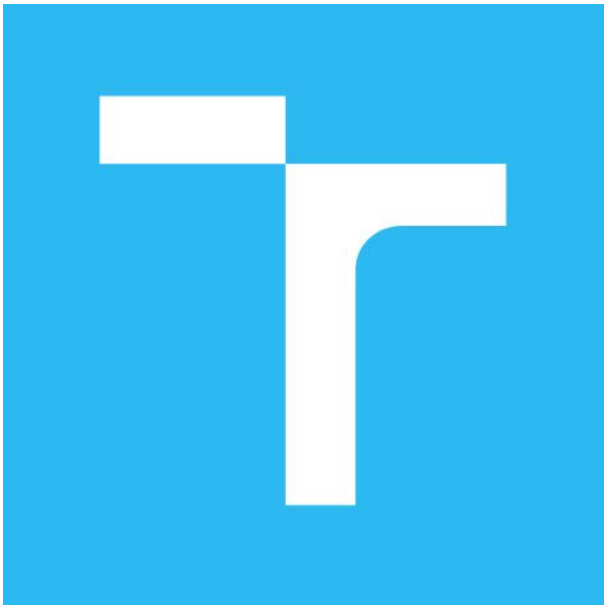
VYSOKÉ UČENÍ TECHNICKÉ V BRNĚ

FAKULTA INFORMAČNÍCH TECHNOLOGIÍ



Dokumentace k projektu pro predmět ISA

ČTEČKA NOVINEK VE FORMÁTU ATOM

S PODPOROU SSL/TLS

Autor: Norbert Ďurčanský, xdurca01@stud.fit.vutbr.cz
Fakulta Informačních Technologií
Vysoké Učení Technické v Brně

Obsah

1 Úvod

Tento dokument slúži ako dokumentácia k projektu do predmetu ISA a zaoberá sa vysvetlením problematiky formátu Atom, SSL/TLS pripojenia, návrhu aplikácie, implementácie a použitia. Táto aplikácia slúži ako čtečka správ vo formáte Atom s podporou SSL/TLS pripojenia.

2 Dôležité pojmy

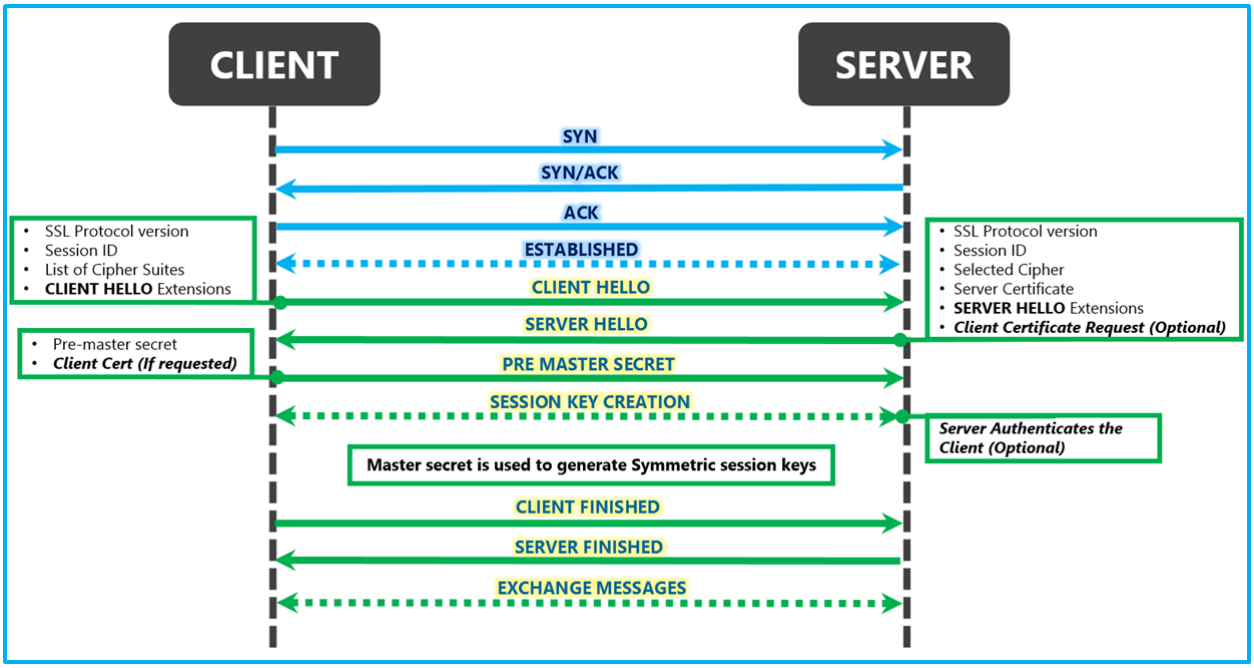
Pre návrh a implementáciu čtečky je potrebne mať znalosti z funkčnosti aplikácie, správneho parsovania dat, pripojenia a zaslania požiadavku. V tejto časti dokumentu je stručne popísaný formát Atom a SSL/TLS pripojenie.

