

- kodem 204, jeśli sprawdzanego adresu nie ma na „czarnej liście" usługi GSB
- kodem 200, po którym mogą wystąpić odpowiedzi:
  - *„phishing"* - dla stron podejrzewanych o wyłudzenia danych,
  - *„malware"* - dla stron podejrzewanych o rozpowszechnianie złośliwego oprogramowania,
  - *„phishing, malware"* – dla stron, które podejrzane są o oba ryzykowne dla użytkownika zachowania.

***Metoda Safe Lookup API - HTTP POST.***

Dzięki własnościom protokołu HTTP, ta metoda pozwala na sprawdzenie do 500 adresów podczas jednego żądania skierowanego do serwera GSB. Sprawdzane adresy URL nie muszą być w żaden sposób kodowane. Adresy do sprawdzenia są przesyłane w sekcji *body* wywołania HTTP.

Klient API stosując tą metodę wysyła żądanie na adres serwera GSB w następujący sposób:

```
https://sb-ssl.google.com/safebrowsing/api/lookup?client=CLIENT
&apikey=APIKEY&appver=APPVER&pver=PVER

4
http://www.pierwszyadres.pl/
http://drugiadres.org/
http://trzeciadres.org/members/
http://czwartyadres.org/
```

Jeżeli forma żądania wysłanego przez klienta API, była poprawna składniowo, serwer odpowiada:

- kodem 204, jeśli żadnego z adresów URL nie ma na „czarnej liście" usługi GSB
- kodem 200, jeśli co najmniej jeden ze sprawdzanych adresów znajduje się na *„czarnej liscie"* GSB. Po kodzie 200 występują odpowiedzi w kolejności w jakiej klient przesłał adresy do sprawdzenia:
  - „ok" – dla adresów, których nie ma na żadnej liście,
  - *„phishing"* - dla stron podejrzewanych o wyłudzenia danych,
  - *„malware"* - dla stron podejrzewanych o rozpowszechnianie złośliwego oprogramowania,
  - *„phishing, malware"* – dla stron, które podejrzane są o oba ryzykowne dla użytkownika zachowania.

### *Zgłaszanie podejrzanych adresów i błędów klasyfikacji*

Ponieważ jednym ze źródeł informacji o podejrzanych adresach zasilających mechanizm GSB są raporty użytkowników, Google Inc. udostępnia użytkownikom Internetu formularze służące zgłaszaniu przypadków niebezpiecznych stron. Błędnie sklasyfikowane adresy URL (tzw. *false positives*) także można zgłosić poprzez formularze dostępne na serwerach Google Inc. lub innych podmiotów.
Adresy URL podejrzane o wyłudzanie danych (*phishing*) należy zgłaszać pod adresem:

http://www.google.com/safebrowsing/report_phish/

Strony, które zostały błędnie sklasyfikowane przez usługę *Google Safe Browsing* jako adresy wyłudzające dane należy zgłaszać pod adresem:

http://www.google.com/safebrowsing/report_error/

Adresy URL podejrzane o rozpowszechnianie złośliwego oprogramowania (*malware*) należy zgłaszać pod adresem:

http://www.google.com/safebrowsing/report_badware/

Strony, które zostały błędnie sklasyfikowane przez usługę *Google Safe Browsing* jako adresy rozpowszechniające malware należy zgłaszać pod adresem:

