Configuring the Squid Internet Proxy Server

Bob Brown

Transentia Pty. Ltd.

bob@transentia.com.au
http://www.transentia.com.au

BES Information Technology SYSTEMS *I.T in Education 2002 Conference*

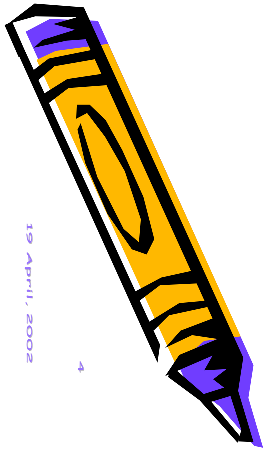
Introduction

- A high-performance proxy caching server for web clients
  - HTTP(S), also FTP and Gopher
  - Can reduce bandwidth consumption
    - thus time and money!
    - if an object is accessed frequently
  - Can facilitate access control/content filtering
    - on its own or in conjunction with squidGuard, DansGuardian, etc.
  - Available on many platforms
    - Linux, *nix, windows, OS/2, etc.
    - originated within US' "Harvest" DARPA project
      - thus is Open-Source
      - now maintained by the National Laboratory for Applied Network Research (NLANR)
      - commercial support exists
  - Well supported by auxiliary tools
    - Calamaris, etc.

Features

- Keeps especially hot objects cached in RAM
- Supports non-blocking DNS lookups
- Caches DNS lookups
- Implements negative caching of failed requests
  - remembers “Not Found” and “Connection Refused” results
- Supports SSL
- Has extensive access controls
- Performs full request logging
- Can be arranged in a hierarchy or mesh for additional bandwidth savings
- Plays well with firewalls

Useful Quotes



- Why the name Squid? “All the good ones are taken.”
- “Simply put, it’s an intermediary (or proxy) computer system between Web browsers and Internet Web servers”
- “…ISPs, educational institutions and corporations all find that it measurably enhances system performance…That’s why you should be interested in running Squid if you’re doing any sort of Web serving.”
- “Bandwidth is expensive, perhaps the most expensive element of an Internet connection.”
- “…at the end of the day, over 100mb of data per day was coming from the cache, and not from the internet.”
- “Squid lets us do two things essential in a school environment…: it lets us force users to identify themselves with a username and password, and it allows us to log and filter the requests they send and (if we wish) the material they receive.”

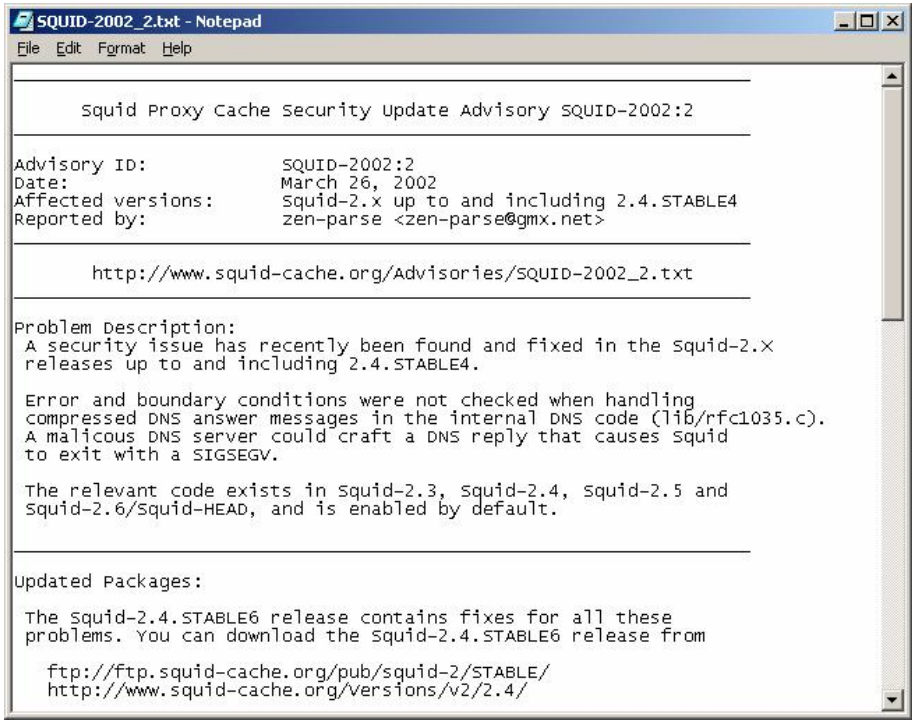
Installation

- Readily available for most Linux-en
  - as source or as an RPM package
    - usually already installed but always ensure that the latest STABLE version is used
      - security is *always* an issue!
      - for RedHat Linux, check http://www.redhat.com/apps/support/errata/
- Reference platform
  - Complete install of Redhat Linux 7.2
  - Intel Pentium Pro 200

```
% uname -a
Linux redhat 2.4.7-10 #1 Thu Sep 6 17:27:27 EDT 2001 i686 unknown
% rpm -qa | egrep '(squid-|apache-)'
apache-1.3.22-2
apache-manual-1.3.22-2
apache-devel-1.3.22-2
squid-2.4.STABLE1-6
```

