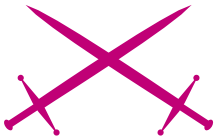


Poznaj swój komputer – oczami intruza

Bartosz Kalinowski

stopień trudności

W Sieci jeszcze bardziej niż w życiu realnym należy chronić wszelkie informacje dotyczące naszej osoby, firmy, działań, kierunków rozwoju... Dbać o te wszystkie dane muszą ludzie, którym je powierzamy. Dlatego zadziwiająca jest beztroska większości użytkowników Internetu w kwestiach bezpieczeństwa i ochrony wrażliwych danych.

Dane, o których mówię, w każdym przypadku mogą być (i z natury są) inne. Dla przeciętnego użytkownika będzie to imię, nazwisko, numery PESEL i NIP, adres zamieszkania – i to właśnie te informacje powinno traktować się ze szczególną troską. Wiele osób zobowiązanych jest także do ochrony – oprócz tak prozaicznych (a ważnych) danych jak te, które wcześniej wymieniłem – różnego rodzaju tajemnic służbowych (patrz głośne ostatnio wycieki danych z komputerów policyjnych przez p2p). Ostatnią grupą ludzi, którzy najsilniej i najskuteczniej powinni chronić dane, są administratorzy i dostawcy usług. Właśnie ci ludzie odpowiedzialni są za składowanie, przesył i zabezpieczenie personaliów swoich klientów oraz za stan bezpieczeństwa usług świadczonych w administrowanej sieci.

Dlaczego o tym piszę? Bo to jest ważne!

Czy stan bezpieczeństwa jest naprawdę tak zły? Tak! Czy dotyczy to wszystkich bez wyjątku? Większości!

W odniesieniu do zwykłych użytkowników Internetu sytuacja wygląda wprost tragicznie – nawet do 87% (na podstawie analizy aktywności sieciowej 2000 użytkowników portali społecznościowych, którzy wyrazili na to zgodę,

Z artykułu dowiesz się

- jak wiele informacji udostępniasz nieświadomie,
- w jaki sposób są one odnajdywane i w jaki sposób mogą być wykorzystane,
- w jaki sposób intruzi przeprowadzają rozpoznanie jednostek komputerowych i sieci,
- w jaki sposób tworzą wirtualny obraz zarządzanych przez Ciebie maszyn,
- jakie stosują narzędzia i metody,
- jak bronić się przed tymi działaniami.

Co powinieneś wiedzieć

- znać system Linux w stopniu pozwalającym na swobodną obsługę,
- znać co najmniej podstawy teoretyczne popularnych protokołów sieciowych,
- znać teorię dotyczącą ataków inter(intra)netowych i być w stanie kojarzyć możliwości ich wykorzystania,
- znać zagadnienia związane z wyszukiwarkami internetowymi.

przeprowadzonej w okresie od lipca do sierpnia 2007 r.) umieszcza w sieci publicznej dane, które nigdy nie powinny w niej się znaleźć, tym samym stwarzając dla siebie zagrożenie. W stosunku do administratorów – zarówno sieci, jak i usług – sprawa nie wygląda tak klarownie, a to ze względu na trudności m. in. w przeprowadzeniu analiz lub braku zgody na ich publikację. Jednak pomimo wszystko muszę stwierdzić – a wielu audytorów na pewno się ze mną zgodzi – że ochrona danych jest często zaniedbywana, a ilość udostępnianych (świadomie i nieświadomie) danych bywa, delikatnie mówiąc, niebezpieczna.

Spis treści

Moim głównym celem i priorytetem jest wskazanie Czytelnikom, jak wielkie zagrożenie niesie za sobą bezmyślne, świadome bądź przypadkowe, udostępnianie danych – zarówno tych personalnych, jak i służbowych czy konfiguracyjnych. Chciałbym w artykule zademonstrować krok po kroku niektóre metody pozyskiwania ważnych i wrażliwych informacji. Cały tekst podzieliłem na trzy główne części. W pierwszej wyjaśniam w sposób teoretyczny, aczkolwiek bardzo ogólnikowy, przyczyny i skutki poszczególnych działań. Opisuję także metody i narzędzia wykorzystywane do przeprowadzania tego typu badań. Staram się to czynić zwięźle i konkretnie, pozostawiając szczegóły na analizy konkretnych przypadków. W drugiej części przedstawiam działania wymierzone w zwykłego użytkownika. Przeprowadzam Czytelnika, w sposób zmuszający do samodzielnych działań i przede wszystkim myślenia, przez znane (a często trywialne! – w celu podkreślenia niebezpieczeństwa) metody i sposoby zbierania informacji, następnie analizuję uzyskane dane oraz finalnie określam stopień zagrożenia i możliwe ataki wynikające z faktu posiadania zdobytych informacji. W części trzeciej natomiast przedstawiam badanie danych udostępnianych świadomie i nieświadomie przez administratora typowej sieci komputerowej za pomocą jej elementów konfiguracyjnych. Określam wszelkie możliwe informacje – od lokalizacji fizycznej serwera po jego konfigurację, stosując metody manualne, jak i automatyczne oraz wykorzystując metody analizy tzw. białych źródeł informacji oraz *Google hacking*.

Rysunek 1. *Przykładowy wyglądu podstrony profilu jednego ze znanych portali społecznościowych*

Część I

Działania, które zaprezentuję, można pojmować i wykorzystywać w kilku celach: po pierwsze – patrząc oczami administratora (audytora lub osoby dbającej o bezpieczeństwo sieci i danych), jako na metody mające na celu zbadanie i ewentualną poprawę pewnych aspektów bezpieczeństwa sieci. W tym przypadku wszelkie uzyskane informacje posłużą do określenia ilości danych udostępnianych na wielką skalę – publicznie w Internecie – oraz jakości sposobu składowania ich na serwerach dostępnych z Internetu. Uzyskana wiedza pomoże w określeniu, jak dokładny obraz naszego serwera i sieci może uzyskać potencjalny intruz. Oprócz tego w bardzo łatwy, a przede wszystkim zgodny z prawem sposób,

Tabela 1. *Dwuwarstwowość rozpoznania*

Rozpoznanie pasywne						
Strony WWW	Grupy dyskusyjne	Social engineering	Bazy dns	Partnerzy biznesowi	...	
Rozpoznanie aktywne						
Skanowanie portów	Transfery stref	Analiza skanerami podatności	Ręczne badanie systemu	Social engineering	wywiad wewnątrzfirmowy	...

niezależnie od przyjętej przez nas polityki bezpieczeństwa oraz regulaminów sieci, będziemy w stanie określić, czy dane przechowywane publicznie przez użytkowników nie naruszają prawa. Z drugiej strony, stawiając się w sytuacji potencjalnego intruza, będziemy w stanie poznać metody i sposoby wykorzystywane do stworzenia obrazu serwera lub nawet całej sieci. Zdobędziemy dane, które z łatwością mogą być wykorzystane do przeprowadzenia wielu rodzajów ataków – od podszywania po kradzieże, i to nie tylko te przeprowadzane w Internecie.

Wszelkie działania, które będziemy wykonywać, mające na celu pozyskanie informacji, można podzielić na dwie kategorie – mogą być one bierne (pasywne) lub aktywne. Rodzaj pierwszy charakteryzuje się bardzo ważną cechą – nie pozostawia śladów mogących bezpośrednio wskazywać na nasze zamiary. Drugi niestety nie posiada wyżej wymienionej cechy – pozostawia ślady np. w plikach dzienników systemowych, lecz jest za to o wiele skuteczniejszy i dostarcza znacznie dokładniejszych i ważniejszych informacji. Mówiąc o działaniach biernych (pasywnych), mam na myśli przede wszystkim analizę informacji pozostawionych przez użytkowników (zarówno przeciętnych internautów, jak i np. administratorów) na różnego rodzaju portalach internetowych, forach lub grupach dyskusyjnych. Obie metody uzupełniają się w działaniach praktycznych, poszerzając znacznie ilość danych uzyskiwaną w początkowych fazach przeprowadzania np. audytu. Fazę tę graficznie przedstawia Tabela 1 *Dwuwarstwowość rozpoznania.*

Tabela 2. *Dane zamieszczone na stronie domowej użytkownika*

Dane	Uzyskany rezultat	Źródło
Imię	Rafał	Strona domowa
Nazwisko	Pokój	Strona domowa
Ulica	Długa 25	Strona domowa
Miasto	Ustrzyki	Strona domowa
PESEL	--	--
NIP	--	--
Telefon domowy	(013) 461-XX-XX	Strona domowa
Telefon komórkowy	888789XXX	Strona domowa
Telefon kom. inny	--	--
Adres e-mail 1	raptowny@jakisadres.pl	Profil portalu społecznościowego
Adres e-mail 2	rafal.pokoj@somemail.com	Strona domowa
Adres e-mail 3	--	--
GG	203XXX	Strona domowa
JID	raptowny@chrome.com	Profil portalu społecznościowego
Inne komunikatory	--	--
Ukończone szkoły	(1986-1994) Szkoła Podstawowa nr 2 ... (1994-1998) Liceum Ogólnokształcące im. Józefa Piłsudskiego ... (1998-2003) Politechnika Gdańska ...	Strona domowa
Aktualne zajęcie	--	--
Zainteresowania	- muzyka - sport (bieganie) - fantastyka	Portal s połecznościowy

Ilość danych prezentowana na wymienione sposoby w Internecie jest nie do ogarnięcia, dlatego też część druga artykułu jest swoistym ostrzeżeniem dla wszystkich i jako taka powinna być przede wszystkim traktowana. W przypadku wyszukiwania informacji w taki sposób stosujemy tzw. *białe* źródła. Do działań pasywnych możemy także zaliczyć wykorzystanie wszelkich narzędzi, które nie pozostawią śladów jawnie wskazujących na cel naszych działań, a za takie uznać można m. in. bazy danych *whois* (*http://www.dns.pl/cgi-bin/whois.pl*, *http://www.internic.net, http://ripe.net, http://whois.afilias.info itp*.), serwery DNS, narzędzia typu *traceroute*, *mtr*, *p0f v2*. Naszymi kolejnymi przyjaciółmi okażą się wyszukiwarki oraz umiejętność zadawania pytań (w celu zwiększenia ilości i polepszenia jakości otrzymywanych danych nie będziemy ograniczać się wyłącznie do *google.com*).