http://www.stopbadware.org/home/reviewinfo

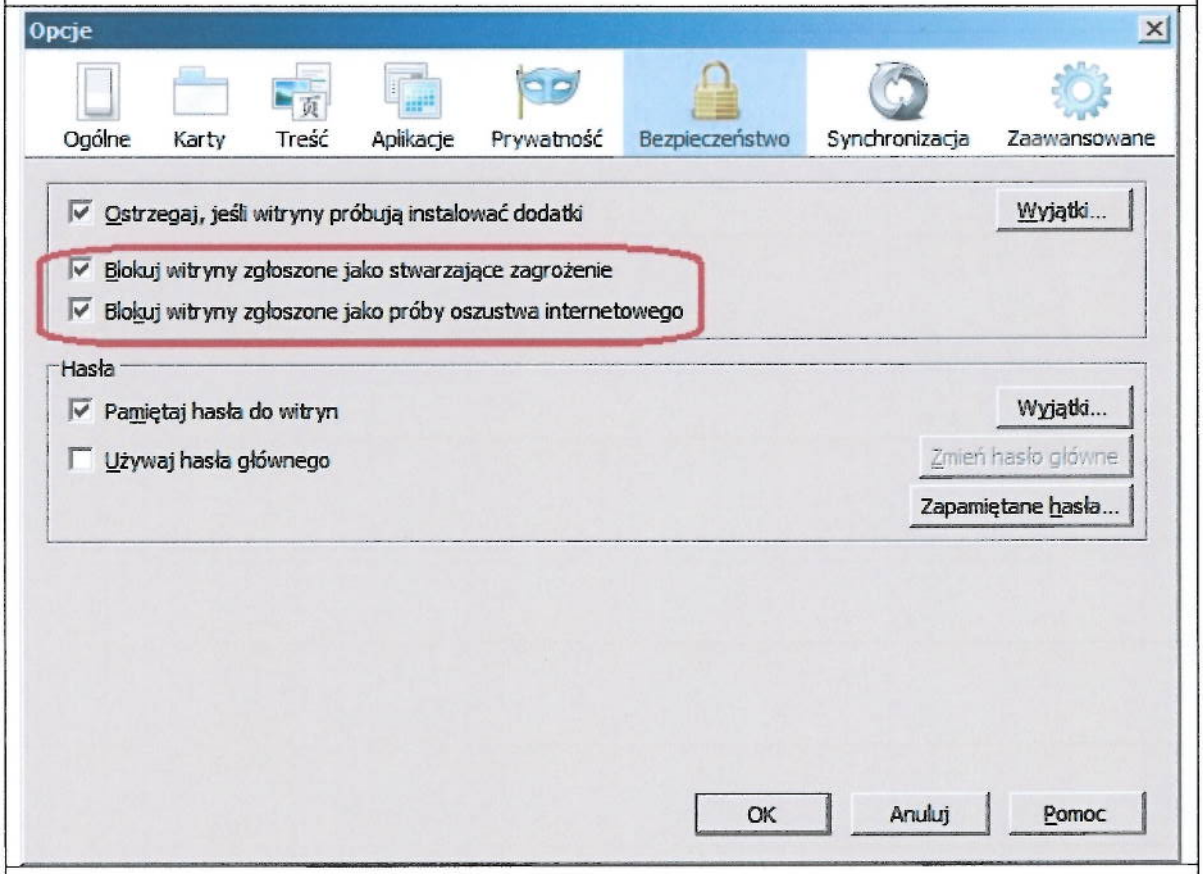
## Klienci Safe Browsing API

### *Przeglądarka internetowa Mozilla Firefox*

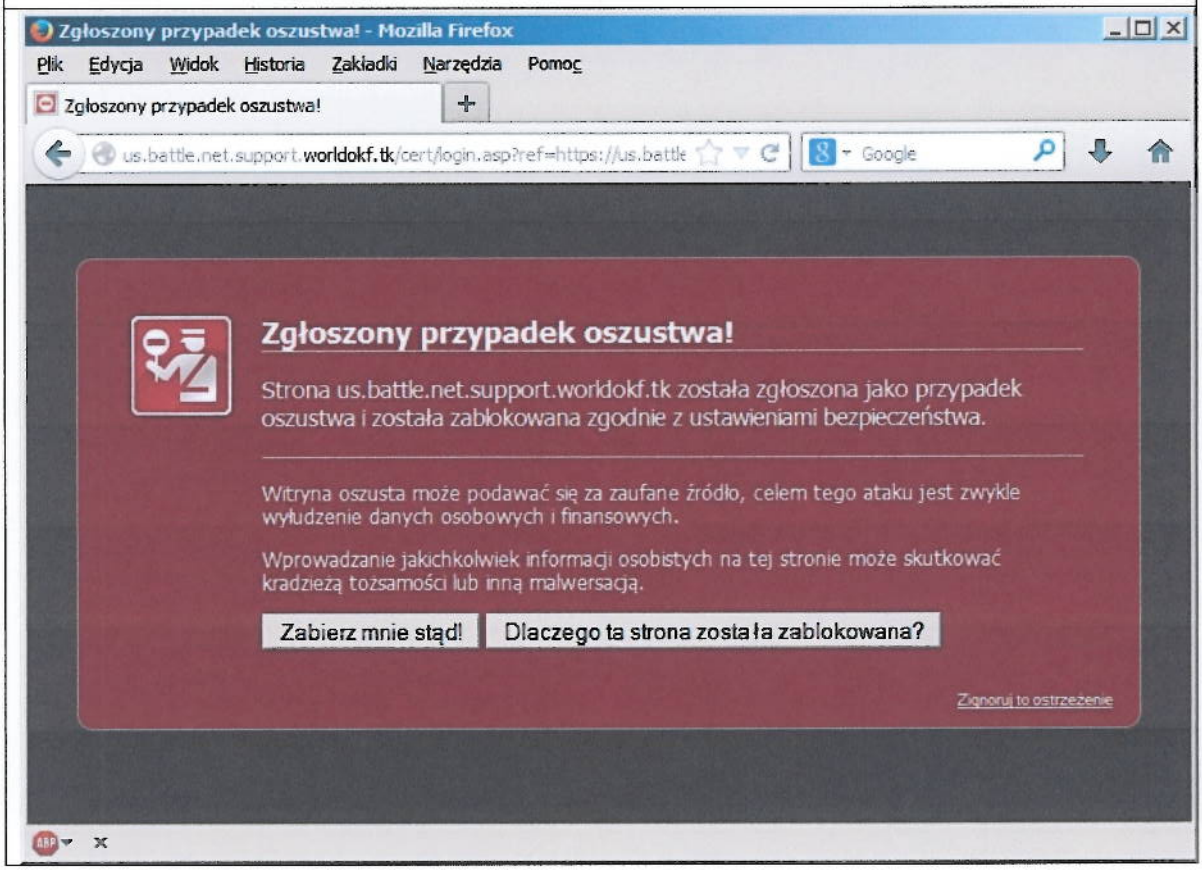
Przeglądarka Mozilla Firefox używa mechanizmu Google *Safe Browsing API* w zakresie ostrzegania użytkowników przed stronami wyłudzającymi dane oraz rozpowszechniającymi złośliwe oprogramowanie. Przyznaje się do tego zarówno firma Google Inc.[36], jak i fundacja Mozilla[37]. **Opcja zapewniania takiej ochrony jest domyślnie włączona po instalacji programu.** Użytkownik może wyłączyć lub włączyć taką ochronę wybierając z menu programu kolejno: *Narzędzia* → *Opcje* → *Bezpieczeństwo.*

[36] http://www.google.com/tools/firefox/safebrowsing/
[37] https://support.mozilla.org/pl/kb/how-does-phishing-and-malware-protection-work

**Mozilla Firefox – włączone opcje Google Safe Browsing**




**Mozilla Firefox – komunikat blokujący wyświetlanie w przypadku phishingu**



Biegły sprawdził działanie przeglądarki ***Mozilla Firefox ver. 25.0.1*** pod kątem ochrony przed wyświetlaniem stron wyłudzających dane. Zdaniem biegłego, przeglądarka Firefox nie spełnia wszystkich warunków wymaganych przez Google Inc. opublikowanych w dokumencie nazwanym *„Warunki korzystania z usługi"*[38] dostępnym elektronicznie w [5].
Punkt nr 1 tych warunków nakazuje, by program klienta API informował użytkownika **przed rozpoczęciem korzystania z usługi, a także przy każdym wyświetleniu komunikatu ostrzegawczego blokującego wyświetlanie podejrzanej strony, że niezawodność oraz dokładność usługi (*Google Safe Browsing*- przyp. biegłego) nie może być zagwarantowana.**
Punkt 1 *„Warunków korzystania z usługi"* wskazuje również, że komunikat ten powinien używać języka wskazanego pod adresem internetowym:

http://code.google.com/apis/safebrowsing/developers_guide_v2.html#UserWarnings

Pod tym adresem, w dziale *„Notice to Users About Phishing and Malware Protection"*[39] znajduje się dodatkowe zalecenie firmy Google Inc., by **poniższa informacja była widoczna dla użytkowników zanim włączą ochronę:**

*„Google pracuje nad tym, aby zapewnić najbardziej dokładne i aktualne informacje o stronach wyłudzających dane i rozprzestrzeniających złośliwe oprogramowanie. Jednak nie może zagwarantować, że jego (Google) informacje są pełne i bezbłędne: niektórych ryzykownych adresów nie można zidentyfikować, a niektóre bezpieczne witryny mogą być zidentyfikowane błędnie jako ryzykowne."*

Przeglądarka *Mozilla Firefox ver. 25.0.1* **nie wyświetla takiego komunikatu ani podczas instalacji, po której opcja ochrony jest domyślnie włączona, ani kiedy użytkownik sam włączy tą opcję** (np. jeśli wcześniej wyłączył ochronę), **ani podczas wyświetlania komunikatu ostrzegającego przed wyświetleniem witryny podejrzanej o wyłudzanie danych.**

Biegły, na stronach 21 i 22 niniejszej opinii, omówił wymagania Google Inc. dotyczące wyświetlania komunikatów ostrzegających użytkownika.

