Politechnika Warszawska

WYDZIAŁ ELEKTRYCZNY

Instytut Elektrotechniki Teoretycznej
i Systemów Informacyjno-Pomiarowych
Zakład Elektrotechniki Teoretycznej
i Informatyki Stosowanej

Praca dyplomowa inżynierska

na kierunku Informatyka stosowana
w specjalności Inżynieria oprogramowania

Aplikacja do zarządzania publikacjami naukowymi z automatyczną analizą PDF

Piotr Jeleniewicz
nr albumu 291072

promotor
dr inż. Bartosz Chaber

WARSZAWA 2021

APLIKACJA DO ZARZĄDZANIA PUBLIKACJAMI NAUKOWYMI Z AUTOMATYCZNĄ ANALIZĄ PDF

Streszczenie

Celem pracy jest przygotowanie systemu do zarządzania publikacjami wraz z automatyczną analizą plików PDF, której celem jest uzyskanie z tego pliku informacji dotyczących między innymi tytułu oraz autorów publikacji. System wykorzystuje architekturę klient-serwer, gdzie do odpowiedzialności serwera należy przetwarzanie i przechowywanie danych dotyczących publikacji, a rolą klienta będzie natomiast prezentacja informacji pochodzących z serwera oraz interakcja z użytkownikiem. Efektem prac są dwie aplikacje: kliencka, która została przygotowana dla systemu Android w języku Kotlin oraz serwerowa, która działa w oparciu o środowisko Node.js. W tej pracy zostały zawarte szczegóły dotyczące implementacji poszczególnych funkcji aplikacji zarówno po stronie serwera oraz klienta, a także testy demonstrujące poprawność pracy całego systemu.

Słowa kluczowe: zarządzanie publikacjami, aplikacja mobilna, usługa sieciowa, Node.js, Android, Kotlin

REFERENCE MANAGER WITH AUTOMATIC PDF ANALYSIS

Abstract

The goal of this thesis is to create a system to prepare publication management system with automatic analysis of PDF files, which purpose is to read from PDF file informations like title or authors of publication. The system uses a client-server architecture, where the server responsibilities are processing and storing publication's data and the client's role is to present information from the server and interact with the user. The result of the work are two applications, client, which was prepared for Android in Kotlin, and server one, which works in the Node.js environment. This thesis contains details of the implementation of individual system functions, as well as tests demonstrating that system works correctly.

Keywords: publication managment, moblie application, network service, Node.js, Android, Kotlin

WARSZAWA, 1 lutego 2021

POLITECHNIKA WARSZAWSKA
WYDZIAŁ ELEKTRYCZNY

OŚWIADCZENIE

Świadom odpowiedzialności prawnej oświadczam, że niniejsza praca dyplomowa inżynierska pt. Aplikacja do zarządzania publikacjami naukowymi z automatyczną analizą PDF:

- została napisana przeze mnie samodzielnie,
- nie narusza niczyich praw autorskich,
- nie zawiera treści uzyskanych w sposób niezgodny z obowiązującymi przepisami.

Oświadczam, że przedłożona do obrony praca dyplomowa nie była wcześniej podstawą postępowania związanego z uzyskaniem dyplomu lub tytułu zawodowego w uczelni wyższej. Jestem świadom, że praca zawiera również rezultaty stanowiące własności intelektualne Politechniki Warszawskiej, które nie mogą być udostępniane innym osobom i instytucjom bez zgody Władz Wydziału Elektrycznego.

Oświadczam ponadto, że niniejsza wersja pracy jest identyczna z załączoną wersją elektroniczną.

Piotr Jeleniewicz.....

Spis treści

Podziękowania

Piotr Jeleniewicz

Rozdział 1

Wstęp

Przy pisaniu artykułów naukowych często sięga się po efekty prac przedstawione w innych publikacjach naukowych. Wraz z pisaniem coraz większej liczby tego typu dokumentów, ilość wykorzystywanych w nich pozycji bibliograficznych może zacząć znacznie wzrastać co może utrudnić odnajdowanie potrzebnych publikacji w stale rozszerzającym się ich zbiorze.

Dlatego też ta praca będzie przedstawiała efekty prac nad aplikacją do zarządzania publikacjami naukowymi z automatyczną analizą plików PDF, której głównym celem jest ułatwienie procesu zarządzania publikacjami naukowymi, które przechowywane są w formie plików PDF. System będzie składał się aplikacji klienckiej przygotowanej dla systemu Android oraz aplikacji serwerowej działającej w kontenerze Dockera.

1.1 Założenia projektowe

System powstający w ramach pracy inżynierskiej będzie opierał się o następujące założenia:

1. Aplikacja serwerowa będzie napisana przy użyciu Node.js oraz języka TypeScript.
2. Baza danych wraz oraz aplikacją serwerową będą uruchamiane jako kontenery Dockera
3. Aplikacja kliencka będzie przeznaczona na system Android oraz do jej napisania wykorzystany zostanie język Kotlin.
4. W celu korzystania z aplikacji użytkownik będzie musiał utworzyć konto w systemie.

