Better End User Visible Web Browsing Performance

Intel Developer Forum Spring 2002

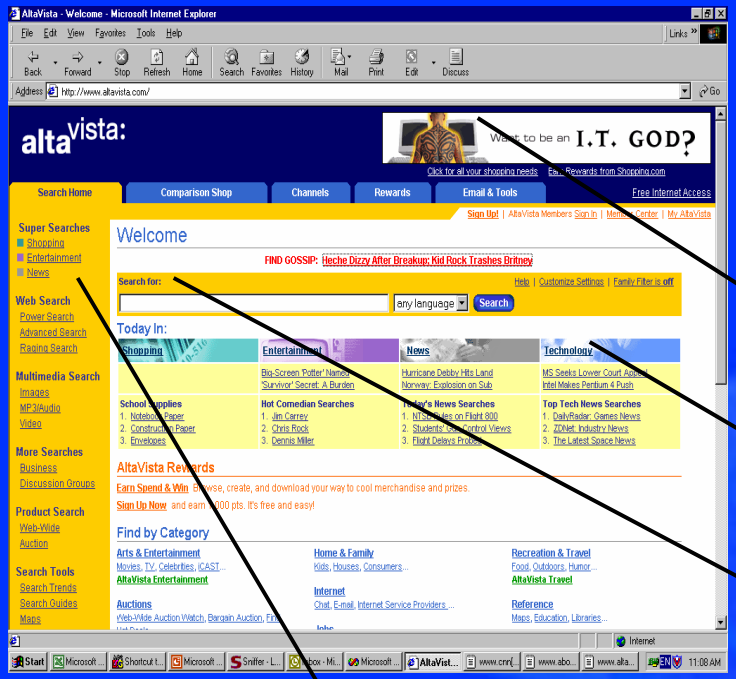
Jennie Yoder
Software Engineer
I/O Architecture
Intel Corporation
February 27, 2002

Agenda

- How we measure Web performance
- Browser performance issues
- Server performance issues
- Web page performance issues

Web Page Download Timeline

- Multiple objects fetched per page
  - Yahoo main page - 5 objects
  - CNN main page - 50 objects
- Single object fetches consist of connect, get, and send times



server

Send page data (html, uncompressed)

Close connection

Send image 1

Close connection

Send image 2

Close connection

Send image n

Close connection

Connect Time

GET latency

Send Time

Connect to server

Request web page (GET)

Render html; extract image links

Connect to image server (doubleclick)

Request image 1

Connect to image server (altivista.com)

Request image 2

Connect to image server (akamai.com)

Request image n

Render image data for each image as it comes in

client

NAPA Charts Answer Performance Questions

- NAPA = Network Application Performance Analyzer
- Where does the time go?
- What limits performance?
- What could be done to improve performance?

NAPA Demo

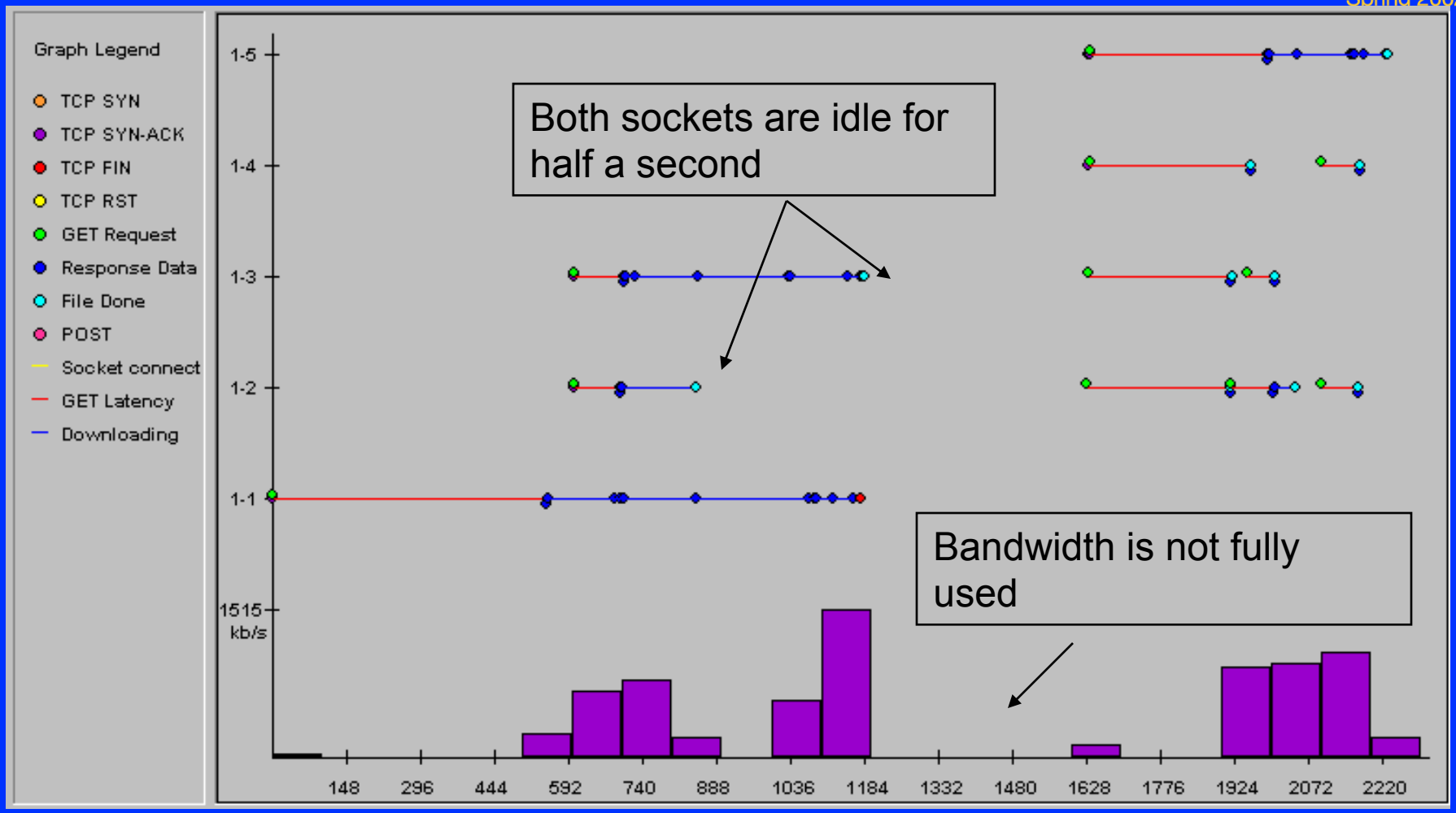
Jim Chu
Senior SW Engineer
I/O Architecture
Intel Corporation

