WYDZIAŁ PODSTAWOWYCH PROBLEMÓW TECHNIKI
POLITECHNIKA WROCŁAWSKA

WIZUALIZACJA POŁĄCZEŃ SIECI INTERNETOWEJ

MARCIN ADAMCZYK
NR INDEKSU: 221 429

Praca inżynierska napisana
pod kierunkiem
Dr. Przemysława Kobylańskiego

Politechnika
Wrocławska

WROCŁAW 2017

Spis treści

Wstęp

Internet. W łatwy i prosty sposób daje dowolnemu użytkownikowi dostęp do usług, czy serwisów z całego świata. W dzisiejszych czasach z Internetu korzysta prawie każdy. Jednak stosunkowo niewielu użytkowników zdaje sobie sprawę, jak właściwie działa Internet. W jaki sposób są w stanie w dowolnym momencie nawiązać kontakt z osobą znajdującą się po drugiej stronie globu i jak wiele urządzeń zaangażowanych jest w przekazanie pojedynczego pakietu do adresata. Do takich właśnie osób skierowane jest oprogramowanie tworzone w ramach tej pracy.

Swoim zakresem praca obejmuje szereg zagadnień powiązanych z Internetem oraz sieciami komputerowymi. Spośród nich szczególnie ważnym jest *routing*, czyli wyznaczanie trasy, za pomocą której pakiet ma dotrzeć do celu. Do analizy sieci wykorzystywane są odpowiednie narzędzia takie jak *Wireshark* czy *traceroute*. Dodatkowo system wykorzystuje usługę geolokalizacji na podstawie adresu IP.

Celem pracy jest stworzenie oprogramowania, które poprzez nasłuchiwanie ruchu sieciowego oraz jego analizę, naniesie na mapę świata punkty, w których znajdują się urządzenia biorące udział w transmisji pakietów użytkownika. W efekcie powstanie wizualizacja, mająca pomóc wspomnianym w pierwszym akapicie osobom w zrozumieniu pewnych podstawowych faktów dotyczących działania sieci, jaką jest Internet. Co najważniejsze, system będzie działał w czasie rzeczywistym. Dzięki temu użytkownik będzie mógł na bieżąco obserwować, jak jego działania wpływają na urządzenia rozlokowane na świecie.

Obecnie nie ma oprogramowania udostępniającego podobne możliwości. Istnieje pewien serwis internetowy pokazujący na mapie trasę pojedynczego pakietu – jednak nie ma to absolutnie zastosowania w przypadku chęci analizowania całego ruchu. Samo odpowiednie połączenie programów *Wireshark* i *traceroute* może dać pewne efekty, jednak w postaci wydruku na konsoli, zupełnie nieczytelnego dla przeciętnego użytkownika Internetu.

Praca zawiera:

- omówienie zagadnień związanych z Internetem oraz sieciami komputerowymi wykorzystanych podczas tworzenia oprogramowania,
- informacje na temat dodatkowych narzędzi wykorzystywanych do działania systemu,
- informacje na temat stworzonego oprogramowania: dokumentację, zasadę działania, schematy UML,
- wyniki działania, instrukcję instalacji i uruchamiania, listę dodatkowego oprogramowania niezbędnego do prawidłowego działania programu, jak również instrukcje instalacji tegoż oprogramowania.

Analiza problemu

W niniejszym rozdziale przedstawiono środowisko pracy systemu oraz omówiono podstawowe pojęcia z tym środowiskiem powiązane. Następnie wyjaśniono procesy zachodzące w programie oraz powiązania pomiędzy tymi procesami. Dalsza część zawiera opis założeń funkcjonalnych jak i niefunkcjonalnych przedstawianego systemu. Na koniec przeprowadzono analizę oprogramowania trzeciego, używanego przez system, jak również porównanie do innych programów dostępnych dla użytkownika, realizujących podobne zadania.