Działania aktywne, w przeciwieństwie do pasywnych, wykorzystują w znakomitej mierze narzędzia stworzone typowo do prowadzenia podstawowych analiz informatycznych. W tym przypadku będziemy stosować gotowe kombajny analityczne, takie jak *Nessus* (*http://nessus.org*) lub skanery portów, takie jak *nmap* (*http://insecure.org/nmap*). Pragnę zauważyć, iż istnieją pewne klasyfikacje, które uznają wykorzystanie baz *whois* lub *traceroute* za działania aktywne. Skuteczność rezultatów prezentowanych przez w/w narzędzia podniesiemy manualnym badaniem

zwróconych przez nie wyników oraz przeanalizujemy inne metody i sposoby działania wymierzone w informacje znajdujące się na serwerze lub mogące go skompromitować.

Część II

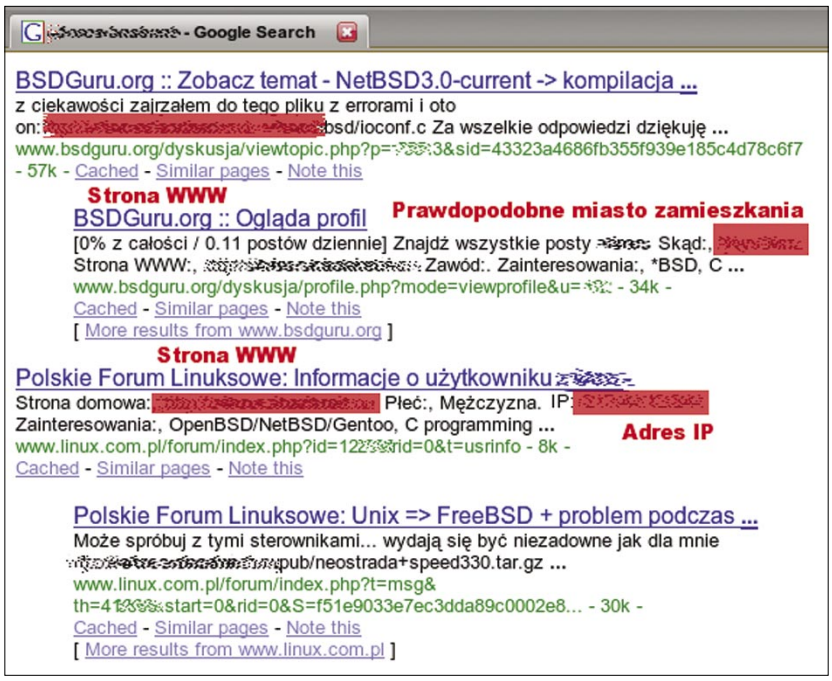
W jaki sposób przedstawić zagrożenia oraz odpowiednio wybrać cel? Uznałem, że wystarczy losowy wybór potencjalnego kandydata do analizy. W takim przypadku nie wiadomo, czy poddawana analizie nie jest np. osoba czytająca ten artykuł, co dodaje wiarygodności badaniom i potwierdza zagrożenia.

Moim celem stał się użytkownik pewnego portalu gromadzącego, krótko mówiąc, ciekawostki z Sieci. W mojej opinii jest to odpowiednie miejsce do wybrania potencjalnego celu, kiedy chce się wygenerować statystycznie dobrą próbę dla ogółu społeczeństwa korzystającego z Internetu – bez specjalistycznego podejścia do zagadnień informatycznych, a jednocześnie twierdzącego/uważającego/żyjącego w przekonaniu, że ukrywanie się w Sieci za pseudonimem gwarantuje jakiś poziom anonimowości. Przekonajmy się, jak wygląda rzeczywistość.

Chciałbym zauważyć, że dla celów artykułu wszelkie adresy, dane i nazwy zostały zmienione, a wszystkie ewentualne zbieżności i podobieństwa są przypadkowe. W celu lepszego poznania problemu zachęcam do przeprowadzenia podobnych działań wymierzonych w siebie lub np. członków rodziny. Da to najlepszy obraz zagrożeń i ich skali, a jednocześnie udowodni fakt łatwości przeprowadzania takich analiz.

Powiedz mi gdzie bywasz, powiem Ci kim jesteś

Na wstępie należy przyjąć pewne wytyczne poszukiwań – musimy określić, czego dokładnie chcemy dowiedzieć się o użytkowniku i jak dokładne muszą to być dane. Na potrzeby dalszych rozważań musimy znaleźć co najmniej imię, nazwisko, miejsce zamieszkania, adresy kontaktowe (telefon, GG, Jabber, Skype – najlepiej co najmniej dwa z wymienionych) oraz zainteresowania danego użytkownika. Jako rozszerzenie listy oczekujemy prywatnych numerów telefonów komórkowych i danych takich jak NIP czy PESEL.



Rysunek 2. *Wyniki przedstawione przez Google.*

Naszą przygodę rozpoczniemy od określenia podstawowych danych personalnych: imienia i nazwiska. W tym celu zastosujemy najprostszą i najbardziej efektywną metodę, mianowicie przeszukamy zasoby WWW. Ale zanim zadamy pytanie wyszukiwarce, ustalmy, co może dać nam listę rezultatów zawierającą przede wszystkim dane o „naszym" użytkowniku?

Będzie to na pewno kombinacja danych, które już posiadamy o użytkowniku, gdyż wystąpienie tuż obok siebie grupy takich danych na przeszukiwanych stronach gwarantuje nam w pewnym stopniu uzyskanie wiarygodnych odpowiedzi. W naszym przypadku użytkownik ukrywa się pod nickiem *Raptowny*. Pragnę zauważyć i przypomnieć, iż większość zarówno forów, jak i portali udostępnia możliwość logowania się oraz posiadania własnego profilu. Dlatego zaraz po określeniu nicku zapoznajemy się z profilem i spisujemy wszystkie istotne dla nas informacje.

Proszę zauważyć, jakie zazwyczaj dane (Rysunek 1. prezentuje typową i często spotykaną w Internecie postać tego rodzaju strony) mogą znajdować się w takim wirtualnym profilu użytkownika:

- imię, nazwisko,
- adres e-mail,
- strona domowa,
- kontakt via GG, Jabber, ICQ etc,
- odnośniki do opublikowanych dokumentów/wątków.

Nie jest to być może długa lista, podejrzewam także, że w większości przypadków w tym miejscu nie spotkamy się z imieniem i nazwiskiem, ale za to często z adresem e-mail i stroną domową – a to jest już dla nas dużo. Warto pamiętać o tym, iż dane, które są wyświetlane w tym miejscu nie muszą być dostępne publicznie. Nie mam tu na myśli oczywiście wprowadzania imienia i nazwiska, ale chociażby adresu e-mail oraz strony domowej. Mógłbym jeszcze zrozumieć zamieszczanie odnośników do strony firmowej w przypadku portalu dla pracowników czy klientów danej firmy, ale w przypadku osób prywatnych

jest to dla mnie, delikatnie mówiąc, nieco dziwne. Proszę sobie odpowiedzieć na pytanie – ile razy w życiu idąc ulicą rozdawaliśmy wszystkim swój adres zamieszkania i numer telefonu? Ktoś powie, że przesadzam? Proszę dotrwać do końca artykułu.

Po analizie profilu uzyskaliśmy następujące dane:

- adres e-mail: *raptowny@jakisadres.pl,*
- strona domowa: raptowny.*jakisblog.pl,*
- abber id: *raptowny@chrome.com.*

Następnym naszym krokiem będzie oczywiście zapoznanie się z zawartością strony domowej. Bardzo często zdarza się – szczególnie ostatnio, w dobie blogów i innych internetowych dzienników, że użytkownicy zamieszczają na nich wszystkie swoje prywatne dane. Niestety, a być może stety, nasz użytkownik zrobił to samo – dane uzyskane wyłącznie po sprawdzeniu profilu i strony domowej prezentuje tabela poniżej. Przedstawiłem informacje specjalnie w taki sposób – dla lepszego uwidocznienia ilości odkrytych danych. Zanim zaczniesz czytać dalej, drogi Czytelniku, proszę odpowiedz sobie na pytanie, czy chciałbyś, aby dowolna osoba w Sieci wiedziała o Tobie aż tyle

Dalsze poszukiwania są właściwie niepotrzebne, gdyż wszystkie ważne dane uzyskaliśmy z dokładnie dwóch najbardziej prozaicznych źródeł.

Jak widać, próba z całkowicie losowym użytkownikiem zakończyła się sukcesem – pomimo użycia tak prostych środków. Przypadek ten przedstawiłem jako przestrogę, gdyż w 79% analizowanych przeze mnie przypadków, dokładnie tyle działania wymaga określenie tożsamości dowolnej osoby korzystającej z zasobów Sieci.

Listing 1. *Sygnatura (podpis) umieszczony przez użytkownika.*

```
** Adam 's3rger' Szerski
** Unix(OpenBSD) System Administrator && C++ Programmer
** JID->s3rger[at]chrome[dot]net && GG->777XXX
** WWW->s3rger.somename.nu && EMAIL->s3rger@viX.pl
```

Kolejnym krokiem będzie wybór celu z grupy osób potencjalnie bardziej świadomych zagrożeń.