5. Do uzyskania informacji z pliku PDF dotyczących publikacji wykorzystane zostanie API dostępne pod adresem `https://api.crossref.org/`

W systemie będą dostępne następujące funkcjonalności:

1. Wyświetlanie listy publikacji.
2. Wyświetlanie opisu publikacji naukowych.
3. Tworzenie nowej publikacji w oparciu o metadane oraz numer DOI, jeśli są dostępne w dodawanym pliku PDF.
4. Edycja publikacji.
5. Pobranie pliku PDF powiązanego z daną publikacją.

Rozdział 2

Steganografia

Wywodzące się z greki słowo „steganografia" oznacza „ukryte pismo" (*steganos* - ukryty, tajny; *graphein* - pisać, malować), co w odniesieniu do kanału informacyjnego oznacza przesyłanie danych w taki sposób, aby osoby postronne mające wgląd do danych nie mogły stwierdzić istnienia w nich ukrytej informacji. Cały mechanizm steganografii opiera się na zasadzie ukrycia informacji w tych częściach wiadomości, które nie służą do przekazywania informacji lub których modyfikacja nie wpływa na treść głównego przekazu.

W celu przesłania informacji za pomocą steganografii należy utworzyć kanał steganograficzny, zdefiniowany [?] jako: „każdy kanał komunikacyjny, który może być wykorzystany przez stronę do przesłania informacji w sposób naruszający politykę bezpieczeństwa systemu". Metoda ta wykorzystuje fakt przesłania danych w sposób i w miejscach, które zgodnie z protokołem do tego nie służą, narażając system na nieautoryzowany przesył informacji.

Steganografia w znaczącym stopniu różni się od kryptografii, która nie dba o zatajenie istnienia przekazu, a jedynie o jego integralność oraz uniemożliwienie stronom trzecim poznanie treści przekazu. Oczywiście najlepszą techniką jest połączenie steganografii z kryptografią. Takie podejście pozwala zabezpieczyć się przed sytuacją, w której strona nadzorująca transmisję, nawet w przypadku odkrycia przekazu steganograficznego nie może go odczytać ze względu na siłę zastosowanej kryptografii.

2.1 Historia

Pomimo, że pierwsze wzmianki o steganografii, a dokładnie o ukrytych kanałach w odniesieniu do systemów informatycznych notuje się na lata siedemdziesiąte XX wieku [?], to przykłady użycia steganografii sięgają starożytności. W literaturze powtarzają się opisy przekazywania tajnej informacji

poprzez wytatuowanie jej na ogolonej głowie posłańca, który po odrośnięciu włosów był wysyłany z mało znaczącą wiadomością do armii swojego dowódcy. Każdy kto natknął się na posłańca miał wgląd do nieważnej wiadomości, niepodejrzewając nawet istnienia sekretnej informacji w postaci tatuażu.

Przykłady z historii odnoszą się także do bardziej współczesnych czasów. Wiele z metod steganografii było stosowanych podczas II Wojny Światowej (np. mikro-kropki) a także w latach Zimnej Wojny. Wiadomo także, że wielu agentów służb wywiadowczych, a szczególnie podwójnych agentów, przekazywało obcym państwom informację wykorzystując steganografię. Przykładem może tu być sprawa szpiega FBI Roberta Hanssena [**?**], który przy pomocy technik steganograficzny przez około dekadę przekazywał tajne informacje służbom KGB.

Rozdziały 2.7 oraz 3 opisują nowoczesne podejście do steganografii wykorzystujące współczesne kanały informacyjne.

2.2 Pojęcia

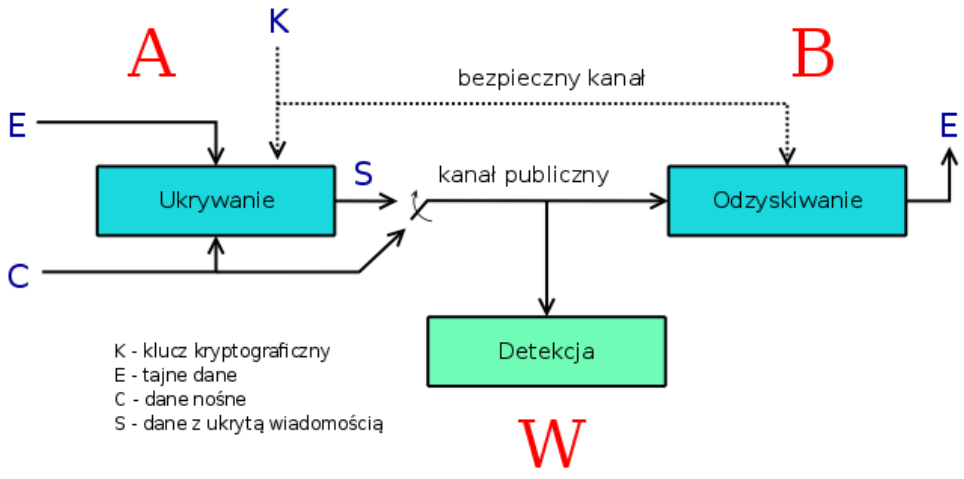
W celu zdefiniowania kanału steganograficznego oraz opisania transmisji z wykorzystaniem takiego kanału należy omówić jego części składowe:

- dane do ukrycia, tajne dane - informacja jaką należy przesłać między uczestnikami komunikacji, tak aby strony trzecie nie miały do niej wglądu,
- dane nośne, wiadomość zakrywająca - wiadomość, w której ukryte zostaną tajne dane; przesyłanie wiadomości zakrywających musi być dozwolone w danym kanale informacyjnym i nie powinno wzbudzać podejrzeń,
- funkcja steganograficzna - funkcja przekształcająca dane do ukrycia oraz wiadomość zakrywającą w jedną połączoną wiadomość,
- dane z ukrytą wiadomością - dane zawierające ukrytą informację a jednocześnie wykazujące cechy danych nośnych,
- nadzorca komunikacji, wartownik - mechanizm mający pełen wgląd do wiadomości przekazywanej między stronami komunikacji, świadomy struktury komunikatów i potrafiący wykrywać występujące w nich anomalie,
- kanał komunikacyjny - kanał zestawiony pomiędzy nadawcą a odbiorcą, zapewniający przepływ informacji, do którego wgląd ma nadzorca komunikacji,

- odwrotna funkcja steganograficzna - funkcja przekształcająca dane z ukrytą wiadomością na tajne dane,
- klucz kryptograficzny - klucz znany tylko obu stronom komunikacji, służący do zabezpieczenia tajnej informacji metodami kryptografii symetrycznej przed ewentualnością złamania funkcji steganograficznej.

2.3 Schemat komunikacji steganograficznej

Podstawowy scenariusz, powszechny w literaturze na temat steganografii, odnosi się do sytuacji opisanej w [?]. Dwóch więźniów (w naszym przypadku Alicja(A) i Bob(B)) zamknięci są w dwóch odrębnych celach. Mogą się ze sobą kontaktować, jednak ich cała korespondencja przechodzi przez ręce Wartownika (W). Ma on pełen wgląd do przekazywanych informacji, więc może przechwycić wszelkie przekazywane tajemnice, a dodatkowo w razie podejrzeń może nie dopuścić do komunikacji[1]. W takim przypadku w celu przekazania ważnych informacji A i B muszą posłużyć się pewnego rodzaju podstępem. Muszą tak sformułować treść przekazu, aby W nie rozróżnił „niegroźnej" wiadomości od wiadomości z ukrytym przekazem. Dlatego też przekazują wiadomość, w której prawdziwa treść możliwa jest do odczytania po złożeniu kolejno każdej np. drugiej litery z każdego wyrazu.



Rysunek 2.1: Schemat komunikacji steganograficznej

[1] podejrzana informacja jest tu analogią do stosowania kryptografii przez więźniów

Przedstawioną tak sytuację pokazuje rysunek 2.1[2]. A próbuje przesłać tajną informację E do B. Cała komunikacja odbywa się przez kanał publiczny, kontrolowany przez W. W celu ukrycia faktu komunikacji A stara się ukryć tajny przekaz w informacji C. W celu uzyskania skutecznej steganografii W nie może rozróżnić informacji poprawnej, nie zawierającej tajnych danych, od informacji S, która zawiera tajną informację. W celu dodatkowego zabezpieczenia przekazu, A i B mogą korzystać z funkcji kryptograficznej zabezpieczającej przekazywane informacje. Można tu wykorzystać metody kryptografii symetrycznej (ustalony klucz kryptograficzny K) lub niesymetrycznej (klucz publiczny K_{pub} i klucz prywatny K_{pryw}).

Stosowanie technik kryptograficznych wpływa na poprawę bezpieczeństwa przesyłanej informacji, jednak należy pamiętać o nieporządnych cechach jakie mogą one wywołać. W większości przypadków umieszczenie tajnej informacji steganograficznej w przekazie wiąże się z zamianą istniejącej już nieważnej części informacji. Jednak każda porcja usuniętej informacji może mieć pewną charakterystyczną postać lub specyficzny histogram. Zastosowanie funkcji kryptograficznej w stosunku do tajnej informacji zmienia ją, a wynikowy rozkład bitów jest nieprzewidywalny i w większości przypadków różny od standardowych histogramów określonych dla podmienianych części wiadomości.