February 27, 2002

Browser Performance:

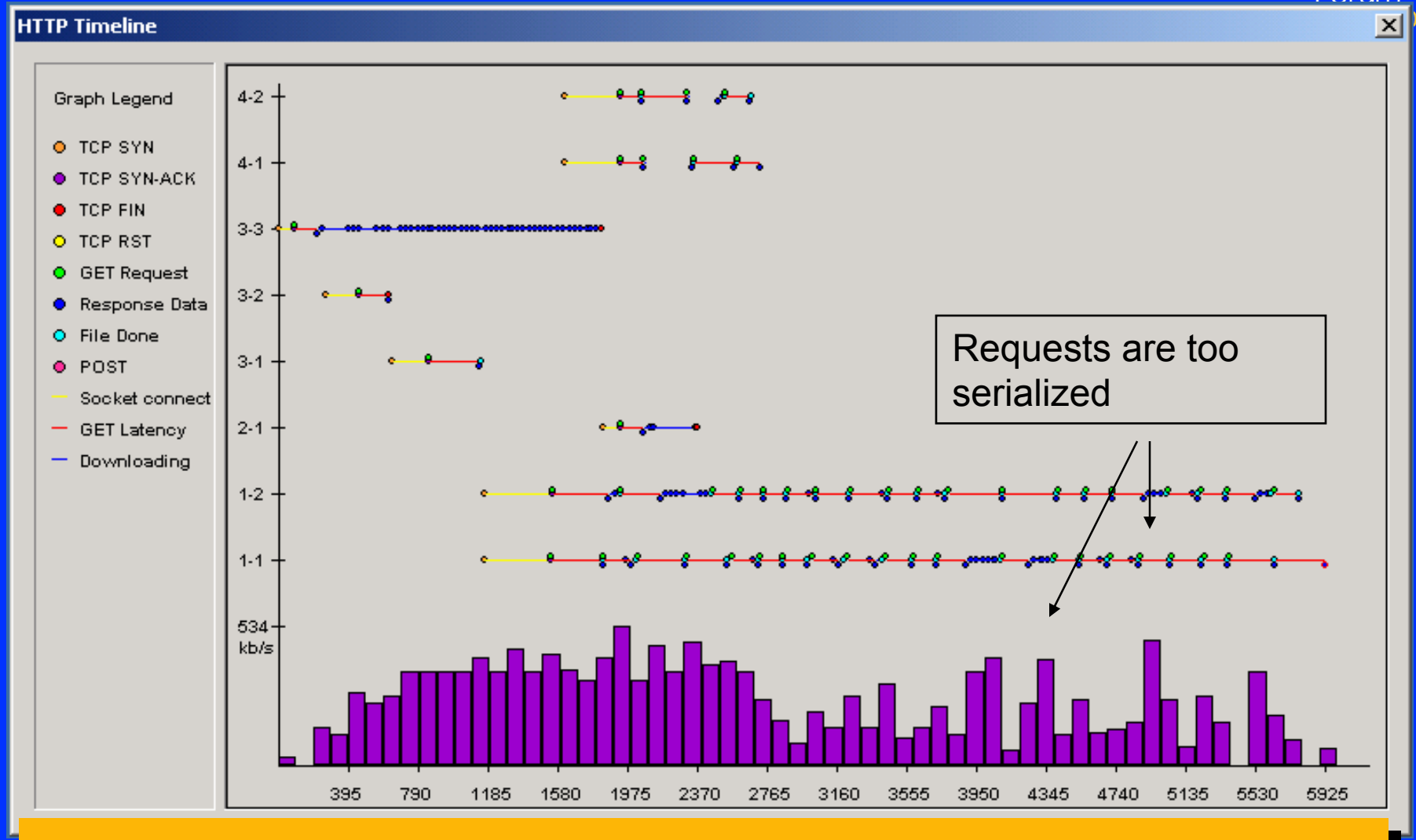
How three leading browsers use the network