Za taką uznałem społeczność moderacyjną forum jednej z dystrybucji Linuksa.

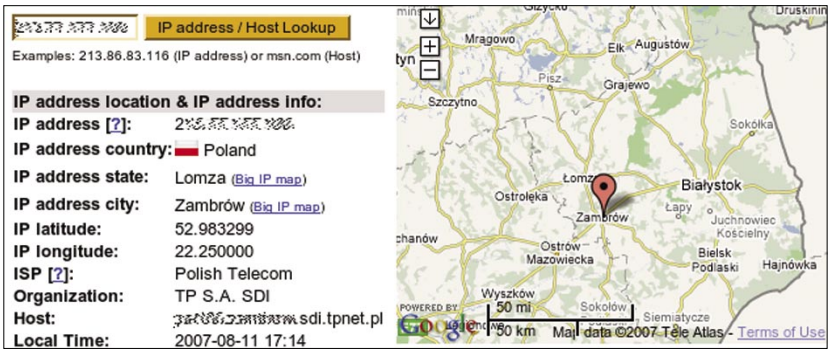
Procedura początkowa wygląda identycznie, jak poprzednio. Najpierw należy ustalić dane oczywiste – nick, e-mail i ewentualnie stronę domową, a w tym celu przeglądamy profil użytkownika. Sprawdzamy także posty, jakie zamieścił na forum. Po tej analizie określamy następujące dane:

- pseudonim: *s3rger,*
- e-mail: *s3rger@hotmail.com.*

Wstępnie widać, iż ten użytkownik prawdopodobnie bardziej chroni swoje dane – określiliśmy jedynie pseudonim i adres e-mail. W dalszym ciągu zależy nam na uzyskaniu danych personalnych – w tym celu wykorzystamy *google.com* i pewne cechy oraz *umiejętności* wyszukiwarki. Nie będę omawiać po raz kolejny, w jaki sposób formułować zapytania, ponieważ temat ten w ostatnim czasie był często poruszany i pojawiło się wiele traktujących o tym publikacji. W celu zapoznania się z zagadnieniem odsyłam na koniec artykułu, gdzie zamieściłem kilka przykładów bardziej zaawansowanych zapytań, oraz do ramki *W Sieci*, w której podałem linki do stron je opisujących. Zaczynając jakiekolwiek działania z użyciem wyszukiwarek (oczywiście Google jest skuteczne, ale wykorzystując także inne zwiększamy szanse uzyskania ciekawszych danych) i mając jedynie pseudonim i adres e-mail ofiary, warto wyświetlić wszystkie strony podchodząc do sprawy tak, jak w poprzednio opisanym przypadku. Da to nam z dobrym przybliżeniem wszystkie strony, na których obserwowany użytkownik prawdopodobnie pojawił się i prowadził jakieś działania. W analizowanej sytuacji po zastosowaniu prozaicznego połączenia dwóch wcześniej uzyskanych informacji otrzymaliśmy 30 wyników. Analiza każdej ze stron nie powinna być ani trudna, ani czasochłonna – dla przypomnienia dodam, że wciąż szukamy imienia i nazwiska. Zanim jednak to nastąpi i wejdziemy na jakąkolwiek stronę, zobaczmy co samo Google prezentuje w swoich wynikach.

Jak widać na Rysunku 2. nie wchodząc na żadną stronę uzyskaliśmy kilka ciekawych rezultatów:

- adresy IP, z których prawdopodobnie łączył się dany użytkownik,
- adres strony prawdopodobnie należącej do użytkownika,
- strony wskazujące na zainteresowania użytkownika,
- strony zawierające inne ważne informacje (m. in. nr telefonu, aktualne miejsce pobytu).



Rysunek 3. *Co kryje adres IP?*

Wszelkie te informacje zostały uzyskane minimalnym nakładem pracy przy zachowaniu maksimum bezpieczeństwa – gdyż przegląd, przynajmniej ten powierzchowny, stron przeprowadził googlebot, całkowicie wyręczając nas w tej niewdzięcznej czynności i zostawiając ślady charakterystyczne dla siebie.

Po pierwsze, zajmijmy się adresem *IP* – sprawdźmy za pomocą polecenia *host*, jaka jest nazwa *DNS* przypisana do odkrytego adresu – być może dostarczy nam ona nowych informacji (niektórzy dostawcy umieszczają w niej nazwę miejscowości na zasadzie indywidualny_numer.miasto.nazwa_dostawcy), a następnie zapytajmy bazy *whois* o inne dane dotyczące adresu. Warto pamiętać o tym, iż wiele stron zawierających księgi gości przedstawia publicznie adresy IP dopisujących się osób, stanowiąc tym samym znakomite źródło danych tego typu. Fakt, że w ostatnim czasie ze względów bezpieczeństwa pojawiły się „ucięte" adresy, np. bez udostępniania zawartości czwartego oktetu, ale jest to naprawdę nikłe zabezpieczenie. Po sprawdzeniu informacji o adresie IP skorzystamy z gotowych narzędzi do geolokacji i stwierdzimy, czy dane przedstawione na jednej ze stron są zgodne z rezultatami badań.

Listing 2. *Finger google.*

```
Links a navegar: 12
        /search?q=%40miasto.pl&num=20&hl=es&lr=&ie=UTF-8&as_qdr=all&start=0&
                        sa=N&filter=0
        "/search?q=%40miasto.pl&num=20&hl=es&lr=&as_qdr=all&ie=UTF-8&start=2
                        0&sa=N&filter=0"

To search: miasto.pl
sekretarz
sekretariat
...
sekretarz
biurorady
powiat
sekretariat
sekretariat
Accounts found: 7
        sekretarz
        sekretariat
        zcaburmistrza
        burmistrz
        biurorady
        info
```

```
$ host 83.26.50.XXX
XXX.50.26.83.in-addr.arpa domain
   name pointer aluXXX.neoplus.
   adsl.tpnet.pl.
```

Niestety – uzyskaliśmy jedynie informację, że użytkownik ten jest abonentem najpopularniejszego ISP w kraju.

Wyszukiwarka przedstawiła na jednej ze stron także inny adres IP. Warto sprawdzić pochodzenie drugiego adresu. Aby było nieco ciekawiej, skorzystamy ze strony *ip-address.com*, dzięki czemu będziemy mieli możliwość sprawdzenia na mapie, skąd prawdopodobnie pochodzi badany adres.

Pomijając otoczkę graficzną, zarówno mapa, jak i schludne podsumowanie generowane są na podstawie informacji z bazy.

Mając te dane możemy śmiało stwierdzić, że użytkownik pochodzi z rejonu Łomży lub Białegostoku, ze szczególnym naciskiem na okolice Zambrowa. Możemy również stwierdzić, że był on abonentem usługi SDI, po czym zmienił sposób dostępu do sieci na DSL. Od tej pory do wszelkich zapytań dodajemy kolejne parametry wyszukiwania, bardziej precyzując (lub rozszerzając) kryteria przeszukiwania o w/w miejscowości – być może użytkownik ten chociaż raz na forum miasta podpisał się imieniem.

Warto sprawdzić także obecność naszego celu na grupach dyskusyjnych – bardzo często użytkownicy stosują sygnatury, w których umieszczają ciekawe dane. W celu przeszukania tych grup

Jak odkryć tożsamość?

Krótki zarys metody.

- wykonujemy przeszukiwanie ze względu na informację, w której posiadaniu już jesteśmy (*nick/email/gg/etc.*),
- odrzucamy strony – śmieci, a następnie analizujemy powiązania informacji z pkt.1 z innymi danymi,
- w przypadku odkrycia powiązań spisujemy je i kontynuujemy poszukiwania jak powyżej dodając do już posiadanych informacji nowo odkryte powiązania,
- z nową listą danych powtarzamy całą procedurę – aż do uzyskania poszukiwanych informacji lub stwierdzenia, że nie istnieją one w publicznych zasobach Sieci.

Miejsca szczególnie obfite w informacje:

- blogi,
- prywatne strony www,
- portale społecznościowe,
- fora,
- grupy dyskusyjne,
- strony firmowe,
- strony i portale z ofertami pracy.

oczywiście można skorzystać z kombajnu sieciowego Google lub przejrzeć archiwum danej grupy. Można zastanawiać się na początku, której – ale odpowiedzi udzielą nam informacje zdobyte o tym użytkowniku na wstępie – trzeba szukać na grupach traktujących o Linuksie i – jak widać po ostatnich efektach szukania – tych o *BSD.

Widzimy teraz dokładnie (Listing 1), że o ile w przypadku stron WWW nasz użytkownik był naprawdę ostrożny, o tyle całkowicie zrezygnował z tego zachowania na grupach dyskusyjnych. Staliśmy się posiadaczami imienia i nazwiska oraz kilku innych cennych danych. Chciałbym zauważyć, iż grupy dyskusyjne są bardzo dobrym źródłem tego typu informacji i są często wykorzystywane właśnie do celów rozpoznawczych. Można na nich prócz takich informacji jak widoczne wyżej, znaleźć m. in. schematy sieci, wykazy urządzeń wykorzystanych w sieciach, dane administracyjne i konfiguracyjne usług dotyczące konkretnych sieci, a dlaczego? Dlatego, że grupy te, jako jeden z najstarszych sposobów wymiany danych, komunikacji i wzajemnej pomocy, przyciągnęły do siebie liczne grono specjalistów, których rady w sytuacjach kryzysowych były i są często bezcenne. Niestety, nie da się udzielić odpowiedzi na żadne pytanie nie znając dokładnie problemu i ewentualnych jego przyczyn – czyli konfiguracji urządzeń i serwerów, które będzie musiał opisać np. administrator będący w potrzebie.