---

[38] https://developers.google.com/safe-browsing/terms - warunki dostępne także w formie tłumaczenia uwierzytelnionego w aktach sprawy VIII GC260/12 – str. 210 akt,

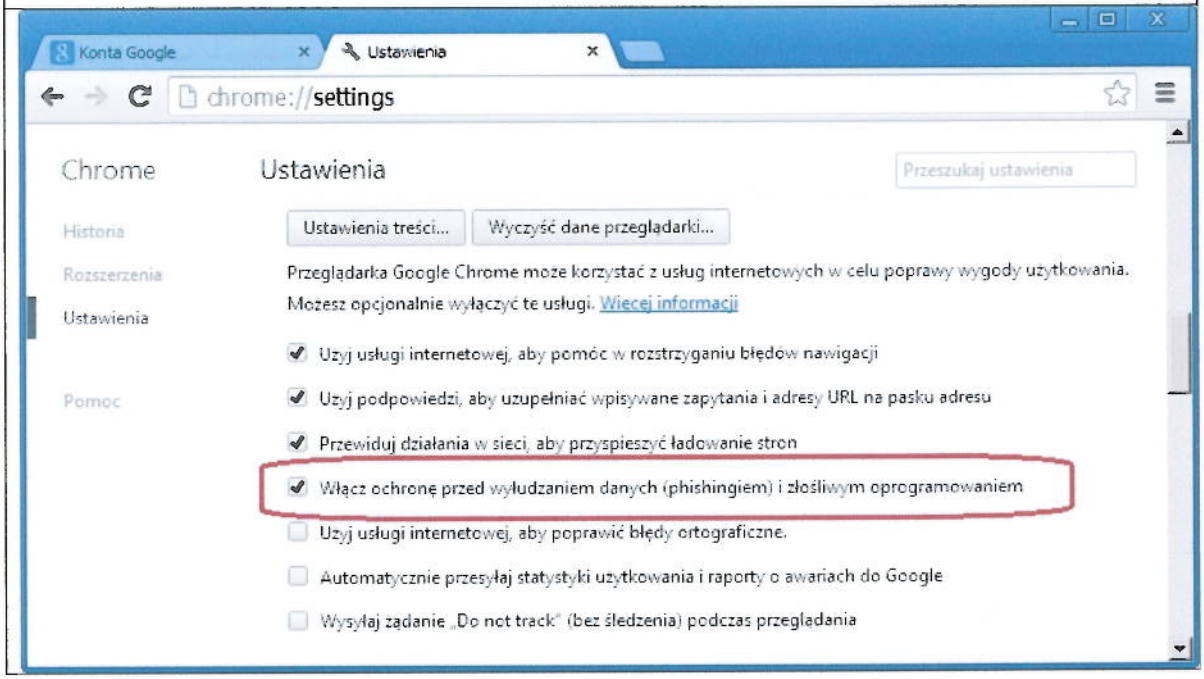
[39] **Notice to Users About Phishing and Malware Protection** - ang. Informacja dla użytkowników o ochronie przed phishingiem i malware

### *Przeglądarka internetowa Google Chrome*

Przeglądarka ***Google Chrome*** jest produktem firmy Google Inc. i co oczywiste w tym przypadku używa mechanizmu Google *Safe Browsing API* w zakresie ostrzegania użytkowników przed stronami wyłudzającymi dane oraz rozpowszechniającymi złośliwe oprogramowanie.

**Opcja zapewniania takiej ochrony jest domyślnie włączona po instalacji programu.** Użytkownik może wyłączyć lub włączyć taką ochronę wybierając z menu programu kolejno: *Ustawienia → Pokaż ustawienia zaawansowane → Prywatność→Włącz ochronę przed wyłudzaniem danych (phishingiem) i złośliwym oprogramowaniem.*