Security

- Important to keep abreast of security
  - squid developers issue security advisories as necessary
    - one was issued as I was preparing these slides



```
______________________________________________________________________

      Squid Proxy Cache Security Update Advisory SQUID-2002:2
______________________________________________________________________

Advisory ID:            SQUID-2002:2
Date:                   March 26, 2002
Affected versions:      Squid-2.x up to and including 2.4.STABLE4
Reported by:            zen-parse <zen-parse@gmx.net>
______________________________________________________________________

       http://www.squid-cache.org/Advisories/SQUID-2002_2.txt
______________________________________________________________________

Problem Description:
 A security issue has recently been found and fixed in the Squid-2.X
 releases up to and including 2.4.STABLE4.

 Error and boundary conditions were not checked when handling
 compressed DNS answer messages in the internal DNS code (lib/rfc1035.c).
 A malicous DNS server could craft a DNS reply that causes Squid
 to exit with a SIGSEGV.

 The relevant code exists in Squid-2.3, Squid-2.4, Squid-2.5 and
 Squid-2.6/Squid-HEAD, and is enabled by default.

______________________________________________________________________

Updated Packages:

 The Squid-2.4.STABLE6 release contains fixes for all these
 problems. You can download the Squid-2.4.STABLE6 release from

   ftp://ftp.squid-cache.org/pub/squid-2/STABLE/
   http://www.squid-cache.org/Versions/v2/2.4/
```

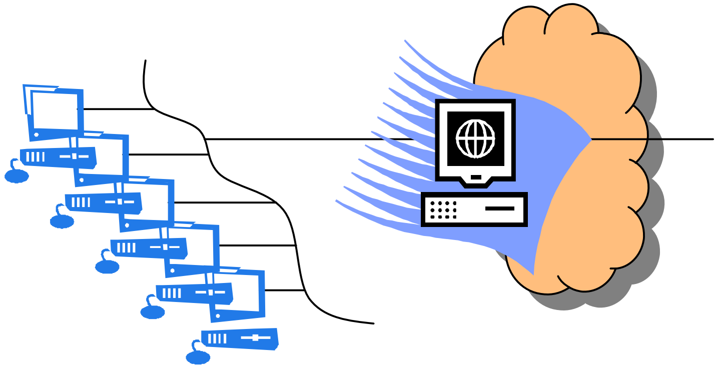
  - [http://www.cert.org should be on all administrator’s bookmarks]

Requirements

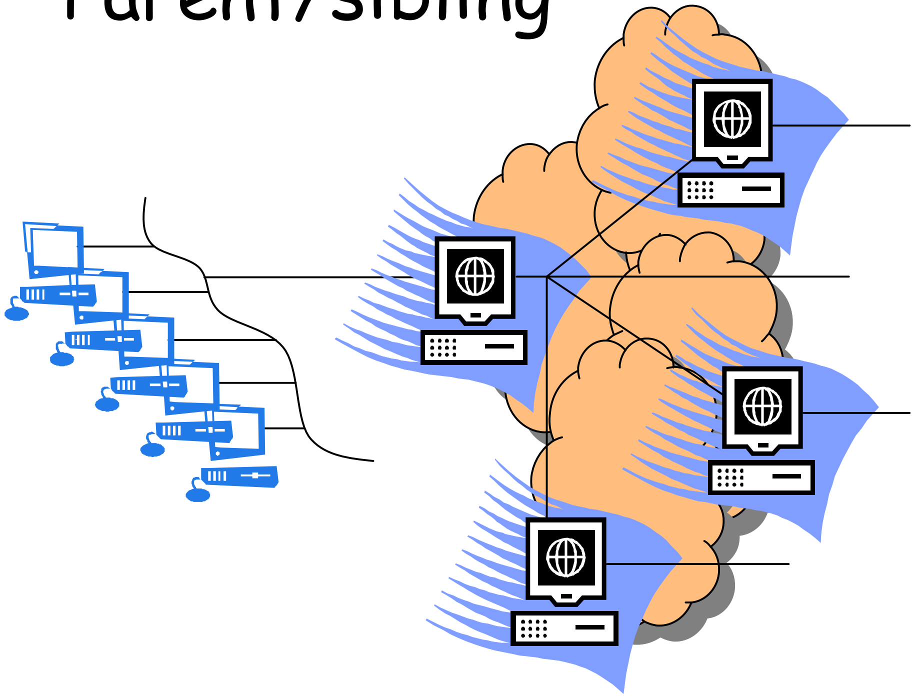
- Fairly large hardware requirements
  - "Squid can compile and run on minimal hardware, but experience shows that a stable Squid cache requires at least 128 MB of RAM and several GB of disk storage"
  - "If you plan to deploy Squid, you'll want to start with fast, robust hardware and tweak your config to get the most out of this open-source solution. Plenty of physical memory and Fast or Ultra Wide SCSI disks are highly recommended."
  - "Squid's performance once it starts swapping is abysmal, and it will drag the rest of the machine to its knees."
  - Highly recommended: *Squid Sizing for Intel Platforms* at http://wwwcache.ja.net/servers/squids.html

Cache Architectures

- Standalone

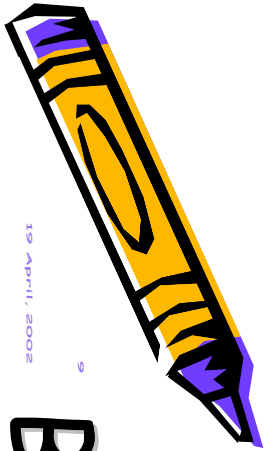


- Parent/sibling



“Much of the sophistication built into Squid, …, has to do with its ability to operate as part of a hierarchical cluster of proxy servers, capable of interrogating other instances of Squid running on other parts of the Internet to see if they can provide a copy of requested material more efficiently than the actual destination sites. But very few schools have a real use for this ability, needing only a single caching proxy on a single site. And this has, from the school sysadmin’s point of view, the advantage that nearly all of the intimidating complexity of Squid's configuration file can be ignored.”