Po tej *podróży* dysponujemy już znaczną ilością informacji na temat naszego celu, jednak jest to wciąż za mało, aby spełnić moje oczekiwania. W tym momencie przystępujemy do dalszej analizy i poszukiwań. Zarówno Google, jak i przeszukanie grup dyskusyjnych dało nam w odpowiedzi na nasze zapytania adres strony WWW – *s3-rger.somename.nu*.Tak jak poprzednio, warto przejrzeć stronę, której adresem podpisywał się nasz użytkownik. Niestety, witryna w czasie pisania artykułu była już niedostępna

Rysunek 4. *Waybackmachine*

(błąd 404). Prawdopodobnie część Czytelników zrezygnowałaby w tym momencie z dalszych działań, my pomimo wszystko tak nie zrobimy – cofniemy się o kilka dni, a może lat, w czasie sieciowym – *http://archive.org* jest miejscem, gdzie możemy to zrobić. Przechowywane są tam archiwa stron WWW zgromadzone w Internecie, a jednocześnie jest to znakomita alternatywa dla „archiwum" *cache Google'a*. (Proszę zauważyć, iż użytkownicy często umieszczają na stronach dane, których normalnie by nie umieścili, jednak czynią to ze względu na chwilową potrzebę. W takich sytuacjach boty podobne do *waybackmachine* czy cache Google'a stają się prawdziwym przekleństwem. Nie wszyscy wiedzą że niby usunięte dane wciąż mogą znajdować się w Sieci właśnie w takich miejscach, dlatego warto pamiętać o tym *narzędziu*.

Sprawdzenie występowania strony w archiwum zajęło dosłownie kilka sekund, a rezultaty osiągnięte dzięki tej czynności są niezwykle zachęcające – uzyskaliśmy potwierdzenie miejsca zamieszkania, informacje o studiach i dane na temat pracy w 2006 roku. Bardzo często popełnianym błędem jest pozostawianie na serwerze stron wchodzących kiedyś w skład portalu, a aktualnie niewykorzystywanych i nie podlinkowanych. Jest to często spotykane zagrożenie, mogące skutkować ujawnieniem wielu ważnych informacji, łatwo wykrywanych m. in. za pomocą archiwów takich, jak *http://archive.org*.

Będąc w posiadaniu znacznych ilości informacji, wciąż nie przestajemy szukać. Wręcz przeciwnie – wytaczamy nieco większe działo, które nazywa się *Maltego* (w trakcie pisania artykułu producent zmienił nazwę oprogramowania z *Evolution* na *Maltego* ze względu na problemy prawne), a zostało stworzone przez firmę Paterva. Jest to program, który przeprowadza działania podobne technicznie do tych, jakie mieliśmy

Listing 3. *Działanie list-urls*

```
$ ./list-urls.py http://xyzmiasto.pl
index.html            \
Burmistrz.htm       | strony które już nie są wykorzystywane
ZcaBurmistrz.htm | przez nowy system portalowy - na pewno warto sprawdzić
UrzadMiasta.htm /
http://www.umwpxxxxpl/bip/xyzmiasto/index.asp?pid=zamowieniap
http://www.umwpxxxx.pl/bip/xyzmiasto/index.asp?pid=inne
http://www.travelpolska.pl
http://www.nadbugiem.com.pl
http://www.podlaskieit.pl/pl/prot.htm
http://longger.info/sms/tv.php
```

możliwość zobaczyć we wszystkich wcześniejszych przykładach, a ponadto poszerza pole wyników przez analizę większej ilości danych i źródeł oraz wykonuje to automatycznie, bez ingerencji użytkownika. Jego oficjalne zastosowanie to określanie relacji między odnośnikami oraz związkami wyrazów (np. nick i imię), występujących w szeroko pojętych zasobach Internetu. Wyszukuje on zależności między ludźmi, grupami osób, sieciami, firmami, stronami WWW, dokumentami i plikami oraz elementami strukturalnymi sieci Internet (np. nazwy DNS, adresy IP itp.). Program wyświetla je w postaci drzewa zależności, zawierającego wszystkie połączenia (również te niewidoczne często z poziomu przeglądarki np. ze względu na zastosowanie takiego samego koloru czcionki i tła) między tymi elementami. Czy jednak działanie i samo narzędzie jest naprawdę tak dobre, jak mogłoby wydawać się po opisie? Czasami podejrzewam, że nawet lepsze – jest to idealne oprogramowanie do przeprowadzania tego typu analiz i nie tylko. Dla potwierdzenia moich słów zachęcam do zapoznania się z w/w oprogramowaniem (*http://www.paterva.com/web/Maltego*), a także do obejrzenia wyników na Rysunku 5. (ustawiony limit 20 relacji) – gdzie na czerwono podkreślone są miejsca przechowywania wrażliwych informacji.

Chciałbym teraz, już bez organizowania danych w sposób tabelaryczny, przedstawić wyniki działań. Poprzez przeszukiwanie ogólnodostępnych zasobów określiliśmy imię, nazwisko, adres zamieszkania, 3 adresy e-mail, numer GG, JID, numer telefonu komórkowego. Po wsparciu oprogramowaniem *Maltego* do listy dołączyły takie informacje, jak fakt przebywania poza granicami kraju, numer telefonu za granicą, dane personalne rodziców i dziadków(!) oraz numer PESEL.

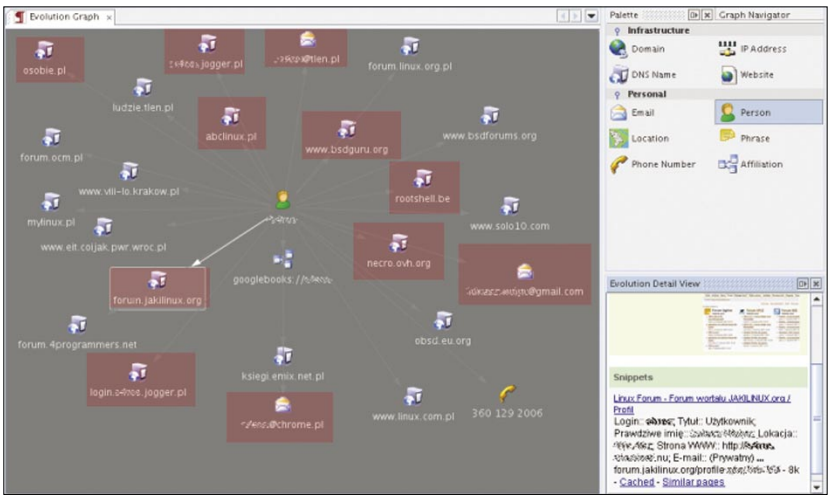
Kończąc tą część, przedstawię jeszcze – jako ciekawostkę – jedną stronę Biuletynu Informacji Publicznej pewnego urzędu – Rysunek 6.

Dysponując opisanym zestawem informacji oraz danymi z tabeli dotyczącymi pierwszego użytkownika, chciałbym przejść do analizy zagrożeń, ale najpierw zwrócę uwagę na dwie rzeczy.

Po pierwsze prostota zdobycia naprawdę istotnych danych jest przerażająca – opisane działania nie wymagają od nikogo specjalistycznej wiedzy, a jedynie chęci i determinacji. Po drugie ilość danych idzie w parze z czasem i łatwością poszukiwań. Co z tego wynika? Rośnie obszar potencjalnego użycia zdobytych informacji, a zatem i zestaw możliwych ataków. Patrząc realnie na aktualny stan rzeczy, działania osoby wykonującej – nazwijmy to – *test*, są ograniczone jedynie jej wyobraźnią. Wszystko to ma znaczenie jedynie teoretyczne, jeśli dotyczy audytora, natomiast gdy wykonującym te działania jest osoba, której celem jest wyrządzenie szkód, sprawa się komplikuje. Mając do dyspozycji takie dane jak te „zdobyte" wcześniej, napastnik może z dużą skutecznością przeprowadzać ataki wykorzystujące inżynierię społeczną – największe zagrożenie bezpieczeństwa, a jest to tylko jeden z przykładów. Kolejną możliwością jest całkowita kradzież tożsamości – zauważmy, że w przypadku drugiej osoby jedynie brak numeru dowodu osobistego utrudnia nam założenie interne-

Listing 4. *Dane prezentowane przez whois.*

```
domain: xyzmiasto.pl
registrant's handle: dmra00XXXXXXXX (CORPORATE)
nservers: ns1.xxnat.pl.[193.239.44.XX]
                ns2.xxnat.pl.[83.18.80.XX]
last modified: 2006.11.02
option: the domain name has not option
REGISTRANT:
company: URZĄD MIEJSKI W XYZMIASTO
street: FABRYCZNA 37
city: XYZMIASTO
location: PL
handle: dmra00XXXXXXXX
phone: +48.XXZZZYYXX
last modified: 005.08.19
REGISTRAR: Rejestrator
ul. K. 37
85-079 Bydgoszcz
Polska/Poland
+48.525XXXXXX
www.rejestratorwwww.pl/info
Whois database last updated:
2007.09.15
```



Rysunek 5. *Działanie Maltego*

towego konta bankowego – a to już nie jest tak duży problem. Wystarczy znaleźć więcej informacji o danej osobie i się jej po prostu zapytać. Z ukradzionej tożsamości agresor może także korzystać w trakcie podróży po Internecie. Może on wykorzystać takiego użytkownika jako kozła ofiarnego w innych swoich działaniach lub w celu odwrócenia uwagi – umiejętnie operując danymi, które posiada. Z wielką łatwością napastnik może się podszywać pod wybraną osobę np. w celu zdobycia potrzebnych danych z atakowanej firmy itd. Konkretnych przykładów inżynierii społecznej przedstawiać tutaj nie będę gdyż jest to bardzo długi i ciekawy temat, a jako uzupełnienie polecam publikacje internetowe, jak i książkowe.

