

# Performance Improvement

- stories from Bing, Hotmail and MSN

**Mukesh Jain**

Principal Engg. Manager  
Bing Advertising R&D  
Microsoft India

[Mukesh.Jain@Microsoft.com](mailto:Mukesh.Jain@Microsoft.com)

# About Myself

- 16+ yrs in SW, worked in Dev, Test, PM, Quality Manager, Process Champ, Engg Mgmt, etc.
- With Microsoft for 12+ yrs
  - 5 years in Microsoft Office Outlook
  - 2 years in Microsoft India Hyderabad heading process and quality division driving Six Sigma, Metrics/Dashboard adoption
  - 5 years Microsoft Online Services driving performance improvements in Bing, Hotmail, MSN, Messenger and Advertising
  - Most recently, moved to Microsoft India heading Bing Ads R&D Quality
- Previously (1995-1999) with MNCs in INDIA implementing SEI CMM level 3-5, Six Sigma, ISO 9000, poka-yoke, etc.
- Author of Books:
  - “Delivering Successful Projects with TSP and Six Sigma: A Practical Guide to Implementing ...”
  - <http://www.amazon.com/Delivering-Successful-Projects-TSP-Sigma/dp/1420061437>
  - “Web Performance Improvements” MS Press
  - <http://www.microsoft.com/MSPress/books/authors/auth12791.aspx>
- Email: [Mukesh.Jain@microsoft.com](mailto:Mukesh.Jain@microsoft.com)



# What do I do/Journey so far...

- As a Tester: I get paid to find mistakes in other people's work
  - Test Lead: I lead a team of people who find mistakes
  - Test Management: I put together plans and process in place for people to find mistakes
  - Process Champ: I put together plans so that people do not make mistakes
  - Quality Manager: Ensure teams follow process and manage using data/metrics
  - Engg Management: put together people and processes in place to build great products
  - Author: write on failures, learnings and how to plan for success
- ...and I get paid \$ for doing all these...

Mentor & Presenter: Give back to community, share learnings



# We will talk about...

- What is Quality?
- How to Measure Web App Performance?
- Traps to avoid for performance improvements
- Overview of Six Sigma
- Using Six sigma to Improve Performance
- Stories from Bing, Hotmail and MSN

# What is Quality?

- Meets expectations
  - Serves the Purpose / Needs
  - Intuitive / Usability
  - Desktop Software → Software + Services
  - Anytime, Anywhere, Any Device
- Reliable
  - Responsive / Performance
  - Security / Privacy
  - High Quality/Low Defect
  - Getting it right the first time, every time
- Think Global

# Business Results to Product Quality





# Quality of Service (QoS)

- Measure Service Quality
  - Performance
  - Availability
  - Reliability
  - Business Metrics
- Monitor critical service capacity
- Proactively identify issues
- Right set of Metrics
- Markets

# Defects are inevitable

- **Fact:** No software can be guaranteed 100% defect-free
- **Action:** No action
- **Result:** We make it horribly true
- **Ask:** Why the defect happened?
- **Do:** Analyze data and improve the process to prevent it

**And Sustain it → Six Sigma**



# Testing @ Microsoft

## Finding Bugs

1 Dev: 2+ Test

Long Cycles 4+ yrs

Testing is Afterthought

Software Tester

(limit by test ability)

## Meet Needs/ Requirements

1 Dev: 1.7 Test

Long Cycles 3+ years

Testing is Planned

Software Tester

(limited by Spec)

## Test Engineering

1 Dev : 1 Test

Medium cycle 1-2 yrs

Testing planned & executed

Scenarios based testing

Software Design Engg in

Test (SDET)

(Test Automation)

## Ensures high quality

1 Dev : 0.5 Test

Short Cycles 6 months

No Plan without Testing

Parallel Testing Org

SDET, VP, Partner Levels

## Very High Test Maturity

1 Dev : 0.25 Test

Very Short Cycles 1 Month

High Maturity Development

Integral Part of Engineering

All are Software Engg

(Data, Usage, Risk, etc.

Just-enough, right set of testing)



# The Performance issue

- Performance top concern (Customer sat. survey)
- Bing, Hotmail, MSN, Messenger, Ads, Mobile, ...
- Loosing mind-share in key markets
- Impact – Market share & revenue
- Solving performance issue traditional way
  - Put more people
  - Do more performance testing
  - Buy Performance Testing tools
  - Add more servers



## Solving the right problem – the right way

- Whenever you encounter an issue/defect, ask yourself:
  - What assumptions were wrong?
  - What rules did I break?
  - How could I have detected this bug earlier?
  - How could I have prevented this bug?

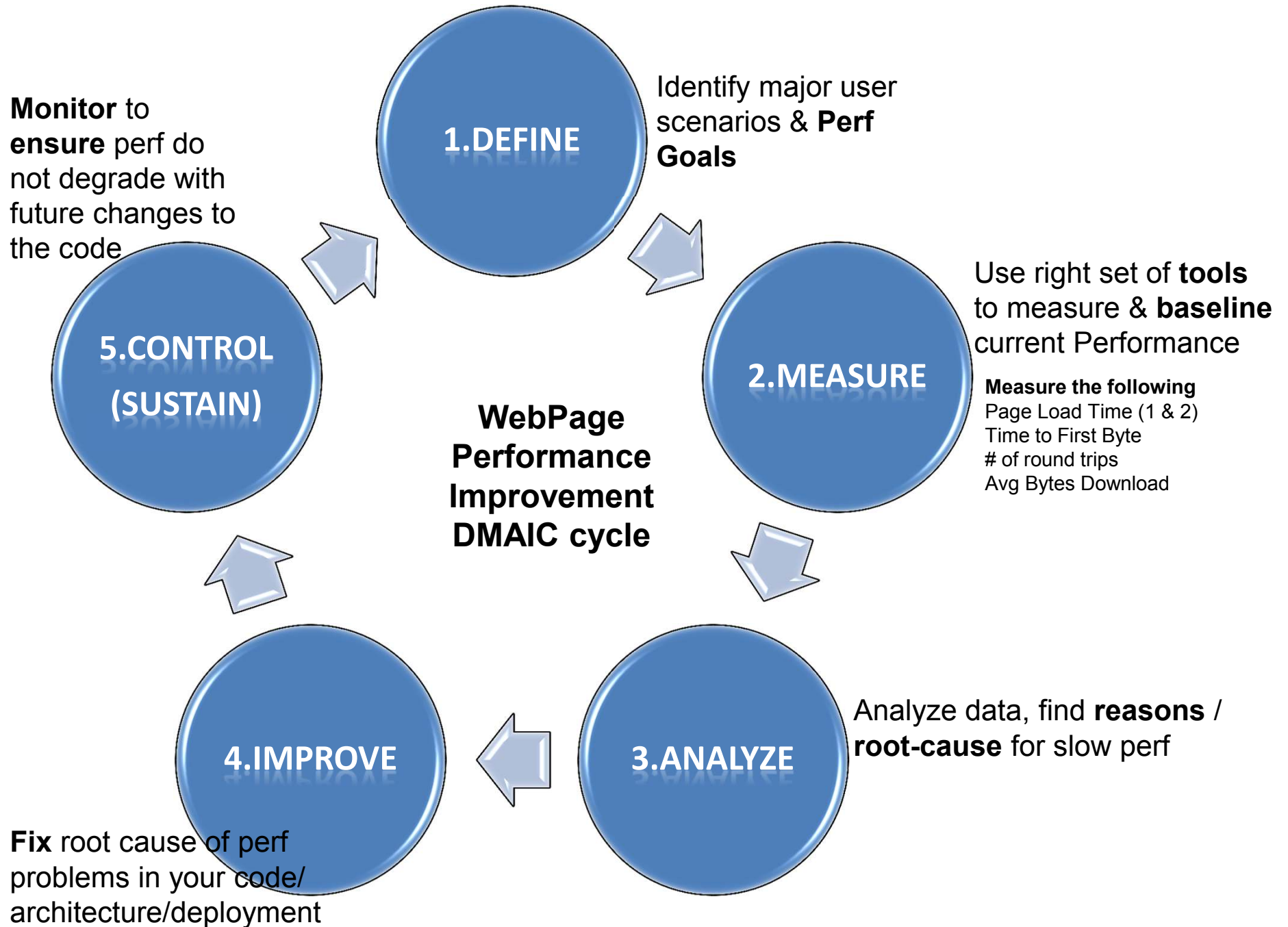
# Six Sigma

- Philosophy, Framework, Methodology
- Roadmap for Continuous improvements
- Metrics Rigor
- Structured problem solving methodology
- Way of doing business
- Reducing variation, defects
- Six Sigma = 3.4 defects / Million opportunities

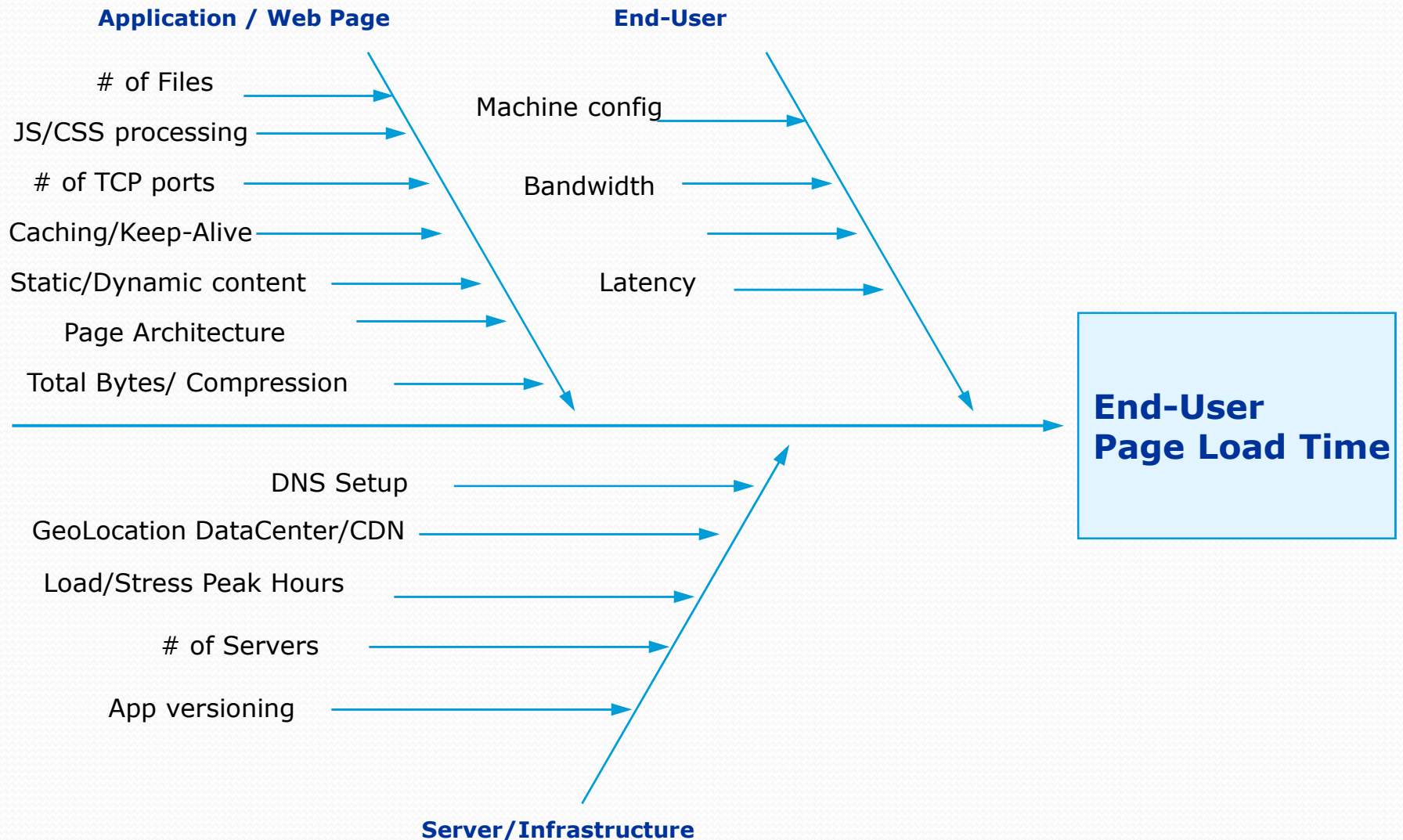
# Six Sigma - DMAIC

- **Define** (D) : Zero in on specific problem with defined return on effort
- **Measure** (M) : Determine current performance of process
- **Analyze** (A) : Validate key drivers of performance (root cause of problem)
- **Improve** (I) : Improved performance and validated realized results
- **Control** (C) : Implement controls to ensure continued performance