Basic Configuration



- Configuration is via a single file
  - squid.conf
    - for Redhat in /etc/squid
      - may be in /usr/etc or /usr/local/squid or … YMMV
  - very well commented

```
#  TAG: http_port
#       Usage:  port
#               hostname:port
#               1.2.3.4:port
#
#       The socket addresses where Squid will listen for HTTP client
#       requests.  You may specify multiple socket addresses.
#       There are three forms: port alone, hostname with port, and
#       IP address with port.  If you specify a hostname or IP
#       address, then Squid binds the socket to that specific
#       address.  This replaces the old 'tcp_incoming_address'
#       option.  Most likely, you do not need to bind to a specific
#       address, so you can use the port number alone.
#
#       The default port number is 3128.
#
#       If you are running Squid in accelerator mode, then you
#       probably want to listen on port 80 also, or instead.
#
#       The -a command line option will override the *first* port
#       number listed here.   That option will NOT override an IP
#       address, however.
```

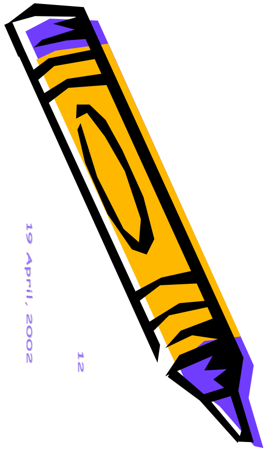
Basic Config...

- Review standard file locations
  - /etc/squid
    - squid.conf, mime.conf
  - /var/spool/squid
    - squid cache directory, subdivided into 2 further levels
  - /var/log/squid
    - access.log, store.log, cache.log
  - /var/run/squid.pid
    - squid's current process id
  - /usr/lib/squid
    - various authenticators and other programs
- Verify user/group configuration
  - squid user/group are usually already configured in /etc/{passwd,group}
- Set cache_mgr
  - so that people know who to gripe at

Basic Config…

- establish acls
  - restrict access by usernames/network, etc.
- configure refresh_patterns
  - retain (e.g.) .zip files for long periods, specify default lifetimes, etc.
- configure squid subsystems
  - dns_children
    - dnsserver performs single, blocking DNS lookups
  - specify unlinkd program
    - deletes cached files in the background
    - diskd
      - squid2.4+; performs asynchronous disk I/O
- execute squid –z to initialize the cache

Simple squid.conf



```
positive_dns_ttl 26 hours

cache_mgr bob@transentia.com.au

http_port 192.168.0.2:3128

icp_port 0
htcp_port 0

cache_dir ufs . 100 16 256

cache_access_log access.log
cache_log cache.log
cache_store_log store.log

mime_table mime.conf

pid_filename pid.txt

refresh_pattern         ^ftp:       1440        20%         10080
refresh_pattern         ^gopher:    1440        0%          1440
refresh_pattern         .           0           20%         4320

acl all src 0.0.0.0/0.0.0.0
acl manager proto cache_object
acl localhost src 127.0.0.1/255.255.255.255
acl local_net src 192.168.0.0/24
acl SSL_ports port 443 563
acl Safe_ports port 80 21 443 563 70 210 1025-65535
acl CONNECT method CONNECT

http_access allow manager localhost
http_access deny manager
http_access deny !Safe_ports
http_access deny CONNECT !SSL_ports
http_access allow localhost
http_access allow local_net
http_access deny all

icp_access allow all

miss_access allow all

icon_directory icons

error_directory errors/english
```

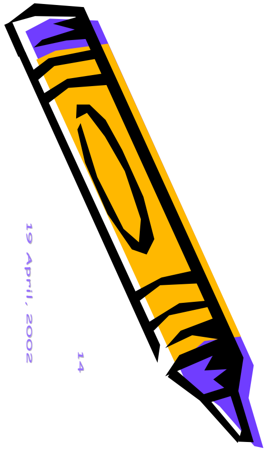
Security Config.

- If needed, squid can authenticate in various ways
  - to an NT domain, to an LDAP service, using a standalone file, etc.
    - shown below is how to use Linux's standard PAM mechanism
      - configurable at a system-wide level

```
# squid.conf
authenticate_program /usr/lib/squid/pam_auth
acl validusers proxy_auth REQUIRED
http_access allow validusers
authenticate_ttl 120 seconds
```

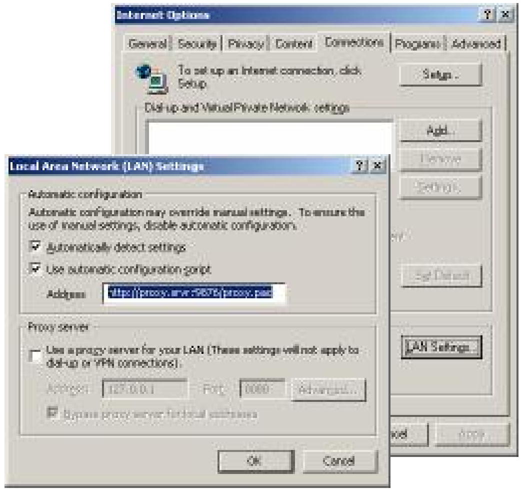
```
# /etc/pam.d/squid
#%PAM-1.0
auth required /lib/security/pam_stack.so service=system-auth
auth required /lib/security/pam_nologin.so
account required /lib/security/pam_stack.so service=system-auth
```

Browser Config.



- May need to point client's browsers at Squid
  - can use manual configuration
  - autoconfiguration also possible

```
// file: proxy.pac
function FindProxyForURL(url, host)
  {
  if (isPlainHostName(host) ||
      dnsDomainIs(host,"proxy.school.edu.au"))
    return "DIRECT";
  else if (shExpMatch(host,"*.com"))
    return "PROXY proxy.for.com:9999";
  else
    return "PROXY proxy.for.others:9999";
  }
```



Internet Options

To set up an Internet connection, click Setup.

Dial-up and Virtual Private Network settings

Local Area Network (LAN) Settings

Automatic configuration

Automatic configuration may override manual settings. To ensure the use of manual settings, disable automatic configuration.