Browser C: LAN



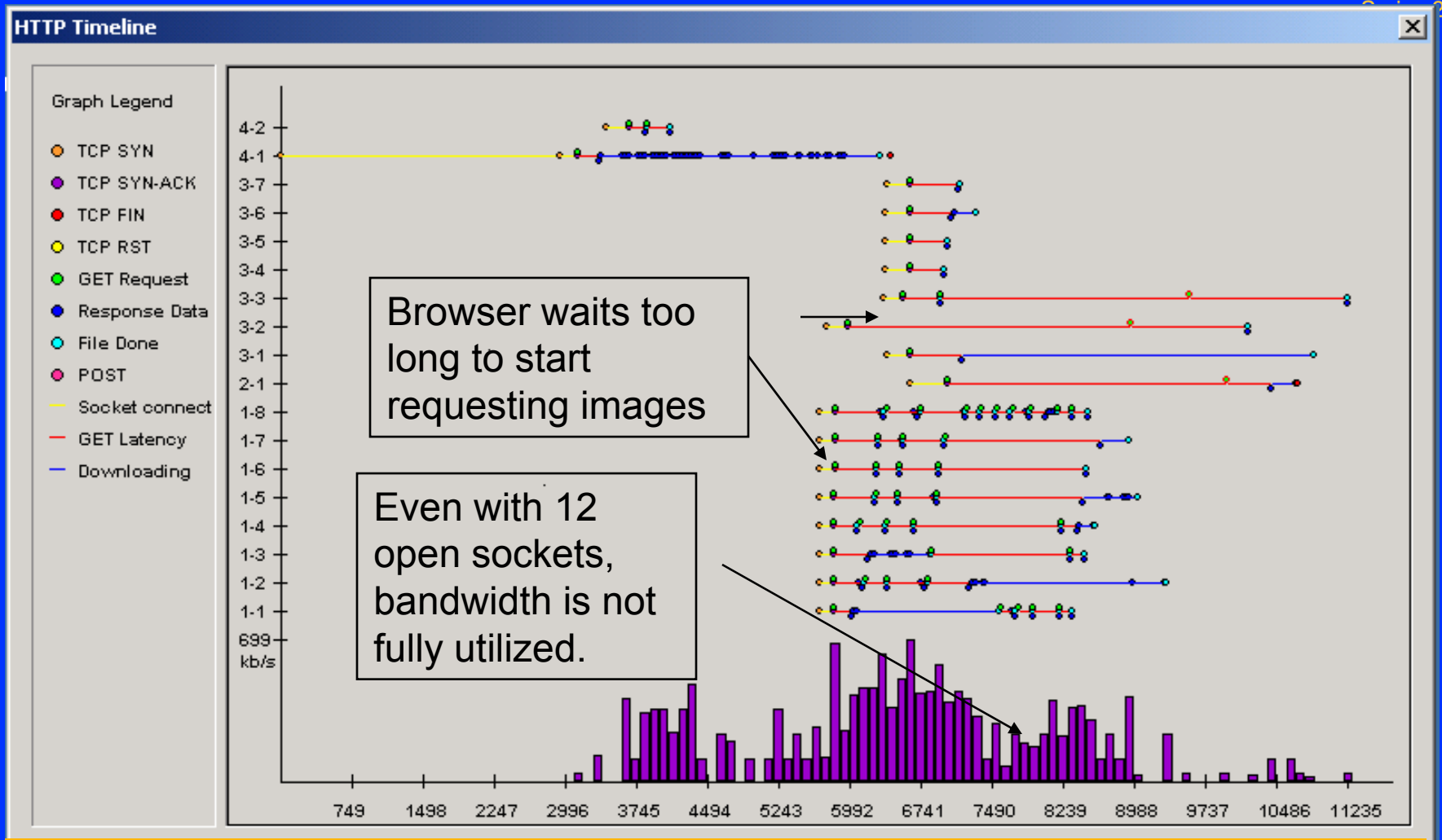
Send data requests as soon as possible

Browser C: DSL



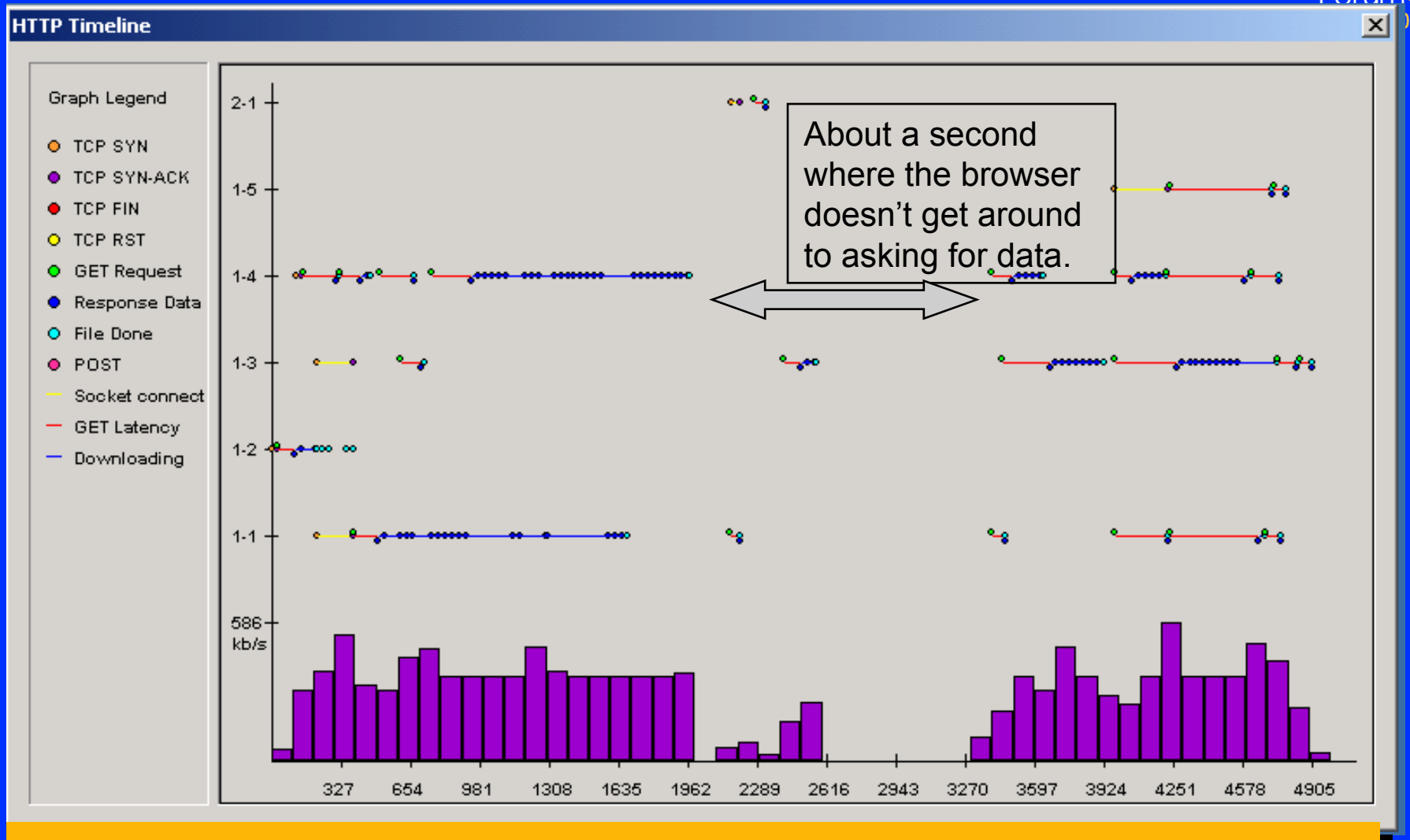
Pipeline GET requests

Browser B: DSL



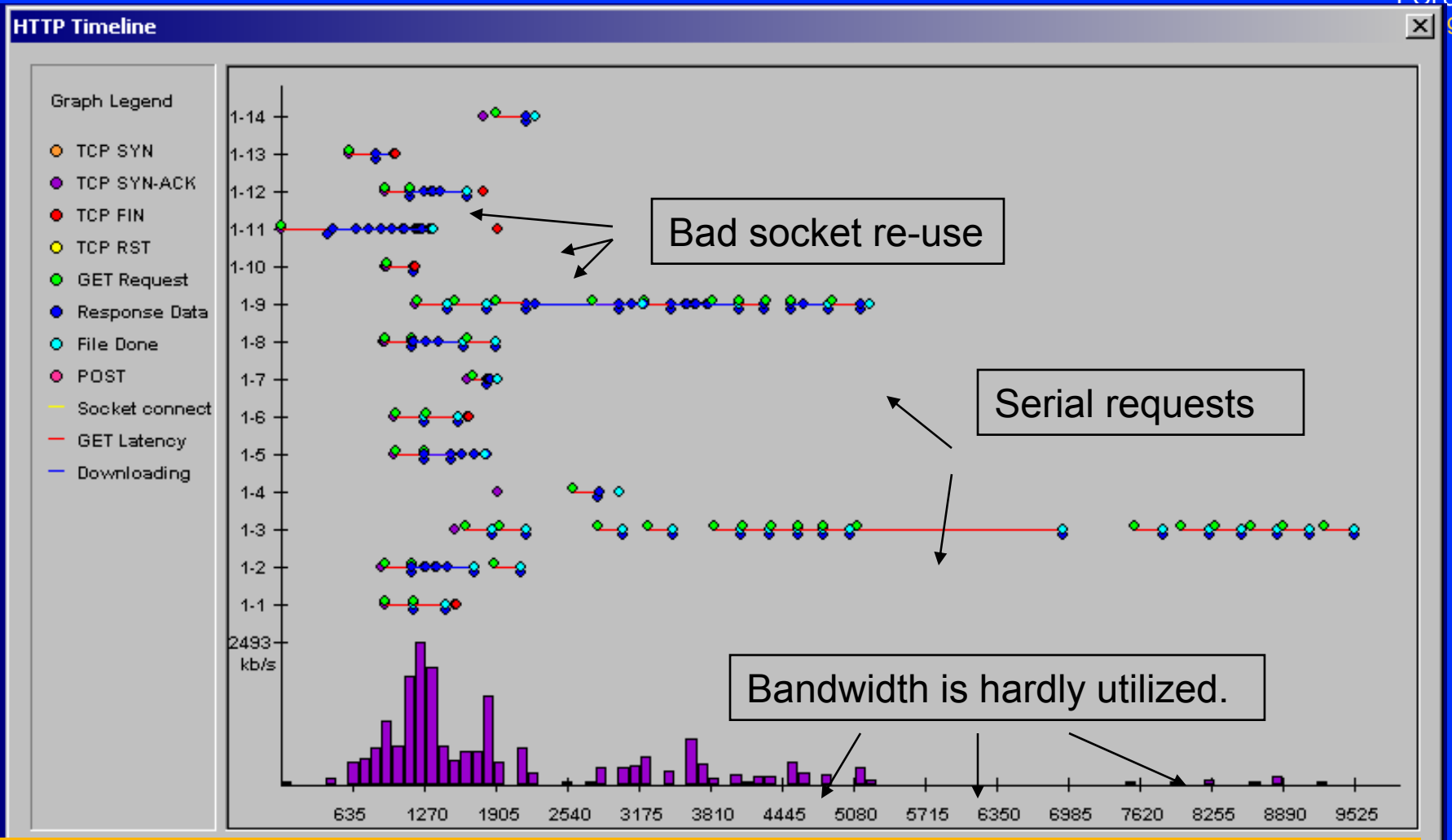
Send data requests as soon as possible