Project Phases and Deliverables				
Define	Measure	Analyze	Improve	Control
<ul style="list-style-type: none"> <li>➤ Project selection</li> <li>➤ Project charter</li> <li>➤ Critical to Customer (CTQ) needs</li> <li>➤ High level process map</li> </ul>	<ul style="list-style-type: none"> <li>➤ Key output variables (metrics or Y's)</li> <li>➤ Possible causes of defects (X's)</li> <li>➤ Data collection and presentation plan</li> <li>➤ Current Performance</li> <li>➤ Internal/ external benchmarking</li> </ul>	<ul style="list-style-type: none"> <li>➤ Key causes (vital few) of defects (X's)</li> </ul>	<ul style="list-style-type: none"> <li>➤ Improvement strategy</li> <li>➤ Prioritize solutions</li> <li>➤ Tested &amp; measured solutions</li> <li>➤ Final solutions</li> </ul>	<ul style="list-style-type: none"> <li>➤ Lock in the results (control plan)                             <ul style="list-style-type: none"> <li>•Mistake proofing</li> <li>•Control points</li> <li>•Monitoring Plan</li> <li>•Positive hand off of control plan</li> </ul> </li> <li>➤ Financial impact of the project</li> </ul>



# Performance – Cause-Effect Diagram



# Perf Monitoring Vs Perf Testing

- Monitor
  - Actual data from users
  - Realtime
  - Helps us find the most important cases/scenarios
  - Real PLT<sub>1</sub> vs PLT<sub>2</sub>
  - Required to maintain
- Testing
  - It does what we tell it to do
  - Simulated environment
  - Required to sign-off before release
  - Best case PLT<sub>1</sub> & PLT<sub>2</sub> measurements



# Monitoring Tools

- Real: Code Instrumentation, toolbar, etc.
- Synthetic:
  - Keynote
  - Gomez
  - WebHancer
  - MobileCompete
- Debugging
  - Visual Studio
  - Fiddler
  - WebRunner
  - HttpWatch
  - WANSim
  - YSlow



# Performance Improvement Stories

- Bing
  - Hotmail
  - MSN
- 
- Mantra: If it appears slow – it is slow, irrespective of what the data shows

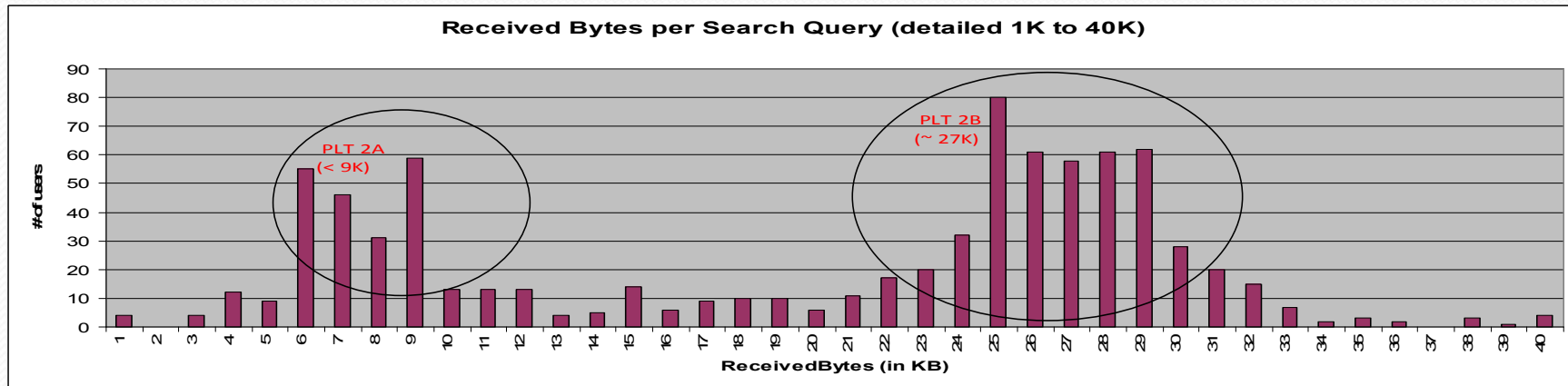
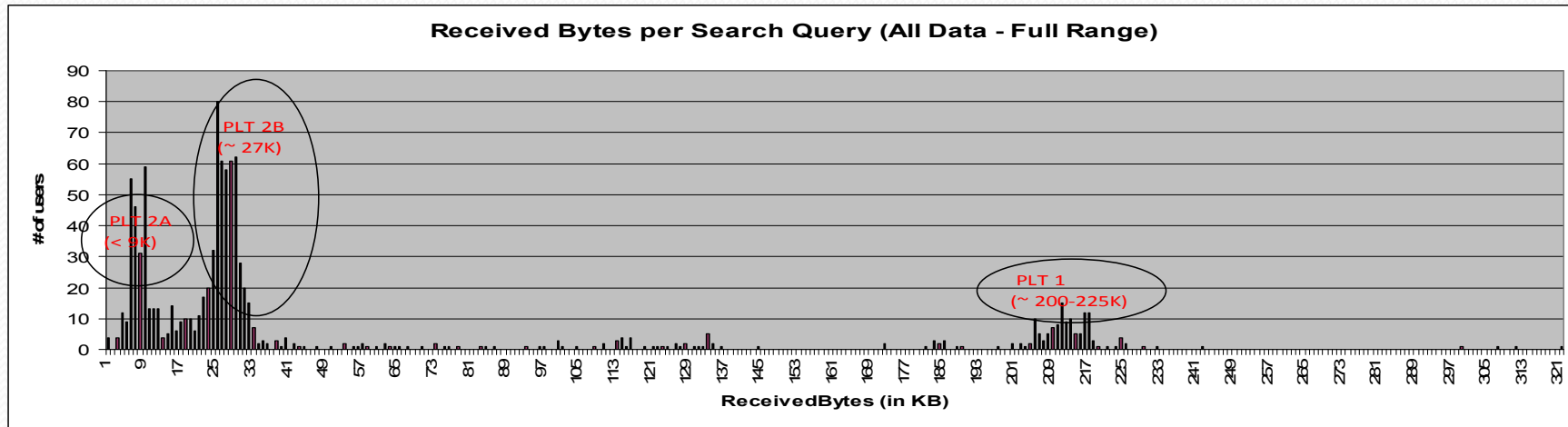
# Performance – Best Practices (1)

- Architecture
  - Use CDN (Akamai, Limelight, etc.)
  - Rearrange the page (css at top, first page loads first)
  - Simplify DNS entries and lookup
  - Geo Location
  - TTL (Time-To-Live)
  - Unnecessary Redirects
  - JS Blocking Issue
  - MAX 2 Parallel Download limit
- Application Optimization
  - Predictive loading & deferred actions (Browsers DOM)
  - Common terms caching
- User Experience
  - If it appears slow – it is slow, irrespective of what the data shows
  - Different page based on the connection speed, location, locale, device, etc.
  - Show something to the user and then download
  - Showing some Progress
  - Ads

# Performance – Best Practices (2)

- Page Components
  - Reduce # of files on the page (reduce round-trips)
  - Combine multiple files (Image Clustering/CSS Sprite)
  - Maintain one copy of the file/directory
  - Explore using CSS instead of images wherever possible
- File Size
  - Reduce file size
  - Compress using gZip
  - Crunch JS, CSS, HTML files
- Settings...
  - Keep Alive
  - Use Expiry Date in file header (enables cache'ing)
  - Etags

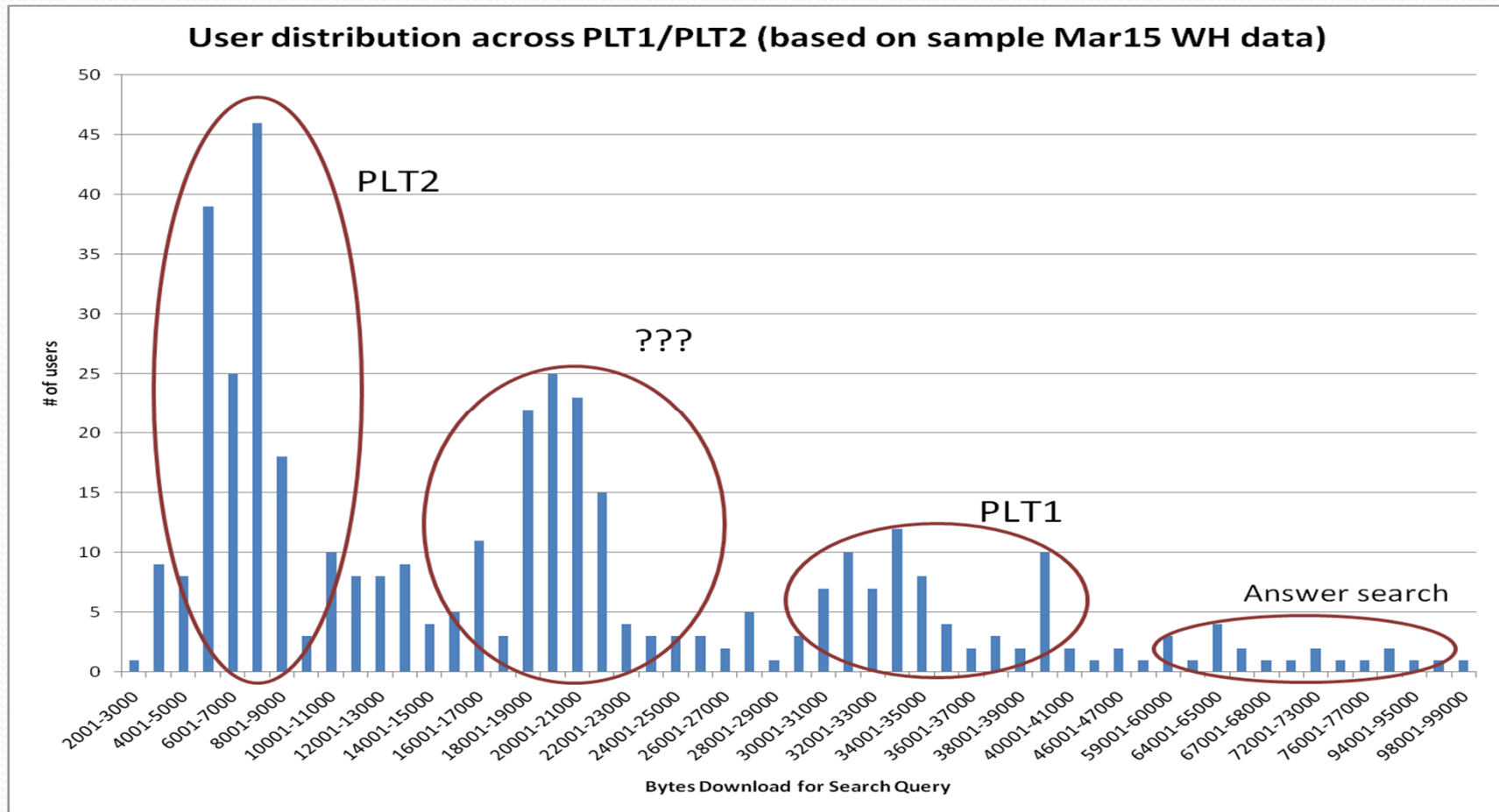
# Live Search PLT1/PLT2 User Distribution (Oct '06)