☑ Automatically detect settings

☑ Use automatic configuration script

Address: http://proxy.srvr:9876/proxy.pac

Proxy server

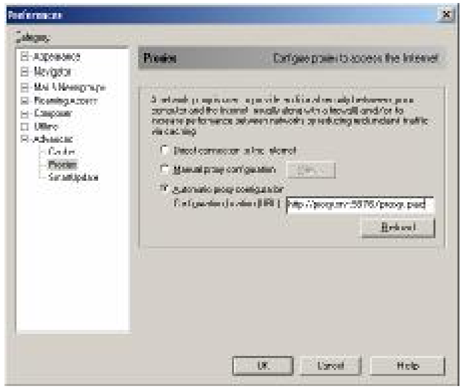
☐ Use a proxy server for your LAN (These settings will not apply to dial-up or VPN connections).

Address: 127.0.0.1 Port: 8080

☑ Bypass proxy server for local addresses

- need to tell web server to serve the file with the appropriate mime-type
  - e.g. Apache's httpd.conf file

```
AddType application/x-ns-proxy-autoconfig      pac
```



Preferences

Category

Proxies Configure proxies to access the Internet

A network proxy is used to provide additional security between your computer and the Internet (usually along with a firewall) and/or to increase performance between networks (by reducing redundant traffic via caching).

○ Direct connection to the Internet

○ Manual proxy configuration

● Automatic proxy configuration

Configuration location (URL): http://proxy.srvr:9876/proxy.pac

Transparent Proxying

- Establishes squid as the only way to the internet
  - requisite squid setup

```
http_port 3128
httpd_accel_host virtual
httpd_accel_port 80
httpd_accel_with_proxy  on
httpd_accel_uses_host_header on
httpd_accel_single_host off
```

  - set squid's Linux host as default gateway
  - forward all traffic for port 80 to squid's port

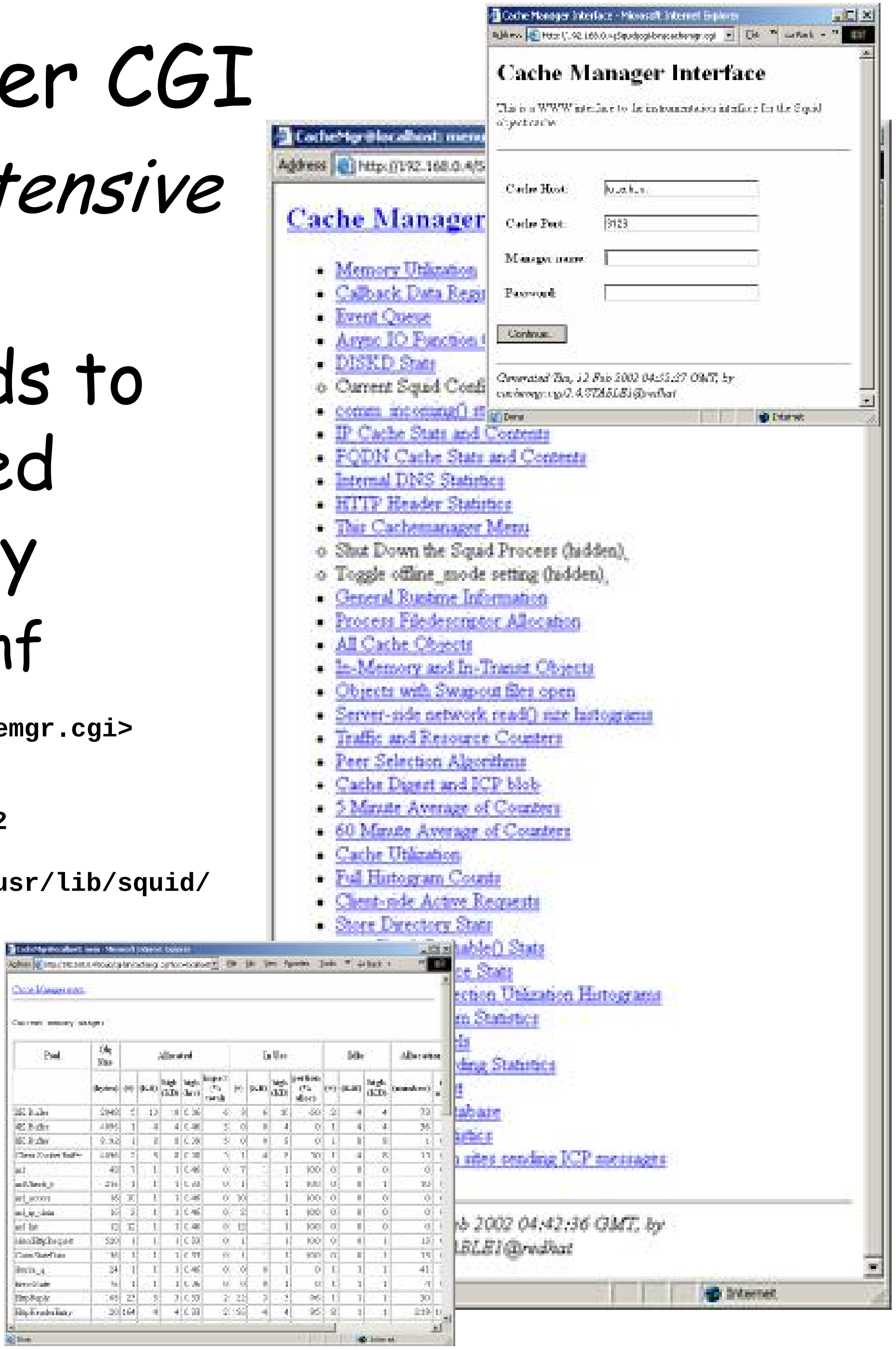
```
# iptables -t nat -A PREROUTING -i interface -p tcp --dport 80
           -j REDIRECT --to-port 3128
# iptables -A INPUT -i interface -p tcp -d your_bridge_ip -s local-network
           --dport 3128
           -m state --state NEW,ESTABLISHED -j ACCEPT
```

http://www.linuxdoc.org/HOWTO/mini/TransparentProxy.html

Monitoring

- Cache manager CGI
  - provides *extensive* statistics
  - Apache needs to be configured appropriately
    - in httpd.conf

```
<Location /usr/lib/squid/cachemgr.cgi>
        order deny,allow
        deny from all
        allow from 192.168.0.2
</Location>
ScriptAlias /Squid/cgi-bin/ /usr/lib/squid/
```



Cache Manager Interface

This is a WWW interface to the instrumentation interface for the Squid object cache.