Browser A - DSL



Send data requests as soon as possible

Browser A: LAN



If you have open sockets, use them

To Speed up Browsers:

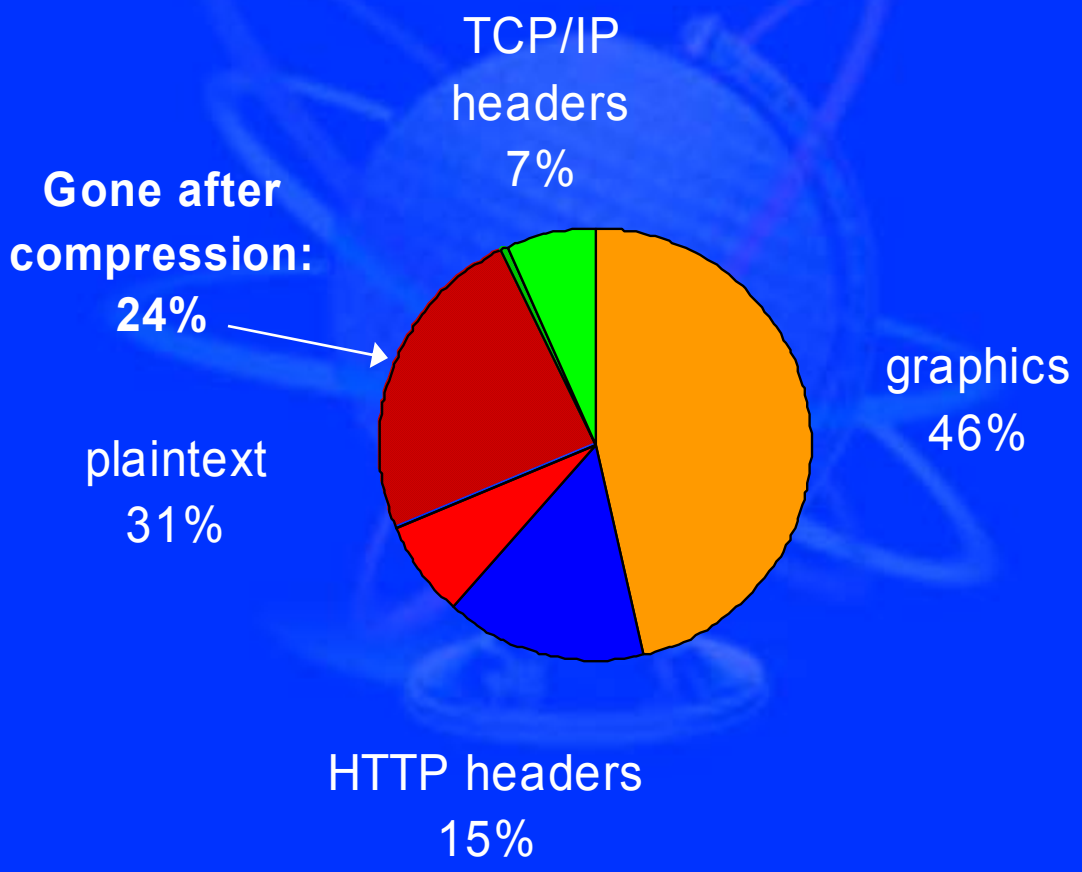
- Send data requests as soon as possible
- Pipeline requests
- If you have open sockets, use them.