## Oct 2006 Data

- About 13% of the user are on PLT1 (Search Results Page Download 200K or more)
- Around 50% of the users are on PLT2B (10K-30K) where they have to download few files again (due to a search cookie issue)
- Around 23% on PLT2A (<10K) experience
- This data uncovered a bug in cookie (expiry issue), which caused download of already cached files – was fixed immediately by the search team

# Search PLT1/PLT2 User Distribution (Mar '07)



## **March 15, 2007 Data analysis**

- Typically 17% of the users are in PLT1 and 83% in PLT2
- There are several users in a middle category (16K to 21K)
- Possible cause- directory naming – (next slide). Also, file image\_cache.htm is not cached
- Answer search (6% of total search) usually download more bytes than regular search

## Multiple version causing additional downloads – impacting PLT2

URL/Transaction - search.msn.com+results.asp+q=food_BBPLT1.2_070314183353_sa14_rta.txt (18:33:53 03/14/07 PST from SA14)												
A	B	C	D	E	F	G	H	I	J	K		
1	Start	Relative T	Ports (Src	URI	Status Code	Content Type	Total By	Content	Compr	Compr	Durati	
2	2	0		Start Timer			0	0				
3	3	0.134	31126 - 80	<a href="http://search.msn.com/results.asp">http://search.msn.com/results.asp</a>	200 -- OK	text/html; chars	9145	8040	0.6	1	1.3	
4	14	1.172	31128 - 80	<a href="http://search.msn.com/sa/3_2_0_100091A_sb_c.css">http://search.msn.com/sa/3_2_0_100091A_sb_c.css</a>	200 -- OK	text/css	5648	4868	0.6	1	0.8	
5	28	2.04	31126 - 80	<a href="http://search.msn.com/sa/3_2_0_100091A_sb_c.js">http://search.msn.com/sa/3_2_0_100091A_sb_c.js</a>	200 -- OK	application/x-ja	8643	7801	0.6	1	0.7	
6	39	2.855	31128 - 80	<a href="http://search.msn.com/s/fewel.png">http://search.msn.com/s/fewel.png</a>	200 -- OK	image/png	2024	1520	0.7	1	0.4	
7	40	2.882	31129 - 80	<a href="http://shared.live.com/~live.themes/~11.6.2119/~live.sear">http://shared.live.com/~live.themes/~11.6.2119/~live.sear</a>	200 -- OK	text/css	2410	1800	0.6	1	0.7	
8	41	2.882	31130 - 80	<a href="http://shared.live.com/~live.themes/~11.6.2119/~live.sear">http://shared.live.com/~live.themes/~11.6.2119/~live.sear</a>	200 -- OK	image/gif	902	371	0.6	1	0.8	
9	56	3.664	31129 - 80	<a href="http://shared.live.com/~live.themes/~11.6.2119/~live.sear">http://shared.live.com/~live.themes/~11.6.2119/~live.sear</a>	200 -- OK	image/gif	1787	1322	0.6	1	0.3	
10	59	3.686	31130 - 80	<a href="http://shared.live.com/~live.themes/~11.6.2119/~live.sear">http://shared.live.com/~live.themes/~11.6.2119/~live.sear</a>	200 -- OK	image/gif	1302	892	1.2	1.8	0.4	
11	60	3.694	31126 - 80	<a href="http://search.msn.com/s/SbBgR.gif">http://search.msn.com/s/SbBgR.gif</a>	200 -- OK	image/gif	588	139	0.5	0.8	0.4	
URL/Transaction - search.msn.com+results.asp+q=food_BBPLT2_070314183422_sa14_rta.txt (18:34:22 03/14/07 PST from SA14)												
A	B	C	D	E	F	G	H	I	J	K		
1	Start	Relative T	Ports (Src	URI	Status Code	Content Type	Total By	Content	Compr	Compr	Duration	TCP
2	1	0		Start Timer			0	0				
3	2	0.126	31134 - 80	<a href="http://search.msn.com/results.asp">http://search.msn.com/results.asp</a>	200 -- OK	text/html; ch	8952	7968	0.6	1	1.405	
4	14	1.223	31135 - 80	<a href="http://search.msn.com/sa/3_2_0_99760A_sb_c.css">http://search.msn.com/sa/3_2_0_99760A_sb_c.css</a>	200 -- OK	text/css	5648	4868	0.6	1	0.823	
5	30	2.081	31134 - 80	<a href="http://search.msn.com/sa/3_2_0_99760A_sb_c.js">http://search.msn.com/sa/3_2_0_99760A_sb_c.js</a>	200 -- OK	application/	8643	7801	0.6	1	0.784	
6	44	3.246	31135 - 80	<a href="http://search.msn.com/sa/3_2_0_99760/image_cache.htm">http://search.msn.com/sa/3_2_0_99760/image_cache.htm</a>	200 -- OK	text/html	926	394	1.1	1.6	0.502	
7	58	16.902		Stop Capture			0	0				
8	TOTAL	16.902					24169	21031				

Files in the circle are not been cached, due to change in directory name (different data center). Search team is trying out few things in one of the datacenter before rolling out to all the data center.

# MSNSearch vs Google (PLT2)

MSN Search 10 files

Started	Time	Size	Method	Result	Type	URL
00:00:00.000	0.605	6044	GET	200	text/html; charset=utf-8	http://search.msn.com/results.aspx?q=quality+conference+in+Seattle&FORM=MSNH
00:00:00.568	0.005	0	GET	(Cache)	text/css	http://search.msn.com/sa/3_6_0_119830/l_sb.css
00:00:00.584	0.001	0	GET	(Cache)	application/x- javascript	http://search.msn.com/sa/3_6_0_119830/l_sb_c.js
00:00:00.611	0.005	0	GET	(Cache)	image/png	http://search.msn.com/s/jewel.png
00:00:00.614	0.003	0	GET	(Cache)	image/gif	http://search.msn.com/s/passport.gif
00:00:00.615	0.003	0	GET	(Cache)	image/gif	http://search.msn.com/s/SrchBtn.gif
00:00:00.694	0.003	0	GET	(Cache)	image/gif	http://search.msn.com/s/HeaderGradientImage.gif
00:00:00.711	0.003	0	GET	(Cache)	image/gif	http://search.msn.com/s/SbBgR.gif
00:00:00.715	0.015	0	GET	(Cache)	image/gif	http://search.msn.com/s/SbBgL.gif
00:00:00.852	0.001	0	GET	(Cache)	application/x- javascript	http://search.msn.com/s/hf/hf_en-US.js

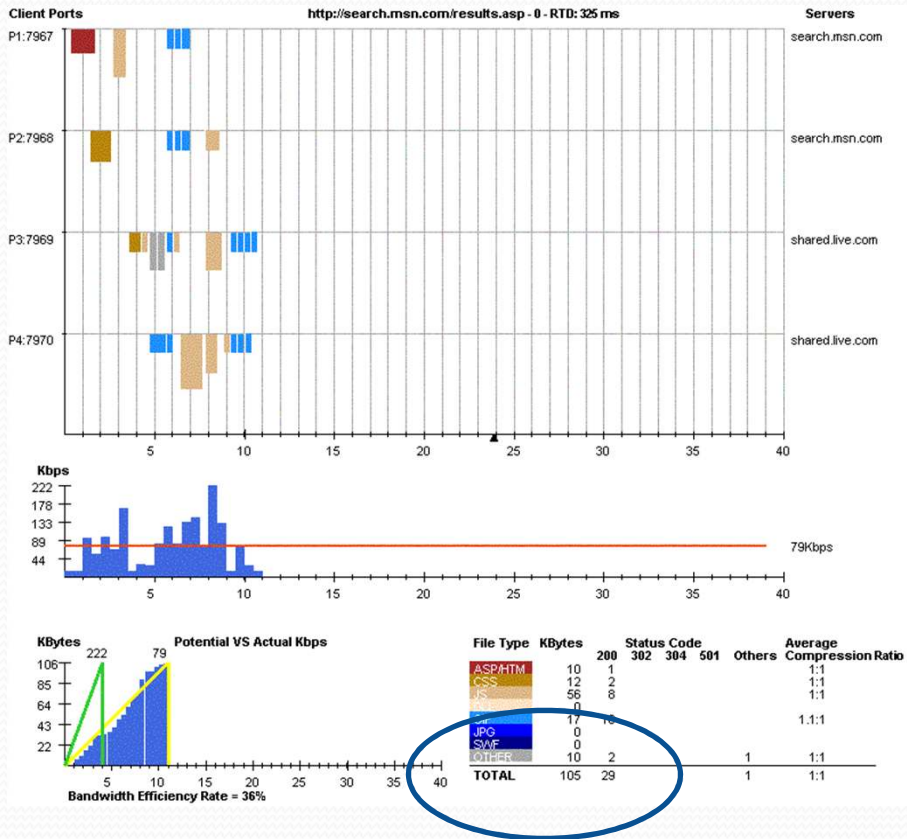
Google

Started	Time	Size	Method	Result	Type	URL
00:00:00.000	0.518	6414	GET	200	text/html; charset=UTF-8	http://www.google.com/search?hl=en&q=Performance+testing+conference&btnG=Google+Search

- In MSN search query – there are 10 GET requests. Even though 9 files are loaded from cache – IE have to process them and render (This adds an average 0.4 – 1 sec delay to display search results)
- The main msn search results file is 6K size and takes 0.6 seconds to download and display.
- Google uses only 1 file and is able to complete the transaction in half second.
- The main google search results file is 6K size and takes 0.5 seconds to download and display.