Zdobyte informacje mogą służyć również do przeprowadzania kradzieży – nie tylko informatycznych, ale także tych zwykłych, znanych z świata realnego. Dysponując tak znaczną ilością informacji, stworzenie wielu rodzajów dokumentów to kwestia minut, a pobranie kredytu *od zaraz* także nie stanowi problemu ze względu na warunki i procedury.

Jak już wcześniej napisałem – a Czytelnik mógł osobiście tego doświadczyć – zdobycie danych w taki sposób jest naprawdę trywialne. Samo powstanie programu Maltego jest oznaką zainteresowania szerszej społeczności internetowej w/w zagadnieniami. Sam, przeprowadzając analizy, skusiłem się na stworzenie drobnego programu automatyzującego wyszukiwanie danych i nie wierzę, że ktoś inny nie zrobił tego samego, być może już w innych celach. Program ten stworzyłem jedynie na potrzeby testu, bez większego wkładu i zaangażowania, ale był w stanie zwrócić dane zgodne z rzeczywistymi w 1389 przypadkach na 2000 (co stwierdziłem na drodze późniejszej, *ręcznej* analizy). Wyszukiwał on jedynie dane personalne: imię, nazwisko, adres e-mail i był w stanie przy rekurencyjnej analizie stron (do 7 od korzenia) znaleźć numery GG, JID i adres zamieszkania, wyodrębniając te dane na podstawie pewnych cech charakterystycznych (numer GG – ciąg cyfr, dla uproszczenia od 6 do 10 znaków, e-mail – ciąg znaków alfanumerycznych zawierający @; poprzednia reguła wraz z listą serwerów Jabbera wyodrębniała źle sklasyfikowane JID jako adresy z listy e-mail itd.). Jak więc widać, był on naprawdę ograniczony – jednak podejrzewam, że przy odpowiedniej determinacji z łatwością można byłoby go rozbudować o inne funkcje. Jako taki mógłby być on wykorzystany do tworzenia baz danych dla automatów spamujących, lub sam zostać poszerzony o taką *usługę*. Oprócz tego mógłby służyć do generowania takich baz danych, które z przyjemnością ktoś mógłby kupić – a o nabywców nietrudno. Wiele firm zajmuje się między innymi profilowaniem – zbieraniem danych o zwyczajach i zachowaniach użytkowników, chociażby dla celów marketingowych, a im bardziej szczegółowe informacje (od danych personalnych po przyzwyczajenia i zainteresowania), tym więcej baza jest warta na sieciowym czarnym rynku. Łatwo więc zauważyć, że zagrożenia istnieją i są rzeczywiście poważne.

Część III

W tej części artykułu postaram się zaprezentować niektóre metody związane z identyfikacją systemu zdalnego. Na początek obierzemy cel, a następnie sukcesywnie będziemy poszerzali nasz zasób wiedzy o danej maszynie i sieci, stosując metody od najmniej inwazyjnych, po takie, których używania teoretycznie nie sposób przeoczyć podczas analizy logów (aczkolwiek praktyka widziała różne rzeczy), a pomimo to bardzo przydatne i często wykorzystywane.

Za cel pozwoliłem sobie wybrać stronę urzędu miasta, które dla potrzeb artykułu zostało przemianowane na Xyzmiasto. Całość swoich działań oprzemy o pewien schemat, mianowicie prześledzimy trasę pakietów od naszego komputera do komputera testowanego, następnie dokładnie zapoznamy się ze stroną WWW i podstawowymi usługami internetowymi (o ile są świadczone). Kolejnym krokiem będzie dokładniejsze i bardziej zaawansowane przeszukanie zasobów sieciowych przy użyciu Google oraz wykrycie użytkowników związanych z danym komputerem. Podążając ścieżką wytyczoną w części drugiej – dowiemy się, co ciekawego ma nam do przekazania program *Maltego*. Mając pewien zasób informacji, przejdziemy do analizy na nieco niższym poziomie – zajmiemy się pakietami produkowanymi i odbieranymi przez serwer. Kolejnym naszym działaniem będzie podwyższenie poziomu anonimowości własnej stacji i dalsza analiza, tym razem jednak z wykorzystaniem metod bardziej inwazyjnych. Nie mówiąc już nic więcej, przystępujemy do działania.

Po pierwsze sprawdzimy, jak wygląda droga pakietów od naszego

http://bip.[illegible].pl/?id=[illegible]

		Rubryka 2 - Organ nadzoru	
1		1. Nazwa organu	RADA NADZORCZA
		Podrubryka 1 Dane osób wchodzących w skład organu	
	1	1. Nazwisko	WR[illegible]RCZYK
		2. Imiona	EW[illegible]TA
		3. Numer PESEL	57[illegible]888
	2	1. Nazwisko	UR[illegible]CZYK
		2. Imiona	RO[illegible]ALEKSANDER
		3. Numer PESEL	53[illegible]491
	3	1. Nazwisko	DY[illegible]SKI
		2. Imiona	GR[illegible]RZ MAREK
		3. Numer PESEL	62[illegible]236
	4	1. Nazwisko	KL[illegible]
		2. Imiona	JĘ[illegible] STEFAN
		3. Numer PESEL	70[illegible]418

Rysunek 6. *Czy trzeba było to umieszczać w takiej postaci?*

komputera do serwera docelowego. Da nam to wyobrażenie o „*odległości* internetowej i pozwoli określić, czy będziemy w stanie poddać dany serwer całkowitemu podsłuchowi. Działanie to można wykonać na przykład wtedy, gdy serwer-cel znajduje się za serwerem stanowiącym bramkę dla niego i będącym jednocześnie słabiej zabezpieczonym – a z doświadczenia wiem, że czasem łatwiej spenetrować serwer pośredniczący, niż ten właściwy. Samo badanie ścieżki możemy wykonać na 3 sposoby: pierwszy, całkowicie manualny, polega na wysyłaniu do naszego celu pakietów z ustawionym czasem życia pakietu (*TTL*) na wartość *1* i nasłuchiwaniu odpowiedzi ICMP serwera, przez który pakiet taki nie przeszedł ze względu na przekroczenie w/w czasu. Spisując po kolei każdy adres serwera, od którego pochodzi odpowiedź, a następnie zwiększając wartość *TTL* i powtarzając tę czynność do czasu, gdy serwerem odpowiadającym będzie docelowy, określamy drogę, jaką muszą pokonać pakiety. Sposobem automatyzującym całą procedurę jest wykorzystanie programu konsolowego – *traceroute*, a sposobem numer trzy – skorzystanie z rozszerzonej i okienkowej wersji *tracerout'a* – *mtr*. Oprócz tego, co napisałem wcześniej – możliwości przeprowadzenia podsłuchu całego ruchu przychodzącego i wychodzącego z naszego celu – badanie to da nam możliwość określenia komputera, który mógłby stać się celem pośrednim. Atakując komputer lub router znajdujący się bezpośrednio na drodze do naszego celu i jednocześnie przed nim, jesteśmy w stanie przechwycić i fałszować cały ruch – np. w celu przeprowadzenia ataku typu phishing, DoS lub innego, którego rezultatem stałoby się faktyczne odcięcie naszego celu od sieci lub sfałszowana komunikacja z klientami. Wykrywanie dróg pakietów ma również bardziej przyziemne zastosowanie – możemy sprawdzić, jak wiele krajów odwiedzają bity generowane przez nasz komputer, zanim dotrą do celu. Wyniki uruchomienia polecenia *traceroute* przedstawia Rysunek 7.

Listing 6. *Co zwrócił nam nmap?*

```
Starting Nmap 4.20 ( http://insecure.org ) at 2007-08-12 12:46 CEST
Interesting ports on xyzmiaso.pl(-):
PORT     STATE  SERVICE  VERSION
21/tcp   open   ftp      ProFTPD 1.3.0a
22/tcp   open   ssh      OpenSSH 4.6 (protocol 1.99)
23/tcp   closed telnet
25/tcp   closed smtp
53/tcp   closed domain
80/tcp   open   http     Apache httpd 2.2.4 ((Unix) DAV/2 PHP/5.2.3)
1433/tcp closed ms-sql-s
3306/tcp closed mysql
5432/tcp closed postgres
No exact OS matches for host (If you know what OS is running on it, see http:
                      //insecure.org/nmap/submit/ ).

TCP/IP fingerprint:
OS:SCAN(V=4.20%D=9/13%OT=22%CT=1%CU=42304%PV=N%DS=0%G=Y%TM=46E977EB%P=i486-

OS:slackware-linux-gnu)SEQ
<CUT>
```

Widzimy więc, że analizowany przez nas serwer znajduje się w odległości 15 urządzeń sieciowych, przez które muszą przejść pakiety. Aby określić występowanie możliwości, o których pisałem wcześniej, należałoby sprawdzić dokładniej, co znajduje się na ostatniej przedstawionej na Rysunku pozycji – my jednak tego teraz robić nie będziemy, a ewentualne próby na urządzeniach *testowych* zostawiam jako pracę domową. Wiemy już na pewno (poprzez analizę trasy), że komputer obsługujący stronę urzędu miasta nie znajduje się bezpośrednio za komputerem, który mógłby służyć jako dodatkowy system logowania – możemy więc czuć się pewniej w działaniach *wywiadowczych*. Proponuję, aby samemu sprawdzić trasy do kilku serwerów lub portali – czasami daje to bardzo ciekawe informacje, ale to proszę potraktować jako ciekawostkę.