Browsers could be 25% faster on average

Server Performance:

Do they really need to send all those bytes?

What data is sent over the network?

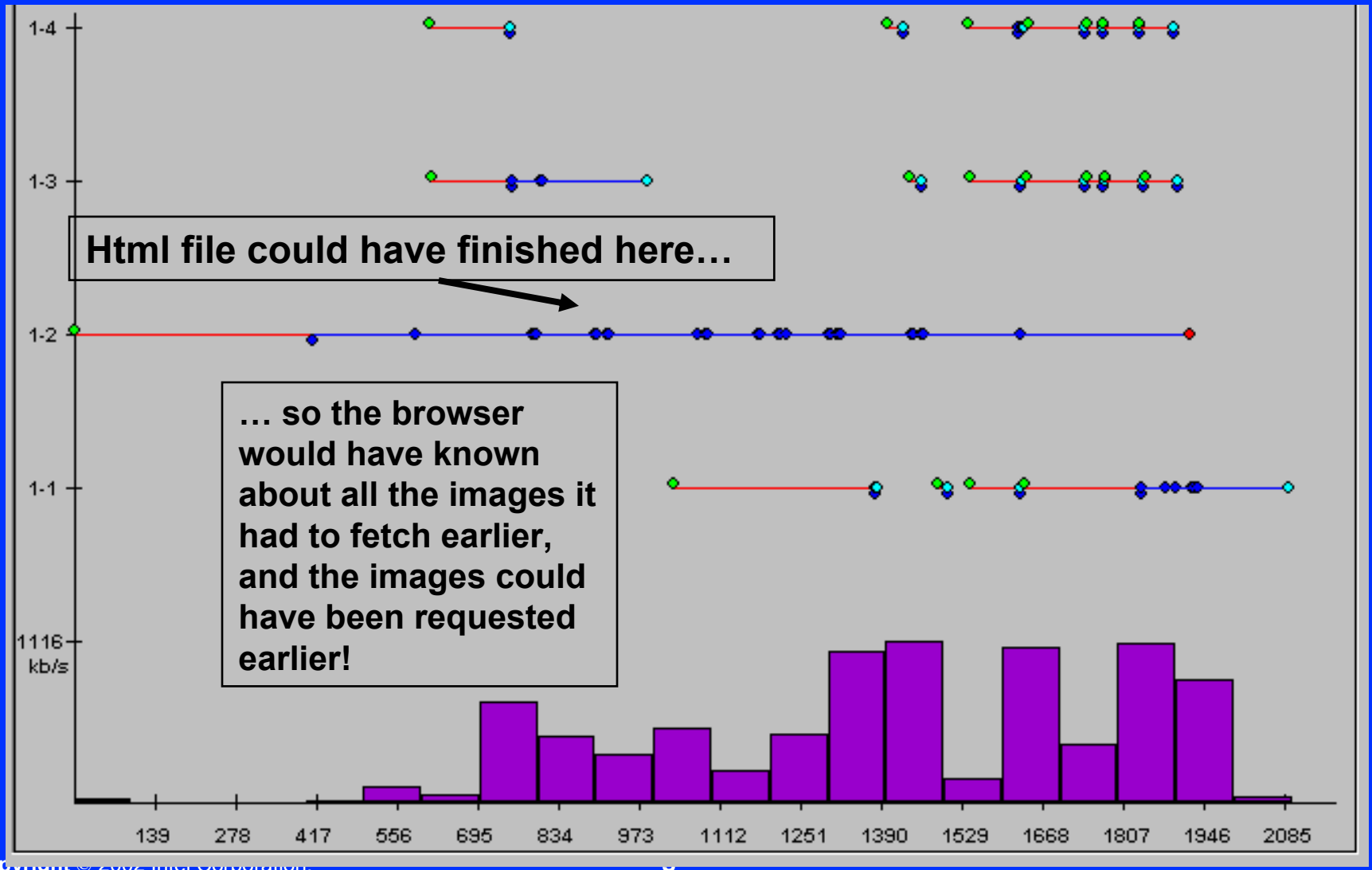


HTML compression

- HTTP traffic reduced 24%
  - Faster download time on slower links
    - 14% at 56k Modem rate
    - 4% at DSL rate
  - Lower network charges for web sites

Compression saves time and money

Even on a high-bandwidth connection, compression can save end user time



Downsides to compression?

