1 Protokol HTTP, server, klient

WWW služby jsou založeny na architektuře klient-server. Klient dostane od uživatele URL (například http://www.linux.cz/blurb.html), provede spojení na server (zde www.linux.cz) a zadá požadavek na stažení stránky (zde /blurb.html). Server příslušnou stránku najde na svém disku a pošle ji klientovi. Komunikace mezi serverem a klientem probíhá pomocí protokolu HTTP (HyperText Transfer Protocol), který je, stejně jako ostatní protokoly na kterých je Internet založen, otevřený a dostupný pro každého. Každý tedy může napsat jak svého klienta, tak svůj server. Komunikace mezi serverem a klientem je nechráněná, jde tedy jednoduše nejen ručně testovat, ale i odposlouchávat. Standartní port pro WWW servery je 80/tcp ale může samozřejmě poslouchat i na jiném nebo jiných portech.

HTTP protokol můžete testovat i „ručně“, například pomocí programu *telnet*. Pokud chcete vyzkoušet například jak funguje ww.linux.cz, zadáte

```
telnet www.linux.cz 80
```

a po navázání spojení bude server čekat na vaše příkazy.

Protokol HTTP se vyvíjel a nadále vyvíjí. První používaná verze bývá označována jako 0.9, nyní se používají verze 1.0 a 1.1. Protokol je zpětně kompatibilní.

Komunikace v první používané verzi (0.9) probíhá jednoduchým požadavkem na konkrétní stránku na daném serveru (pouze stránku, server už je vybraný při navázání spojení) a následnou odpovědí vlastními daty (HTML, obrázek,. . .).

`GET /blurb.html` *požadavek*

`<HTML>` *odpověď*

```
<HTML>
<HEAD>
<TITLE>Co je Linux?</TITLE>
</HEAD>
<BODY BGCOLOR=WHITE>
...
```

V HTTP verze 1.0 a vyšším je možné používat více různých přenosových metod (kromě GET i HEAD, POST, PUT a jiné) a je možné přiložit i další specifikace přenosu (oběma směry). To může zahrnovat vyžadované a vrácené upřesnění MIME typu dat, použité kódování či jazyk, informace o browseru a serveru, délku přenášených dat a další.

Klient u požadavku specifikuje i verzi HTTP protokolu, poté pošle své hlavičky (jedna na řádek, ve formátu „hlavička: údaj“) ukončené prázdným řádkem. Odpověď serveru má na prvním místě informaci o chybě (OK, permission denied,. . .), pak následují hlavičky (stejní forma jako u klienta) opět ukončené prázdným řádkem. Po něm server pošle vyžádanou stránku.

```
GET /blurb.html HTTP/1.0 požadavek
Connection: Keep-Alive
User-Agent: Mozilla/4.08 [en] (X11; U; SunOS 5.7 sun4u; Nav)
Host: www.linux.cz
Accept: image/gif, image/jpeg, image/pjpeg, image/png, */*
Accept-Encoding: gzip
Accept-Language: cs
Accept-Charset: iso-8859-1,*,utf-8

HTTP/1.1 200 OK odpověď
Date: Mon, 22 May 2000 16:23:05 GMT
Server: Apache/1.3.9 (Unix) (Red Hat/Linux) mod_ssl/2.4.10 PHP/3.0.14
Vary: accept-charset
Connection: close
Content-Type: text/html; charset=ISO-8859-2
Expires: Mon, 22 May 2000 16:23:05 GMT

<HTML> data pro klienta
<HTML>
<HEAD>
<TITLE>Co je Linux?</TITLE>
</HEAD>
<BODY BGCOLOR=WHITE>
...
```

Z metod je kromě GET dobré se seznámit ještě s POST, která slouží k předávání většího množství údajů serveru (např. upload velkého souboru). Metoda PUT slouží k umísťování nových stránek na server (ale ve většině instalací je zakázána a preferuje se jiný postup, například pomocí FTP). Metoda HEAD slouží pouze ke zjištění informací o souboru (jestli nebyl od poslední návštěvy změněn,...) a server při něm odpoví pouze hlavičkami, nikoliv vlastním obsahem souboru.

Kromě protokolu HTTP existuje ještě i protokol HTTPS, neboli HTTP over SSL. Vlastní komunikační protokol je naprosto totožný s HTTP ale při přenosu se nepoužívá „otevřené“ spojení ale šifrované spojení pomocí protokolu SSL. Tím je zamezeno odposlechu komunikace[1]. Můžete proto bez obav posílat po síti i svá hesla. Díku použití **certifikátů** je zabráněno „podvrhnutí“ serveru nějakým falešným.

[1] Odposlechu vlastně ne, pouze jsou odposlechnutá data nečitelná a dešifrování je časově *velmi* náročné.