2.4 Stegoanaliza

Stegoanaliza to nauka zajmująca się wykrywaniem istnienia ukrytych informacji w kanałach komunikacyjnych. Nie zawsze prowadzi to do odkrycia dokładnej treści ukrytego przekazu, a w większości przypadków polega jedynie na wskazaniu istnienia ukrytego kanału steganograficznego.

Możliwość wykrycia kanału steganograficznego sprowadza się do analizy różnych części wiadomości lub strumienia danych w celu wykrycia anomalii. Takie podejście wynika z faktu, że tajna informacja ukryta jest w miejscach nie przeznaczonych do przesyłania informacji lub na miejscu danych, które są w pewien sposób nadmiarowe (np. dla zmysłów człowieka). Można wskazać dwa podstawowe sposoby wykrywania anomalii:

- pierwsze podejście opiera się na przebadaniu wszystkich części informacji (np. pól nagłówka TCP/IP), których struktura jest w pełni przewidywalna lub których wartości są zdefiniowane przez standardy lub powszechne praktyki; ważne jest także sprawdzenie czy występują war-

[2]sporządzony na podstawie [?], rysunek 1, strona 3

tości nadmiarowe oraz czy elementy sygnalizujące wystąpienie dodatkowych danych mają faktyczne pokrycie w danych,

- drugą metodą jest porównanie wartości części wiadomości (np. pól nagłówka TCP/IP) i zaklasyfikowanie ich jako prawdopodobnych lub nie dla danego systemu bądź protokołu; takie podejście może być stosowane do wartości ściśle określonych, takich jak wymienione w pierwszym punkcie, jednak można je także stosować do wartości które są pseudolosowe lub których histogram jest charakterystyczny; w celu realizacji tej metody warto posłużyć się sieciami neuronowymi takimi jak SVM i RSVM, zdolnymi rozpoznawać wzorce i separować dane.

2.5 Metody tworzenia steganografii oraz rodzaje ukrytych kanałów

Przesłanie danych za pomocą przekazu steganograficznego wiąże się w większości przypadków z umieszczeniem dodatkowej informacji w wiadomości. Odbywa się to za pomocą podmiany tej części wiadomości (nagłówka TCP/IP), która wykazuje cechy nadmiarowości lub której (kontrolowana) zmiana nie prowadzi do przerwania transmisji. Pewną podgrupą może być w tym przypadku wykorzystanie pól oryginalnie pustych (zerowych) lub niewykorzystywanych w istniejących implementacjach.

Kanały steganograficzne można podzielić na dwa zasadniczne typy[**?**]:

- kanał pojemnościowy (ang. storage channel) - informacja zawarta w częściach wiadomości, polach nagłówka,
- kanał czasowy (ang. timing channel) - informacja zawarta w czasach wystąpienia danych zdarzeń, np. przesłania pakietu TCP/IP.

W przypadku sieci pakietowych można także połączyć dwa typu kanałów steganograficznych, tworząc kanał mieszany, w którym jeden z typów (np. pojemnościowy) będzie wykorzystywany do przekazywania informacji, a drugi (np. czasowy) do sygnalizacji tego zdarzenia.

Większość opracowanych programów służących do przesyłania danych z wykorzystaniem steganografii opiera się na kanałach pojemnościowych. Wynika to z faktu, że kanały czasowe narzucają pewne ograniczenia na generację pakietów TCP/IP przez co ich wykrycie staje się prostsze.

Dodatkowo należy zauważyć, że w sieciach pakietowych można skonstruować abstrakcyjny kanał steganograficzny, w którym do przesyłania tajnych danych lub/i obsługi protokołu steganograficznego wykorzystywane są różne pola nagłówka. Zmiana wykorzystania danego pola może być dynamiczna,

zależna od wymaganej przepustowości lub w celu zminimalizowania wykrycia kanału steganograficznego.

2.6 Cechy kanału steganograficznego

Każdy kanał steganograficzny posiada trzy cechy, które decydują o jego przydatności w danej sytuacji:

1. pojemność (przepustowość) - określa jaką porcję informacji możemy przesłać w danej wiadomości nośnej; w przypadku steganografii w TCP/IP, wyrażana jest w bitach na sekundę, bitach na pakiet lub bitach na sesję TCP; przepustowość odgrywa ważną rolę w przypadku konieczności przekazania dużej ilości informacji, jednak należy pamiętać, że to przeważnie prowadzi do ułatwionej detekcji steganografii,
2. bezpieczeństwo - określa jak łatwo jest uzyskać dostęp do przekazywanej tajnej informacji w przypadku poznania mechanizmu tworzenia przekazu steganograficznego; dodatkowym mechanizmem zwiększającym bezpieczeństwo może być używanie znanych tylko sobie zmiennych pseudolosowych lub modyfikacji algorytmu[3],
3. krzepkość (ang. robustness) - określa stopień w jakim możemy zmodyfikować przekaz nie uszkadzając zawartej w nim informacji steganograficznej; niestety w przypadku steganografii naruszenie kanału (pola) zawierającego przekaz steganograficzny przeważnie wiąże się z utratą tajnego przekazu.