- Is the client slower?
  - No
- Is the server slower?
  - Static pages: No
  - Dynamic pages: possibly
- Why don't people turn it on?
  - Server default is compression off

HTTP headers are too big

- HTTP Headers can't be compressed
  - Each file fetched has ~800 bytes of HTTP headers attached

- Browsers and servers pass redundant or unnecessary information

  Example: "Cache-Control: no-cache
  Pragma: no-cache
  Cache-Control: no-cache
  Expires: Thu, 07 Jun 2001 17:48:48 GMT
  Date: Thu, 07 Jun 2001 17:48:48 GMT"

Keep HTTP headers small

To Speed up Servers:

- Turn compression on at the server!
- Keep HTTP headers small

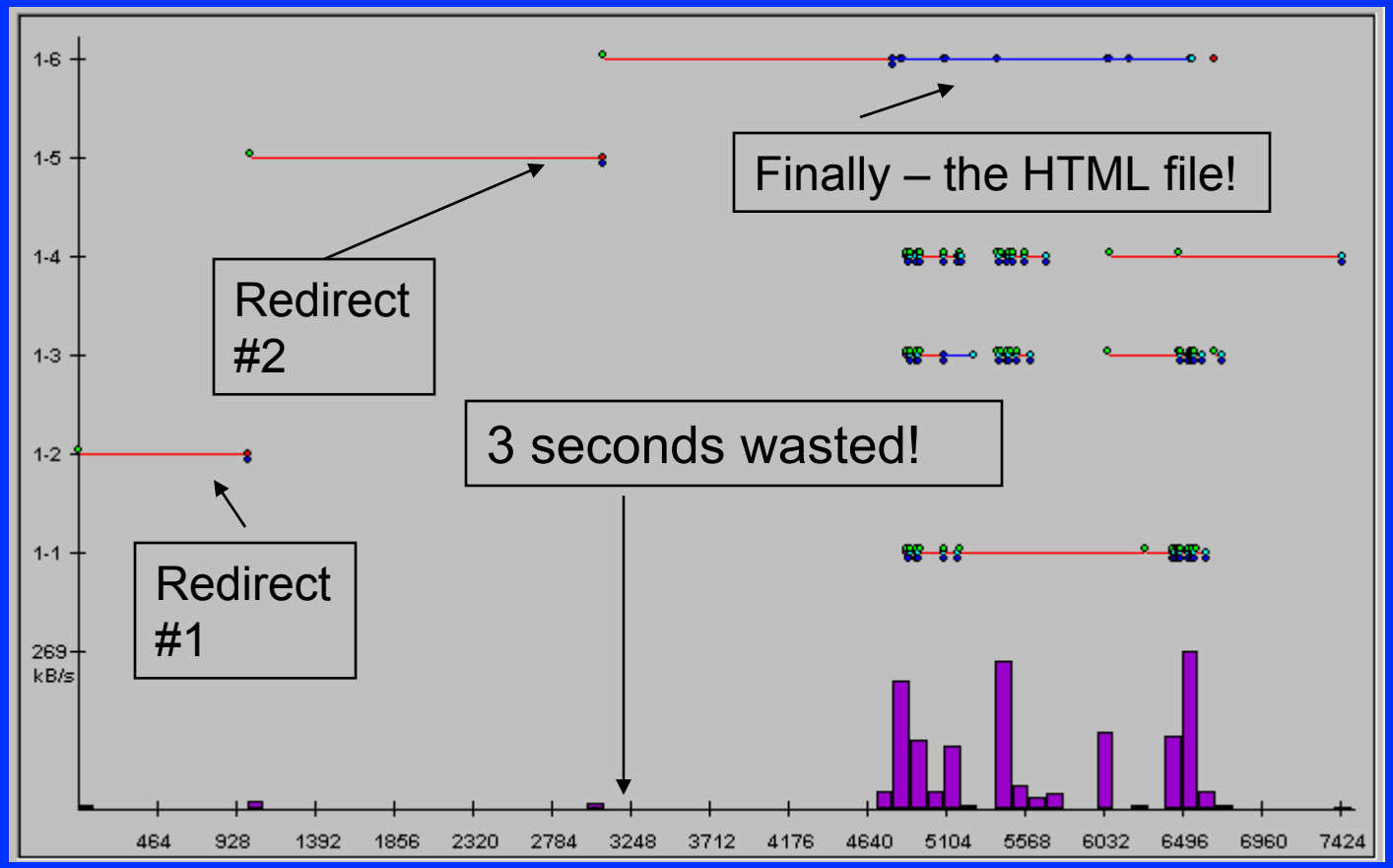
Web Page Performance Issues:

Keep Web pages small

- Smaller webpages are faster – Make sure you really need all those tiny graphics.
- Large numbers of small graphics files are efficiency killers
  - Each file gets about 800 bytes of HTTP header
  - Browsers do not aggressively pipeline, so each file on a page will incur a wait time from your server.

Smaller is Faster

Web page redirects can waste a lot of time



Minimize redirects

Summary

- Improve the end user responsiveness of the Web by:
- Improving browser efficiency
  - Ask for images sooner
  - Pipeline requests
- Improving server efficiency
  - Turn compression on
  - Minimize HTTP header size
- Improving web page efficiency
  - Send fewer files
  - Avoid redirects

Call To Action:

Intel Developer Forum Spring 2002

Tune your applications and web pages to use the network efficiently

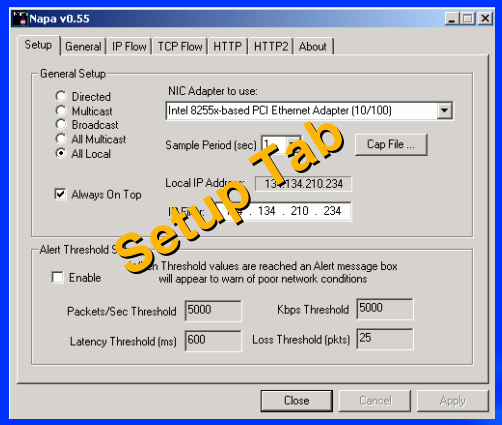
Collateral

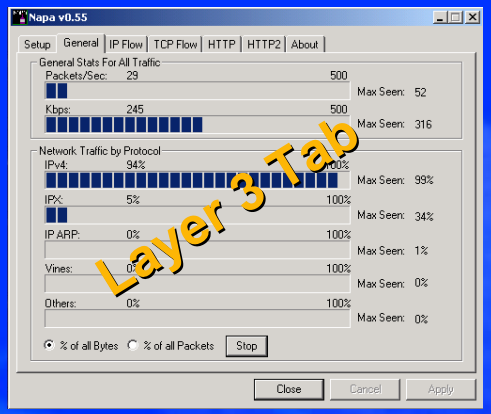
- NAPA Executable & Source Code
  - CDROM – Come and get ‘em
  - Jennie.yoder@intel.com
  - Jim.chu@intel.com
  - URL to download from is http://www.intel.com/software/products/opensource/