2.1 Opis środowiska

Na początek wspomniany we wstępie Internet. To właśnie on stanowi swego rodzaju środowisko pracy systemu. Na potrzeby opisu funkcjonalności oprogramowania, wystarczy przytoczyć jedynie kilka podstawowych faktów na temat struktury i działania Internetu. Przede wszystkim jest to sieć złożona z wielu mniejszych sieci. Te z kolei składają się z dalszych, jeszcze mniejszych sieci. Idąc dalej w ten sposób można dojść do pojedynczego urządzenia znajdującego się w pewnej lokalnej sieci, należącej do potencjalnego użytkownika Internetu. Aby zapewnić komunikację pomiędzy wszystkimi urządzeniami, potrzeba zarówno tysięcy urządzeń sterujących przepływem danych, kilometrów kabli i przewodów, jak również niezwykle zaawansowanych algorytmów. Algorytmów zajmujących się między innymi wyliczaniem *routingu*, czyli opracowywaniem trasy jaką pakiet wysłany przez nadawcę ma dotrzeć do adresata. Wizualizacją właśnie tego zjawiska zajmować ma się system opracowany w ramach tej pracy. Ponieważ wyświetlenie listy adresów IP urządzeń „odwiedzonych” przez dany pakiet niewiele powie przeciętnemu użytkownikowi Internetu (a właśnie do takich osób kierowany jest stworzony system), w projekcie wykorzystywana jest usługa geolokalizacji. Pozwala ona określić współrzędne geograficzne obiektu na podstawie jego adresu IP. Sama usługa geolokalizacji do działania używa baz danych zawierających adresy IP oraz lokalizacje do których te adresy są przypisane.

2.2 Opis procesów

System składa się z trzech głównych procesów: **wczytywanie**, **analiza**, **wyświetlanie**. Ogólny sposób działania procesów został opisany poniżej.

A. Wczytywanie danych

   Bazą tego procesu jest program *Wireshark* - popularny sniffer, czyli program nasłuchujący ruch sieciowy. Dla wygody, wykorzystywana jest wbudowana w ów sniffer opcja filtrowania danych, dzięki której na standardowym wyjściu pojawiają jedynie interesujące system rekordy. W tym przypadku są to adresy IP urządzeń do których użytkownik wysyła pakiety. Dane ze sniffera przekazywane są przy pomocy unixowego pipe’a na wejście głównego programu. Ten po sprawdzeniu poprawności danych przekazuje je do następnego procesu.

B. Analiza danych

   Proces ten (jak sama jego nazwa wskazuje) odpowiada za przeprowadzenie analizy otrzymanych danych. Jest to najważniejszy element całego systemu, bowiem to właśnie na nim spoczywa cała logika. Podstawą jego działania jest drzewiasta struktura, przechowująca informacje na temat odebranych z wejścia danych. Konkretnie jest to drzewo czerwono – czarne, którego kluczem jest łańcuch

zawierający adres IP, a wartością - wskaźnik na obiekt przechowujący statystyki dotyczące odpowiadającego połączenia. Statystyki zawierają dane na temat łącznej ilości pakietów wysłanych w ramach konkretnego połączenia oraz ilości pakietów wysłanych w określonym odcinku czasu. Dzięki przechowywaniu statystyk jako obiektu osobnej klasy, uzyskujemy izolację logiki związanej z przetwarzaniem danych statystycznych, a co za tym idzie, łatwą rozszerzalność typów informacji jakie chcemy gromadzić i przetwarzać.

Kolejnym zadaniem analizatora jest określanie trasy jaką przebywa pakiet w drodze do adresata. W tym celu system ponownie korzysta z oprogramowania trzeciego. Tym razem jest to program *traceroute*. Oczywiście nie ma pewności, że trasa wskazana przez wspomniany program, będzie w pełni zgodna z rzeczywistą trasą jaką przebył pakiet przechwycony przez sniffer. Jednak warto pamiętać, że system ma pełnić rolę jedynie edukacyjną, więc tego typu niedokładności w niczym nie przeszkadzają.

Po określeniu trasy, potrzebne jest już tylko ustalenie lokalizacji (współrzędnych geograficznych) poszczególnych punktów odwiedzonych przez pakiet użytkownika. W tym celu system wykorzystuje API serwisu geolokalizacyjnego. W zapytaniu wysłanym przy pomocy programu *curl* zawarty jest adres IP urządzenia, którego lokalizacja jest potrzebna. Jako odpowiedź system otrzymuje potrzebne mu współrzędne w postaci długości i szerokości geograficznej.

Na koniec wszelkie potrzebne dane zostają opakowane w odpowiednią strukturę i przekazane do następnego procesu.