2 Apache

Nejrozšířenější WWW server (nejen) pro Linux je server Apache. Jde o open-source program s „příjemnou“ licencí, takže je možné ho používat pro osobní i komerční účely. Podporuje HTTP protokol až do (momentálně nejvyšší) verze 1.1. V současné době existuje stabilní verze 1.3.12 a vývojová alfa verze 2.0a3.

Architektura Apache je modulární, umožňující snadný vývoj vlastních modulů díky jeho (dobře dokumentovanému) API. Apache je tvořen několika základními moduly a rozšiřujícími moduly které jsou přímo v jeho distribuci. Dále pro něj existuje veliké množství nezávislých uživatelských modulů které zajšťují nadstandartní a atypické služby.
Co nám WWW server Apache přináší?

- Podpora **statických stránek** a **CGI skriptů**. Jako CGI skripty lze použít jakýkoliv spustitelný program (binárka, skript,…) který na standartní výstup generuje (alespoň minimální) hlavičky a vlastní data. Apache předává skriptu potřebné údaje pomocí proměnných. Lze nastavit kde mohou být CGI skripty uloženy (všude, jen ve vybraných adresářích,…).

- Modularizovaná architektura, podpora **DSO** (Dynamic Shared Object). Díky tomu je možné při spuštění serveru určit jaké moduly mají být nahrány a tak redukovat velikost serveru v paměti.

- Podpora **virtuálních serverů**. Apache může obsluhovat požadavky na všech nebo na vybraných síťových rozhraních serveru. Na těchto rozhraních se může tvářit buď jako jeden a tentýž server, nebo jako několik různých (a s různými nastaveními). Nabízí i možnost provozovat více virtuálních serverů nad jedním rozhraním. Takto je například provozován www.linux.cz, www.cstug.cz, mysql.linux.cz a několik dalších serverů které běží na jednom počítači s jednou IP adresou.

- Omezování přístupu. Apache umí omezovat přístup k jednotlivým adresářům (a tedy i celým virtuálním serverům) podle IP adresy, uživatelského jména a hesla, typu souboru či jiných metod. Nejen pro tyto případy obsahuje i mocný přepisovací nástroj pracující s regulárními výrazy.

- Zobrazení informací o serveru (zatížení, počet klientů, jaké jsou aktuální požadavky na server, nahrané moduly,…) jako virtuální WWW stránky.

- Podpora uživatelských domovských stránek. Tyto stránky jsou přístupné jako `~uživatel/`. Navíc nabízí možnost přistupovat k uživatelským stránkám (a CGI skriptům a podobně) pod identifikací tohoto uživatele (chrání tak vzájemně uživatele) („`suexec`“).
- Podpora vícejazyčných stránek. Možnost vlastních chybových stránek.
- Podpora pro „přátelské“ zobrazování obsahu adresáře a pro určení defaultních souborů které se mají zobrazovat (`index.html`).
- Možnost změny chování (či konfigurace) serveru pro jednotlivé adresáře. Uživatelé tak mají možnost modifikovat chování serveru na svých stránkách (`.htaccess`).
- Zabudovaná **WWW proxy**. Apache tak kromě WWW serveru může (i zároveň) pracovat jako proxy cache.
- **SSI** (Server Side Includes) neboli serverem zpracované dokumenty. Nabízí možnost dynamických HTML stránek které jsou před odesláním zpracovány serverem a umožňují tak provést do výstupu některé funkce (vložení souboru, generování data, jednoduché podmínky if…then, spuštění externích rogramů,…).

Z rozšiřujících modulů jsou nejznámější moduly:

- **PHP** nabízející serverem zpracovávaný jazyk PHP. Pracuje obdobně jako SSI nebo ASP, umožňuje do HTML kódu vkládat kusy programu, který server před odesláním dokumentu zpracuje. PHP nabízí bohatý jazyk umožňující například i spolupráci s externími databázemi.
- **mod_perl** zabudovávající do Apache interpret jazyka Perl. Ten vám umožní psát skripty zpracovávané zabudovaným interpretrem, možnost předkompilování programů (mnohem vyšší rychlost zpracování) a navíc ovlivňovat konfiguraci Apache – ať už voláním API Apache přímo ze skriptu nebo možností vložit Perlový kód do konfiguračních souborů Apache.
- **moduly pro češtinu** nabízejí možnost za běhu převádět kódování češtiny z toho co je uloženo na disku v HTML souborech nebo co generují CGI skripty na kódování které vyžaduje klientský prohlížeč. Umožňují jak automatické rozhodování co klient chce, tak vynucené kódování např. přidáním klíčových slov do URL. Nejznámější jsou *mod_html, mod_czech* a *C-SaCzech*.