Jak widać na Rysunku 7, od pewnego momentu zaczynają pojawiać jednostki z nazwami DNS o końcówkach z pochodzenia zagranicznych. Może mieć to dwa, a w niektórych przypadkach trzy powody, bardzo do siebie zbliżone – po pierwsze, ISP świadczy usługi wykorzystując jako jednostki nadrzędne dostawców z zagranicy, ale pomimo to serwer znajduje się w budynku urzędu. Po drugie, strona korzysta z hostingu, który wykorzystuje łącza zagraniczne i po trzecie – są to zwykłe

Wielkość śladów

Bez precyzowania kryteriów:

```
-rw-r--r--  1 root root 1280427 2007-08-12 11:57 access_log
-rw-r--r--  1 root root 1312252 2007-08-12 11:57 error_log
-rw-r-----  1 root    root       9957 2007-08-12 11:55 messages
-rw-r-----  1 root    root       1684 2007-08-12 11:53 syslog
```

Po sprecyzowaniu kryteriów:

```
-rw-r--r--  1 root root 8230 2007-08-12 12:01 access_log
-rw-r--r--  1 root root 9956 2007-08-12 12:01 error_log
-rw-r-----  1 root    root       3257 2007-08-12 11:55 messages
-rw-r-----  1 root    root        174 2007-08-12 11:53 syslog
```

trasy pakietów jednego z największych dostawców w Polsce i to jest ta ciekawostka, o której pisałem wcześniej.

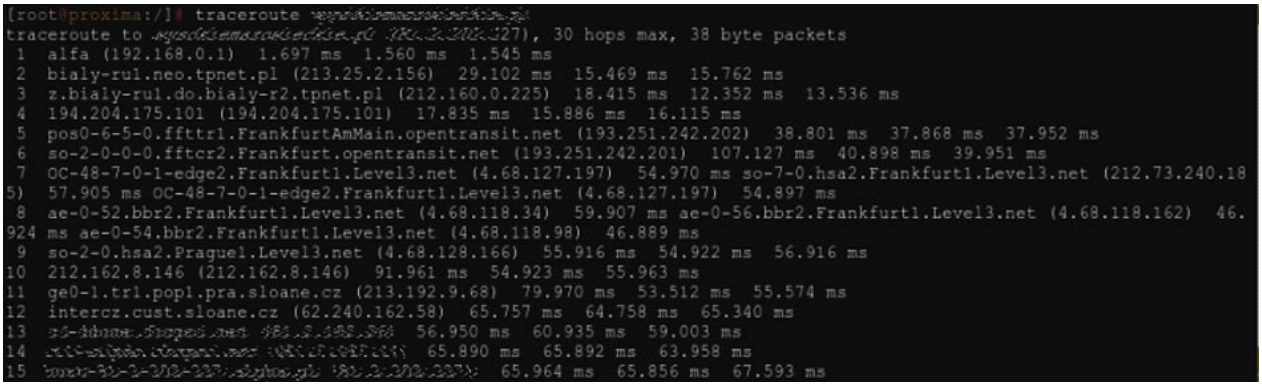
Bardziej zaawansowaną metodą detekcji i analizy serwera jest pasywny przegląd cech charakterystycznych pakietów generowanych przez komputer poddawany badaniu. Przede wszystkim musimy być świadomi, że każda implementacja obsługi protokołów sieciowych jest inna – pozwala nam to na detekcję m. in. rodzaju systemu pracującego na danej maszynie. Celowe wymuszanie obciążenia komputera lub zmuszanie go dowolnymi metodami do pracy w sposób odbiegający od założeń może spowodować np. wycieki pamięci z komputera odpowiadającego na sztuczny ruch – a wszystko przez błędy implementacji. Oprócz tego, dzięki analizie pakietów pochodzących z danej maszyny możemy określić fakt istnienia oraz budowę sieci za danym serwerem. Ale po kolei.

W identyfikacji pasywnej opartej na analizie cech charakterystycznych głównym i najważniejszym źródłem, a jednocześnie gwarantem sukcesu, jest odpowiednio duża i bogata baza danych zawierająca ów zbiór sygnatur. Pierwszą możliwością jest wykrycie łącza – możemy spróbować analizy adresu IP i zadania pytania dla baz danych *whois*, a następnie wydedukować typ możliwego łącza, ale również możemy poszerzyć swoją wiedzę sprawdzając łączną długość pakietu IP (16-bitowe pole w nagłówku datagramu IP). Długość ta, pomimo że z góry ograniczona jest wartością 65535, w praktyce bywa różna – w zależności od ograniczeń narzuconych przez protokoły niższej warstwy. Podążając tą drogą i znając ograniczenia np. *MTU*, możemy określić że np. typowe ustawienie parametru *MTU* dla łącza *DSL* wynosi 1492. Odrzucając możliwość wystąpienia łącza DSL na serwerach pośrednich (sprawdzonych *traceroute'm*), możemy stwierdzić, że takie łącze jest wykorzystywane przez serwer. Następnie wkraczamy na nieco głębsze wody – protokół TCP. Na podstawie jego charakterystycznych ustawień możemy określić system pracujący na danym komputerze. Mając przed oczami budowę pakietu TCP (Rysunek 8), chciałbym przedstawić kilka faktów:

- ustawiony znacznik DF w pakietach pochodzących od serwera może wskazywać, że zostały one wygenerowane w systemach: Windows, Linux, *BSD, Solaris,
- rozmiar okna w zależności od swojej wartości może nasuwać następujące wnioski:
  - dwu- lub czterokrotność maksymalnego rozmiaru segmentu –

Listing 5. *Wynik działania dig dla wp.pl*

```
$ dig wp.pl

; <<>> DiG 9.4.1 <<>> wp.pl
;; global options:  printcmd
;; Got answer:
;; ->>HEADER<<- opcode: QUERY, status: NOERROR, id: 60633
;; flags: qr rd ra; QUERY: 1, ANSWER: 1, AUTHORITY: 3, ADDITIONAL: 3

;; QUESTION SECTION:
;wp.pl.                         IN      A

;; ANSWER SECTION:
wp.pl.                  1832    IN      A       212.77.100.101

;; AUTHORITY SECTION:
wp.pl.                  1857    IN      NS      dns.task.gda.pl.
wp.pl.                  1857    IN      NS      ns2.wp.pl.
wp.pl.                  1857    IN      NS      ns1.wp.pl.

;; ADDITIONAL SECTION:
dns.task.gda.pl.        205394  IN      A       153.19.250.100
ns1.wp.pl.              1948    IN      A       212.77.102.200
ns2.wp.pl.              2551    IN      A       153.19.102.182

;; Query time: 43 msec
;; SERVER: 194.204.152.34#53(194.204.152.34)
;; WHEN: Sat Sep 15 17:57:07 2007
;; MSG SIZE  rcvd: 150
```



Rysunek 7. *Wyniki działania programu traceroute*

praca Linuksa,
    - wartość 64512 – na system z rodziny Windows.
- wartość pola TTL:
    - zbliżona do wartości 64 – wskazuje na system Linux, HP/UX, IRIX, AIX,
    - zbliżona do wartości 128 lub 32 – systemy z rodziny Windows,
    - 255 – może wskazywać na Solarisa,
    - zbliżona do wartości 60 – MacOS.
- zerowy identyfikator pakietu może wskazywać na Linuksa w wersji 2.4.,
- ustawienie datownika określa czas aktywnej pracy systemu.

Jest to tylko niewielka część danych, które można wydobyć analizując pakiety dochodzące do nas np. podczas przeglądania strony internetowej – a widać jasno, że przy takiej identyfikacji przede wszystkim bierzemy pod uwagę następujące pola:

- dla protokołu IP*: TTL, ID, TOS, bit DF,*
- dla protokołu TCP: kombinację ustawień i opcji *Window Size, MSS, NOP, SA, Datownik*.

Łatwo też dostrzec, że wszelkie dane ustalane są na podstawie zgromadzonych lub zaobserwowanych cech charakterystycznych. Dlatego też wykorzystanie analizy pasywnej wiąże się z koniecznością posiadania bogatej bazy tych informacji. Jak widać, jest to temat rzeka.

Następnym elementem naszej analizy będzie dokładne zapoznanie się ze stroną WWW. W jaki sposób, nie będę opisywać – jest to intuicyjne. Ciekawszym wykorzystaniem serwera HTTP w tym przypadku jest sprawdzenie wersji samego serwera i być może systemu operacyjnego. Wystarczy wejść na stronę, która nie istnieje, a w wielu przypadkach pojawi się (czy to dla serwera Apache czy dla tego produkcji Microsoftu) strona z błędem 404 i poszukiwanymi danymi (w tym przypadku dla Apache): *Not Found. The requested URL /url was not found on server. Apache/1.3.29 Server at miasto.pl Port 80*. Jest to naprawdę często spotykane, a na tej podstawie bardzo łatwo wyznaczyć cel ataku – jeżeli wiemy że Apache w wersji 1.3.29 jest podatny na jakiś błąd, przy czym nie zależy nam na konkretnej maszynie, a jedynie na samym fakcie włamania, wystarczy poprosić Google, żeby wyświetliło nam kilka stron z adresami, gdzie takie serwery się znajdują.

Kolejnym krokiem będzie analiza przeprowadzana właśnie za pomocą Google'a. Po pierwsze należy ustalić, co chcielibyśmy znaleźć:

- hasła,
- loginy,
- prywatne dane,
- dane konfiguracyjne,
- wrażliwe informacje.

Aby to uczynić, należy odpowiednio przeszukać zasoby udostępnione przez serwer. Ze swojej strony proponuję sprawdzić, czy na serwerze występują pliki *robots.txt* – odpowiedzialne za nadzór ruchu botów internetowych na stronie. Często pliki i foldery zablokowane przed indeksowaniem przez boty zawierają ważne informacje. Sprawdzamy prozaicznym pytaniem: `intitle: index.of robots.txt`.

