Filtr protokołu SMTP

Praca inżynierska

Zbigniew Artemiuk
Z.Artemiuk@stud.elka.pw.edu.pl

14 lipca 2008

Spis treści

1 Wstęp

W dobie obecnego Internetu, pomimo szeregu rozwijających się form komunikacji, chociażby technologii takich jak tekstowa komunikacja czasu rzeczywistego (czyli IM - instant messaging i popularne chaty) czy też jeszcze bardziej zaawansowanych technologii, które umożliwiają przesyłanie głosu oraz wideo (chociażby Skype), pospolite maile cały czas jednak znajdują zastosowanie. Używamy ich chyba obecnie najczęściej do kontaktów biznesowych, rodzinnych, przyjacielskich, głównie w celach informacyjnych (czyli sam tekst), jak również do przekazywania niewielkich plików. Chyba każdy obecny internauta legitymuje się przynajmniej jedną skrzynką mailową (a czasami jest ich znacznie więcej). Każdy też internauta spotkał się także z niechcianą pocztą (niechcianymi mailami), nazywaną bardzo ogólnym określeniem Spam, i narzędziami (chociażby tymi wbudowanymi w klientów pocztowych dostępnych z poziomu przeglądarki internetowej), które umożliwiają odseparowanie takiej poczty, od tej która niesie interesującą nas zawartość merytoryczną. Separacja ta to tzw. filtrowanie ze względu na obiekt filtracji zwane filtracja poczty. Dokonanie filtracji wymaga jednak poznania dokładnie w jaki sposób transportowane są nasze maile, a to wszystko zawarte jest w protokole Simple Mail Transfer Protocol (w skrócie SMTP).

2 Cel i zakres pracy

Celem pracy jest zaprojektowanie systemu w języku Java, który umożliwiałbym przy pomocy dosyć prostej konfiguracji w postaci pliku tekstowego, jego uruchomienie oraz dokonywałby filtracji, na podstawie reguł zawartych w konfiguracji, przesłanych do niego protokołem SMTP wiadomości.

3 Protokół SMTP

3.1 Historia powstania

3.1.1 ARPANET czyli początki Internetu

Historia powstania protokołu SMTP jest ściśle związana z początkami Internetu. Internet zaś i jego kreowanie związane jest bezpośrednio ze swoim przodkiem czyli ARPANETEm.

ARPANET (Advanced Research Projects Agency Network) został stworzony przez przez jedną z agencji United States Department of Defense (departament bezpieczeństwa Stanów Zdjednoczonych) o nazwie ARPA (Advanced Research Projects Agency). Nazwa agencji została później przekształcona na DARPA (D od Defence). Agencja ta miała zając się rozwojem nowych technologii na potrzebę amerykańskiego wojska.

W miarę wchodzenia w życie komputerów wykorzystywanych w ramach agencji powstała idea stworzenia sieci pomiędzy nimi, która to umożliwiłaby komunikacje pomiędzy ich użytkownikami. Idea ta została po raz pierwszy zaproponowana przez Josepha Carla Robnetta Licklidera z firmy Bolt, Beranek and Newman (obecnie BBN Technologies) w sierpniu 1962 w serii notatek na temat koncepcji "Międzygalaktycznej Sieci Komputerowej". Zawierała ona prawie wszystko czego możemy doświadczyć w dzisiejszym Internecie.

W październiku 1963 roku Licklider został mianowany szefem programu Behavioral Sciences and Command and Control w ARPA. Przekonał on wtedy Ivana Sutherlanda i Boba Taylora, że jego wizja jest czymś naprawdę istotnym. Sam Licklider nie doczekał jednak żadnych konkretnych prac w kierunku jej urzeczywistnienia, gdyż opuścił ARPA.

ARPA i Taylor cały czas byli zainteresowanie stworzeniem sieci komputerowej, ażeby zapewnić naukowcom pracującym w ramach ARPA w różnych lokalizacjach, dostęp do innych komputerów, które firma oferowała. Istotne było także, aby nowe oprogramowanie i rezultaty badań były jak najszybciej widoczne dla każdego użytkownika sieci. Sam Taylor posiadał 3 różne terminale, które dawały mu podłączenie do 3 różnych komputerów - jeden do SDC Q-32 w Santa Monica, drugi w ramach projektu Project Genie do komputera na Uniwersytecie w Kalifornii (Berkley) i ostatni do komputera z Multicsem w MIT (The Massachusetts Institute of Technology).

Taylor w taki sposób opowiadał od podłączeniu do tych komputerów: " Dla każdego z tych terminali miałem inny zestaw poleceń. Dlatego też kiedy rozmawiałem z kimś z Santa Monica, a poźniej chciałem ten sam temat skonsultować z kimś z Berkley albo MIT, musiałem przesiąść się do innego terminala. Oczywistym wtedy wydało mi się, że musi być 1 terminal, który obsłuży te 3 połączenia. Idea ta to właśnie ARPANET"

Do połowy 1968 roku kompletny plan sieci został stworzony i po zatwierdzeniu przez ARPA, zapytanie ofertowe RFQ (Request For Quotation) zostało posłane do 140 potencjalnych wykonawców. Większość potraktowała propozycję jako dziwaczną. Tylko 12 firm złożyło oferty z czego 4 zostały uznane za najważniejsze. Do końca roku wyłoniono 2 firmy, z których ostatecznie 7 kwietnia 1969 roku została wybrana firma BBN.

Propozycja BBN była najbliższa planom ARPA. Pomysłem ich było stworzenie sieci z mały komputerów zwanych Interface Message Processors (bardziej

znanych jako IMPs), które to obecnie nazywamy routerami. IMPsy z każdej strony zapewniały funkcje przechowywania i przekazywania pakietów, a połączone były między sobą przy użyciu modemów podpiętych do łączy dzierżawionych (o przepustowości 50 kbit/sekundę). Komputery podłączone były do IMPsów poprzez specjalny bitowy interfejs. W ten sposób stawały się one częścią sieci ARPANET.