- **mod_ssl** přidávající do Apache možnost používat pro komunikaci SSL (protokol HTTPS). Pro instalaci je potřeba mít balík OpenSSL.

Konfigurace Apache

Apache používá jeden konfigurační soubor `httpd.conf` který se nalézá (podle distribuce a nastavení při kompilaci) v `/etc/httpd/conf`. V adresáři `/etc/httpd/` se také dají nalézt logy přístupů a chyb (serveru i CGI skriptů).

V konfiguračním souboru je možné vytvářet sekce podle určitých omezení (virtuální server, vybrané soubory, vybrané adresáře,...) a těm přidělovat určitá nastavení. S distribucí Apache se dodává i funkční vzor. Navíc většina distribucí Linuxu už dodává Apache předkonfigurovného tak, že s největší pravděpodobností bude vyhovovat (alespoň pro začátek;-) i vám.

Výhody a nevýhody

Apache je modulární WWW server vhodný i pro velice rozdílné typy serverů. Je hodně rozšiřitelný a velice konfigurovatelný. Nabízí (i díky rozšiřujícím modulům) možnost vyhovět všem vašim potřebám. Je aktivně vyvíjen a komunita nabízí dobrou podporu.

Na druhé straně je jeho obsáhlost i nevýhodou. Kromě nepřehledného kódu ve kterém mohou být bezpečnostní díry (i když v případě objevení rychle odstranitelné) může být Apache také kanónem na vrabce pokud nepotřebujete všechny jeho možnosti. Díky tomu nepatří k nejrychlejším například při obsluze jak statických stránek tak i CGI skriptů.

Nicméně je nejuniverzálnějším WWW serverem s největším nasazením v Internetu a i já ho mohu jen doporučit.

3 Další WWW servery pro Linux

Kontroverzní **khttpd** server nacházející se přímo v jádře Linuxu. Tento server nabízí pouze omezené možnosti – jako je poskytování *pouze* statických stránek. To znamená nemožnost zpracování stránek serverem nebo CGI skripty. Výhodou je vám neuvěřitelná rychlost tohoto serveru (obsah poskytovaných stránek se vůbec nedostane do user-space, takže se šetří i na kopírování v paměti).

Komerční **Zeus Web Server** o kterém jeho tvůrci tvrdí že je to nejrychlejší a nejrozšiřitelnější web server. Podporuje ISAPI, Java Servlets, secure CGI, Fast CGI, NSAP a další. Podporuje publikování pomocí FrontPage.

Otec serveru Apache je **NCSA HTTPd**. Konfigurace je podobná Apachi, server je volně ke stažení ale možnostmi za Apachem zaostává.

Dalšími komerčními servery jsou **Netscape Fast Track Server** a **Netscape Enterprise Server**. Tyto servery jsou dodávány do komerčních aplikací. Jsou samozřejmě k dispozici i samostatně a pro Linux.

Pod GNU licencí je uvolněn server **WN**. Tento server je menší, rychlejší a bezpečnější než Apache. Neovládá sice všechny jeho možnosti ale má poměrně rozsáhlé bezpečnostní funkce a umožňuje i spouštění CGI skriptů. Také podporuje SSI a vyhledávání dokumentů podle názvu, klíčového slova a dalších atributů.

Kromě těchto výše uvedených serverů existuje i neuvěřitelné množství jiných, které jdou od plně funkčních serverů až po „hračky“ a „důkazy že to jde“. Do této kategorie můžeme zahrnout různé minimalistické WWW servery spouštěne z inetd nebo WWW server PS-HTTPD napsaný v jazyce PostScript.

Tento výčet samozřejmě není úplný a výběr vhodného WWW serveru musíte přizpůsobit účely k jakému má být použit. K dispozici je také software (opět je ho více, ne pouze jeden) pro rozložení zátěže na více serverů a pro správu serverových farem. Ve světě Linuxu je největší předností možnost volby.

4 Užitečné odkazy

Linux – www.linux.cz
Apache – www.apache.org
mod_perl – perl.apache.org
mod_php – php.apache.org
mod_ssl – www.modssl.org
mod_html – http://www.fee.vutbr.cz/~lampa/mod_html/
C-SaCzech – http://www.csacek.cz/
Russian Apache – http://apache.lexa.ru:8100/
mod_czech – http://www.intersoft.cz/mod_czech/
OpenSSL – www.openssl.org
HTTP protokol – http://www.fi.muni.cz/usr/brandejs/P005/
HTTP protokol – RFC 1945 a 2068
Zeus Web Server – www.zeus.com
NCSA HTTPd – http://hoohoo.ncsa.uiuc.edu
Netscape Server – http://home.netscape.com
WN – http://hopf.math.nwu.edu/
PS-HTTPD – http://www.pugo.org:8080/
FreshMeat – http://freshmeat.net