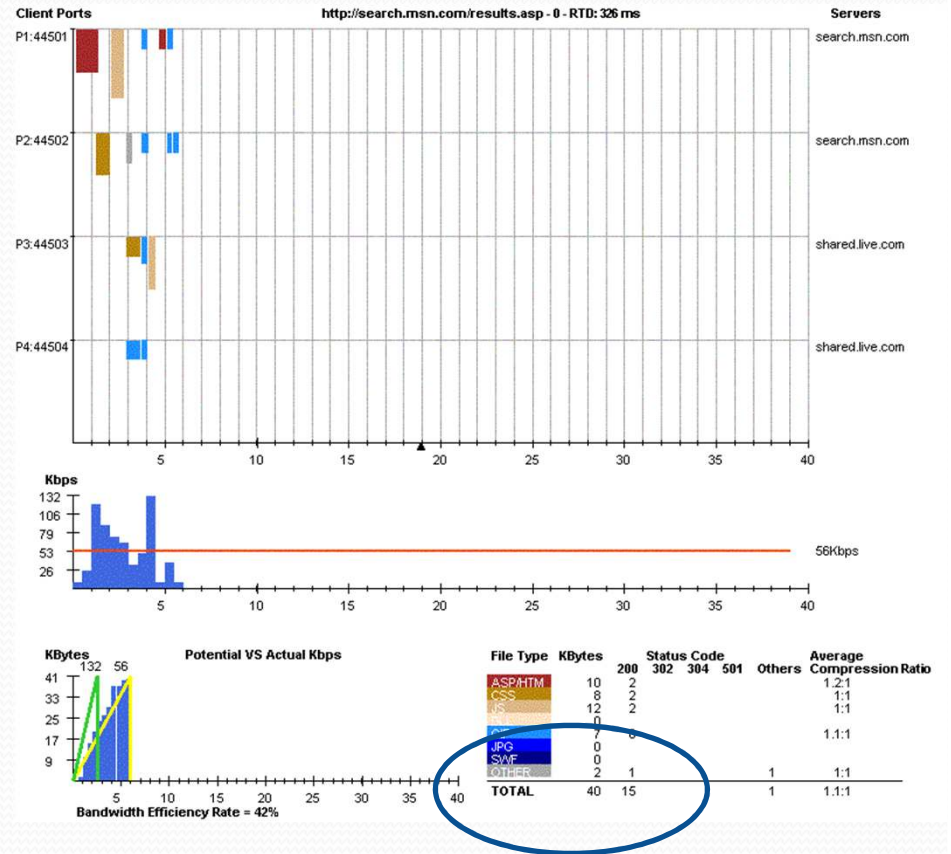


# MSN Search PLT1 VRTAs (Before/After Mar07 changes)



## Feb28 PLT1 VRTA

Files: 29  
KBytes: 105  
Time: 11 Sec

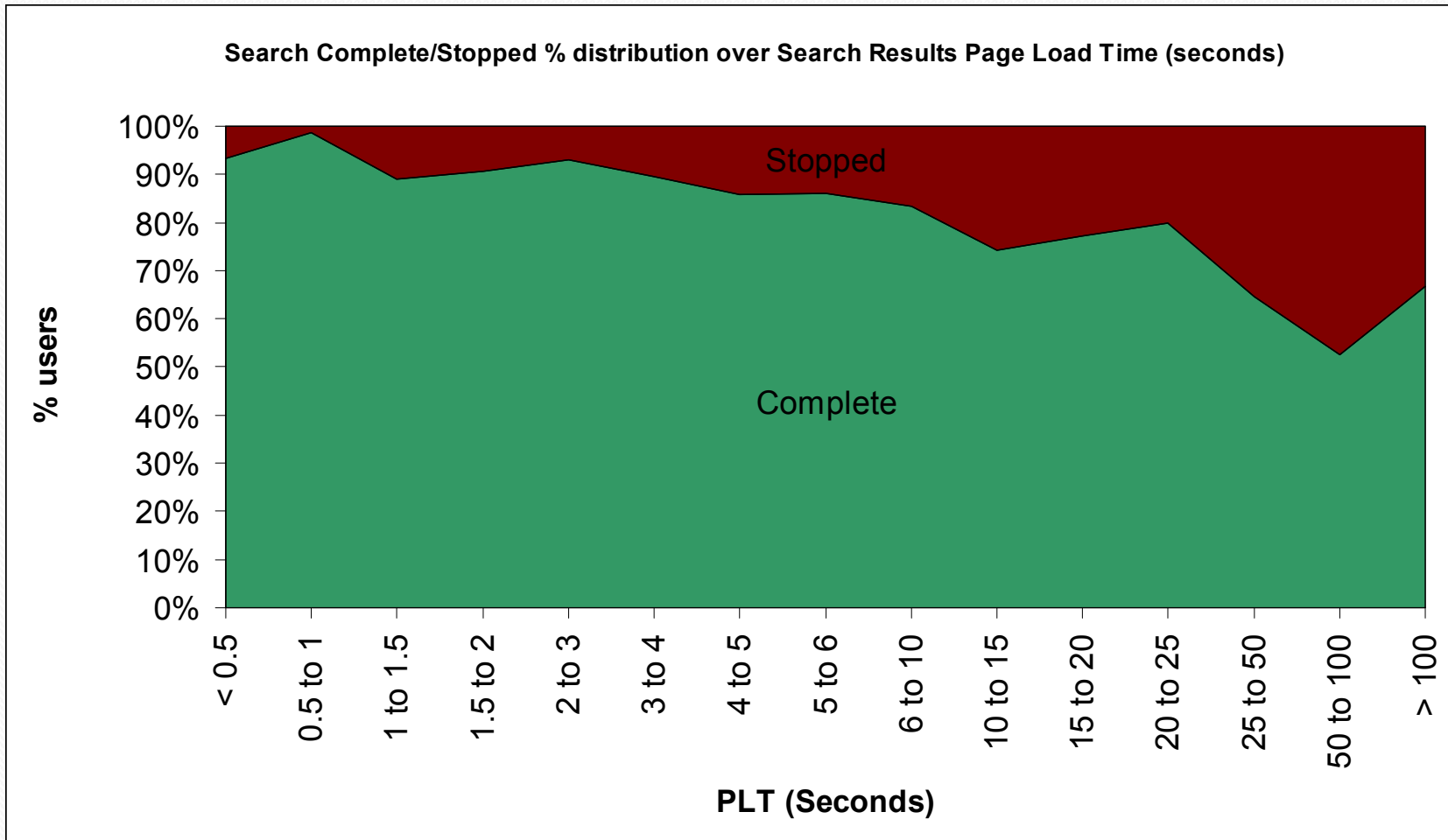


## Mar13 PLT1 VRTA

Files: 15  
KBytes: 40  
Time: 6 Sec

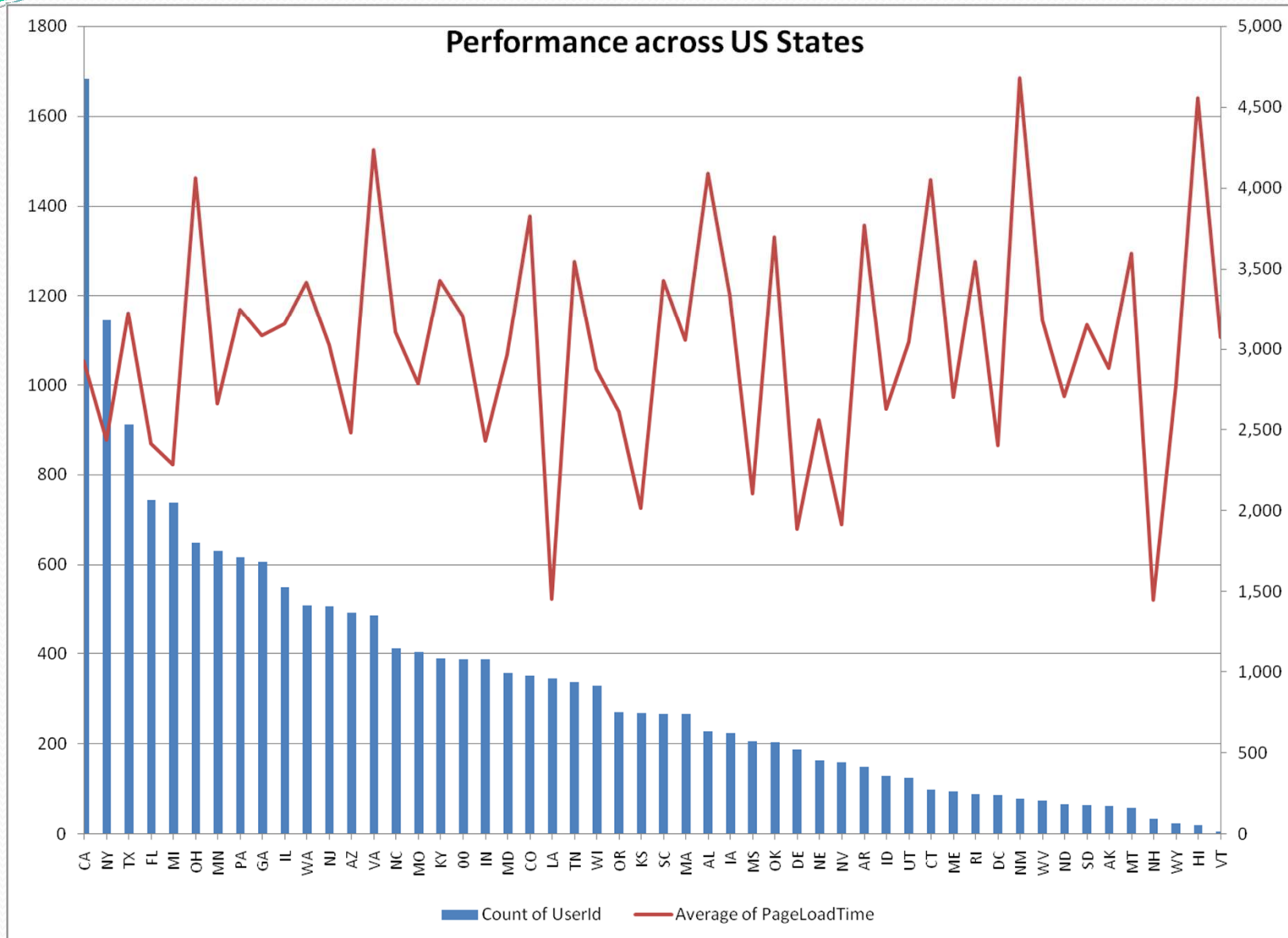
Major Perf changes (Mar 07) in PLT1 – around 50% reduction in # of Files & Size have improved Page Load Time

# Impact of slow performance on user

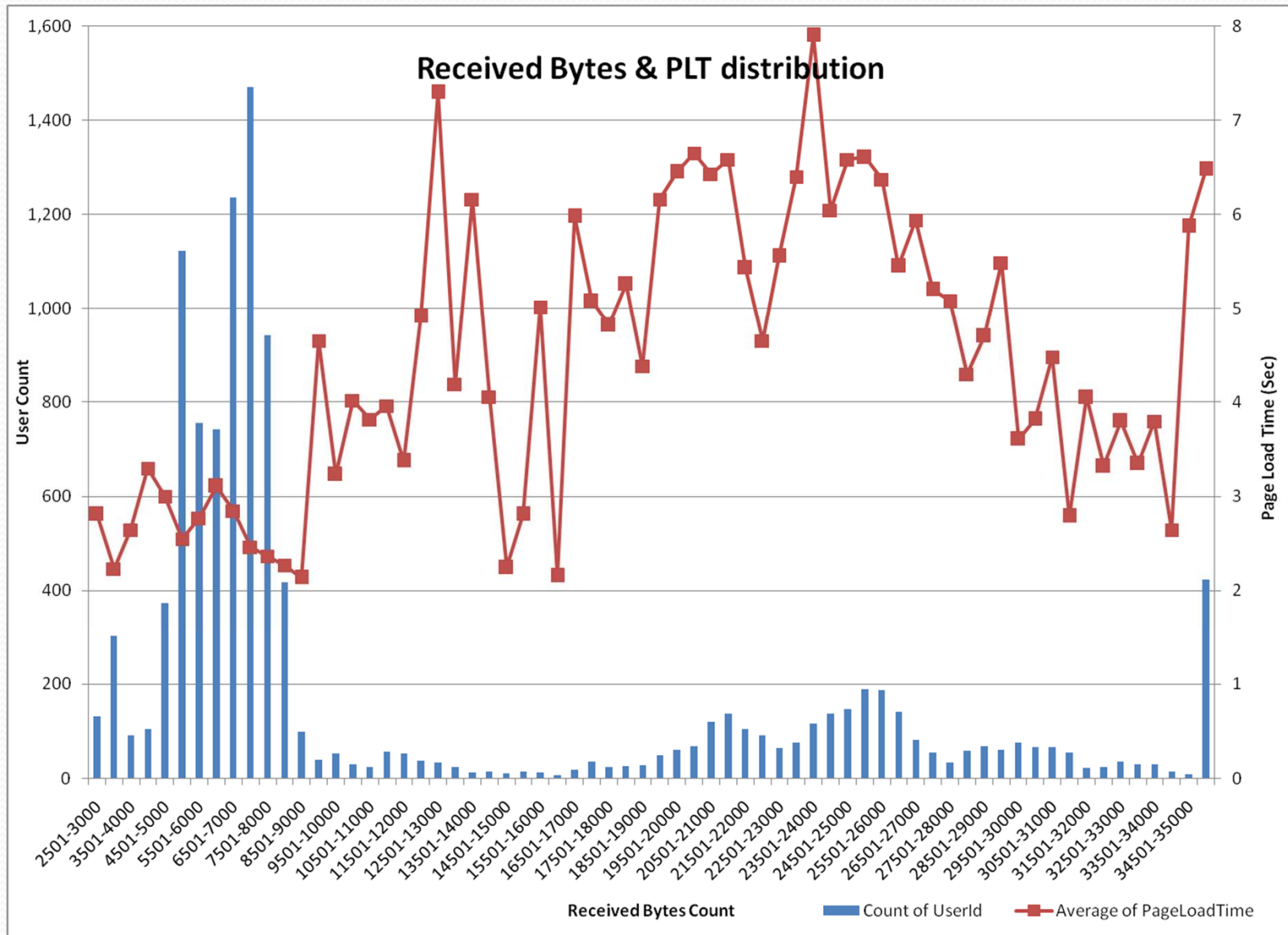


As it takes more time to display page – users STOP the page  
Typically after 5 seconds of wait, 15% user stops the page from loading and the % grows with the time