Cache Host:	localhost
Cache Port:	3128
Manager name:	
Password:	

Continue...

Generated Tue, 12 Feb 2002 04:51:27 GMT, by cachemgr.cgi/2.4.STABLE1@redhat

Cache Manager

- Memory Utilization
- Callback Data Registry
- Event Queue
- Async IO Function Counters
- DISKD Stats
- Current Squid Configuration
- comm_incoming() stats
- IP Cache Stats and Contents
- FQDN Cache Stats and Contents
- Internal DNS Statistics
- HTTP Header Statistics
- This Cachemanager Menu
- Shut Down the Squid Process (hidden),
- Toggle offline_mode setting (hidden),
- General Runtime Information
- Process Filedescriptor Allocation
- All Cache Objects
- In-Memory and In-Transit Objects
- Objects with Swapout files open
- Server-side network read() size histograms
- Traffic and Resource Counters
- Peer Selection Algorithms
- Cache Digest and ICP blob
- 5 Minute Average of Counters
- 60 Minute Average of Counters
- Cache Utilization
- Full Histogram Counts
- Client-side Active Requests
- Store Directory Stats
- ...able() Stats
- ...ce Stats
- ...ection Utilization Histograms
- ...n Statistics
- ...ls
- ...ding Statistics
- ...
- ...tabase
- ...istics
- ...n sites pending ICP messages

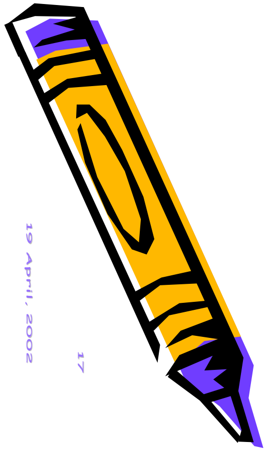
...b 2002 04:42:36 GMT, by ...BLE1@redhat

Cache Manager menu

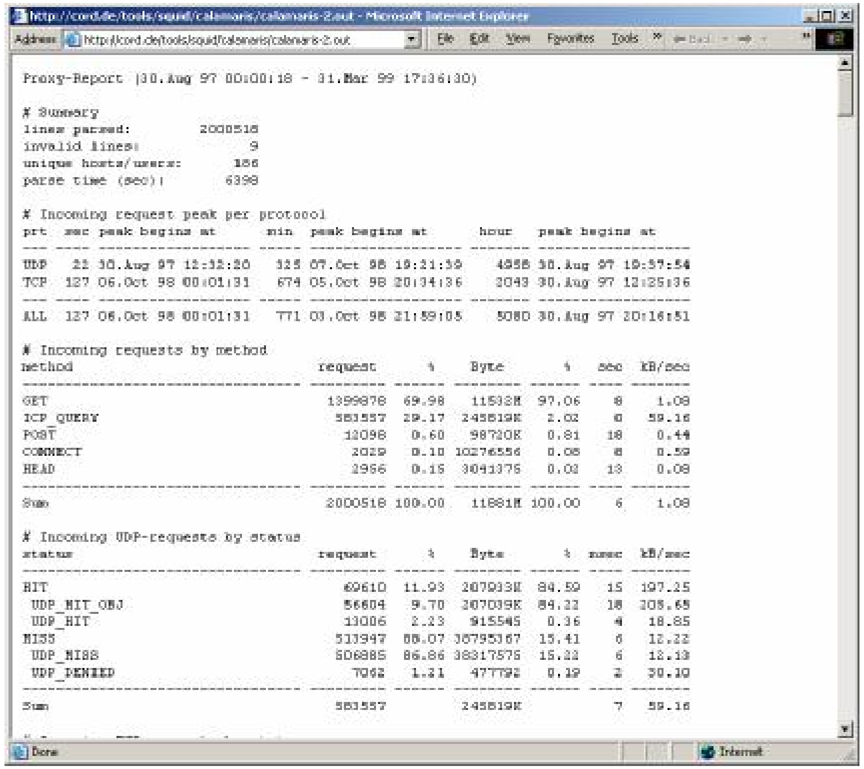
Current memory usage:

[illegible]

Calamaris



- Parses logfiles from Squid, NetCache, Inktomi Traffic Server, Oops! proxy server, Novell Internet Caching System, Compaq Tasksmart or Netscape/iplanet Web Proxy Server and generates a report.
  - Written in perl5.