2.7 Steganografia w obiektach multimedialnych

Pomimo, że steganografia ma zastosowanie prawie w każdej formie komunikacji, w latach 90-tych zyskała ona powodzenie jako technika ukrywania informacji w obiektach multimedialnych. Wynika to przede wszystkim z powszechności tego rodzaju przekazu, jego rozmiarów oraz prostoty obsługi programów do ukrywania informacji w obiektach multimedialnych, takich jak obraz, dźwięk i wideo. Dodatkowym atutem przy zastosowaniu tych metod jest stosunek ukrytej informacji do oryginalnego przekazu, sięgający w ekstremalnych sytuacjach 50%, bez zauważalnego pogorszenia się jakości przekazywanych danych.

[3]jest to znane jako „bezpieczeństwo przez zatajenie" (ang. security through obscurity) i powinno być używane tylko jako dodatkowy element systemu zabezpieczeń

Użycie steganografii w treściach multimedialnych sprowadza się do takiego manipulowania danymi, aby plik wynikowy zawierał dodatkowe informacje, a jednocześnie nie był rozróżniany przez zmysły człowieka w porównaniu z oryginałem.

Jedną z najszerzej omawianych form steganografii w obiektach multimedialnych jest ukrywanie informacji w plikach graficznych. Istnieje wiele rozwiązań, zarówno bezpłatnych, o otwartym kodzie jak i komercyjnych. Przykładami mogą tu być takie programy jak Outguess, JPHide, StegHide. Istnieją różne techniki ukrywania informacji w plikach graficznych. Najprostszym rozwiązaniem jest podmiana najmniej znaczących bitów opisujących kolor danego piksela. Możliwe jest też zastosowanie dyskretnej transformaty kosinusowej.

W przypadku wybrania jako wiadomości nośnej pliku audio, możemy także zastosować metodę podmiany najmniej znaczących bitów. Dodatkowo stosowane są metody ukrywania tajnych wiadomości poprzez rozszerzanie spektrum danego nagrania, czy też dodawanie echa. Przykładem narzędzia do tworzenia wiadomości steganograficznych może być UnderMP3Cover, MP3Stego czy S-Tools[4].

Kolejnym przykładem wykorzystania jako pliku nośnego obiektu multimedialnego jest plik wideo. Dodatkowa informacja może być przekazana przy użyciu dyskretnej transformaty kosinusowej. Jako przykładowe implementacje można podać StegoVideo.

Istnieje kilka technik umożliwiających wykrycie lub usunięcie steganografii zastosowanej w obiektach multimedialnych. Pierwszym podejściem, choć przeważnie trudnym do zastosowania, jest użycie oryginalnego pliku jako wzorca do porównania z przechwyconą wersją. W przypadku plików graficznych lub wideo możliwe jest użycie analizatorów statystycznych, które mogą wykryć anomalie występujące w histogramach tych wiadomości.

Zamiast wykrywać istnienie steganografii, częstym podejściem jest jej ograniczanie lub „ślepe" usuwanie z wiadomości tych danych, które mogą być nośnikiem kanału steganograficznego. W przypadku plików multimedialnych najlepszym sposobem uzyskania takiego efektu jest przekodowanie pliku na inny standard i powrót do standardu wejściowego. Przeważnie zmiany w jakości plików są niezauważalne, a użycie konwersji sprowadza się do takiej zmiany bitów, która niszczy zawartą w nich steganografię.

W przypadku plików multimedialnych użycie steganografii jest pomocne w ochronie praw autorskich, przez stosowanie jej jako cyfrowych znaków wodnych. Niestety, tak jak zostało to wcześniej przedstawione w trakcie konwersji wiele z zakodowanej informacji ginie bezpowrotnie. Skutkiem tego może być

[4] `http://www.stegoarchive.com`

pogorszenie jakości pliku multimedialnego, ale także usunięcie z niego cyfrowego znaku wodnego.

Itd., itd., itd. ...

Rozdział 3

Steganografia w ruchu TCPIP

Itd., itd., itd ...

Rozdział 4

Wnioski

Protokół TCP/IP jest najbardziej rozpowszechnionym i używanym protokołem komunikacji między systemami w sieci Internet oraz w sieciach intranet. Niestety został on opracowany na początku lat siedemdziesiątych, gdy problemy bezpieczeństwa informacji nie stały na pierwszym miejscu. Ciągły wzrost działań przestępczych w sieci Internet, w tym wymiana nielegalnych treści, prowadzi do stosowania coraz to nowszych technik zabezpieczających. Z tego względu obserwuje się działania mające na celu wprowadzenie tajnej komunikacji między przejętymi systemami, tak aby nie wzbudzić ostrzeżeń w analizatorach sieciowych. Taka ukryta komunikacja odbywa się z wykorzystaniem steganografii.