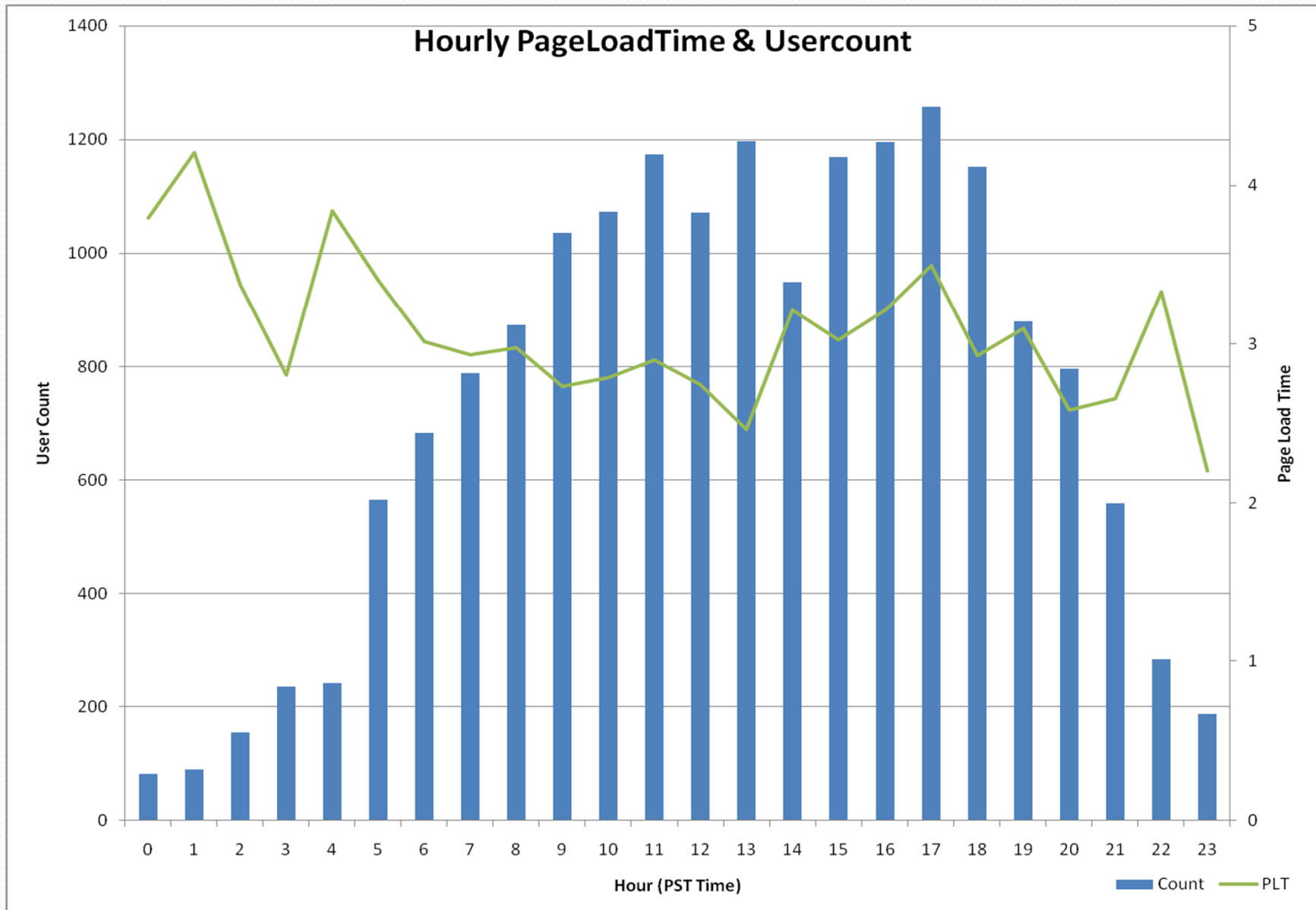
# Performance across states



# Bytes & Page Load Time distribution



# Hourly Users & PLT distribution



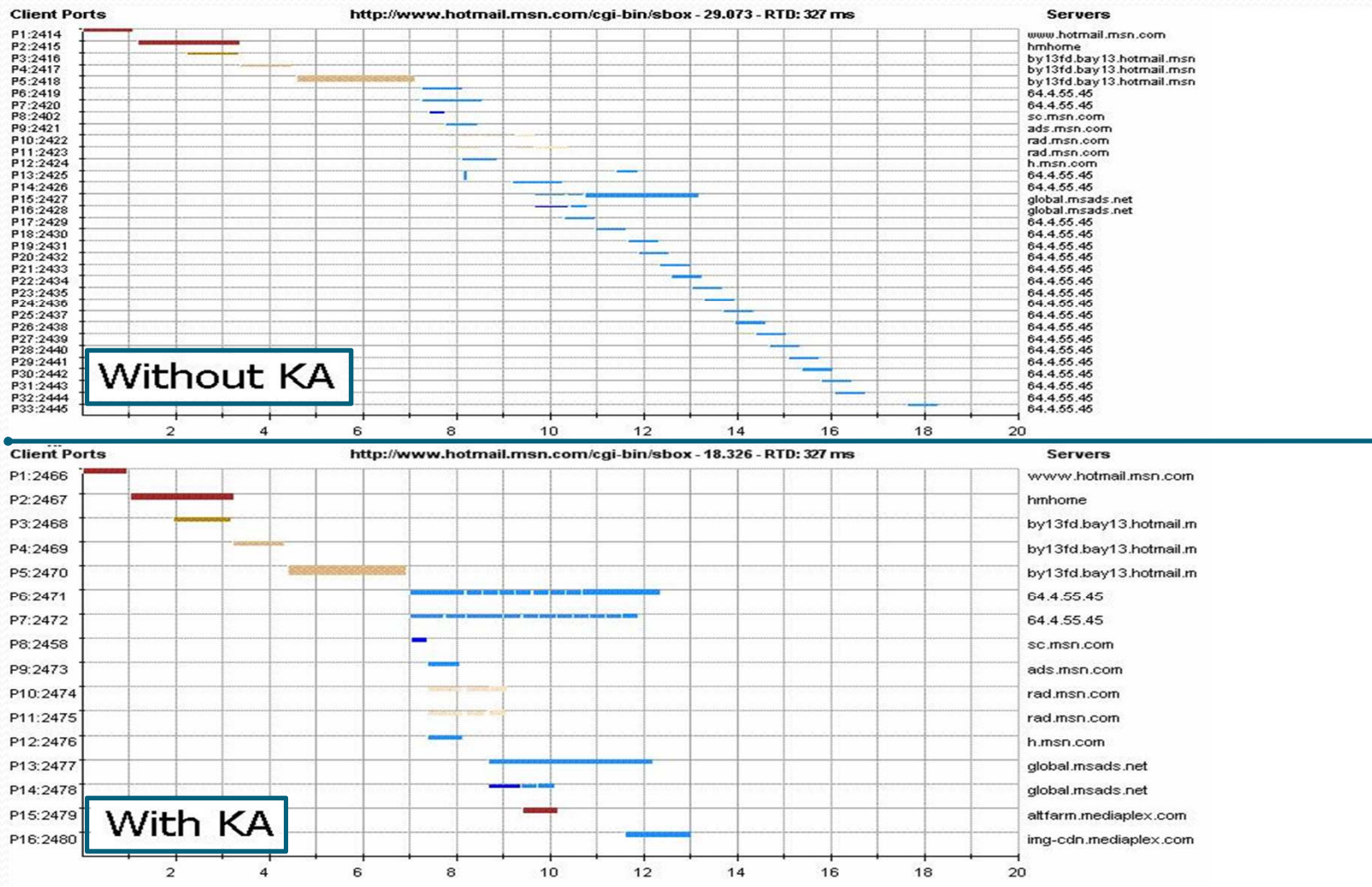
# Expiration dates example

## PLT2

- No 304s should be seen
- Set dates to 3 years
- Use dir, arg string, or file name to break cache

Rel Time	URI	Content Len	Status Code
0.00	<a href="http://groups.msn.com/people/">http://groups.msn.com/people/</a>	27659	200 -- OK
0.70	<a href="http://c.msn.com/c.gif">http://c.msn.com/c.gif</a>	42	200 -- OK
4.53	<a href="http://groups.msn.com/global/css.htm">http://groups.msn.com/global/css.htm</a>	0	304 -- Not Modified
4.88	<a href="http://groups.msn.com/spacer.gif">http://groups.msn.com/spacer.gif</a>	0	304 -- Not Modified
4.89	<a href="http://groups.msn.com/home_icons_chat_48x40.gif">http://groups.msn.com/home_icons_chat_48x40.gif</a>	0	304 -- Not Modified
4.90	<a href="http://www.match.com/msnprofile/profile.asp">http://www.match.com/msnprofile/profile.asp</a>	2481	200 -- OK
5.22	<a href="http://groups.msn.com/home_icons_IM_48x40.gif">http://groups.msn.com/home_icons_IM_48x40.gif</a>	0	304 -- Not Modified
5.55	<a href="http://groups.msn.com/msnmess_themes_65x60.gif">http://groups.msn.com/msnmess_themes_65x60.gif</a>	0	304 -- Not Modified
5.57	<a href="http://groups.msn.com/home_icons_heart_42x39.gif">http://groups.msn.com/home_icons_heart_42x39.gif</a>	0	304 -- Not Modified