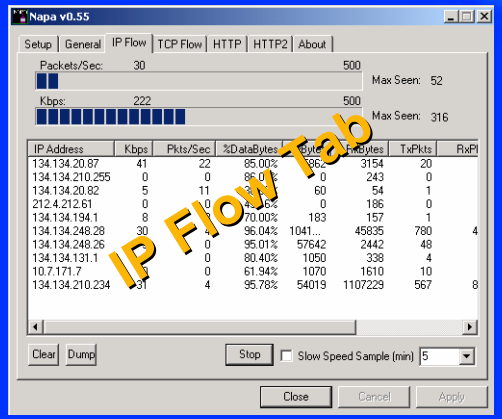
Backup

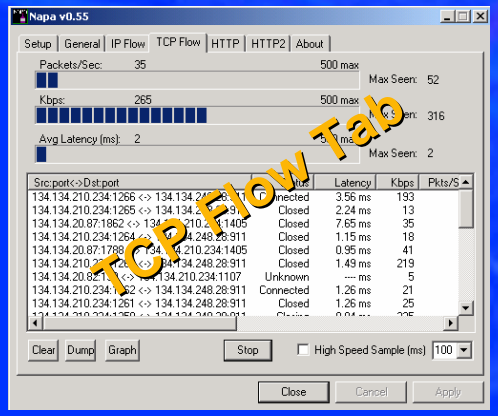
Intel Developer Forum Spring 2002

Network App Perf Analyzer





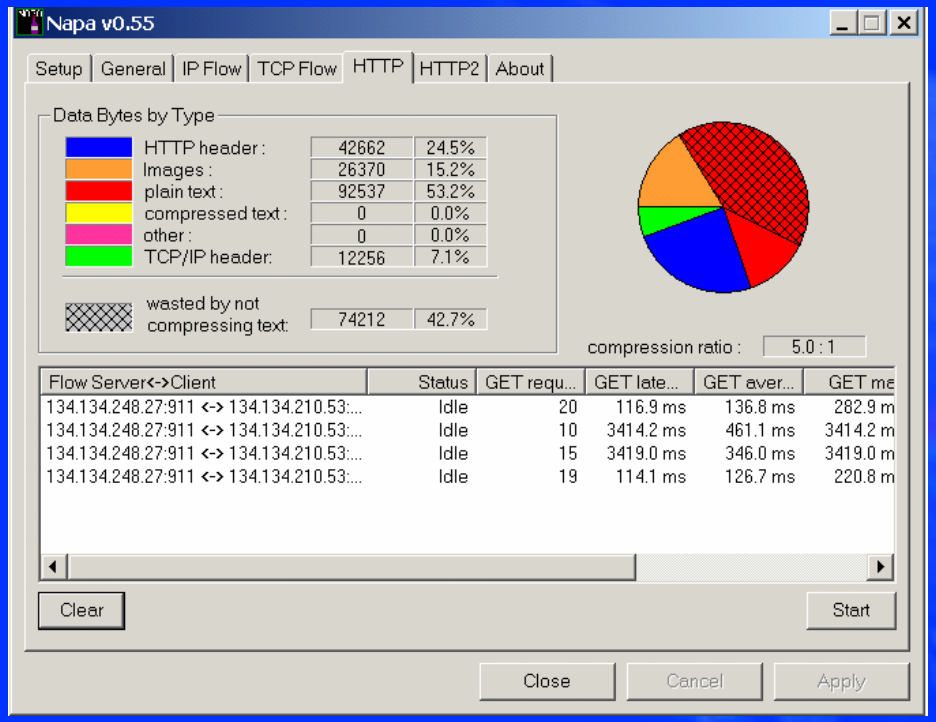




- Ring3 app on top of packet sniffing driver
- Looks at local client & server traffic
- Supports IP, TCP, & HTTP protocols

