Internet-Access mit Squid

August Hörandl

Das Problem

Es gibt (endlich) Internetzugang im EDV-Saal. Natürlich haben die Schüler jetzt nichts Besseres zu tun, als zu surfen, zu chatten, MP3-Dateien zu laden etc. Natürlich ist es ein Leichtes, den Zugang wieder zu sperren, doch damit ist wiederum auch der Einsatz im Unterricht nicht möglich.

Mögliche Lösungen

- man erlaubt den Zugriff für alle - damit ist aber ein sinnvolles Unterrichten (teilweise) nicht möglich
- man sperrt den Zugriff - damit ist aber eine Recherche und z.B. der Zugriff auf Datenblätter nicht möglich
- man installiert einen Proxy und vergibt Zugriffspasswörter - damit hat man eine gute Kontrolle, aber
- es müssen Passwörter für jeden Benutzer verwaltet werden
- der Zugriff ist nicht wirklich verhindert sondern nur erschwert
- man kann den Zugriff auch zu bestimmten fixen Zeiten erlauben, damit ergibt sich auch keine flexible Lösung
- man verwendet die hier verwendete Lösung der Zugriffskontrolle:
- man erlaubt bzw. verhindert den Zugriff auf den Proxy über eine einfache Firewall
- einfache Konfiguration via Webinterface - für die Änderungen (Ein/Aus) ist ein Passwort notwendig
- gezieltes Ein- und Ausschalten: ein Rechner, eine Gruppe (Reihe) oder ganzer Saal - und auch jede Kombination

Warum ein Proxy?

Durch den Proxy werden alle Zugriffe (HTTP und FTP) auf das Internet »abgefangen«. Der Proxy nimmt die Anfragen der Clients entgegen und schickt sie im eigenen Namen weiter. Die Antworten werden an die Clients zurückgeschickt, zusätzlich werden die Daten aber auch auf der Festplatte gespeichert. Dadurch ist es möglich, den nächsten Zugriff direkt von der Festplatte zu beantworten und man erreicht damit höhere Datenraten bzw. eine bessere Ausnützung der zur Verfügung stehenden Bandbreite. Zusätzlich erlaubt diese Art des Internetzugangs eine einfache zentrale Zugangskontrolle und der Proxy kann auch als Filter verwendet werden d.h. unerwünschte Inhalte bzw. Anbieter können einfach gesperrt werden.

Obwohl fast alle modernen Browser einen lokalen Cache unterstützen bzw. anlegen, ist die Lösung mit einem zentralen Proxy besser, da die Daten dort nur einmal gespeichert werden und durch die gemeinsame Nutzung durch mehrere Benutzer die Trefferwahrscheinlichkeit erhöht wird.

Aber auch durch einen Proxy darf man sich natürlich keine Wunder erwarten. Wenn alle Benutzer verschiedene Seiten lesen bzw. verschiedene große Softwarepakete herunter laden kann die zur Verfügung stehende Bandbreite natürlich nicht größer werden, und die Daten werden entsprechend langsam vom Proxy zum Client übertragen.

Eine weitere Verbesserung kann der Einsatz eines oder mehrerer übergeordneter Proxies beim Provider bringen.

Anmerkung: Eine alte Frage »Proxy oder nicht Proxy - was ist schneller?«

Die Verwendung eines Proxies durch die Kunden eines Providers ist ein Kooperationsproblem (es entspricht dem bekannten Gefangenendilemma):

- wenn alle Surfer den Proxy verwenden ist der Einzelne, der keinen Proxy verwendet, schneller (die Leitungen zum Internet sind weniger ausgelastet und er erspart sich den Overhead durch den vielleicht sogar überlasteten Proxy).
- wenn niemand den den Proxy verwendet, ist der Einzelne, der den Proxy doch verwendet, schneller da es doch eventuell zu einem Zugriff aus dem Festplattencache kommt.

Konfiguration der Clients (Browser)

Als Proxy ist der Rechner mit dem Namen **`Proxy`** bzw. die entsprechende IP-Adresse und das Port 3128 einzutragen.

Internet Explorer

Extras - Internetoptionen Verbindungen - LAN Einstellungen

☒ Proxyserver verwenden
Proxy Adresse: proxy
Anschluss: 3128

☒ Proxyserver für lokale Adressen umgehen

Navigator

Bearbeiten/Einstellungen/Erweitert/Proxies bzw. Edit / Preferences / Advanced / Proxies

☒ Manual Proxy Configuration - View

3x eintragen (HTTP, Security, FTP):
Adresse: proxy Port: 3128
Do not use Proxy : proxy