# Keep-Alive TCP ports



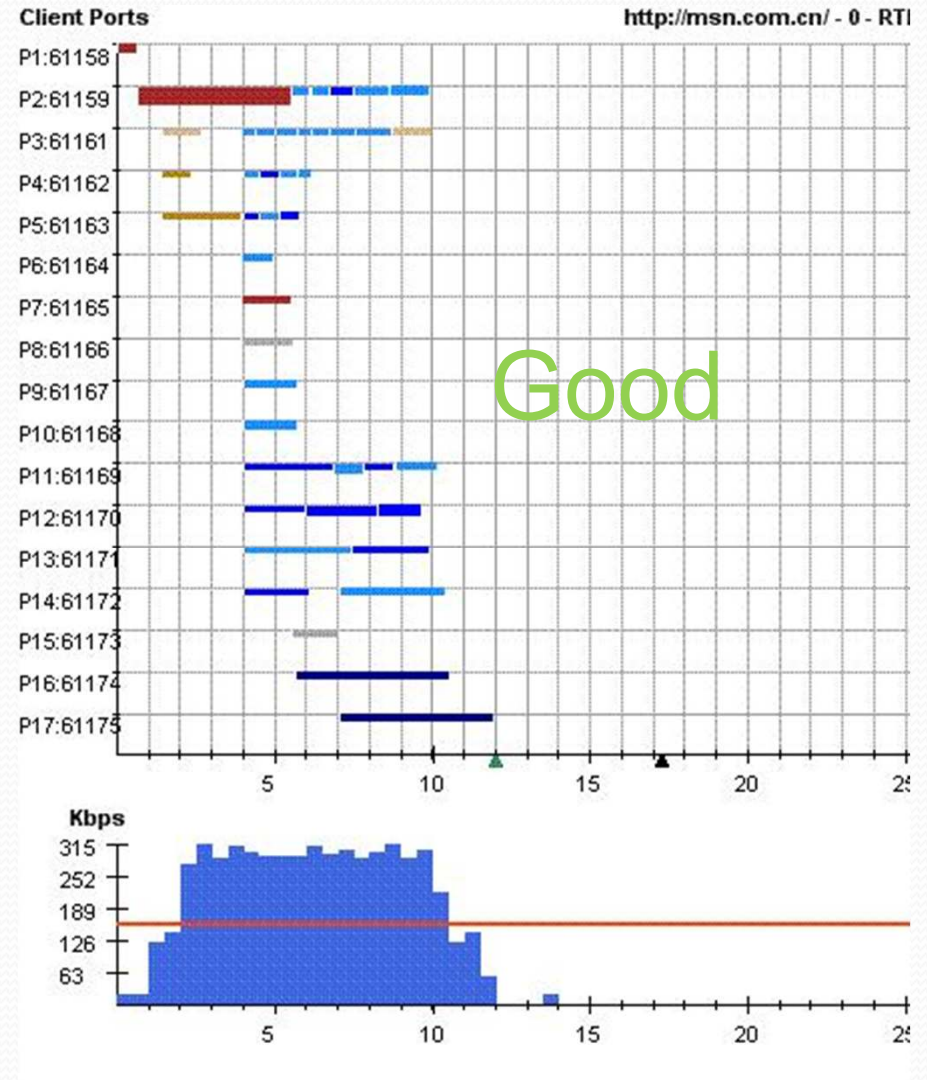
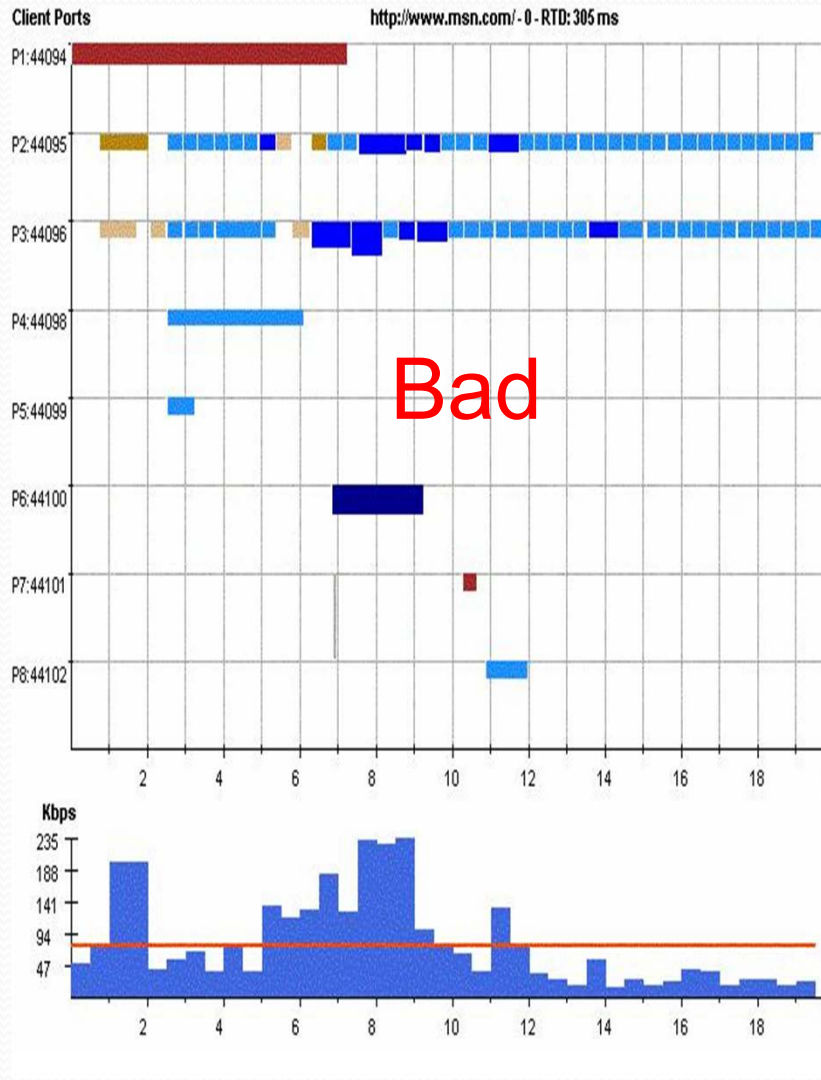
# Compression example

potential compression ratio  
<1 = already compressed

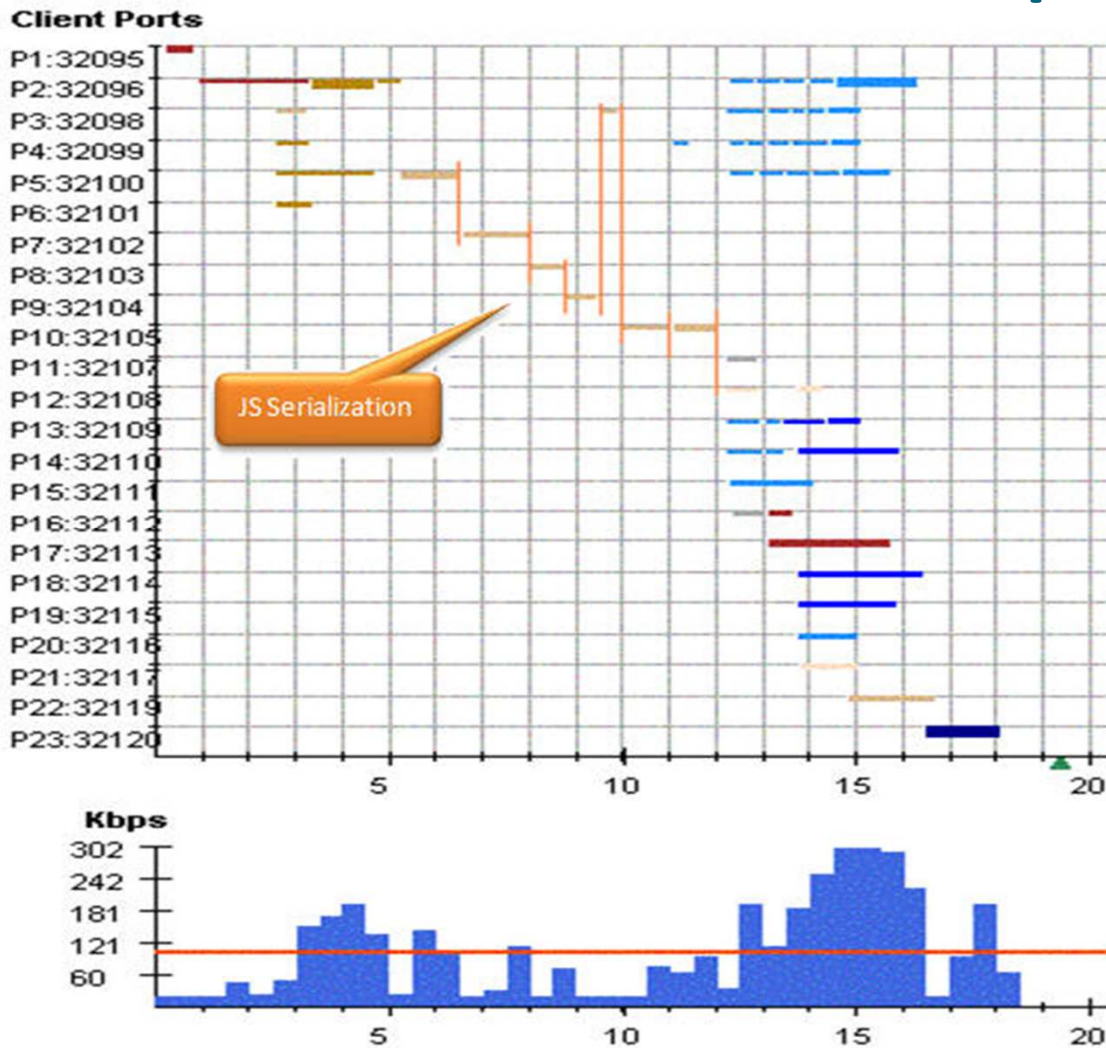
Relative Time	Ports (...)	URI	Total Bytes	Compressability	Compressability <sub>2</sub>
0.000	1367 - 80	http://www.msn.com/	30288	2.9	3.8
0.031	1368 - 80	http://i.msn.com/m/8/j/helppane17b.js	2697	1.8	2.2
0.109	1367 - 80	http://i.msn.com/m/8/c/scheme0b.css	5855	3.4	4.5
0.109	1368 - 80	http://i.msn.com/m/8/c/site-win-ie6.css	2589	2.1	2.6
0.140	1367 - 80	http://i.msn.com/m/8/c/home-win-ie6.css	1187	1.5	2.1
0.140	1368 - 80	http://c.msn.com/c.gif	577	0.3	0.7
0.187	1368 - 80	http://sc.msn.com/c/portal/logo/full/msft.gif	1755	0.6	0.9
0.187	1367 - 80	http://global.msads.net/ads/1/0000000001_0000000...	658	0.4	1.0
0.328	1368 - 80	http://msimg.com/m/8/tab-back_line_tall.gif	595	0.4	0.8
0.328	1367 - 80	http://sc.msn.com/c/portal/tabs/tabFrontOn.gif	1575	0.6	0.9
0.343	1367 - 80	http://sc.msn.com/c/portal/tabs/tabMidOn.gif	1227	0.6	0.9
0.343	1368 - 80	http://sc.msn.com/c/portal/tabs/tabEndOff.gif	952	0.6	0.9
0.359	1367 - 80	http://www.passportimages.com/1033/signin.gif	1174	0.6	0.9
0.359	1368 - 80	http://global.msads.net/ads/1/0000000001_0000000...	3714	0.6	1.0
0.375	1367 - 80	http://sc.msn.com/c/portal/misc/search_arrow.gif	864	0.6	0.8
0.390	1368 - 80	http://sc.msn.com/3K/Y4EV94+2,Z3LFRU~Y{`LV3.gif	2768	0.6	1.0
0.390	1367 - 80	http://global.msads.net/ads/363/0000000363_00000...	3446	0.7	1.0
0.406	1368 - 80	http://sc.msn.com/c/portal/misc/tp.gif	487	0.2	0.6
0.421	1367 - 80	http://sc.msn.com/1J/+@V{508WB[1Z+,Q,3QY}QU.gif	8811	0.6	1.0
0.421	1368 - 80	http://sc.msn.com/2D/{2+[NA[[_2`MMW9Y+JT3.jpg	7087	0.7	1.0
0.437	1367 - 80	http://sc.msn.com/5X/QQFN-8{!}92KEN-DK,-CT4.jpg	2604	0.6	1.0
0.452	1368 - 80	http://msimg.com/m/8/tab-back_line_on.gif	761	0.7	1.2