**Google Chrome – włączona opcja Google Safe Browsing**



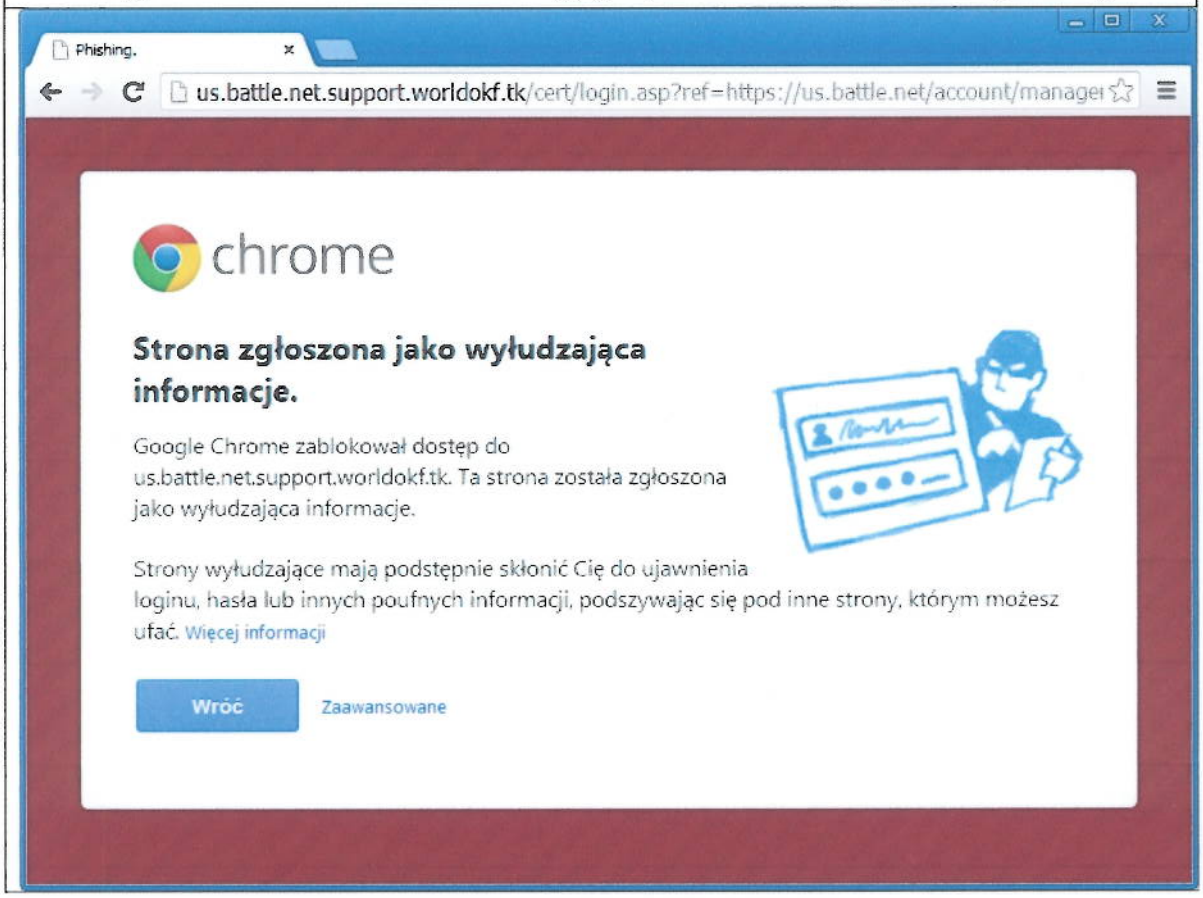

Biegły sprawdził działanie przeglądarki ***Google Chrome ver. 31.0.1650.57 m*** pod kątem ochrony przed wyświetlaniem stron wyłudzających dane i doszedł do wniosku, że podobnie jak w poprzednim przypadku (opis działania Mozilla Firefox), przeglądarka *Chrome* nie spełnia wszystkich warunków wymaganych przez Google Inc. opublikowanych w dokumencie nazwanym *„Warunki korzystania z usługi"*[40] dostępnym elektronicznie w [5]. **Wszystkie uwagi biegłego, dotyczące zachowania w zakresie poinformowania użytkownika końcowego o możliwości pojawienia się błędów**

[40] https://developers.google.com/safe-browsing/terms - warunki dostępne także w formie tłumaczenia uwierzytelnionego w aktach sprawy VIII GC260/12 – str. 210 akt,

**klasyfikacji stron, które dotyczyły przeglądarki *Firefox*, dotyczą również przeglądarki *Google Chrome*.**

Przeglądarka *Google Chrome ver. 31.0.1650.57 m* **nie wyświetla takiego komunikatu ani podczas instalacji, po której opcja ochrony jest domyślnie włączona, ani kiedy użytkownik sam włączy tą opcję** (np. jeśli wcześniej wyłączył ochronę), **ani podczas wyświetlania komunikatu ostrzegającego przed wyświetleniem witryny podejrzanej o wyłudzanie danych.**

**Google Chrome – komunikat blokujący wyświetlanie w przypadku phishingu**

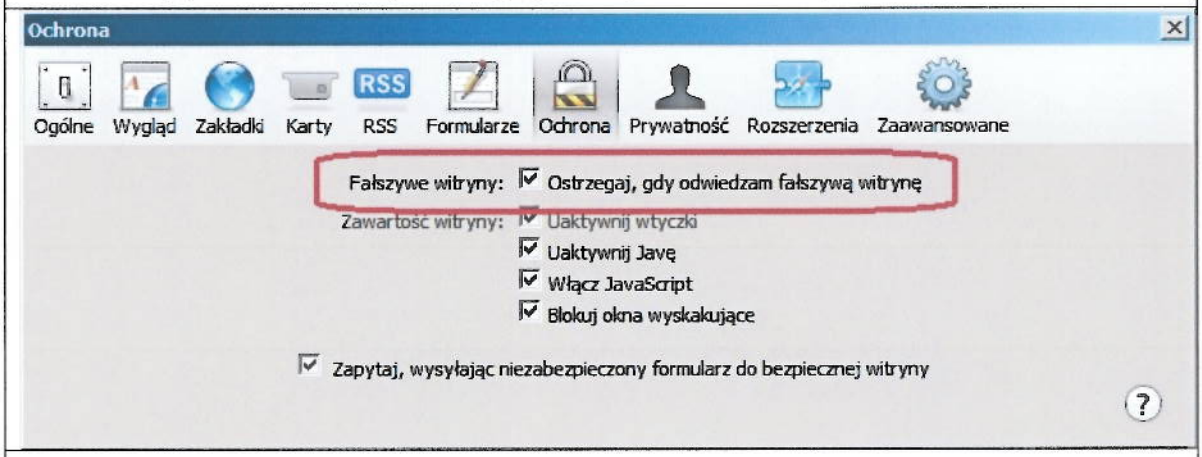



### *Przeglądarka internetowa Apple Safari*

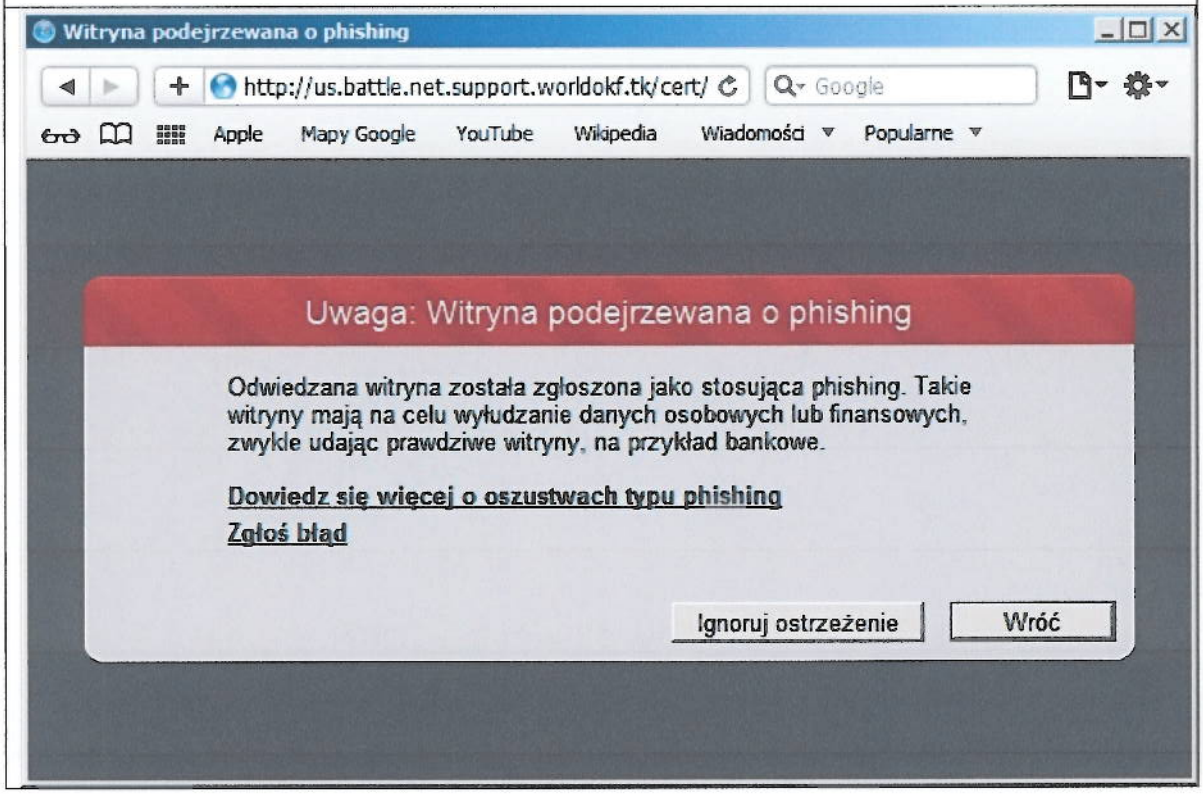
Przeglądarka ***Apple Safari*** używa mechanizmu Google *Safe Browsing API* w zakresie ostrzegania użytkowników przed stronami wyłudzającymi dane oraz rozpowszechniającymi złośliwe oprogramowanie. **Opcja zapewniania takiej ochrony**