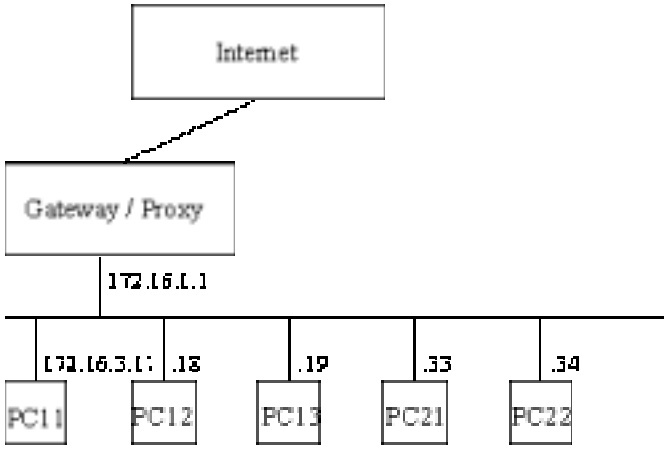
Das Netzwerk

Für die Rechner des internen Netzwerks (EDV Saal) empfiehlt sich die Verwendung von internen (privaten) IP Adressen (***Siehe z.B. PCNEWS 68 - Seite 73***).

Im Rest des Artikel verwende ich das Klasse B Netzwerk 172.16.x.x - dadurch kann auch eine große Anzahl von Rechnern großzügig mit IP-Adressen versorgt werden. Damit können z.B. die Zugriffsrechte einfach über die einzelnen Netzwerke verwaltet werden

Die Verbindung in das Internet wurde im Beispiel mit einen Linux-PC realisiert.

Damit ergibt sich folgender Aufbau des Netzwerks:



Am Gateway wird Squid als Proxysoftware installiert. Eine detaillierte Anleitung gibt es z.B. im AON Howto (*http://members.aon.at/heimo.schoen/AON-HOWTO.html* bzw. ***PCNEWS 66, Seite 98***). Squid ist bei aktuellen Linux Distributionen enthalten und sollte eigentlich »out-of-the-box« funktionieren. Weiterführende Links zu Squid gibt es unter

http://links.ee.htlw16.ac.at/linux/Netzwerk/WWW/Proxy_-_Squid/ .

Inbetriebnahme-Test durch Surfen

- Direkter Internetzugriff am Rechner Proxy funktioniert? Da am Proxy keine grafische Oberfläche installiert ist bzw. installiert werden sollte, empfiehlt sich z.B. der Einsatz des Programms **lynx** als Browser.
- Zugriff vom Rechner Proxy unter Verwendung des Squid-Proxies.
- Zugriff von einem der PCs über den Proxy.

Die IP-Adressenvergabe via DHCP

Die IP Adressen können fix bei jedem Client eingetragen werden. Bei vielen Rechnern bzw. bei Änderungen ist diese Variante aber aufwendig bzw. fehleranfällig.

Eine bessere Möglichkeit ist die Vergabe der Adressen mittels DHCP. Im Schulbetrieb und für die angestrebte Zugriffskontrolle sind aber fixe IP-Adressen von Vorteil. Ein DHCP-Server bietet die Möglichkeit, jedem PC immer wieder die gleiche Adresse zuzuteilen. Als Erkennungsmerkmal dient dabei die eindeutige MAC Adresse, die jede Netzwerkkarte vom Hersteller zugeteilt bekommt.

Die Konfiguration des DHCP-Servers geschieht in der Datei **/etc/dhcpd.conf**. Diese Datei besteht aus Zeilen mit Optionen. Bei aufwendigeren Optionen werden die Parameter in **{ }** eingeschlossen. Das Zeichen # leitet Kommentarzeilen ein.

```
# /etc/dhcpd.conf
server-identifier name.des.servers.ac.at;
# option definitions common to all supported networks...
option domain-name "unsere.domain.ac.at";
option domain-name-servers 172.16.1.1;
option subnet-mask 255.255.0.0;
option broadcast-address 172.16.255.255;
option routers 172.16.1.1;
default-lease-time 360000;
max-lease-time     720000;
use-host-decl-names on; # nur für Linux Clients
```

Linux-Rechner können auch den Rechnernamen via DHCP empfangen. MS-Windows bietet diese Möglichkeit leider nicht.

Da wir fixe Adressen verwenden können, diese auch für lange Zeit (*lease-time*) an die Rechner vergeben werden. Bei dynamischen Adressen sollte diese Zeit kürzer gewählt werden.