# Use more parallel TCP ports



# Unblock Java Script

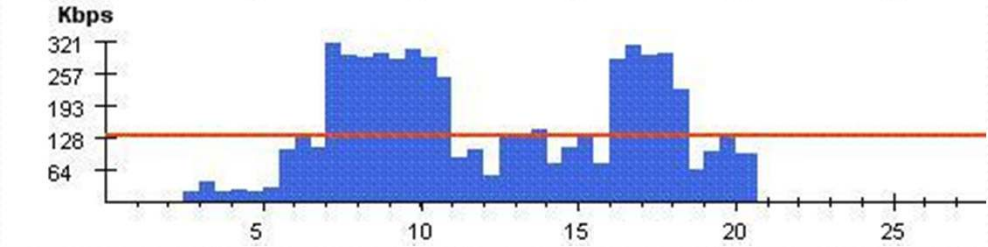
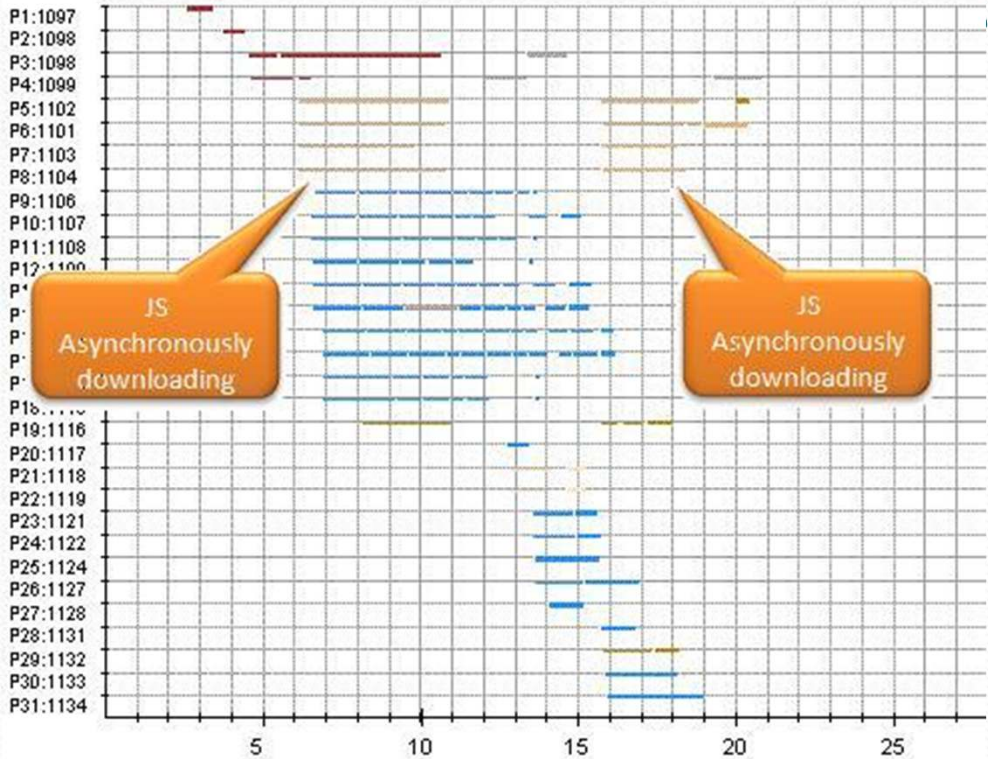


Standard JS downloads in Serial and creates bandwidth bottlenecks

Use a binding methodology to get around this issue



Client Ports



From: WR-Client VRTA

# JS Solution I

This method has been successfully used by the Windows Live Hotmail Team

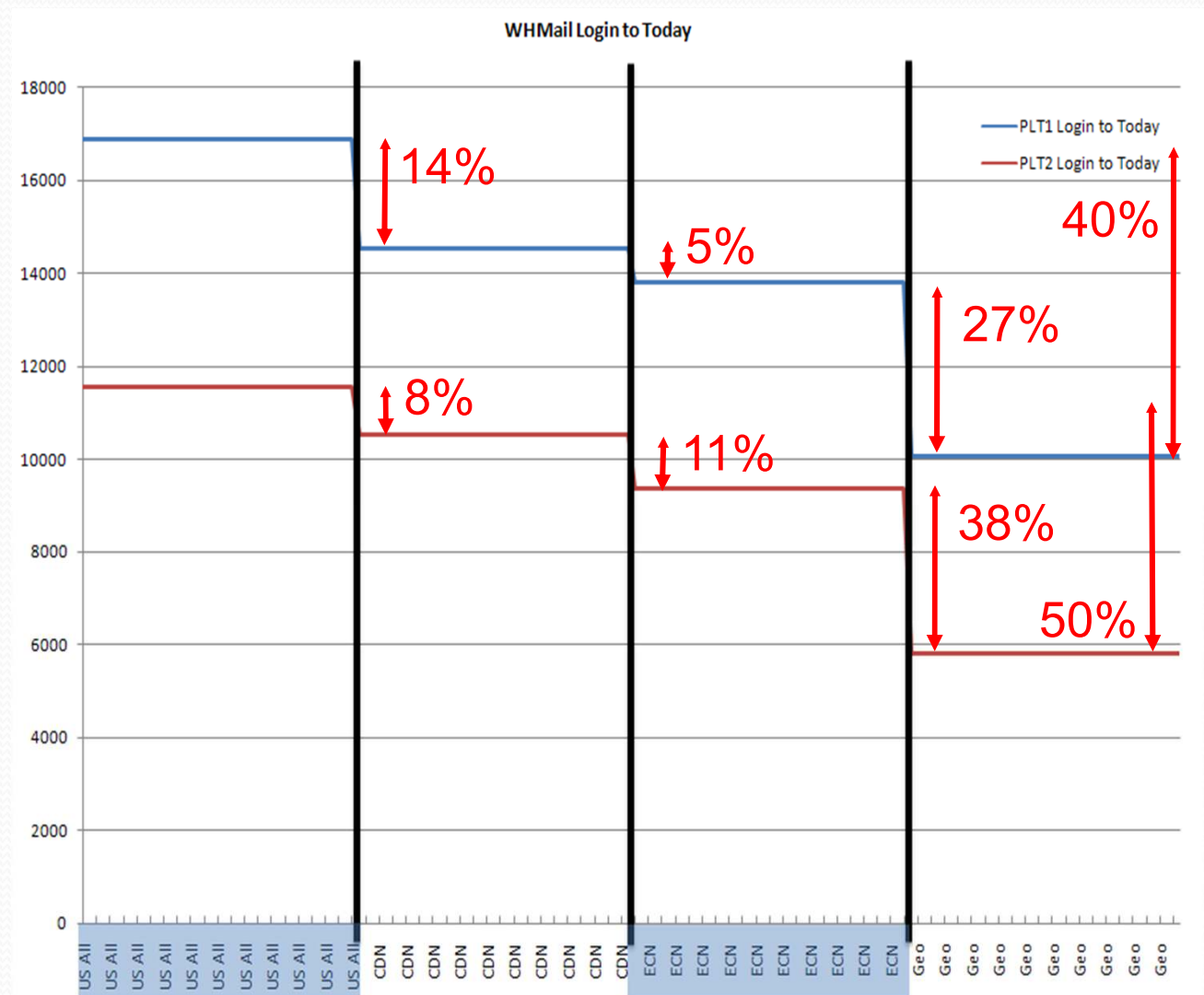
```

=====
function AsyncLoad()
{
var l = arguments.length;
for (var i=0;i<l;i++)
{
document.write("<script src='" + arguments[i] + "'></"
+ "script>");
}
}
AsyncLoad(
"file1.js",
"file2.js",
"file3.js");
=====

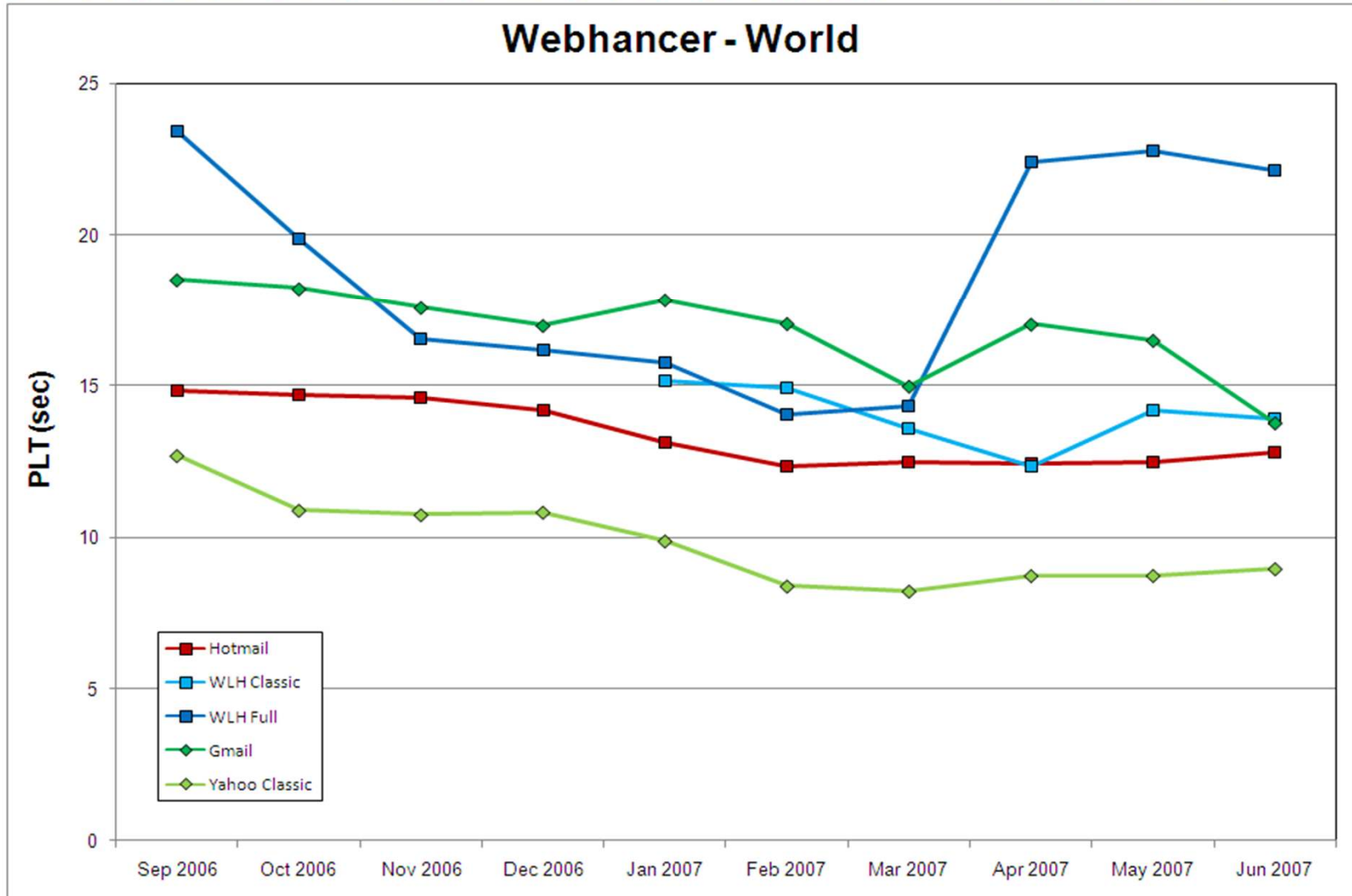
```