Helps identify inefficient network usage

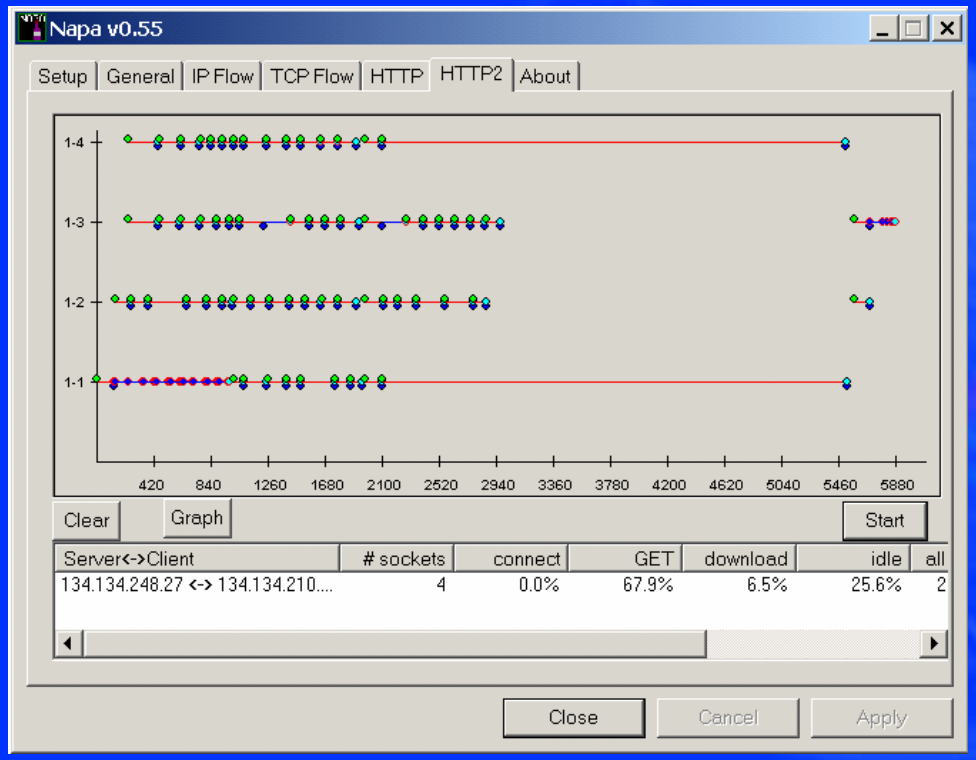
HTTP Data Breakdown



- Categorizes HTTP data bytes sent
- Shows benefits of compressing text
- Helps analyze Get latencies

Good analysis for webpage developers

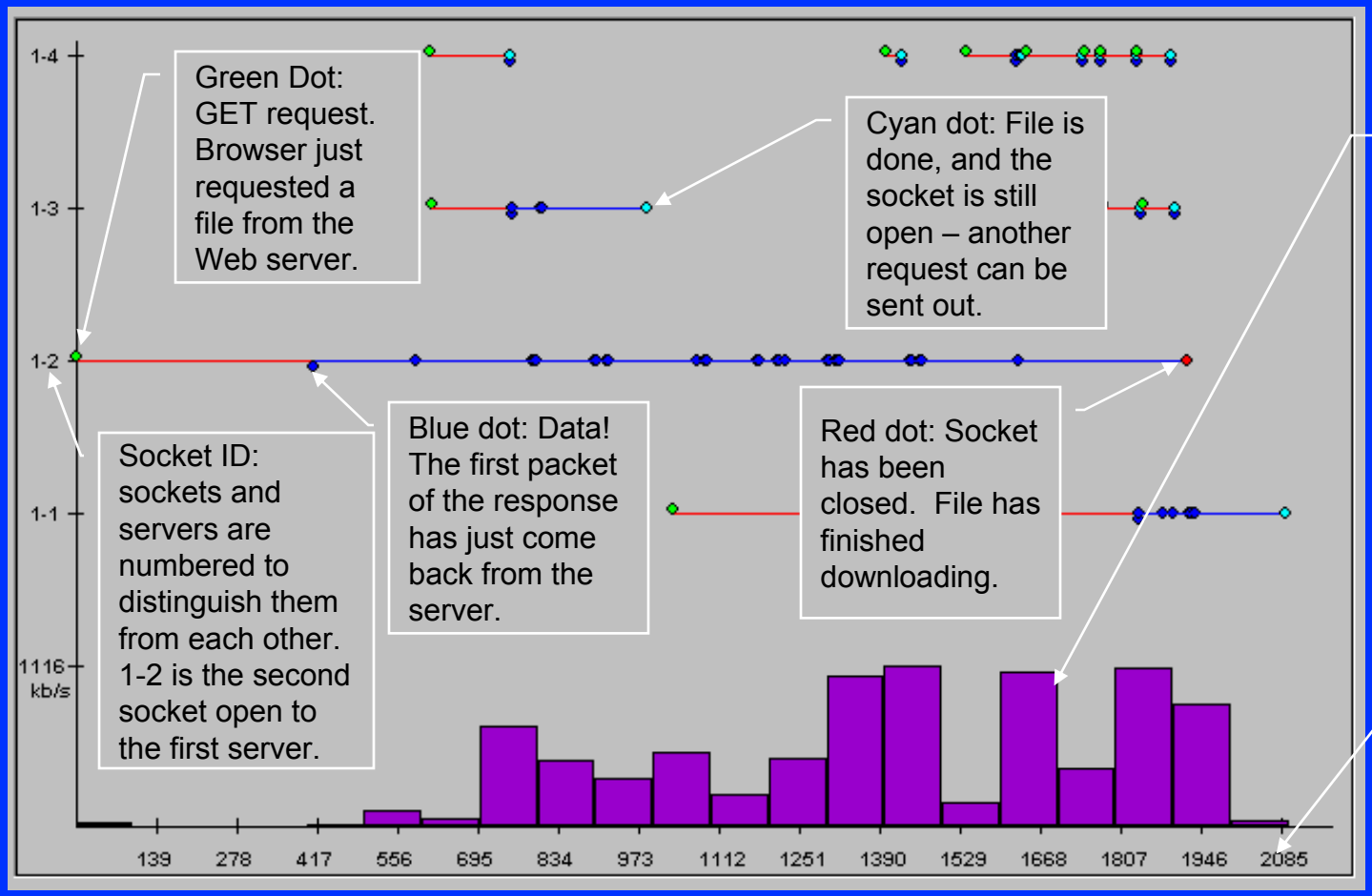
HTTP Time Breakdown



- Application socket usage
- Shows packets wrt time
- Easily identify latency & pipelining issues

Good analysis for network app developers

HTTP Timeline Chart



Visualizing network traffic helps identify issues

HTTP Timeline Chart Key

- X-axis: time (ms)
- Y-axis: shows all sockets open on the machine.
  - N-M: N = server number, M = socket number
- Y-axis: shows total bandwidth used: the maximum bandwidth seen is used as the axis label.
  - Each purple bar sums the bytes seen over 100ms.
- Each dot is an event associated with a packet
  - In practice, each dot is a packet
- Lines connecting dots indicate if the socket is waiting for more packets
  - Red lines mean the socket is waiting for the initial response to a request
  - Blue lines indicate that the response has started coming in and is downloading

Collateral

- NAPA Executable & Source Code
  - CDROM – Come and get ‘em
  - Jennie.yoder@intel.com
  - Jim.chu@intel.com
  - URL to download from is:
    - http://www.intel.com/software/products/opensource/