PRO-COW: Protocol Compliance on the Web

BALACHANDER KRISHNAMURTHY | MARTIN ARLITT
bala@research.att.com | **arlitt@hpl.hp.com**
AT&T Labs - Research | HP Labs

State of the Web

- HTTP is the dominant protocol (75% of backbone traffic)

- HTTP/0.9, HTTP/1.0 versions never formally standardized

- 4+ years spent developing HTTP/1.1 with many requirements on clients, proxies, and servers

- Several intermediate implementations

- As of June '99 HTTP/1.1 at draft standard

- Lots of servers claiming to be HTTP/1.1 compliant

Motivations for PROCOW study

- Measure of protocol adoption, repeat to get rate

- Web site admins can see if they should run HTTP/1.1

- Learn why people might be turning off some HTTP/1.1 features

- Protocol designers can see if all the hot air expended in endless discussions in WG actually led anywhere

- Help quantify benefits of protocol changes

Study: Methodology

- Requests from few client sites around world to hundreds of popular origin servers

- Popularity gleaned from many sources: MediaMetrix, Netcraft, Hot100, Fortune500, Global200

- 517 server sites selected based on popularity of $request$ traffic (not on response size)

- Not enough pornographic sites included (self-censorship by site raters?)

- 6 client sites (.au, .cl, .fr, uky.edu, nj.att.com, ca.hp.com)

Study: What

3 categories of tests:

1. Some of the MUST features of HTTP/1.1
   GET, HEAD, `Host` header

2. Features that are important additions to HTTP/1.1
   Persistent connections, pipelining, range requests

3. Non-mandatory features deemed useful

   OPTIONS, TRACE, POST
   Expect/100-Continue
   If-None-Match, If-Unmodified-Since...

Top server vendors seen in our test

Server vendor	Percentage
Netscape	34.8
Microsoft	32.8
Apache	28.2
Lotus	2.7
Zeus	0.4
Oracle	0.2
Others	0.8

Note: Apache has around 61% of $total$ server market share.

Category 1: Unconditional Compliance Results

Client Site	GET(%)	HEAD(%)	Host(%)	Pass All(%)	Fail All(%)
AT&T	82.1	72.4	64.6	59.8	7.4
Australia	82.3	72.7	64.4	60.0	7.3
Chile	82.3	70.3	64.4	60.3	7.9
France	82.4	72.4	64.1	59.7	7.4
HPL	83.5	72.9	64.5	60.6	7.1
Kentucky	82.4	72.7	64.2	60.1	7.5

Location didn't matter - minor differences due to load balancing front-ends

$7+$% failure rate of all tests - bad!

Breakdown of Category 1 Test Results (CA-HPL)

	GET (%)	HEAD (%)	Host(%)
Unconditionally compliant	83.5	72.9	64.5
Conditionally compliant	16.1	9.4	28.6
Not compliant	0.4	17.7	6.9

Conditional compliance: headers like `Content-Length`,
`Transfer-Encoding:chunked` are absent.

Failure in `HEAD`: headers in response different than GET
17% either didn't return expected metainformation, or returned message
body as well.

Absence of `Host:` *header*: 6.9% of servers accepted such 1.1 client
requests. Bad. Very bad. This is a MUST.

Category 2 Unconditional Compliance (CA-HPL)

Server	Persistence(%)	Pipelining(%)	Range(%)	Pass All(%)	Fail All(%)
Apache/1.3	87.0	87.0	51.1	47.8	9.8
Apache/1.2	89.1	89.1	52.7	43.5	10.9
IIS/4.0	87.9	87.3	52.4	52.4	12.7
Netscape/3.5	41.1	38.4	67.2	37.5	30.6
Netscape/3.6	41.5	35.4	47.7	35.4	52.3

Not very encouraging considering major improvements in HTTP/1.1 are correctly implemented in less than half of tested servers.

Note Netscape/3.6 is worse than Netscape/3.5.

Category 3 Unconditional Compliance (CA-HPL)

Feature	% Servers Unconditionally Compliant	% Servers Not Compliant
OPTIONS	59.8	0.8
TRACE	97.3	0.2
FOO	54.7	7.1
POST, Expect	63.2	32.0
Incorrect URL	80.5	7.1
Long URL	62.7	2.0
If-None-Match	14.8	0.8
If-Unmodified-Since (1123)	41.7	57.1
If-Unmodified-Since (1036)	41.7	57.1
If-Unmodified-Since (ANSI C)	41.7	57.1

`If-Unmodified-Since` with `Date` in RFC 1123/1026/ANSI-C formats.

Responding to `FOO` method is violation of SHOULD: such a method might be introduced!

Security, DOS, and other problems

- Some servers melt instead of sending 414 Request-URI Too Large (maybe SHOULD should become a MUST?)

- Devices terminating a HTTP/TCP connection (e.g., L7 switch) should identify themselves (i.e., MUST add Via); and undergo HTTP compliance testing.

- Servers should fully identify version numbers/configuration (IIS)

Reasons for non-compliance + speculations

- Subtle: reasons not always known to implementors (One lone MS-IIS/4.0 failing `Host` test, uses ISAPI dll filter)

- "Intelligent" switches/load balancers transparently terminate connections may not support persistent connections though server can.
  (`Server:` field was different in responses from the same IP address)

- Since these devices don't identify themselves it looks like server is misbehaving. Anecdotal evidence that switch vendors don't support persistent connections.

- Turning off features (persistent connections/pipelining or range requests): performance concerns? Hallway conversations?

Conclusion of study

- Many sites are moving to HTTP/1.1 but not necessarily in a compliant way

- Maybe some SHOULDs in 2616 should change to MUST – most implementors pay attention to the MUSTs

- Maybe spec should state requirements for L7 switches

- Lots of 0.9, 1.0 proxies in path (some implementing selective HTTP/1.1 features!)

- Measurement is not aided by protocol – keep in mind for future?

- End to end 1.1 compliant traffic: RSN

What happened after study

- Threat of lawsuits, nastygrams...

- Fix of DOS attack in major server

- Persistent connection now default in major browser

- Paper being submitted to WWW-9, test of proxies next

- Updated results: (done Monday 11/8) similar results (+-3%)

- Thanks to client sites who let us do the study!

Bibliography

- RFC 2616 HTTP/1.1 draft standard

- RFC 1945 HTTP/1.0 Informational RFC (best current practice)

- Differences between HTTP/1.0 and HTTP/1.1
  Krishnamurthy, Mogul, Kristol (WWW-8, June '99)

- PRO-COW paper (being submitted to WWW-9)

- Predicting HTTP/1.1 from HTTP/1.0 traffic (Global Internet '99)

Papers available from **www.research.att.com/~bala/papers**