**jest domyślnie włączona po instalacji programu**[41]**.** Użytkownik może wyłączyć lub włączyć taką ochronę wybierając z menu programu kolejno: *Preferencje* → *Ochrona* → *Ostrzegaj gdy odwiedzam fałszywą witrynę.*

**Apple Safari – włączona opcja korzystania z Google Safe Browsing**




**Apple Safari – komunikat blokujący wyświetlanie w przypadku phishingu**




Biegły sprawdził działanie przeglądarki ***Apple Safari ver. 5.1.7*** pod kątem ochrony przed wyświetlaniem stron wyłudzających dane i doszedł do wniosku, że podobnie jak w poprzednich przypadkach (opisy działania Mozilla Firefox i Google Chrome),

[41] biegły testował przeglądarkę Apple Safari dla platformy Windows 7.

przeglądarka *Safari* nie spełnia wszystkich warunków wymaganych przez Google Inc. opublikowanych w dokumencie nazwanym *„Warunki korzystania z usługi"*[42] dostępnym elektronicznie w [5]. **Wszystkie uwagi biegłego, dotyczące zachowania w zakresie poinformowania użytkownika końcowego o możliwości pojawienia się błędów klasyfikacji stron, które dotyczyły przeglądarki *Firefox i Chrome* dotyczą również przeglądarki *Apple Safari*.**

Przeglądarka ***Apple Safari ver. 5.1.7* nie wyświetla takiego komunikatu ani podczas instalacji, po której opcja ochrony jest domyślnie włączona, ani kiedy użytkownik sam włączy tą opcję** (np. jeśli wcześniej wyłączył ochronę), **ani podczas wyświetlania komunikatu ostrzegającego przed wyświetleniem witryny podejrzanej o wyłudzanie danych.**

### *Podsumowanie*

Zdaniem biegłego, twórcy wymienionych przeglądarek internetowych nie przyłożyli odpowiedniej wagi do poinformowania użytkowników o możliwości wystąpienia błędów klasyfikacji podejrzanych stron. Żadna z przeglądarek korzystających z mechanizmu *Google Safe Browsing* nie informuje użytkowników, w sposób sugerowany przez *„Warunki korzystania z usługi"* o tym, że usługa nie gwarantuje 100 % pewności w zakresie wydawanych klasyfikacji podejrzanych stron, ani o możliwości wystąpienia błędów polegających na tym, że cyt.: *„niektórych ryzykownych adresów nie można zidentyfikować, a niektóre bezpieczne witryny mogą być zidentyfikowane błędnie jako ryzykowne"*. Ponadto, komunikaty ostrzegające użytkowników przed wyświetleniem podejrzanej strony są sformułowane w sposób niezgodny z *Safe Browsing API v 2*. Tylko przeglądarka *Apple Safari* używa w oknie komunikatu sugerowanego[43] sformułowania: *„Strona **podejrzewana** o phishing"*. Przeglądarka *Mozilla Firefox* stosuje sformułowanie: *„Zgłoszony przypadek oszustwa !"*, a program Google Chrome: *„Strona zgłoszona jako wyłudzająca informacje"*, nie informując ani gdzie zgłoszono ten przypadek, ani jaki jest stopień pewności tej informacji.

---

[42] https://developers.google.com/safe-browsing/terms - warunki dostępne także w formie tłumaczenia uwierzytelnionego w aktach sprawy VIII GC260/12 – str. 210 akt,

[43]https://developers.google.com/safe-browsing/developers_guide_v2?csw=1#UserWarnings - *w komunikacie odnoszącym się do strony, która została zidentyfikowana jako potencjalnie ryzykowna dla użytkownika, jako ostrzeżeń należy używać terminów takich jak:* ***podejrzenie, potencjalnie możliwe, prawdopodobne, może być***

# Analiza materiału dowodowego

Biegły zapoznał się z materiałem dowodowym w postaci akt sprawy ***VIII GC 260/12***. Po zapoznaniu się z aktami sprawy biegły uznał, że do wydania opinii konieczne jest zapoznanie się z kodem źródłowym stron sklepów internetowych w wersji ze spornego okresu 05-08.01.2012 r.

W związku z tym, w dniu 21.10.2013 roku o godzinie 16:00 biegły udał się do siedziby powodowej spólki IAI S.A w celu zabezpieczenia takiego materiału dowodowego do dalszych badań i skopiowania na zewnętrzny dysk twardy biegłego. W trakcie rozmowy z Prezesem spółki, Panem Pawłem Fornalskim biegły ustalił, że strona powodowa nie posiada zarchiwizowanych stron sklepów (tzw. backup) ze spornego okresu blokady. W posiadaniu powodowej spółki jest natomiast kod źródłowy sklepów przechowywyany w systemie GIT. Ponadto, stosowana przez spółkę technologia uniemożliwia uruchomienia kopii stron w środowisku wirtualnym biegłego.

