

Grundlagen

Squid, ein Proxy-Cache-Server für Internet-Inhalte, kann folgende Aufgaben übernehmen:

- zwischenspeichern (Caching) von Internet-Seite und hierdurch Beschleunigung des Zugriffs auf häufig besuchte Internet-Seiten und Programme;
- Proxy-Funktion und die somit erfolgte Abkopplung Client-Anfragen und Server-Antworten, hierdurch kann die Sicherheit erhöht werden und die Clients müssen keinen vollständigen Zugang zum Internet haben;
- Filterung von Internet-Inhalten, die allgemein oder für bestimmte Gruppen unerwünscht sind;
- Dokumentation und Protokollierung des Internetverkehrs;

Eigenschaften

- Größe des Festplatten-Caches kann frei definiert werden (je nach Bedürfnissen und erwartetem Volumen);
- Scalierbarkeit durch Aufbau von Cache-Hierarchien (mehrere Squid-Server stehen in bestimmten Beziehungen zueinander);
- eigener DNS-Cache für schnellere Namensauflösung;
- Authentifizierung des Clients am Server über zusätzliche Module möglich (z.B. Samba/WinNT/Win2k PDC, LDAP, NCSA, PAM ...);
- Squid ist für alle modernen UNIX-Varianten (Linux, BSD, AIX, Solaris ...) sowie für OS/2 und WinNT/Win2k verfügbar;
- Ausfilterung und Anonymisierung von Client-Anfragen (z.B. Browsertyp maskieren);
- sehr flexible Access Control Lists, auch mit regulären Ausdrücken und ausgelagerten Sperrlisten;
- Filterung über externe Programme (z.B. Squid-Guard);

Protokolle

Squid unterstützt die gebräuchlichsten Protokolle für die Kommunikation. Dies sind für die Client-Kommunikation:

- HyperText Transfer Protocol (HTTP);
- File Transfer Protocol (FTP);
- Gopher;
- Wide Area Information Service (WAIS);
- Secure Socket Layer (SSL)

Für die Kommunikation zwischen Proxy-Cache-Servern und zum Management:

- HyperText Transfer Protocol (HTTP);
- Internet Cache Protocol (ICP (3130));
- Cache Digests (Index von anderen Caches abfragen);
- Simple Network Management Protocol (SNMP);
- HyperText Caching Protocol (HTCP (4837));
- Web Cache Coordination Protocol (WCCP);

Systemvoraussetzungen

Für Squid ist die CPU-Leistung weniger relevant, mehr zählt die Größe und Zugriffszeit des Arbeitsspeichers (RAM) sowie die Geschwindigkeit der Festplatte(n). Müssen Teile vom verwendeten Arbeitsspeicher auf die Festplatte ausgelagert werden, wird Squid extrem langsam. Den Squid-Durchsatz kann durch Verteilung des Festplatten-Caches auf mehrere Festplatten, entweder durch Definition mehrere Caches-Verzeichnisse oder durch Verwendung von RAIDs mit Level 0, sowie durch heraufsetzen der Hash-Struktur erhöht werden.

Rechenbeispiel für Arbeitsspeicherverbrauch:

Jedes gecachte Objekt (Web-Seite ...) belegt im Arbeitsspeicher ca. 75 Byte für den Index. Die durchschnittliche Objektgröße beträgt 13 kByte. Bei einem Festplattencache von 1 GB können ca. 78.000 Objekte abgespeichert werden. Die Index-Größe (bei Vollauslastung) benötigt ca. 5.7 MByte. Bei einem Festplattencache von 10.4 GB steigt die Größe des Indexes auf ca. 60 MByte. Hierbei handelt es sich nur um den Index, um die Objekte finden zu können. Zusätzlich verbrauchen DNS-Cache, die gerade bearbeiteten Objekte selbst und natürlich das eigentliche SQUID-Programm und Betriebssystem noch RAM.

Konfiguration

Wie die meisten Programme im Unix-Umfeld wird Squid über eine ASCII-Konfigurationsdatei gesteuert (squid.conf, z.B. in /etc oder in /usr/local/squid/etc). Festplatten-Cache und Log-Dateien befinden sich häufig unter /var/squid.

wichtige Konfigurationseinstellungen

http_port <Port(s)>: TCP-Port, auf dem Squid auf Anfragen wartet; Standardport ist 3128, häufig wird auch 8080 verwendet;

cache_dir <Typ Pfad MBytes Level-1 Level-2>: Gibt das Festplattencache-Verzeichnis mit Zugriffstyp, max. Größe und Anzahl der Unterverzeichnisse (Level-1 und Level-2) an. Werden mehrere Cache-Verzeichnisse angegeben (einzelne Direktiven), wird die Last zwischen beiden Verzeichnissen verteilt;

cache_access_log <Pfad/Dateiname>: Zugriffslogbuch;

cache_log <Pfad/Dateiname>: generelle Cache-Informationen;

log_fqdn <on|off>: Auflösung der IP-Adressen in DNS-Namen im Zugriffslogbuch (performanzbremsend);

ftp_user <String>: eMail-Adresse, die Squid bei Anonymous FTP verwendet;

authenticate_program <Program>: externes Authentifizierungsprogramm;

refresh_pattern <Optionen>: Gibt an, wie lange Objekte im Festplatten-Cache verbleiben, bevor sie gelöscht werden;

netative_ttl <Zeit Einheit>: Lebensdauer für nicht beantwortbare Anfragen;

negative_dns_ttl <Zeit Einheit>: Lebensdauer für nicht beantwortbare DNS-Anfragen;

acl <aclname acltype string>: Definiert eine Access Liste, um eine Zugangskontrolle sowie eine Sperrliste aufzubauen; Einige wichtige ACL-Typen sind:

- src: Client IP-Adresse/Adressbereich;
- dst: Zielservers IP-Adresse/Adressbereich;
- srcdomain: Client (Domain-)Name;
- dstdomain: Zielservers (Domain-)Name;
- srcdom_regex: Client (Domain-)Name als RegEx;
- dstdom_regex: Zielservers (Domain-)Name als RegEx;

- port: Port oder -bereich für Verbindungen;
- proto: Protokoll (z.B. HTTP, FTP ...) für die Verbindung;
- method: Methode der Verbindung (z.B. GET, POST ...);
- proxy_auth: Benutzernamen bei Authentifizierung;
- urlpath_regex: Teil einer URL als RegEx;

http_access <allow|deny [!]aclname>: Erlaubt oder verweigert den Zugriff basierend auf der definierten Access-Liste;

Was Squid nicht kann

Squid kann nicht den Inhalt einer Seite analysieren und somit ist auch keine Sperrung nach Textinhalten oder in HTML-eingebetteten JavaScript und ActiveX möglich. Hierfür müssen zusätzliche Anwendungs-Proxies implementiert werden (z.B. HTTP-GW/SQUID-GW aus dem TIS Firewall-Toolkit). Zusätzlich kann Squid keine Proxy-Dienste für SOCKS anbieten, auch hier müssen andere Proxies verwendet werden.

Ressourcen

squid.conf Konfigurationsdatei, (fast) alle Konfigurationsparameter sind dort ausführlich erklärt

www.squid-cache.org Squid-Homepage mit Squid-FAQ und User Guide