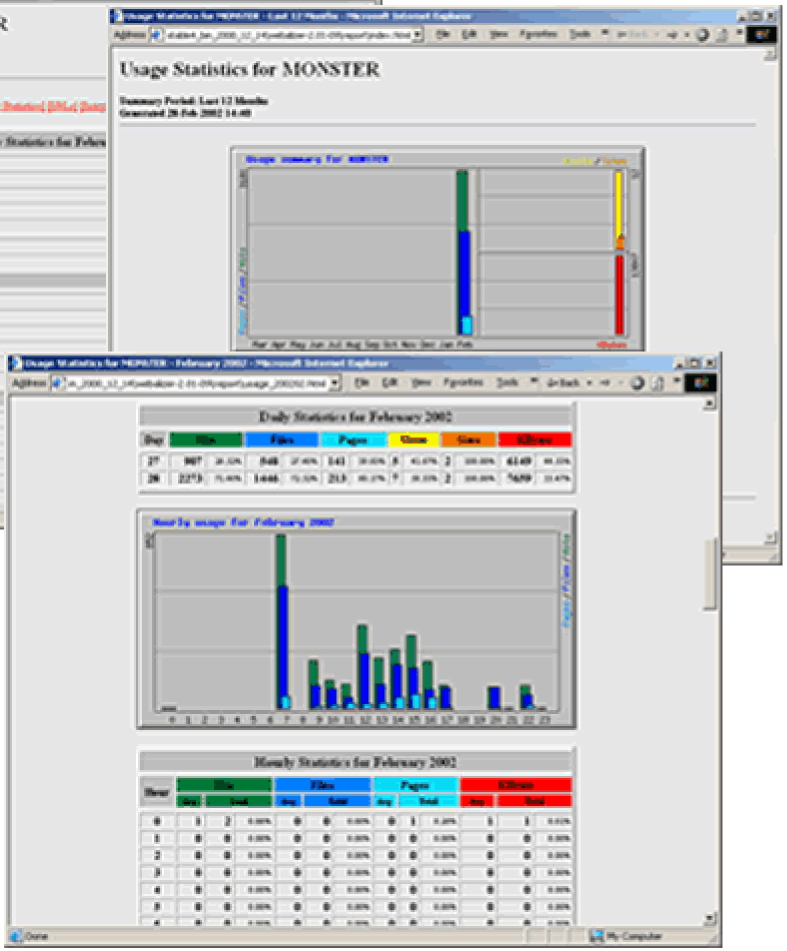


http://cord.de/tools/squid/calamaris/calamaris-2.out - Microsoft Internet Explorer
Address
http://cord.de/tools/squid/calamaris/calamaris-2.out
File
Edit
View
Favorites
Tools
Back
Done
Internet

```
Proxy-Report (30.Aug 97 00:00:18 - 31.Mar 99 17:36:30)

# Summary
lines parsed:          2000518
invalid lines:               9
unique hosts/users:        186
parse time (sec):         6398

# Incoming request peak per protocol
prt  sec peak begins at      min  peak begins at       hour  peak begins at
--- ---- ------------------ ----- ------------------ ------- ------------------
UDP   22 30.Aug 97 12:32:20   325 07.Oct 98 19:21:39    4958 30.Aug 97 19:57:54
TCP  127 06.Oct 98 00:01:31   674 05.Oct 98 20:34:36    2043 30.Aug 97 12:25:36
--- ---- ------------------ ----- ------------------ ------- ------------------
ALL  127 06.Oct 98 00:01:31   771 03.Oct 98 21:59:05    5080 30.Aug 97 20:16:51

# Incoming requests by method
method                              request      %     Byte      %  sec  kB/sec
----------------------------------- -------- ------ -------- ------ ---- -------
GET                                  1399878  69.98   11532M  97.06    8    1.08
ICP_QUERY                             583557  29.17  245819K   2.02    0   59.16
POST                                   12098   0.60   98720K   0.81   18    0.44
CONNECT                                 2029   0.10 10276556   0.08    8    0.59
HEAD                                    2956   0.15  3041375   0.02   13    0.08
----------------------------------- -------- ------ -------- ------ ---- -------
Sum                                  2000518 100.00   11881M 100.00    6    1.08

# Incoming UDP-requests by status
status                              request      %     Byte      % msec  kB/sec
----------------------------------- -------- ------ -------- ------ ---- -------
HIT                                    69610  11.93  207933K  84.59   15  197.25
 UDP_HIT_OBJ                           56604   9.70  207039K  84.22   18  205.65
 UDP_HIT                               13006   2.23   915545   0.36    4   18.85
MISS                                  513947  88.07 38795367  15.41    6   12.22
 UDP_MISS                             506885  86.86 38317575  15.22    6   12.13
 UDP_DENIED                             7062   1.21   477792   0.19    2   30.10
----------------------------------- -------- ------ -------- ------ ---- -------
Sum                                   583557         245819K              7   59.16
```