Wprowadzenie steganografii do niskich warstwach stosu TCP/IP umożliwia obejście wielu filtrów nałożonych na warstwy wyższe. Większość sieci oparta jest na protokołach rodziny TCP/IP, przez co nie można zabronić ich używania. Możliwa jest jedynie kontrola poprawności semantyki protokołów TCP/IP, a także ewentualna ingerencja w przekazywane wartości, z uwzględnieniem stanowości niektórych pól.

Opracowany schemat generacji początkowych numerów sekwencyjnych w jak najlepszy sposób odzwierciedla oryginalny proces zachodzący w stosie sieciowym systemu Linux. W większości przypadków występujących w rzeczywistych sieciach i systemach, numery wygenerowane przy pomocy `Shushi` nie byłyby rozróżnialne od numerów wygenerowanych przez stos sieciowy systemu.

Jeżeli proces generacji wartości użytych do przekazania danych steganograficznych zostanie oparty o oryginalne mechanizmy używane do ich generacji, to pasywny analizator sieciowy nie będzie w stanie wykryć istnienia anomalii. Różnice możliwe są do zaobserwowania w przypadku zaistnienia specyficznych sytuacji występujących dla danej implementacji protokołu. W przypadku zastosowania pasywnego analizatora wymaga to jednak oczekiwa-

nia na taką sytuację. Z przeprowadzonych testów wynika, że lepszym podejściem jest zastosowanie analizatorów aktywnych, które posiadają wiedzę na temat testowanych systemów oraz ich chrakterystycznych cech implementacji. Skonstruowanie takiego analizatora jest zadaniem stosunkowo prostym a daje bardzo wysoką skuteczność.

Z przeprowadzonych testów wynika, że celowe jest prowadzenie dalszych prac w następujących obszarach:

- dokładniejszy mechanizm generacji wartości mikrosekund
- wprowadzenie algorytmów zdolnych wykryć i uniemożliwić działanie analizatora aktywnego

Jeżeli powyższe punkty nie zostaną spełnione, analizatory aktywne będą w stanie wykryć istnienie modułu steganograficznego opartego na początkowych numerach sekwencyjnych.

pierwsza kolumna	druga	trzecia
1	2	3
a	b	c

$$E = mc^2 \tag{4.1}$$

Rozwój opracowanego rozwiązania steganograficznego jest możliwy poprzez wprowadzenie elementów – patrz wzór (4.1) – jak:

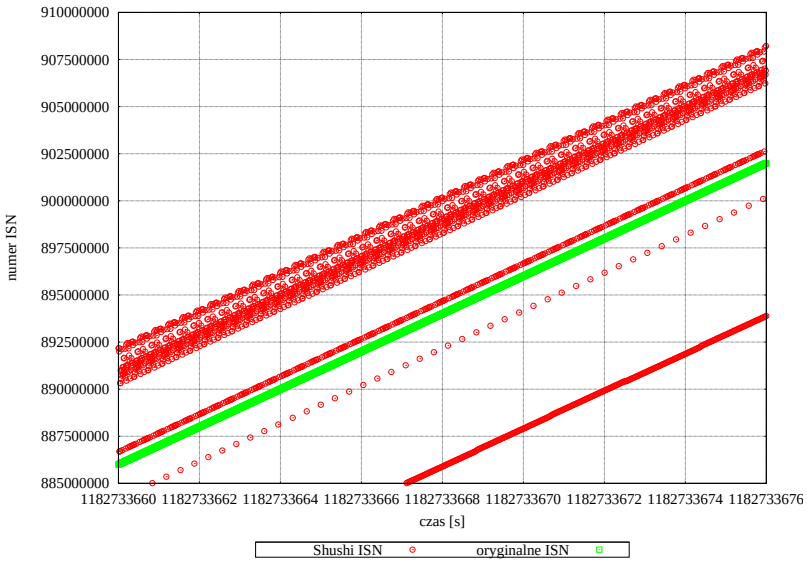
- obsługa innych, przyszłościowych protokołów sieciowych, takich jak SCTP (ang. Stream Control Transmission Protocol)[?]
- zapewnienie dwustronnej komunikacji z wykorzystaniem numerów potwierdzenia `ACK`
- przeniesienie implementacji do innych systemów operacyjnych

Wraz ze wzrostem przepustowości urządzeń sieciowych (obecnie 10Gb/s i więcej) wzrasta problem analizy przepływających danych w czasie rzeczywistym. Analizatory sieciowe muszą w coraz krótszym czasie zbadać coraz większy strumień danych (miliony pakietów na sekundę). Jednak problem wzrostu prędkości sieci utrudnia zadanie także osobom implementującym kanały steganograficzne w protokole TCP/IP. Coraz więcej operacji wyższych warstw stosu sieciowego przenoszonych jest do układów scalonych interfejsów sieciowych. Taka technologia znana jest pod skrótem TOE (ang. TCP Offload Engine) i odnosi się przede wszystkim do sprzętowej generacji sum kontrolnych oraz mechanizmu TSO (ang. TCP segmentation offload). W następnych latach spodziewane jest przenoszenie kolejnych elementów stosu sieciowego TCP/IP do implementacji sprzętowych.