Für die ersten Tests oder für zusätzliche Rechner kann auch ein Bereich mit wirklich dynamischen d.h. frei vergebenen Adressen definiert werden:

```
subnet 172.16.2.0 netmask 255.255.0.0 {
  default-lease-time 6000;
  max-lease-time 7200;
  range 172.16.2.2 172.16.2.250;
}
```

Für die Rechner in den EDV-Sälen wird eine Gruppe mit fixen Adressen verwendet. Die Kommentare werden wir später noch brauchen.

```
# saal edv1 172.16.3.0/24
group {
  # reihe edv1-a 172.16.3.16/28
  # rechner edv1-a1 172.16.3.17
  host edv1-a1 {
    hardware ethernet 00:50:04:ad:60:52;
    fixed-address 172.16.3.17;
  }
  # rechner edv1-a1 172.18.3.18
  host edv1-a2 {
    hardware ethernet 00:50:04:ad:60:53;
    fixed-address 172.16.3.18;
  }
  # reihe edv1-b 172.16.3.32/28
  # rechner edv1-a1 172.16.3.33
  host edv1-a2 {
    hardware ethernet 00:50:04:ad:60:54;
    fixed-address 172.16.3.33;
  }
  ...
}
# saal edv2 172.16.4.0/24
group {
  ...
```

Hier gibt es für jede Gruppe bzw. für jeden Rechner einen Eintrag.

Wichtig ist die Netzmaske d.h. die Zahl hinter dem **»/«**.

/24 24 Bit müssen übereinstimmen d.h. Die ersten 3 Zahlen sind fix

/28 28 Bit müssen übereinstimmen, d.h. die letzten 4 Bits können sich ändern, damit kann eine Gruppe von 14 Rechner (die erste und die letzte Adresse der 16 Möglichkeiten kann nicht verwendet werden).

Mit dieser Konfiguration sollte jetzt die Konfiguration der Rechner einfacher sein: die IP-Adresse kann jetzt »dynamisch« zugewiesen werden.

Zugriffskontrolle mittels »Firewall«

Mit dieser Konfiguration können alle Rechner ungehindert surfen. Jetzt gilt es, den Proxy vor den »bösen« Schülern zu schützen. Hier empfiehlt sich der Einsatz einer Firewall. Mit einigen Regeln kann der Zugriff auf den Proxy bzw. auf das entsprechende Port (3128) gesperrt werden.

Dazu definieren wir eine eigene Regelgruppe:

```
ipchains -I input 1 -y -p tcp eth0 --destination-port 3128 -j
squid
ipchains -F squid
ipchains -A squid -l -p tcp -j DENY
```

Erklärungen

- **Regel**: erste Regel in der Gruppe **»input«** (eingehende Pakete): alle Zugriffe bzw. der Verbindungsaufbau (**-y**) von der Netzwerkkarte **eth0** auf das Port **3128** wird auf die Regelgruppe »**squid**« umgeleitet.
- **Regel**: alle Regeln in der Gruppe »**squid**« löschen
- **Regel**: eine Regel in der Gruppe »**squid**« anhängen (**-A**): Zugriff verweigern (**-j DENY**)

Damit ist der Proxy jetzt gut gesichert: kein Zugriff wird gestattet. Durch zusätzliche Regeln kann jetzt einem Rechner bzw. einer Gruppe der Zugriff erlaubt werden.

```
ipchains -I squid 1 -s 172.3.1.16/28 -p tcp
--destination-port 3128 -y -j ACCEPT
```

mit dieser Regel wird den Rechnern **172.3.1.16** bis **172.16.3.31** der Zugriff gestattet. Durch die geschickte Wahl der Adressen bzw. Maske kann jetzt der Zugriff sehr genau definiert werden. Die Adresse besteht aus 4 Bytes (üblicherweise in dezimaler Darstellung), die Maske gibt die Anzahl der übereinstimmenden Bits an

Beispiel

172.3.1.0/24 alle Rechner im EDV Saal EDV1

172.3.1.16/28 alle Rechner in der ersten Reihe im EDV Saal EDV1

172.3.1.17/32 der erste Rechner in der ersten Reihe im EDV Saal EDV1

Nach getaner Arbeit wird der Zugriff wieder gesperrt d.h. die entsprechende Regel wieder gelöscht:

```
ipchains -D squid -s 172.3.1.16/28 -p tcp --destination-port
3128 -y -j ACCEPT
```

Verbesserungen

Die direkte Eingabe erlaubt eine gute Kontrolle, für eine wirklich schöne Lösung fehlen aber noch einige Punkte:

- die Kommandozeile ist nicht wirklich einfach bzw. leicht zu merken;
- nur der Administrator kann das Kommando **ipchains** aufrufen - eigentlich sollte nicht jeder Lehrer das root-Passwort kennen.
- am Ende des Unterrichts müssen die Regeln wieder gelöscht werden (Lehrer sind sehr vergesslich).
- und nicht zuletzt: unsere tollen Kommentare in der Datei **/etc/dhcpd.conf** werden nicht verwendet

Webinterface

Zur einfachen Konfiguration bietet sich eine einfache WWW-Seite d.h. ein Formular an. Die notwendigen Daten können aus der Datei **/etc/dhcpd.conf** entnommen werden.