Others

- webalizer
- squid log analyzer



Usage Statistics for MONSTER

Summary Period: February 2002
Generated 28-Feb-2002 14:49

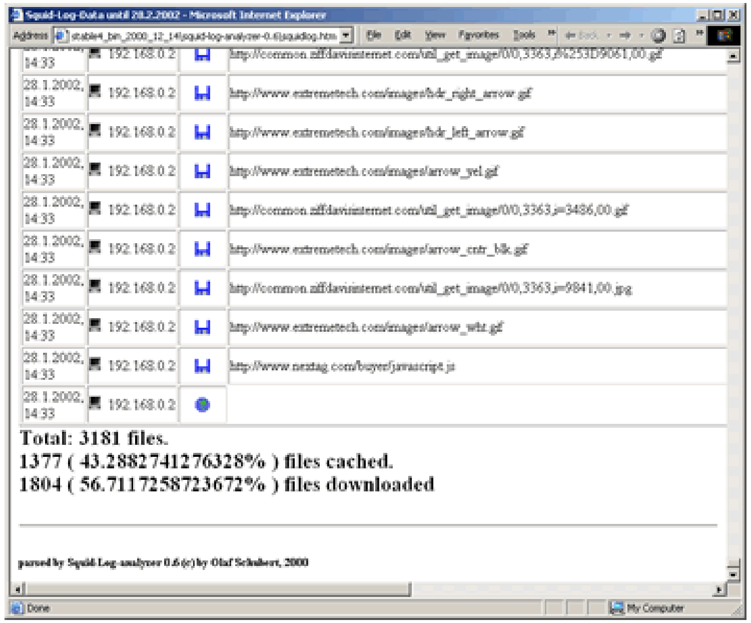
Usage Statistics for MONSTER

Summary Period: Last 12 Months
Generated 28-Feb-2002 14:49

Daily Statistics for February 2002

Hourly usage for February 2002

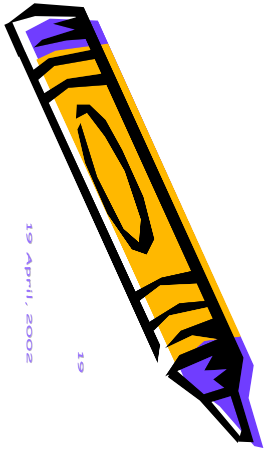
Hourly Statistics for February 2002



```
14:33        192.168.0.2  http://common.ziffdavisinternet.com/util_get_image/0/0,3363,i%253D9061,00.gif
28.1.2002,   192.168.0.2  http://www.extremetech.com/images/hdr_right_arrow.gif
14:33
28.1.2002,   192.168.0.2  http://www.extremetech.com/images/hdr_left_arrow.gif
14:33
28.1.2002,   192.168.0.2  http://www.extremetech.com/images/arrow_yel.gif
14:33
28.1.2002,   192.168.0.2  http://common.ziffdavisinternet.com/util_get_image/0/0,3363,i=3486,00.gif
14:33
28.1.2002,   192.168.0.2  http://www.extremetech.com/images/arrow_cntr_blk.gif
14:33
28.1.2002,   192.168.0.2  http://common.ziffdavisinternet.com/util_get_image/0/0,3363,i=9841,00.jpg
14:33
28.1.2002,   192.168.0.2  http://www.extremetech.com/images/arrow_wht.gif
14:33
28.1.2002,   192.168.0.2  http://www.nextag.com/buyer/javascript.js
14:33
28.1.2002,   192.168.0.2
14:33
```

Total: 3181 files.
1377 (43.2882741276328%) files cached.
1804 (56.7117258723672%) files downloaded

parsed by Squid-Log-analyzer 0.6 (c) by Olaf Schubert, 2000



Administration

- squid -k
  - send signal to running copy and exit
    - vital for maintaining a 24x7 service
  - various messages:
    - reconfigure
      - reread squid.conf
    - rotate
      - useful for setting up daily logfiles
    - shutdown
    - interrupt
    - kill
    - debug
    - check
- set up /etc/rc*n*.d entry
  - to ensure that squid runs at startup

Performance Tips

- Choose right filesystem
  - anecdotal evidence suggests that reiserfs performs better than the efs2/3 used in most Linux distributions

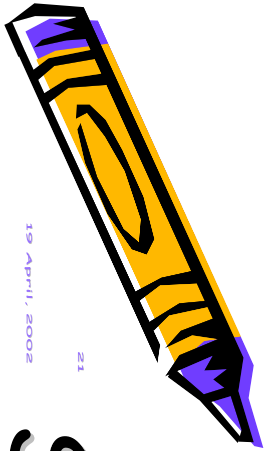
"We spent less time installing Squid than we'd expected, but far more time tuning Squid than we'd planned."

- Memory
  - simple formula to predict squid process' memory requirements given available disk space

$$\left(\frac{disk\_space}{13000}\times 130\right)+cache\_mem+2.5E+6$$

    - gives ~11.5Mb for a 100Mb disk cache
- Use Web Polygraph for load testing
  - realistic traffic generation and content simulation
  - ready-to-use standard workloads
- Preloading with wget
  - squid cannot pre-cache the internet but it is a simple job to script access to a predefined set of "interesting places."

Some Dos & Don'ts



- DO
  - put a name server on the machine with Squid. It's an extra level of caching, and minimizes choke points.
  - aim to have 20-30 dnsservers
  - increase the size of your fqdncache and ipcache. Bigger is better. Stale addresses are less important than many entries and long TTLs. Cache addresses for at least 24 hours, and negative cache for at least 5 minutes.
  - split your cache over several physical drives. Four 20-Gbyte drives are better than one 80-Gbyte drive—you save time using multiple spindles.
  - keep your logs on a non-cache drive, and preferably on a different chain or controller.
- DON'T
  - cache big objects. Next to CPU and RAM, disk I/O is your biggest bottleneck. Try not to cache anything over about a megabyte.
  - put two cache drives on the same IDE controller
  - use ICP if you have a single parent you always use.

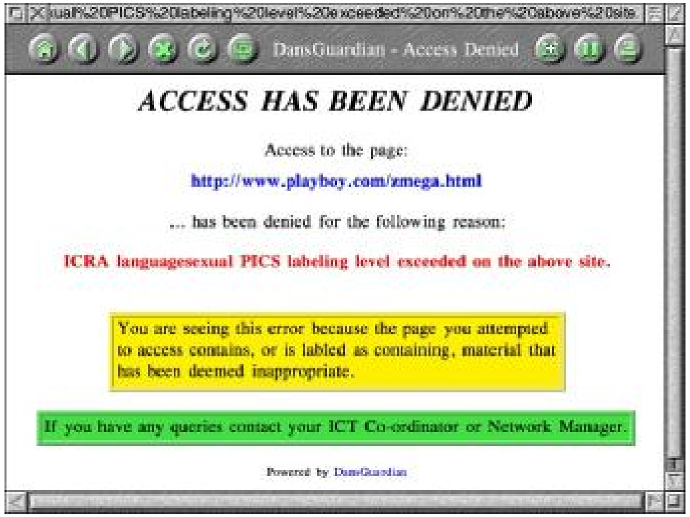
squidGuard

- a free, flexible and efficient filter and redirector program for squid
  - looks for URL patterns
  - configured via a number of databases specified by squidGuard.conf
    - easy to update
      - create local ‘diff’ file and merge into main database

```
dbhome /usr/local/squidGuard/db
logdir /usr/local/squidGuard/log

dest gambling{
            log             gambling
            domainlist      gambling/domains
            urllist         gambling/urls
            redirect        https://www.centrebet.com.au/english/en_resbet.html
            }

dest warez{
            log             warez
            domainlist      warez/domains
            urllist         warez/urls
            redirect        http://www.bsaa.com.au/pirates
            }

acl {
    default {
            pass !gambling !warez all
            redirect http://www.yahooligans.com/
            }
    }
```