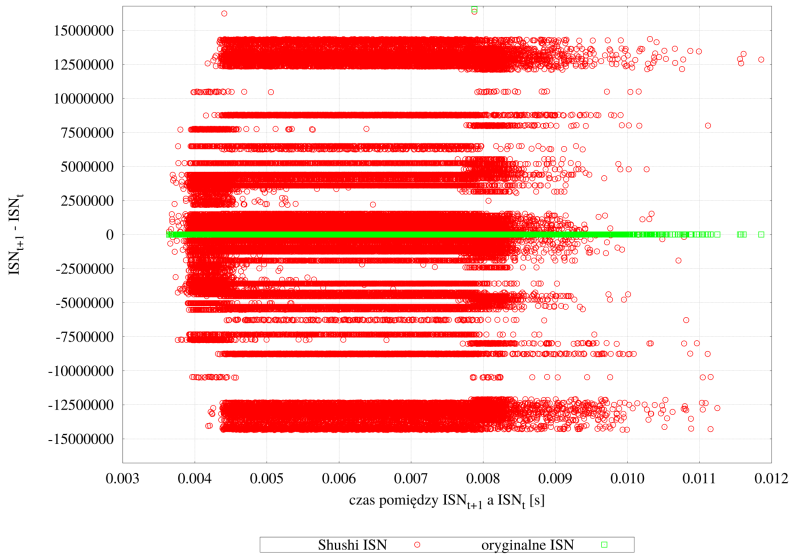
Ze względu na rozwój systemów zabezpieczających ruch sieciowy oraz wzrost bezpieczeństwa systemów operacyjnych, w kolejnych latach wzrośnie także wykorzystanie technik steganograficznych przez grupy przestępcze działające w ramach Internetu. Z tego powodu poznanie technik steganograficznych oraz wypracowanie metod obrony i wykrywania takiej komunikacji jest bardzo ważne.

Dodatek A

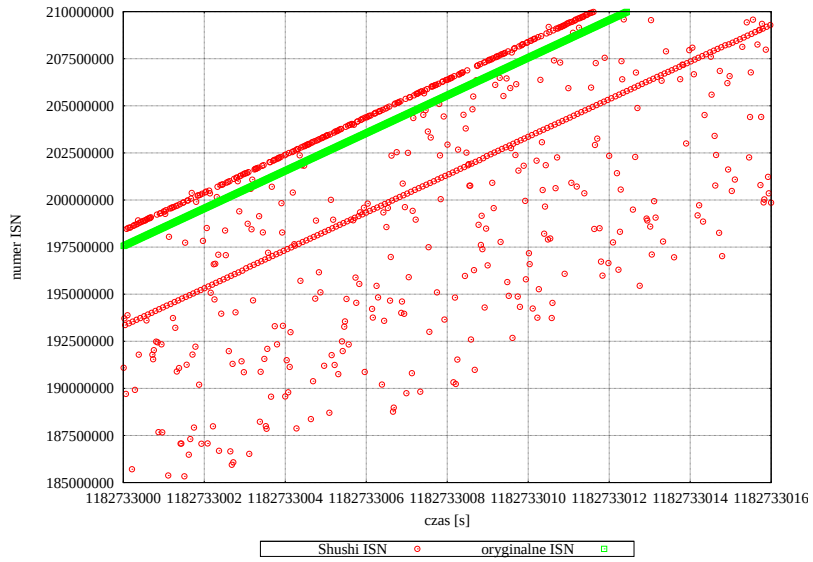
Porównanie numerów ISN jądra Linux i modułu Shushi