C. Wyświetlanie obrazu wynikowego

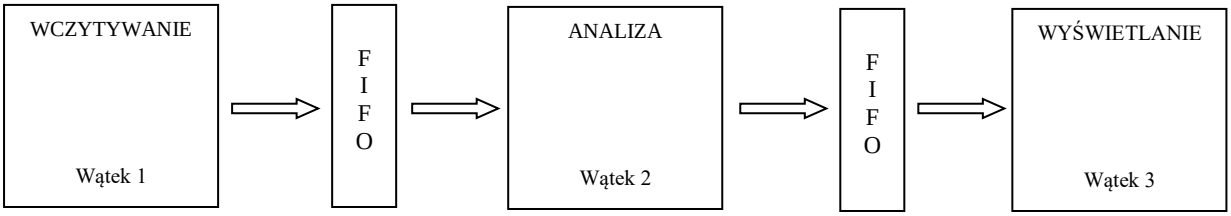
Zadaniem tej części systemu jest przetworzenie danych odebranych z wejścia celem wyświetlenia ich w odpowiedniej formie na ekranie. Co ważne, ten proces nie przeprowadza żadnej analizy – zakłada, że to co odebrał, jest informacją prawidłową i gotową do wyświetlenia. Dokonywana jest tu przede wszystkim konwersja współrzędnych geograficznych na położenie na ekranie (piksele). Dodatkowo na podstawie opcjonalnych danych przekazanych przez analizator, mogą zostać nałożone dodatkowe efekty wizualne (przykładowo: pogrubienie lub zmiana koloru wyświetlanej linii w przypadku połączenia o wysokim natężeniu ruchu).

Po ustaleniu wszystkich potrzebnych danych, system przystępuje do wizualizacji, czyli nanoszenia na wcześniej przygotowaną mapę świata punktów oraz krzywych je łączących.

2.3 Założenia systemu

Podstawowym założeniem dotyczącym funkcjonalności systemu, jest jego działanie w czasie rzeczywistym. Ma na celu to przede wszystkim zapewnienie użytkownikowi wrażenia wyższej responsywności używanego oprogramowania. Gdyby system najpierw zbierał wszystkie dane, a dopiero potem je przetwarzał, użytkownik nie wiedziałby, która z jego akcji wywołała taki, a nie inny efekt. Dzięki zbieraniu danych na bieżąco, efekt interakcji użytkownika z siecią widoczny jest w ciągu ułamków sekund. Oczywiście taki sposób pracy systemu wiąże się z określonymi problemami. Przede wszystkim program nie może wstrzymywać swojej pracy w przypadku nieplanowanych opóźnień w działaniu którejkolwiek z funkcjonalność. Dlatego też, z założenia, napisany system ma działać wielowątkowo. Konkretnie, każdy z opisanych w poprzednim podrozdziale procesów ma działać jako osobny wątek. Komunikacja pomiędzy nimi, odbywa się przez umieszczanie wyjścia jednego procesu w buforze postaci nieblokującej kolejki FIFO, z której następny proces pobiera dane, gdy tylko

jest na to gotowy. Uproszczony schemat przedstawiający komunikację między procesami przedstawiony jest na *rysunku 2.1*.



Rysunek 2.1 – komunikacja między procesami

2.4 Porównanie do istniejącego oprogramowania

Sam system wykorzystuje szereg oprogramowania trzeciego, w celu realizacji pewnych zadań, bez konieczności „wymyślania koła na nowo”:

A. *Wireshark* ma za zadanie przechwycić ruch sieciowy i wydobyć z niego informacje na temat adresów urządzeń z którymi komunikuje się użytkownik podczas korzystania z Internetu,
B. *Traceroute* wyznacza trasę pomiędzy komunikującymi się urządzeniami,
C. Serwis i*p-api.com* pozwala określić lokalizację urządzeń pośredniczących w komunikacji.

Jednak żadne z wymienionych oprogramowań nie jest w stanie samo z siebie dać przeciętnemu użytkownikowi łatwo odczytywalnych i zrozumiałych danych.

Dostępne jest pewne oprogramowanie (wspomniane we wstępie do pracy), występujące w postaci serwisu internetowego, który umożliwia zwizualizowanie (w postaci punktów na mapie Google) połączenia pomiędzy dwoma adresami IP. Jednak wymaga on od użytkownika ręcznego wprowadzenia adresów zarówno nadawcy jak i adresata. Prezentowany system po uruchomieniu nie wymaga od użytkownika żadnej ingerencji. Automatycznie pobiera i przetwarza dane na temat wszystkich nawiązywanych połączeń.

[Koniec fragmentu]