Sprawdzić, czy wśród danych publicznych lub zabezpieczonych przez ukrycie, nie występują interesujące dane. W tym celu skorzystamy z wyszukiwania w nazwie *domeny xyzmiasto.pl* (`site: xyzmiasto.pl`) stron zawierających indeks folde-

Tabela 3. *Usługi oraz porty, na jakich nasłuchują ich aplikacje*

Nazwa usługi	Port	Opis
ftp-data	20/tcp/udp	Protokół przesyłu plików [dane]
ftp	21/tcp/udp	Protokół przesyłu plików [sterowanie]
ssh	22/tcp/udp	Zdalny dostęp do serwera – usługa Secure Shell Login
telnet	23/tcp/udp	Zdalny dostęp do serwera – usługa telnet
smtp	25/tcp/udp	Simple Mail Transfer Protocol – wysyłanie poczty
domain	53/tcp/udp	Domain Name Server – obsługa zapytań DNS
finger	79/tcp/udp	Usługa finger
http	80/tcp/udp	Usługa World Wide Web HTTP
pop3	110/tcp/udp	Post Office Protocol - Version 3 – odbiór poczty
sftp	115/tcp/udp	Simple File Transfer Protocol
netbios-ssn	139/tcp	NETBIOS Session Service
sqlsrv	156/udp	SQL Service
ms-sql-s	1433/tcp	Microsoft SQL Server- baza danych SQL firmy Microsoft
mysql	3306/tcp	MySQL – baza danych SQL
postgresql	5432/tcp	Baza danych PostgreSQL

rów ze specyficzną nazwą. Zrobimy to stosując zapytanie „`index.of._nazwa_folderu_`". Proponuję zbadać istnienie folderów o nazwach: prywatne, ukryte, `temp`, `secure`, `pass`, `moje`, wszelkie kombinacje w/w nazw angielskich i polskich, `backup`, `kopia`, `tmp etc`. – wszystko według uznania poszukującego.

Sprawdzić, czy na serwerze nie znajdują się raporty *Nessusa*, co ułatwiłoby bardzo zadanie: `intitle: "Nessus Scan Report", "This file was generated by Nessus"`. Szukamy, jak poprzednio, w domenie *xyzmiasto.pl*.

Sprawdzić, czy na serwerze nie znajdują się pliki zawierające nazwy użytkowników lub hasła dostępne z poziomu wyszukiwarki – `passwd etc`.

Sprawdzić, czy serwer nie obsługuje systemu statystyk takiego jak np. *phpsysteminfo* lub inny podobny – dałoby nam to ogromny zasób informacji o maszynie: *+intext: Info +intext:Usage Statistics for*.

Więcej przykładów zapytań znajduje się w *GHD* – zajrzyj do ramki *W Sieci*.

Innym ciekawym zastosowaniem wyszukiwarki jest zadawanie pytań zwracających wyniki podobne do rezultatów działania programu *finger*. Wyszukiwarka wyświetli nam wszystkich użytkowników, których działalność jest bezpośrednio kojarzona z daną domeną według zindeksowanych przez nią stron. W celu wykorzystania tej funkcjonalności możemy jak zwykle zadawać pytania ręcznie lub wykorzystać gotowe skrypty, np. autorstwa Sergio Alvareza. Listing 2. prezentuje uzyskane wyniki, przy czym pierwsze linijki to wygląd zapytania, a następne to zwracane rezultaty.

Prowadząc w dalszym ciągu działania tego typu, wyszukiwarka po zastosowaniu innego skryptu zwróciła nam aktualne adresy e-mail na podstawie listy z Listingu 2. – do każdej nazwy należy dodać *@xyzmiasto.pl*. Istnieje inne interesujące zastosowanie – ekstrakcja adresów URL z podanej domeny. W tym celu dla wygody i automatyzacji możemy zastosować narzędzie o nazwie *list-urls*. Uzyskane adresy pozwolą nam określić inne strony, na których być może znajdują się dane dotyczące interesującej nas maszyny. Często taka sytuacja ma miejsce i istotnie wpływa na obniżenie poziomu bezpieczeństwa. Przykładowy wynik działania programu prezentuje Listing 3.

Przechodząc na nieco wyższy poziom sprawdzania i poszukiwania, warto zapoznać się z wynikami dostarczanymi przez bazy *whois* – działania funkcjonalne *DNS* opisane są w *RFC* o numerach 1034 i 1035 w wolnych chwilach można się z nimi zapoznać. Należy pamiętać, że istnienie poddomen dla danej domeny z nazwami np. *smtp*, *pop*, *pop3*, *www*, *ftp* itp. może świadczyć o istnieniu tych usług na konkretnej maszynie. Ułatwia nam to pracę przy analizie hosta, ponieważ wygenerujemy mniej ruchu, który zostanie zalogowany na serwerze, a ponadto nie musimy już sprawdzać istnienia usług we wstępnej fazie działań, co znacznie je przyspieszy. Mamy także świadomość, że serwer może być narażony na ataki wymierzone w te usługi, dlatego też zawęża się nasze pole poszukiwań w przypadku testów penetracyjnych. Wśród danych wyświetlanych z baz danych DNS znajduje się kilka innych wartych uwagi rzeczy – przykładowy wycinek przedstawia Listing 4.

Jesteśmy w stanie zdobyć dane administratora domeny, co może być bardzo pomocne w działaniach *social engineering* oraz zdobywamy informacje pomocne w przypadku prób dostępu do strefy DNS lub

Tabela 4. *Rezultaty działania Nessusa*

Wykrycie	Status/Opis
System	Linux
Poziom Zagrożenia	Wysoki
Otwarte porty	9, z czego 2 jako zagrożone Port palace-5 (9996/tcp) działające PsyBNC Port ftp (21/tcp) Anonymous FTP – dostępne Port ssh (22/tcp) 1.33 1.5 1.99 2.0 SSHv1 host key fingerprint : 75:97:cc:70:2c:87:56:c0:b0:77:38:f8:56:20:b1:2b *SSHv2 host key fingerprint : 16:90:40:fb:9e:f0:6e:ea:51:bc:51:9d:2b:74:49:2e* Port http (80/tcp) mod_ssl hook functions format string vulnerability Synopsis : Arbitrary code can be executed on the remote host Description : The remote host is using a version of mod_ssl which is older than 2.8.18. This version is vulnerable to a flaw which may allow an attacker to disable the remote web site remotely, or to execute arbitrary code on the remote *host*. Apache mod_proxy content-length buffer overflow
ltd.	

jej transferu (za pomocą np. *Sam Spade*). Inną możliwością uzyskania przydatnych informacji dotyczących nazw internetowych danej maszyny jest wykorzystanie programu *DiG*, a efekty jego wykonania prezentuje Listing 5. Dane uzyskane w ten sposób, pomimo że same w sobie nie są groźne, mogą zostać wykorzystane w bardzo niebezpieczny sposób. W tym momencie nasuwa się pytanie o problem ukrycia danych personalnych w rejestrach i bazach DNS. Jak się przedstawia sytuacja? Możliwość ukrycia danych na pewno poprawiła by bezpieczeństwo i zapewniła większą prywatność właścicielom domen. W końcu sam przez cały artykuł mówię o odpowiednim poziomie bezpieczeństwa oraz rozwadze przy publikacji jakichkolwiek danych. Z drugiej strony pojawiają się zwykli użytkownicy korzystający z Sieci – ukrycie danych w bazach przede wszystkim przeszkodziłoby właśnie im, gdyż personalia te są jedynym sposobem zidentyfikowania prowadzącego usługi. Jak widać, w tej kwestii problem jest trudny do rozwiązania. Nadeszła pora wykorzystania metod aktywnych i silniej ingerujących w działanie serwera. Aby móc odczuwać pewien komfort pracy, warto zadbać przynajmniej w minimalnym stopniu o własne bezpieczeństwo. W tym obszarze mamy kilka możliwości, a do bardziej popularnych należy wykorzystanie metody *onion routingu* (aplikacja *TOR*) lub użycie oprogramowania *JAP*.

Jaka jest ich zasada ich działania w pigułce? *JAP* wykorzystuje serwery pośredniczące w taki sposób, że użytkownik jest ukryty za wszystkimi połączeniami (tworzone są wirtualne tunele). Nikt z zewnątrz sieci pośredniczącej, a nawet z samej sieci nie jest w stanie przyporządkować konkretnego połączenia do użytkownika. Usługę serwerów pośredniczących udostępniają niezależne instytucje, które oficjalnie deklarują nieprzechowywanie logów połączeń.

Mając już nieco podniesiony stopień anonimowości połączeniowej (należy pamiętać, że pomimo tak wspaniałych deklaracji i cudownych opisów działań, w Sieci ciężko być anonimowym), przystępujemy do dalszych analiz.

Kolejnym naszym krokiem będzie ręczna analiza usług świadczonych przez serwer. Zaczniemy od poszerzenia informacji dotyczących usług pracujących na zdalnej maszynie. Po badaniach przeprowadzonych wcześniej jesteśmy w 90% przekonani, iż system obsługuje pocztę, serwer *WWW* i *FTP* (lub w najbliższym czasie zacznie). Możemy to wnioskować z tak wysokim poziomem pewności, gdyż mamy świadomość istnienia odpowiednich nazw *DNS* związanych z testowaną domeną. Aby nie generować od razu wielkiej ilości logów, sprawdzimy jedynie usługi najczęściej występujące. Jakie to usługi? Przedstawia je tabelka usług przygotowana na podstawie pliku services.