Pan Fornalski poinformował biegłego, że kod źródłowy stron jest generowany w taki sam sposób dla wszystkich sklepów, których strony różnią się tylko szablonem wpływającym na ich wygląd i zawartością bazy danych. W związku z tym, pomimo braku backup-u możliwe będzie odtworzenie zawartości stron sklepów, na podstawie repozytorium wersji kodu źródłowego przechowywanego w spółce. Biegły zobowiązał Prezesa spółki Pana Pawła Fornalskiego do przesłania mailem opisu technicznego środowiska generującego strony sklepów internetowych oraz do udostępnienia biegłemu pod publicznie dostępnym adresem w Internecie 2 sklepów działających na podstawie kodu „cofniętego w czasie" do okresu 05-08.01.2012.

W mailu od Pana Grzegorza Szukalskiego z dnia 25.10.2013, strona powodowa przedstawiła następującą informację opisującą środowisko produkcyjne:

**System operacyjny**

debian 5.0.6

kernel: Linux vm595 2.6.32-4-686-bigmem

**Serwer http**

Zend Engine v2.2.0, Copyright (c) 1998-2009 Zend Technologies
with Zend Core v2.5.5, Copyright (c) 1998-2008, by Zend Technologies
with eAccelerator v0.9.5.3, Copyright (c) 2004-2006 eAccelerator, by eAccelerator
with Zend Extension Manager v1.2.2, Copyright (c) 2003-2007, by Zend Technologies
Apache Apache/2.2.14
nginx/0.7.65

**Moduły, modyfikacje lub nietypowe skrypty**

Konfigurację serwera php, listę włączonych modułów, itp. przesyłam jako załącznik. Jest to zrzut bezpośrednio z serwera.

**Generowanie stron www**

Strony sklepu generowane są w technologii XSLT (warstwa prezentacji oddzielona od warstwy danych). Skrypty php (wersja 5.3) generują dane w formacie XML następnie dane te są transformowane razem z plikiem XLS (szablonem wyglądu strony, w nomenklaturze naszej firmy zwanym „maską sklepu") w efekcie czego powstaje kod html widoczny na stronie.

Każdy sklep posiada tą samą wersję kodu php oraz indywidualną maskę (plik xsl) i bazę danych. Wszystkie formularze są przetwarzane przez kod php, który jest wspólny dla wszystkich naszych sklepów.

**Kopie archiwalne**

Codziennie tworzone są kopie baz danych klientów oraz plików znajdujących się na serwerze. Kopie bazy danych są trzymane są przez 7 dni. Pliki (np. zdjęcia) przechowywane są maksymalnie 2 wersje wstecz, nie starszej jednak niż rok.

**Wersjonowanie kodu**

W styczniu 2012 roku kod był wersjonowany za pomocą SVN (systemu kontroli wersji), obecnie cały kod podlega wersjonowaniu za pomocą GIT. Zmiana została dokonana na przełomie 2012 i 2013 roku. Do repozytorium GIT została zaimportowana cała historii SVN (dotychczasowego system), dzięki czemu możemy odtworzyć działanie sklepu z tego okresu.

Ponadto strona powodowa, udostępniła biegłemu:

1. kopię sklepu http://www.xtreme-style.pl/ z okresu 05-08.01.2012, która dostępna jest w sieci Internet pod adresem: http://vmgr1.iai-shop.com
2. kopię strony http://iai-shop.com z okresu 05-08.01.2012, która dostępna jest w sieci Internet pod adresem: http://vmgr2.iai-system.com/
3. wskazania do witryny http://web.archive.org, która przechowuje archiwalne wersje spornych stron internetowych ze spornego okresu blokady:

- dla witryny http://www.iai-shop.com , adres z kopią z dn. 15.08.2012 r.: http://web.archive.org/web/20120815000000*/http://www.iai-shop.com
- dla witryny http://www.xtreme-style.pl, adres z kopią z dnia 04.01.2012 r.: http://web.archive.org/web/20120104114034/http://www.xtreme-style.pl/
- dla witryny http://www.netranova.pl/, adres z kopią z dnia 04.01.2012 r.: http://web.archive.org/web/20120104214518/http://www.netranova.pl/
- dla witryny http://www.djbox.pl/ , adres z kopią z dnia 30.12.2011 r.: http://web.archive.org/web/20111230231836/http://www.djbox.pl/
- dla witryny http://jpmax.pl/ , adres z kopią z dnia 29.12.2011 r.: http://web.archive.org/web/20111229071001/http://jpmax.pl/

Biegły zapoznał się aktami sprawy oraz z elektronicznym materiałem dowodowym, dostarczonym przez powodową spółkę i stwierdził, co następuje:

1. W przedstawionym przez stronę powodową elektronicznym materiale dowodowym, **biegły nie znalazł żadnych śladów wskazujących na proceder rzeczywistego wyłudzania danych**. Biegły przeanalizował kod źródłowy kopii stron dostępnych pod adresem http://vmgr1.iai-shop.com, oraz http://vmgr2.iai-system.com/, a także stron dostępnych w witrynie http://web.archive.org wskazanych powyżej i nie znalazł śladów takich działań zarówno na stronach głównych tych witryn, jak i na stronach służących do logowania na konto, lub zakładania konta użytkownika.
2. Strona http://www.xtreme-style.pl, w wersji z dnia 04.01.2012 roku, zachowanej w witrynie http://web.archive.org, stosuje odwołania do innej domeny: https://magnuspolska.iai-shop.com/signin.php w elementach służących do logowania, takich jak przycisk i odnośnik *„Zaloguj"*, odnośniki *„Ustawienia konta"*, *„Zarejestruj się"*. **Jest to zachowanie, które mogło być oceniane jako działanie wskazujące na wyłudzanie danych chociaż takim w rzeczywistości nie było** – domena https://magnuspolska.iai-shop.com jest tzw. domeną techniczną tej strony.
3. Strona http://www.netranova.pl/ w wersji z dnia 04.01.2012 r., zachowana w witrynie http://web.archive.org, nie używa bezpiecznego, szyfrowanego połączenia ***https*** na stronie ***http***://www.netranova.pl/login.php , do której prowadzi odnośnik *„Zaloguj się"* chociaż na stronie znajdują się pola do wprowadzania nazwy użytkownika i hasła.