Do zbudowania pierwszej generacji IMPsów BBN wykorzystała komputer Honeywell DDP-516. Został on wyposażony w 24kB pamięci rdzenia (z możliwością rozszerzenia) oraz 16 kanałów Direct Multiplex Control (DMC) do bezpośredniogo dostępu do tej pamięci. Poprzez DMC podłączane były komputery użytkowników (hosty) i modemy. Dodatkowo 516 otrzymał ezstaw 24 lamp, które pokazywały status kanałów komunikacyjnych IMPa. Do każdego IMPa można było podłączyć do czterech hostów i mógł się on komunikować z 6 zdalnymi IMPami poprzez współdzielone łącza.

Zespół z BBN (początkowo 7 osób) szybko stworzył pierwsze działające jednoski (IMPy). Cały system, który zawierał zarówno sprzęt jak i pierwsze oprogramowania zarządzające pakietami, został zaprojektowany i zainstalowany w ciągu 9 miesięcy.

Początkowo ARPANET składał się z 4 IMPów. Zostały one zainstalowane w:

- UCLA (University of California, Los Angeles), gdzie Leonard Kleinrock założył centrum pomiaru sieci (Network Measurment Center)
- The Stanford Research Institute's Augmentation Research Center, gdzie Douglas Engelbert stworzył system NLS, który między innymi wprowadził pojęcie hypertextu
- UC Santa Barbara
- The University of Utah's Graphics Department, gdzie przebywał ówcześnie Ivan Sutherland

3.1.2 Piersze wiadomości w ARPANETcie

Pierwsza komunikacja host-host w sieci ARPANET wykorzystywała protokół 1822, który definiował sposób w jakis host przesyłał wiadomość do IMPa. Format wiadomości był tak zaprojektowany, żeby bez problemu mógł pracować z szerokim zakresem architektur. Zasadniczo wiadomość składała się z:

- typu wiadomości
- adresu hosta
- pola z danymi

W celu wysłania wiadomości do innego hosta, host wysyłający powinien sformatować wiadomość tak, aby ta zawierała adres hosta docelowego oraz dane, a następnie dokonać transmisji wiadomości przez interfejs sprzętowy 1822. IMP dostrzeże dostarczenie wiadomości albo poprzez dostarczenie jej bezpośrednio do hosta docelowego albo poprzez przekazanie jej do kolejnego IMPa. Kiedy wiadomość została odebrana przez docelowego hosta, IMP do którego host był

podłączony wysyła potwierdzenie odbioru (zwane Ready for Next Message or RFNM) do hosta wysyłającego.

W przeciwieństwie do obecnych protokołów datagramowych w Interecie (takich jak no IP), ARPANETowy porotokół 1822 zapewniał niezawodność w taki sposób, że informował o niedostarczonej wiadomości. Niemniej protokół 1822 nie był odpowiedni do żonglowania wieloma połączeniami w różnych aplikacjach uruchomionych na pojedynczym hoście. Problem ten został rozwiązany dzięki wprowadzeniu na hostach Network Control Program (NCP), dzięki któremu możliwe było niezawodne, z kontrołą przepływu, dwukierunkowe połączenia pomiędzy różnymi procesami na różnych hostach. NCP implementował kolejną warstwę znajdującą się na górze stosu protokołów. Dzięki niemu aplikacje, które miały mieć już jakąś konkretną funkcjonalność, mogły wykorzystywać spójny interfejs i korzystać swobodnie z dobrodziejstw ARPANETu czyli wykonywać połączenia do innych aplikacji przez sieć.

Już niedługo, bo na początku roku 1970, powstał pierwszy program (a w zasadzie 2 oddzielne) do wysyłania i odbierania wiadomości. Zaimplementował go Ray Tomlinson podczas pracy w niewielkiej grupie nad systemem operacyjnym TENEX. Programy te to SNDMSG i READMAIL. Pierwsza wersja tych programów służyła jednak do wymiany informacji między użytkownikami jednej maszyny. Już jednak w 1971 Tomlinson stworzył pierwszą aplikację ARPANETową, która umożliwiała wysyłanie wiadomości do dowolnych hostów. Tomlinson dokonał usprawnień w programie SNDMSG przy okazji

3.2 Szczegóły protokołu

Szczegóły protokołu

3.2.1 Komunikacja klient-serwer

Komunikacja

3.2.2 Konstrukcja wiadomości

Konstrukcja

3.3 Obecne wykorzystanie protokołu i jego forma

Obecne wykorzystanie

4 Dzisiejsze narzędzia do filtracji protokołu SMTP

Dzisiejsze

4.1 Konieczność wprowadzenia filtracji

Konieczność

4.2 Produkty komercyjne

Produkty komercyjne

4.2.1 Clearswift

Clearswift

4.2.2 Aladdin eSafe

Aladdin eSafe

4.2.3 Surfcontrol Email Filter

Surfcontrol Email Filter

4.3 Produkty open-source

5 Własny projekt do filtracji poczty

5.1 Założenia projektu

5.2 Moduły projektu

5.2.1 Parser wiadomości

5.2.2 Parser reguł

5.2.3 Analizator wiadomości

5.2.4 Kolejka

5.3 Kompilacja, konfiguracja i uruchomienie projektu

5.4 Testy wydajnościowe

6 Spostrzeżenia, wnioski

Literatura

[Cro82] David H. Crocker. *RFC822 - STANDARD FOR THE FORMAT OF ARPA INTERNET TEXT MESSAGES*. 1982.