Dafür benötigt man eigentlich drei Programme:

- ein Formular zum Ausfüllen (siehe Bild)
- eine Ergebnisseite
- da der Web Server als einfacher User läuft, braucht man auch noch ein Programm, das als Administrator (**root**) die entsprechenden Kommandos ausführt.

Durch die Verwendung einer entsprechenden Bibliothek **cgi-lib.pl** (*http://cgi-lib.berkeley.edu/*) können alle drei Funktionen durch ein Perl-Skript verwirklicht werden:

Beim Aufruf durch den WWW-Server ohne Parameter wird das entsprechende Formular angezeigt; mit Parameter werden die notwendigen Kommandos in eine Datei (*named pipe*) geschrieben. Die notwendigen Daten über die EDV-Säle, Reihen und Rechner werden direkt aus der Datei **/etc/dhcpd.conf** gelesen. Man kann aber natürlich auch eine andere Datei verwenden. In diesem Fall muss nur eine Zeile am Beginn des Skripts angepasst werden.

Beim Aufruf als **root** werden die Kommandos aus der Datei **ws.pipe** gelesen. Neben dem entsprechenden Aufruf von **»ipchains«** wird auch mittels **»at«** das Entfernen der Zugriffsberechtigung zu einem bestimmten Zeitpunkt veranlasst. Der Aufruf des Skripts als Root muss einmal (am einfachsten beim Start des WWW-Servers) erfolgen.

Der entsprechende Teil im Skript (Perl):

```
if ($< == 0) {
  &readpipe;
}
if (&ReadParse(*input)) {
  &ProcessForm;
} else {
  &PrintForm;
}
```

Das vollständige Skript gibt es unter *http://elina.htlw16.ac.at/~hoerandl/squid/*. Die möglichen Zeiten werden am Beginn des Perlskripts definiert und können damit leicht angepasst werden.

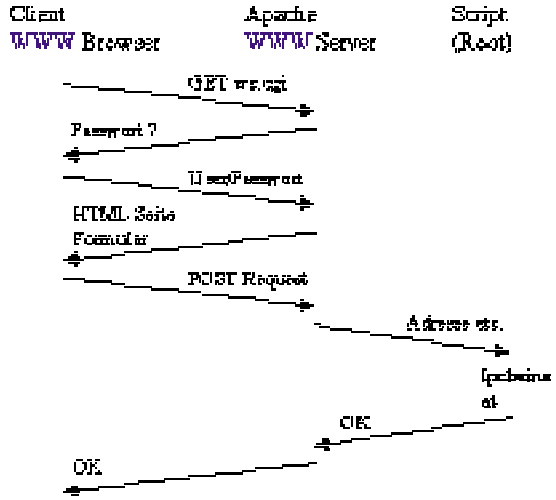
Der Apache Web Server muss entsprechend konfiguriert werden.

Wenn er mit dem Namen **proxy** angesprochen wird, soll er eine entsprechende Seite angezeigt werden:

in **/etc/httpd.conf**:

```
NameVirtualHost 172.16.1.1
<VirtualHost  proxy>
DocumentRoot /usr/local/httpd/squid
CustomLog /var/log/httpd/squid.access_log common
<Directory "/usr/local/httpd/squid">
AuthType Basic
AuthName anything
AuthUserFile /etc/squid.passwd
require valid-user
Options +ExecCGI
DirectoryIndex ws.cgi
</Directory>
</VirtualHost>
```

Durch diese Konfiguration kann das Formular zur Konfiguration direkt als **http://proxy** aufgerufen werden. Es ist wichtig, dass für diesen Zugriff der Proxy nicht verwendet wird, denn dieser Zugriff ist im Normalfall gesperrt. Es muss deshalb bei der Konfiguration des Browsers die Option ***»kein Proxy für«*** angegeben werden. Damit ergibt sich folgender Ablauf



Mit dem Kommando **»htpasswd«** können vom Supervisor entsprechende Benutzer angelegt werden, das Passwort wird interaktiv abgefragt

```
htpasswd /etc/squid.passwd username
```

Alternativ kann das Passwort direkt beim Aufruf angegeben werden:

```
htpasswd /etc/squid.passwd username passwort
```

Ein Passwort ist nur für die Lehrer notwendig, denn nur diese sollen das Internet einschalten können. Zusätzlich kann man natürlich einigen Schülern bzw. einer Schülergruppe ein Passwort zuteilen.

Kurzanleitung

Aktivieren

- Folgende URL anwählen: *http://proxy*
- Im erscheinenden Formular auswählen:
  Aktion: Einschalten / Ausschalten
  Dauer: Zeitpunkt der automatischen Deaktivierung
  Rechner: einen Saal, eine Reihe oder einen Einzelrechner auswählen
- Senden-Knopf anklicken
- Passwort

Zum Ein- und Ausschalten ist ein Passwort notwendig - Bitte wenden Sie sich an die ITZ-Verantwortlichen