Strona ta również obecnie stosuje ten sposób rejestrowania użytkowników. W ocenie biegłego **jest to zachowanie, które mogło być oceniane jako działanie niebezpieczne dla danych użytkownika i nadal nim pozostaje.**

4. Strona http://www.djbox.pl/, w wersji z dnia 30.12.2011 r., zachowana w witrynie http://web.archive.org, nie używa bezpiecznego, szyfrowanego połączenia ***https*** na stronach do których prowadzą przyciski i odnośniki *„Zaloguj się"* (***http***://www.djbox.pl/signin.php) i *„Rejestracja konta"* (***http***://www.djbox.pl/client-new.php?register), chociaż na stronach do których prowadzą te odnośniki użytkownicy wprowadzają dane osobowe takie jak imię, nazwisko, data urodzenia, adres zamieszkania, numer telefonu, adres e-mail. Strona ta również obecnie stosuje ten sposób rejestrowania użytkowników. W ocenie biegłego **jest to zachowanie, które mogło być oceniane jako działanie niebezpieczne dla danych użytkownika i nadal nim pozostaje.**
5. Strona http://jpmax.pl/ w wersji z dnia 29.12.2011 r., zachowana w witrynie http://web.archive.org, nie używa bezpiecznego, szyfrowanego połączenia ***https*** na stronach do których prowadzą odnośniki *„Zaloguj się"* (***http***://jpmax.pl/signin.php) i *„Zarejestruj się"* (***http***://jpmax.pl/client-new.php?register), chociaż na stronach do których prowadzą te odnośniki użytkownicy wprowadzają dane osobowe takie jak imię, nazwisko, adres zamieszkania, numer telefonu, adres e-mail. Strona ta również obecnie stosuje ten sposób rejestrowania użytkowników. W ocenie biegłego **jest to zachowanie, które mogło być oceniane jako działanie niebezpieczne dla danych użytkownika i nadal nim pozostaje.**
6. W związku z tym, że wskazywaną w aktach sprawy (str. 25 i 26 akt VIII GC 260/12) przyczyną blokady wyświetlania stron, był odnośnik wyświetlający logo z adresu http://iai-shop.com/panel/gfx/pb-iai.png, biegły przeanalizował kod źródłowy oraz zawartość tej strony udostępnionej przez powoda pod adresem https://vmgr1.iai-shop.com/panel/index.php . **Biegły nie znalazł żadnych śladów wskazujących na proceder wyłudzania danych na tej stronie.**
7. Biegły przeanalizował porady [6] przeznaczone dla twórców stron internetowych, które opublikowała firma Google Inc. w portalu *Google Webmasters Central Blog* w celu informowania o tym, w jaki sposób należy projektować strony internetowe, aby uniknąć pomyłkowego sklasyfikowania witryny jako niebezpiecznej przez mechanizm Google Safe Browsing. **Biegły stwierdził, że niektóre wzorce projektowe badanych stron powodowej spółki nie spełniają tych zaleceń.** Pierwsza porada z [6] brzmi:

*„Nie pytaj o nazwy użytkowników i hasła, które nie należą do Twojej witryny. Takie zachowania z definicji uważamy za phishing, więc nie rób tego! Jeśli chcesz zapewnić na swojej stronie dodatek do usługi na innej stronie, rozważ zamiast tego używanie publicznego API lub metody OAuth.*" Tymczasem, jak biegły wykazał w punkcie 2 niektóre strony przekierowywały użytkowników do innej domeny w celu zalogowania, prosząc o podanie nazwy użytkownika i hasła w domenie technicznej typu: https://nazwasklepu.iai-shop.com/signin.php , czyli w innej domenie niż domena sklepu.

8. Kolejna zasada projektowa zalecana przez Google Inc. brzmi: *„Ogranicz liczbę domen wykorzystywanych przez witryny, zwłaszcza dla logowania. Pytanie o nazwę użytkownika i hasło dla strony X wygląda bardzo podejrzanie na stronie Y. Nie dość, że trudniej nam ocenić Twoją stronę, to takie postępowanie uczy odwiedzających ignorowania podejrzanych adresów, co czyni ich bardziej podatnymi na rzeczywiste próby phishingu. Jeśli musisz mieć swoją stronę logowania w innej domenie niż główna strona, należy rozważyć użycie przezroczystego proxy, aby umożliwić użytkownikom dostęp do tej strony z domeny podstawowej".* Z lektury akt sprawy (np. str. 145) wynika tymczasem, że przynajmniej część sklepów miała przekierowanie strony logowania dla użytkowników sklepu do domeny technicznej, oraz przekierowanie strony administracyjnej *http://nazwasklepu/panel/* do domeny technicznej typu *https://nazwasklepu.iai-shop.com/panel/*,
9. Analiza znalezionego przez użytkowników i administratorów powodowej spółki rozwiązania problemu blokady stron (str. 25 i 26 akt sprawy VIII GC 260/12), które polegało na usunięciu odnośnika do grafiki z logo spólki, dostępnej pod adresem http://iai-shop.com/panel/gfx/pb-iai.png wskazuje na to, że **z dużym prawdopodobieństwem można stwierdzić wystąpienie *błędu I rodzaju* („fałszywy alarm") w procesie klasyfikacji niebezpiecznych stron usługi Google Safe Browsing, co doprowadziło do zablokowania sklepów powodowej spółki.** Zdaniem biegłego, z dużym prawdopodobieństwem można założyć, że przedstawione strony nie wyłudzały danych, chociaż mogły zawierać elementy wskazujące na takie zachowanie. Świadczy o tym także opisane w aktach sprawy niewytłumaczalne zachowanie systemu weryfikacji „reputacji" stron Google Safe Browsing, który wyświetlał dla poszczególnych stron komunikaty o tym, że strony nie są uznawane za podejrzane przez GSB, a jednocześnie nie usuwał ich z „czarnej listy" przez sporny okres kilku dni. **W tym świetle wprowadzanie komunikatu o wyłudzaniu danych przez strony**

**internetowe powódki oraz obsługiwane przez nią sklepy, zdaniem biegłego było nieuzasadnione.** Biegły zwraca jednak uwagę na fakt, że w sensie technicznym to nie usługa *Google Safe Browsing* blokowała wyświetlanie tych stron, choć niewątpliwie błąd klasyfikacji leżał po stronie tego mechanizmu. Wyświetlanie blokowały przeglądarki internetowe *Mozilla Firefox, Google Chrome* i *Apple Safari*, korzystające z tej usługi do ochrony przed phisingiem. Przeglądarka *Microsoft Internet Explorer* wyświetlała te strony bez komunikatów o zagrożeniu, bo korzysta z innych mechanizmów ochrony. Nie można więc twierdzić, że strony były w jakikolwiek sposób zablokowane, bo w sensie technicznym ruch sieciowy do tych stron był możliwy.