2.1 Formát Atom

Formát Atom je založený na XML jazyku používaný pre webový „feed“, ktorý nám umožňuje jednoducho sťahovať a analyzovať informácie vďaka štandardu tohto formátu. Vlastník stránky môže použiť špecializovaný software, ktorým zverejní zoznam posledných článkov v štadardizovanej, programovo prečitateľnej forme a následne sa uživateľ pomocou programu môže pripojiť na odber „feed“ a sledovať jeho obsah. Vo formáte Atom je možné nadefinovať mnoho atribútov a vlastností, pre naše zadanie je to názov „feed“ , názov článku, meno autora, posledná aktualizácia a url adresa.

2.2 SSL/TLS

Pretože zdroje sa často nachádzajú na adrese, kde je potrebné pristupovať cez šifrované HTTP naša aplikácia musí podporovať protokol na šifrovanie dát. Tieto protokoly nám nasledne slúžia na bezpečnú komunikáciu cez internet. TLS klient používa obojstrannú autorizáciu vo forme „podania rúk“, počas ktorého sa dohodnú na rôznych parametroch na vytvorenie bezpečného spojenia. Pripojenie začne vtedy, keď sa klient pripojí na server s povoleným TLS spojením a začne žiadať o bezpečné spojenie. V našej aplikácií sme využili aj možnosť klienta kontaktovať server, a overiť si či je certifikát ktorý sme použili platný.



Obrázok 1: HTTPS handshake *Zdroj: http://blogs.msdn.com/cfs-filesystemfile.ashx/__key/communityserver-blogs-components-weblogfiles/00-00-01-38-03-metablogapi/7608.080613_5F00_0416_5F00_SSLHandshak4.png*

3 Návrh aplikácie

Aplikácia musí podporovať http a https protokoly. Po vytvorení pripojenia sa pošle požiadavka na súbor vo formáte atom. Následne odpoveď spracujeme s využitím libxml knižnice a získané data vypíšeme na výstup. Po výpise program skončí alebo pri použití *feedfile* pokračuje na ďalšiu adresu. Bolo potrebné naštudovať RFC, aby sme formát vedeli spracovať a najsť odpovedajúce elementy v xml odpovedi.

4 Implementácia

Program je implementovaný v jazyku C/C++.Je navrhovaný objektovo z dôvodu lepšej prehľadnosti. Aplikácia je vytvorená pre OS Linux a bola vyvyjaná a testovaná na Linux/Ubuntu *http://nes.fit.vutbr.cz/isa/ISA2015.ova*

4.1 Pripojenie a podpora SSL/TLS

Program sa spúšťa vytvorením objektu Connection Atom(MyCommand). MyCommand je štruktúra uchovavájúca rozparsované informácie, napr. hostname , port ... Po správnom rozparsování sa podľa zadaného protokolu zavolá funkcia SSLdownload alebo TCPdownload. Obe funkcie používajú BIO štruktúry z knižnice openssl, ktoré slúžia k jednoduchému pripojeniu a získaniu dat. Kódy pripojenia sú v častiach prevzáte z IBM stránok od autora Kenneth Ballard. Pri použiti protokolu http sa nevytvára bezpečné pripojenie, nedochádza

k overovaniu certifikátov, nastavovaniu kontextu pripojenia ako v prípade protokolu https vo funkcii SSLdownload.

4.2 Parsovanie Atom

Po prijatí odpovede zo servera sa zovalá funkcia Feedparser(), ktorá spracuje odpoveď použitím knižnice libxml. Vytvorili sme si pole štruktúr kde si ukladáme informácie ku každému elementu <entry> a <feed>. Pri výpise prechádzame pole a vypisujeme odpovedajúce údaje vyžadované uživateľom v poradí, v akom ich zadal. Aplikácia je v štandarde voľnejšia, čo nám umožňuje aj pri nedodržaní všetkých pravidiel RFC daný výpis uživateľovi poskytnúť.

4.3 Použité knižnice

```
#include <cstring>

#include <string>

#include <unistd.h>

#include <fstream>

#include <vector>

#include <openssl/rand.h>

#include <openssl/ssl.h>

#include <openssl/err.h>

#include <openssl/bio.h>

#include <ctype.h>

#include <iostream>

#include <libxml/xmlmemory.h>

#include <libxml/parser.h>

#include “Timefun.hpp”

#include “arfeed.hpp”
```

5 Použitie aplikácie

Program sa spúšta cez prikazový riadok vo formáte:

./arfeed http://tools.ietf.org/agenda/atom | -f feedfile [-c certfile] [-C certaddr] [-l] [-T] [-a] [-u]

Argument -f špecifikuje súbor ktorý obsahuje adresy zdrojov.

Argument -c špecifikuje súbor s certifikátmi.

Argument -C špecifikuje adresár ktorý obsahuje súbory s certifikátmi.

Prepínač -l vypíše len najaktualnejší zdroj.

Prepínač -T vypíše aktualizáciu zdroja.

Prepínač -a vypíše meno autora.

Prepínač -u vypíše url adresu zdroja.

6 Záver

Program vypisuje informácie o zdrojoch ziskaných z odpovede vo formáte Atom. Program je prekládaný prekladačom g++, pre preklad slúži Makefile príkazom make. Aplikácia bola úspešne otestovaná na operačnom systéme Linux.

A Metriky

Počet súborov: 4 súbory

Počet riadkov zdrojového textu: 1040

Veľkosť statických dat: 24024B

Veľkosť spustiteľného kódu: 51353B

Zdroje

[1] RFC Atom

https://tools.ietf.org/html/rfc4287

[2] Openssl

http://www.ibm.com/developerworks/library/l-openssl/