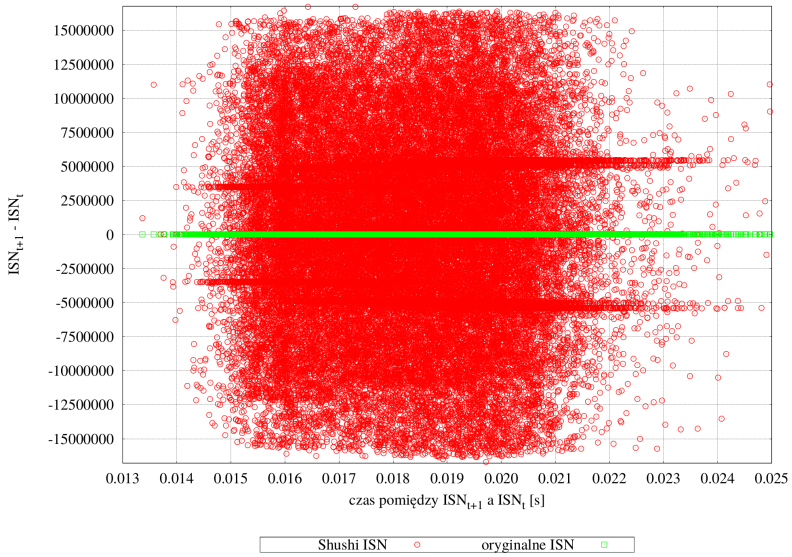
Rysunek A.1: Numery ISN wygenerowane przez jądro oraz `Shushi`, stałe numery IP oraz porty TCP, stałe dane dla `Shushi`, serie po około 2800 próbek.



Rysunek A.2: Różnice pomiędzy kolejnymi numerami ISN wygenerowanymi przez jądro oraz `Shushi`, stałe numery IP oraz porty TCP, stałe dane dla `Shushi`, serie po około 60000 próbek.



Rysunek A.3: Numery ISN wygenerowane przez jądro oraz `Shushi`, stałe numery IP oraz porty TCP, losowe dane dla `Shushi`, serie po około 860 próbek.



Rysunek A.4: Różnice pomiędzy kolejnymi numerami ISN wygenerowanymi przez jądro oraz `Shushi`, stałe numery IP oraz porty TCP, losowe dane dla `Shushi`, serie po około 60000 próbek.

Bibliografia

[1] daemon9, „LOKI2”, Phrack Magazine, Issue 51. http://phrack.org

[2] van Hauser, Reverse WWW Shell, THC, The Hacker's Choice. www.thc.org

Opinia

o pracy dyplomowej magisterskiej wykonanej przez dyplomanta
Zdolnego Studenta i Pracowitego Kolegę
Wydział Elektryczny, kierunek Informatyka, Politechnika Warszawska
Temat pracy
TYTUŁ PRACY DYPLOMOWEJ

Promotor: **dr inż. Miły Opiekun**
Ocena pracy dyplomowej: **bardzo dobry**

Treść opinii

Celem pracy dyplomowej panów dolnego Studenta i Pracowitego Kolegi było opracowanie systemu pozwalającego symulować i opartego o oprogramowanie o otwartych źródłach (ang. Open Source). Jak piszą Dyplomanci, starali się opracować system, który łatwo będzie dostosować do zmieniających się dynamicznie wymagań, będzie miał niewielkie wymagania sprzętowe i umożliwiał dalszą łatwą rozbudowę oraz dostosowanie go do potrzeb. Przedstawiona do recenzji praca składa się z krótkiego wstępu jasno i wyczerpująco opisującego oraz uzasadniającego cel pracy, trzech rozdziałów (2-4) zawierających opis istniejących podobnych rozwiązań, komponentów rozpatrywanychjako kandydaci do tworzonego systemu i wreszcie zagadnień wydajności wirtualnych rozwiązań. Piąty rozdział to opis przygotowanego przez Dyplomantów środowiska obejmujący opis konfiguracji środowiska oraz przykładowe ćwiczenia laboratoryjne. Ostatni rozdział pracy to opis możliwości dalszego rozwoju projektu. W ramach przygotowania pracy Dyplomanci zebrali i przedstawili w bardzo przejrzysty sposób duży zasób informacji, co świadczy o dobrej orientacji w nowoczesnej i ciągle intensywnie rozwijanej tematyce stanowiącej zakres pracy i o umiejętności przejrzystego przedstawienia tych wyników. Praca zawiera dwa dodatki, z których pierwszy obejmuje wyniki eksperymentów i badań nad wydajnością, a drugi to źródła skryptów budujących środowisko.

Dyplomanci dość dobrze zrealizowali postawione przed nimi zadanie, wykazali się więc umiejętnością zastosowania w praktyce wiedzy przedstawionej w rozdziałach 2-4. Uważam, że cele postawione w założeniach pracy zostały pomyślnie zrealizowane. Proponuję ocenę bardzo dobrą (5).

(data, podpis)

Recenzja

pracy dyplomowej magisterskiej wykonanej przez dyplomanta
Zdolnego Studenta i Pracowitego Kolegę
Wydział Elektryczny, kierunek Informatyka, Politechnika Warszawska
Temat pracy
TYTUŁ PRACY DYPLOMOWEJ

Recenzent: **prof. nzw. dr hab. inż. Jan Surowy**
Ocena pracy dyplomowej: **bardzo dobry**

Treść recenzji

(data, podpis)