10. Biegły przeanalizował techniczną stronę działania mechanizmu *Google Safe Browsing* w celu ustalenia zakresu kompetencji pozwanej Google Poland sp. z o.o., w odniesieniu do zarządzania katalogiem stron, na których znajduje się informacja o podmiotach wyłudzających dane. Biegły przeprowadził analizę adresów internetowych występujących w opisie technicznym Safe Browsing API, i ustalił, że klient API usługi GSB korzysta z poniższych adresów:
    - *http://safebrowsing.clients.google.com*,
    - *https://sb-ssl.google.com* w domenie google.com, zarejestrowanej na firmę Google Inc. pod adresem *1600 Amphitheatre Parkway, MountainView, USA*. Również dostępne w Internecie serwisy geolokalizacyjne wskazują na powyższą lokalizację adresów IP powiązanych ze wskazanymi adresami URL. **Z technicznego punktu widzenia klient API komunikuje się podczas swojej pracy i wymienia dane z serwerami umieszczonymi w USA, więc zdaniem biegłego pozwana spółka Google Poland sp. z o.o. nie zarządza katalogiem zablokowanych stron.**

Dane geolokacyjne adresów usługi GSB

# Geolocation by IP

IP Address Lookup. Domain Lookup.

**Your local IP address is 62.69.19**

safebrowsing.clients.google.c Track IP or host

Examples: 74.125.93.147 or google.com

**General IP Information**

| | |
|---|---|
| IP Address: | 173.194.39.229 |
| ISP: | Google Inc. |
| Domain: | safebrowsing.clients.google.com |

**Geolocation Information**

| | |
|---|---|
| Country: | UNITED STATES |
| Country code: | US |
| Region: | CALIFORNIA |
| City: | MOUNTAIN VIEW |
| ZIP code: | 94043 |
| Latitude: | 37.406 |
| Longitude: | -122.079 |
| Timezone: | -07:00 |


# Geolocation by IP

IP Address Lookup. Domain Lookup.

**Your local IP address is 62.69.19**

sb-ssl.google.com Track IP or host

Examples: 74.125.93.147 or google.com

**General IP Information**

| | |
|---|---|
| IP Address: | 173.194.39.238 |
| ISP: | Google Inc. |
| Domain: | sb-ssl.google.com |

**Geolocation Information**

| | |
|---|---|
| Country: | UNITED STATES |
| Country code: | US |
| Region: | CALIFORNIA |
| City: | MOUNTAIN VIEW |
| ZIP code: | 94043 |
| Latitude: | 37.406 |
| Longitude: | -122.079 |
| Timezone: | -07:00 |

# Wnioski

1. Biegły wyjaśnił zasadę działania usługi Google Safe Browsing zarówno w zakresie mechanizmu klasyfikacji stron internetowych, jak i zasady udostępniania listy podejrzanych stron i metod współpracy aplikacji klienta API z serwerami usługi.
2. Zdaniem biegłego, **z dużym prawdopodobieństwem można założyć, że przedstawione strony nie wyłudzały danych, chociaż mogły zawierać elementy wskazujące na takie zachowanie.** W świetle posiadanej wiedzy biegły uważa, że strony powodowej spółki padły ofiarą wystąpienia *błędu I rodzaju* („fałszywy alarm") w procesie klasyfikacji niebezpiecznych stron usługi Google Safe Browsing, co doprowadziło do zablokowania wyświetlania sklepów powodowej spółki w większości przeglądarek internetowych. **Wprowadzanie komunikatu o wyłudzaniu danych przez strony internetowe powódki oraz obsługiwane przez nią sklepy, zdaniem biegłego było nieuzasadnione.** Biegły zwraca jednak uwagę na fakt, że w sensie technicznym to nie usługa *Google Safe Browsing* blokowała wyświetlanie tych stron, choć niewątpliwie błąd klasyfikacji leżał po stronie tego mechanizmu. Wyświetlanie blokowały przeglądarki internetowe *Mozilla Firefox, Google Chrome* i *Apple Safari*, korzystające z tej usługi do ochrony przed phisingiem. Przeglądarka *Microsoft Internet Explorer* wyświetlała te strony bez komunikatów o zagrożeniu, bo korzysta z innych mechanizmów ochrony. Nie można więc twierdzić, że strony były w jakikolwiek sposób zablokowane, bo w sensie technicznym ruch sieciowy do tych stron był możliwy.
3. Biegły ustalił, że z technicznego punktu widzenia klient API komunikuje się podczas swojej pracy i wymienia dane z serwerami umieszczonymi w USA, więc zdaniem biegłego pozwana spółka Google Poland sp. z o.o. nie zarządza katalogiem zablokowanych stron.
4. W związku z powyższym, biegły nie mógł ustalić działań podjętych przez pozwaną spółkę Google Poland sp. z o.o. zmierzających do usunięcia spornego komunikatu ostrzegającego, ponieważ zdaniem biegłego, w świetle poczynionych ustaleń nie ma do tego żadnych podstaw.

## *Bibliografia*


[1] CERT Polska, „Analiza incydentów naruszających bezpieczeństwo teleinformatyczne w roku 2011," 2011. [Online].

[2] Colin Whittaker, Brian Ryner, Marria Nazif, „Large-Scale Automatic Classification of Phishing Pages," 2010. [Online]. Dostępny: http://research.google.com/pubs/pub35580.html.

[3] Brian Ryner, Noe Lutz, „SafeBrowsing Design," 2009. [Online]. Dostępny: http://code.google.com/p/google-safe-browsing/wiki/SafeBrowsingDesign.

[4] Garrett Casto, Oliver Fisher, Raphaël Moll, Marria Nazif, Dan Born, „Protocolv2Spec," 2009. [Online]. Dostępny: http://code.google.com/p/google-safe-browsing/wiki/Protocolv2Spec#3.6._List_Contents.

[5] Google Inc., „Safe Browsing API," 2011. [Online]. Dostępny: https://developers.google.com/safe-browsing/.

[6] Colin Whittaker, Anti-Phishing Team, „Will the Real <Your Site Here> Please Stand Up?," 30 03 2010. [Online]. Dostępny: http://googlewebmastercentral.blogspot.com/2010/03/will-real-site-here-please-stand-up.html.


15.11.2013

BIEGŁY SĄDOWY
SĄDU OKRĘGOWEGO W SZCZECINIE
*inż. Radosław Jedyński*

**Wykonano w 4 egzemplarzach:**

Egz. nr 1,2,3 – Sąd Okręgowy w Szczecinie Wydział VIII Gospodarczy

Egz. nr 4 – a/a