DansGuardian

- Web content filter that works with squid
  - effectively an augmented squidGuard
  - filters using multiple methods:
    - URL and domain filtering, content phrase filtering, PICS filtering, MIME filtering, file extension filtering, POST limiting.
    - able to handle huge filter lists
      - regularly updated



xual%20PICS%20labeling%20level%20exceeded%20on%20the%20above%20site.

ACCESS HAS BEEN DENIED

Access to the page:

http://www.playboy.com/zmega.html

... has been denied for the following reason:

ICRA languagesexual PICS labeling level exceeded on the above site.

You are seeing this error because the page you attempted to access contains, or is labled as containing, material that has been deemed inappropriate.

If you have any queries contact your ICT Co-ordinator or Network Manager.

Powered by DansGuardian

  - can check for pages that contain profanity and phrases often associated with pornography and other undesirable content
  - allows you to block or limit web upload
  - "The default settings are geared towards what a primay [sic] school might want but DansGuardian puts you in control of what you want to block."
  - works with webmin to allow remote administration

Squid on Windows

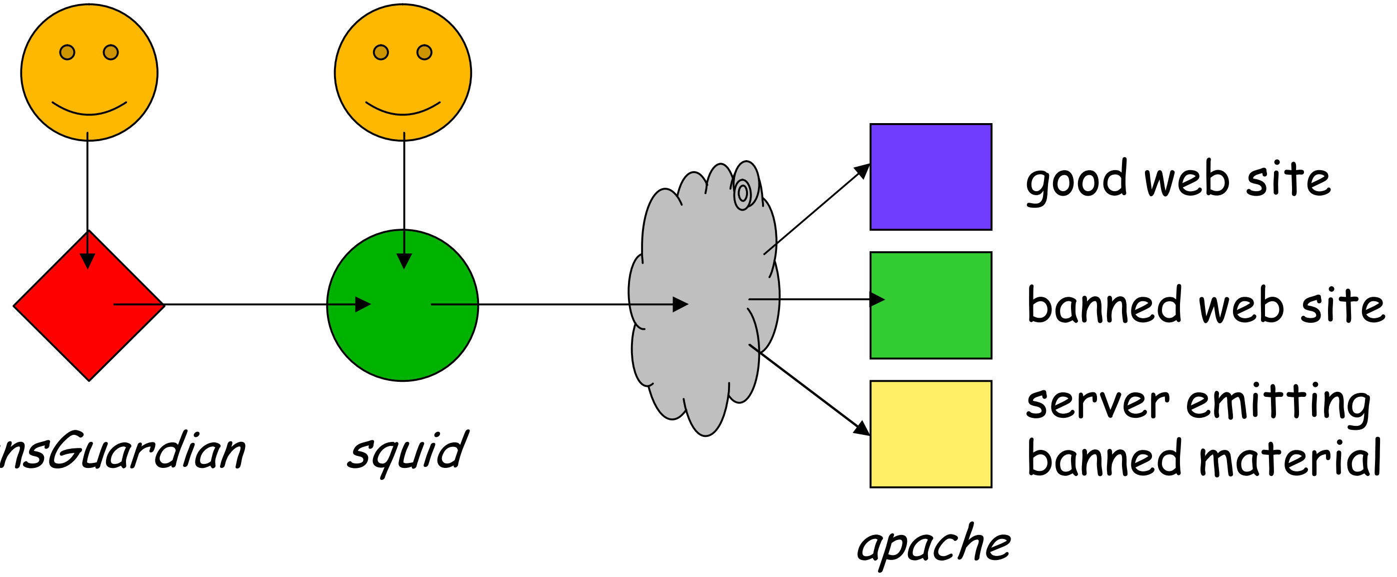
- Basically two Open-Source alternatives
  - Both a little limited
    - neither properly supports authentication, for example
    - “work in progress” but still useful
      - also guiding the ‘real’ development activities
  - Cygwin
    - actually a whole (very useful!) suite of GNU utilities and POSIX operating environment
      - http://sources.redhat.com/cygwin/
  - SquidNT
    - I use this one on win2k for my home network
      - http://www.serassio.it/SquidNT.html

Links/Resources

- Linux in Schools pages
  - http://linux.lexilog.org.uk/squid.html
- Installing and Configuring Squid
  - http://linux.oreillynet.com/lpt/a//linux/2001/07/26/squid.html
- The Squid FAQ
  - http://www.squid-cache.org/Doc/FAQ/FAQ-1.html
- Do-It-Yourself Caching: Squid 2.3
  - http://www.bsdtoday.com/2000/February/Tutorials28.html
- Open Source Filtering
  - http://opensourceschools.org/article.php?story=20011251 82207522
- Web Polygraph
  - http://polygraph.ircache.net/
- Squid security
  - http://www1.securityfocus.com/focus/linux/articles/squid.html
- squidGuard
  - http://ftp.ost.eltele.no/pub/www/proxy
- dansGuardian
  - http://dansguardian.org

Demo Time!

- Apache
- Squid
- DansGuardian



DansGuardian

squid

good web site

banned web site

server emitting banned material

apache