Sprawdzenie, chociażby przez połączenie za pomocą telnetu z każdym portem, prócz upewnienia nas o istnieniu danej usługi, zwróci nam często jej baner. Informacja taka jest prezentowana jako powitanie, które nierzadko zawiera wersję serwera, lub nawet systemu na którym ten pracuje.

Co dalej? Sprawdźmy, co powie nam o systemie *nmap*: *nmap -sS -sV -p 13,21,22,80,1433, 3306,5432 -P0 xyzmiasto.pl* (opcje i sposoby skanowania to temat na oddzielną publikację, dlatego nie będę ich tłumaczyć nawet powierzchownie – zapraszam do zapoznania się z odnośnikami).

W rezultacie uzyskaliśmy dane takie, jak na Listingu 6. z których możemy odczytać wersję *demona ftp, ssh i http*. Do informacji zaprezentowanych przez nmapa należy także zaliczyć dane o systemie, prezentowanym jako nieznany. Do wyników można podejść na kilka sposobów, ale najlepiej przyjrzeć się sekcji *fingerprint*. Znajduje się w niej informacja, która może być niezwykle ważna – slackware–linux–gnu. Możemy wnioskować z zaprezentowanych danych, iż serwer pracuje na systemie Linux Slackware, a ze względu na niemożliwość dokładnego rozpoznania, że jest to świeża instalacja np. wersji 12.0 lub z *currenta*.

W każdym przypadku należy pamiętać, iż podczas wykonywania jakichkolwiek analiz, zawężenie obszaru poszukiwań i testów zmniejsza znacznie ilość danych które zostaną zapisane w logach serwera. Na dowód tych słów proszę zapoznać się z ramką *wielkości śladów*.

Kolejnym narzędziem, które zastosujemy jest *Nessus* – znakomity skaner podatności. Jest on w zasadzie podstawowym narzędziem, które jest wykorzystywane w wielu sytuacjach przez specjalistów różnej kategorii. Najważniejszym elementem wykorzystania *Nessusa* jest umiejętność doboru kryteriów testu. Będąc już po licznych krokach analizy serwera, jesteśmy w stanie określić jego konfigurację i budowę strukturalną – po analizach pasywnych możemy śmiało powiedzieć, że komputer nie pracuje ani na systemie produkcji *Microsoftu*, ani na **BSD* czy *Solarisie*. Potrafimy nawet wskazać, że systemem zarządzającym maszyną jest *Linux Slackware*. W trakcie badania strony nie mogliśmy stwierdzić działania aplikacji *CGI*, ani też faktu istnienia bardziej zaawansowanych i znanych systemów portalowych w *PHP*. Wiemy także, że serwer na pewno nie jest serwerem sprzętowym, zatem możemy wyłączyć z procesu analizy kolejne grupy wtyczek *Nessusa*. Śmiało usuwamy wszystkie pluginy związane z systemami innymi niż *Linux* w ogólności, oraz inne niż *Slackware*, możemy także usunąć te związane z *CGI* oraz routerami sprzętowymi *CISCO*. W tym stadium analizy nie interesują nas także podatności na ataki *DoS* ani możliwości zdalnego przejęcia uprawnień *roota*. Przeglądając dalej listę pluginów *Nessusa* oraz analizując ją w kontekście posiadanych informacji, jesteśmy w stanie znacznie zmniejszyć potencjalną ilość logów wygenerowanych na analizowanym systemie. Do skutecznego wykorzystania *Nessusa* powyższa ścieżka jest najlepsza – im więcej dowiemy się o serwerze bez silnej ingerencji, tym mniej śladów zostawimy w trakcie działań aktywnych. Sam proces skanowania nie jest dla nas aż tak istotny, ważniejsze są natomiast wyniki, które częściowo prezentuje tabela rezultatów. Warto

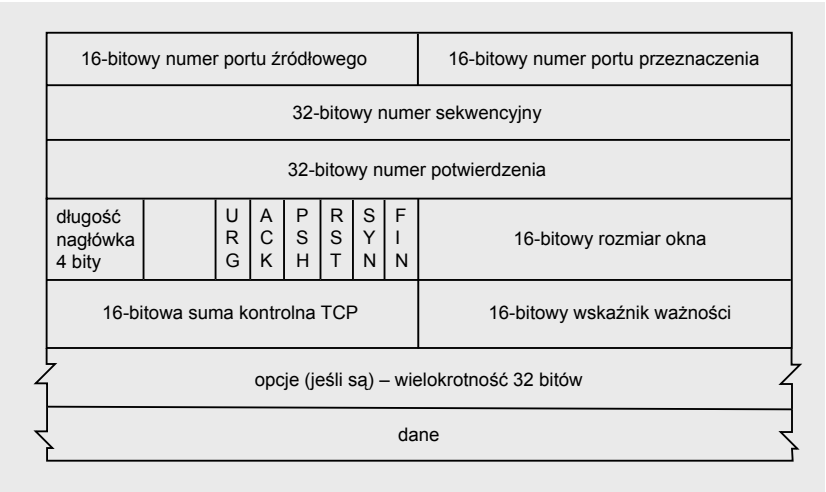
pamiętać, że *Nessus* oprócz wskazania potencjalnych błędów zamieszcza do nich opis i sposób usunięcia, oraz że czytanie o czymś bez robienia tego nie ma sensu – zachęcam do zabawy tym programem i poznawaniem świata bezpieczeństwa informatycznego.

Zbierając informacje ze wszystkich etapów analizy, możemy stworzyć wirtualny obraz serwera – określić, na jakim systemie pracuje, jakie usługi świadczy i za pomocą jakiego oprogramowania. Potrafimy także określić charakterystykę działalności serwera, potencjalnych właścicieli i administratorów. Mamy wystarczający zasób wiedzy, aby wskazać partnerów systemu poddawanego analizie, a prócz tego jesteśmy w stanie odtworzyć jego wygląd z rożnych etapów istnienia. Problemem nie jest również stworzenie profili psychologicznych osób związanych z maszyną w sposób pośredni, jak i bezpośredni. Mamy możliwość zbudowania podobnej maszyny wyglądającej z zewnątrz w sposób identyczny, jak nasz cel i prowadzenia na niej prób penetracyjnych. Zasób informacji, które jesteśmy w stanie zgromadzić w taki sposób, jest praktycznie nieograniczony – stosujemy metody nieinwazyjne, a dopiero mając świadomość wyglądu, poziomu bezpieczeństwa i zaawansowania konfiguracji przystępujemy do działań aktywnych – takie rozpoznanie daje nam dużą przewagę nad administratorem i całym personelem technicznym.

Podsumowanie

Nie bez przyczyny wzięło się powiedzenie *kto pyta, nie błądzi* – rodzajem informatycznego pytania jest umiejętność wyszukiwania informacji w Sieci. Przeszukując ją jesteśmy w stanie natrafić na dane personalne milionów ludzi i właśnie w tym miejscu tworzy się bardzo groźny problem. Nie jest ważne, czy skierujemy swoją uwagę w konkretnym kierunku – na jedną osobę, czy też spojrzymy na ogół społeczności informatycznej. Zarówno z jednej, jak i z drugiej perspektywy widzimy użytkowników będących w niebezpieczeństwie, które sami prowokują. Podążając dalej ścieżką poszukiwań przedstawiłem podstawowe sposoby zdobywania danych o serwerach. Metody, które zaprezentowałem, należą do elementarnych, lecz ich skuteczność jest na tyle wysoka, że w zupełności wystarcza do stworzenia wstępnego obrazu analizowanego środowiska. Łącząc oba działania w całość i stosując konkretnie przeciwko określonej jednostce – jesteśmy, w zależności od naszych dalszych intencji, w stanie znaleźć się w posiadaniu danych niebezpiecznych i ważnych dla firmy, otwierając sobie drogę do całej palety ataków: od social engineeringu zaczynając, kończąc na kradzieży. Znając stan Sieci i wagę problemu zaprezentowałem wiele metod – w sposób może mało dokładny, jednak działanie to było zamierzone. Bo przede wszystkim moim głównym celem było wywołanie w Czytelniku chęci samodzielnego zgłębienia szczegółów – co uczyni jedynie praktyką, oraz uwrażliwienie wszystkich użytkowników na niebezpieczeństwo kradzieży danych i tożsamości oraz zwrócenie szczególnej uwagi na rangę informacji personalnych, które w Sieci umieszczane są gdzie popadnie.

Pamiętajmy, że skutecznym atakiem jest atak wymierzony w niestrzeżone punkty (jawnie pozostawione wrażliwe dane) i przeprowadzony niespodziewanymi drogami. ●



Rysunek 8. *Nagłówek TCP*

W Sieci

- *http://www.paterva.com/web/Maltego/* – znakomity program do analizy relacji w Sieci,
- *http://www.googleguide.com/advanced_operators_reference.html* – Google w pigułce,
- *http://johnny.ihackstuff.com/ghdb.php* – baza zapytań Google stworzona w celach penetracyjnych,
- *http://tor.eff.org/* – oprogramowanie do zapewnienia anonimowości w Sieci,
- *http://www.securityfocus.com/infocus/1527* – podstawy social engineeringu,
- *http://www.securityfocus.com/infocus/1860* – social engineering: inne spojrzenie,
- *http://nessus.org* – strona domowa Nessusa.

O autorze

Autor od wielu lat interesuje się informatyką – swoje zainteresowania skupił głównie na zagadnieniach i problematyce sieci komputerowych oraz bezpieczeństwa teleinformatycznego. Jest samoukiem i pasjonatem. Studiuje informatykę na wydziale Cybernetyki Wojskowej Akademii Technicznej.
Kontakt z autorem: *bartosz.kalinowski@gmail.com*