# Hotmail Login to Today page

- 40% drop in PLT<sub>1</sub>
- 50% drop in PLT<sub>2</sub>



# Windows Live Hotmail – Global



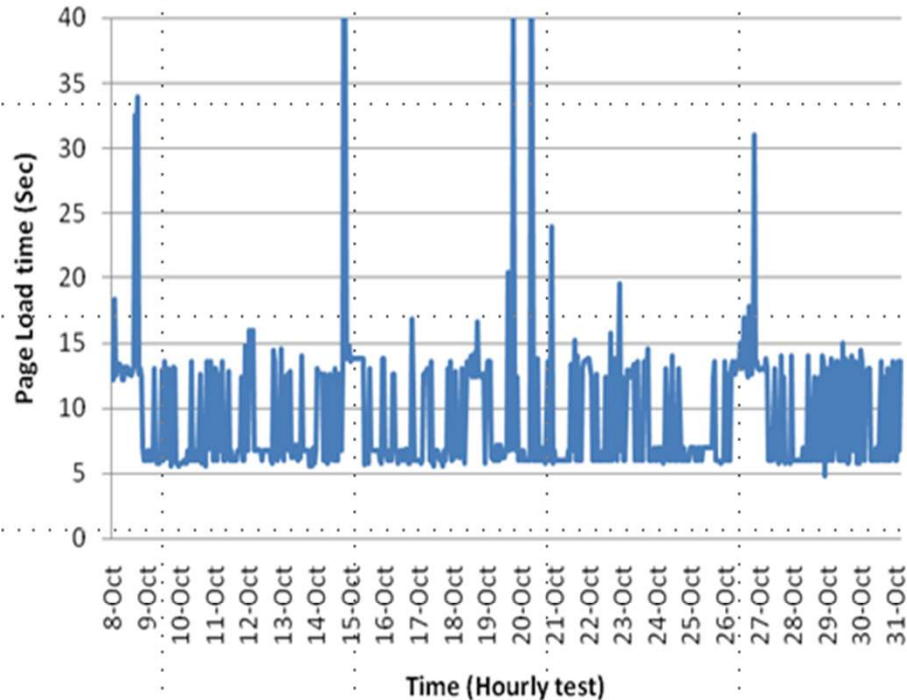
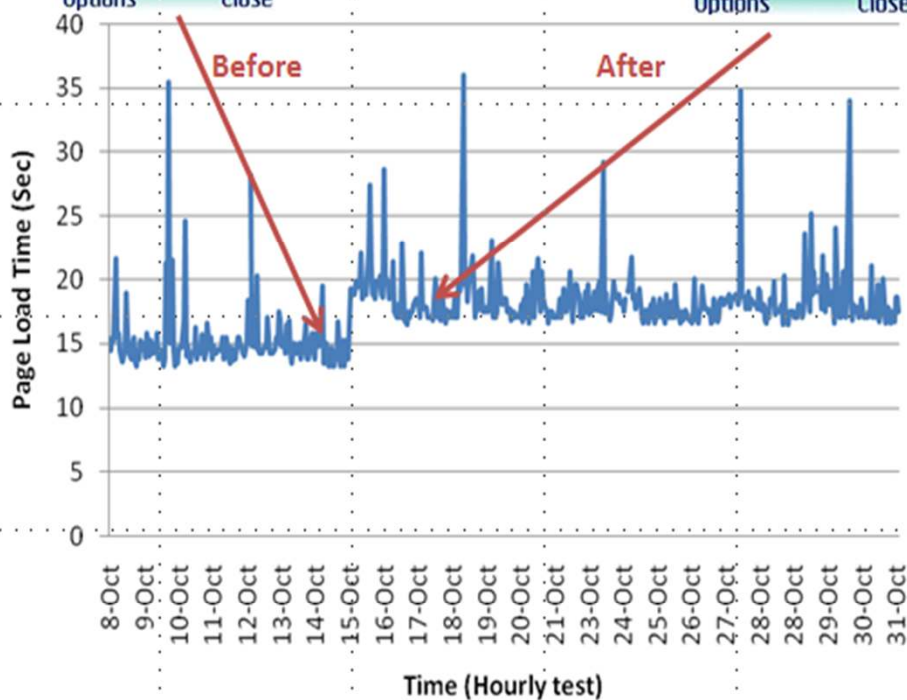
# WAP Portal Vodaphone/Nokia UK (London)



MSN Portal (UK)



BBC Portal (UK)



Reliability : 99.4%

75PLT: 18.5 sec  
(best time 13.2 sec)

StdDev: 2.8

Reliability: 99.8%

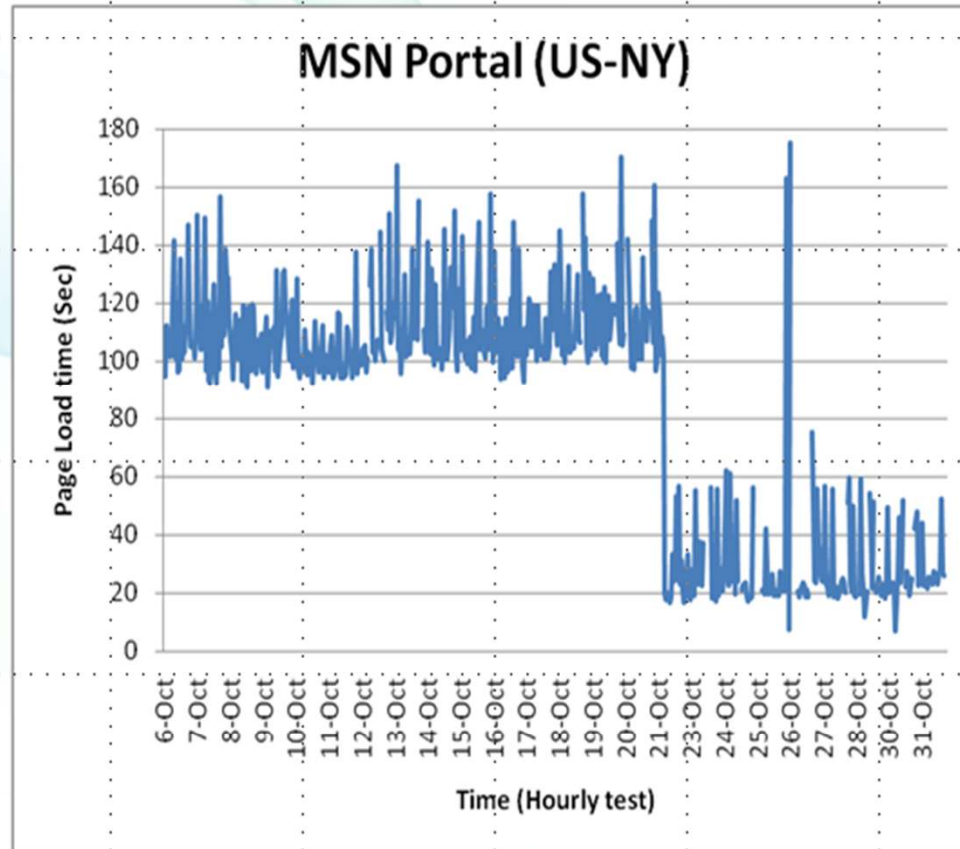
75PLT: 13.0 sec

StdDev: 5.1

**Observations:**

- On Oct 15 MSN UK Mobile portal 2.0 was released (that increased page weight from 25K to 33K) and resulted in sharp shift in performance
- For BBC Portal – the PLT is 6 sec or 13 sec (because of different route taken by the carrier - **carrier was notified**)

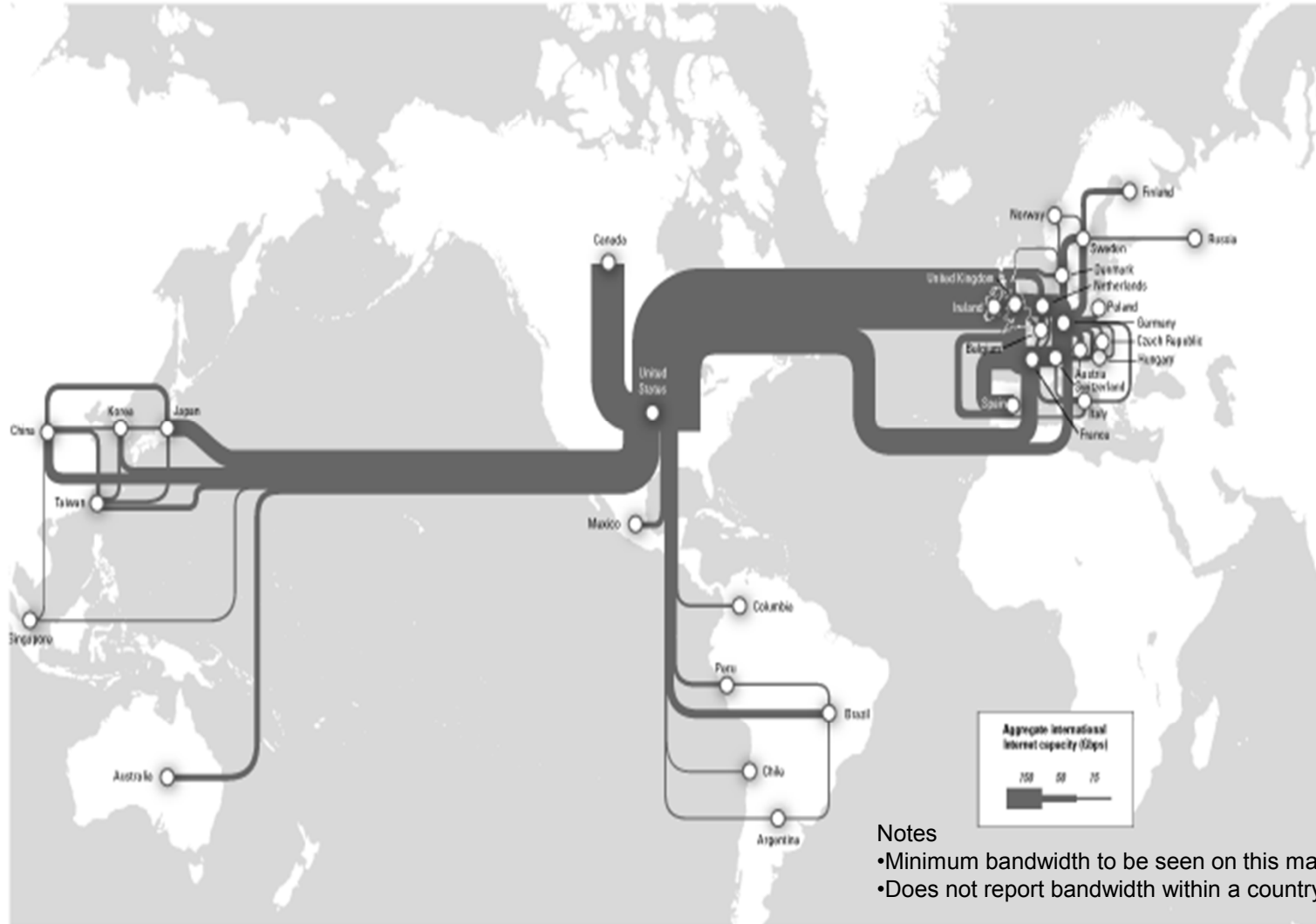
# MSN WAP Portal on ATT/NYC Moto Razr v3xx



## Observations:

- Moto Razr v3xx sends an user agent string IE6– due to which anybody hitting the PC version of MSN.com page instead of the mobile MSN page. The test was adjusted on Oct 22 to hit the MSN Mobile page

# International Internet Routes

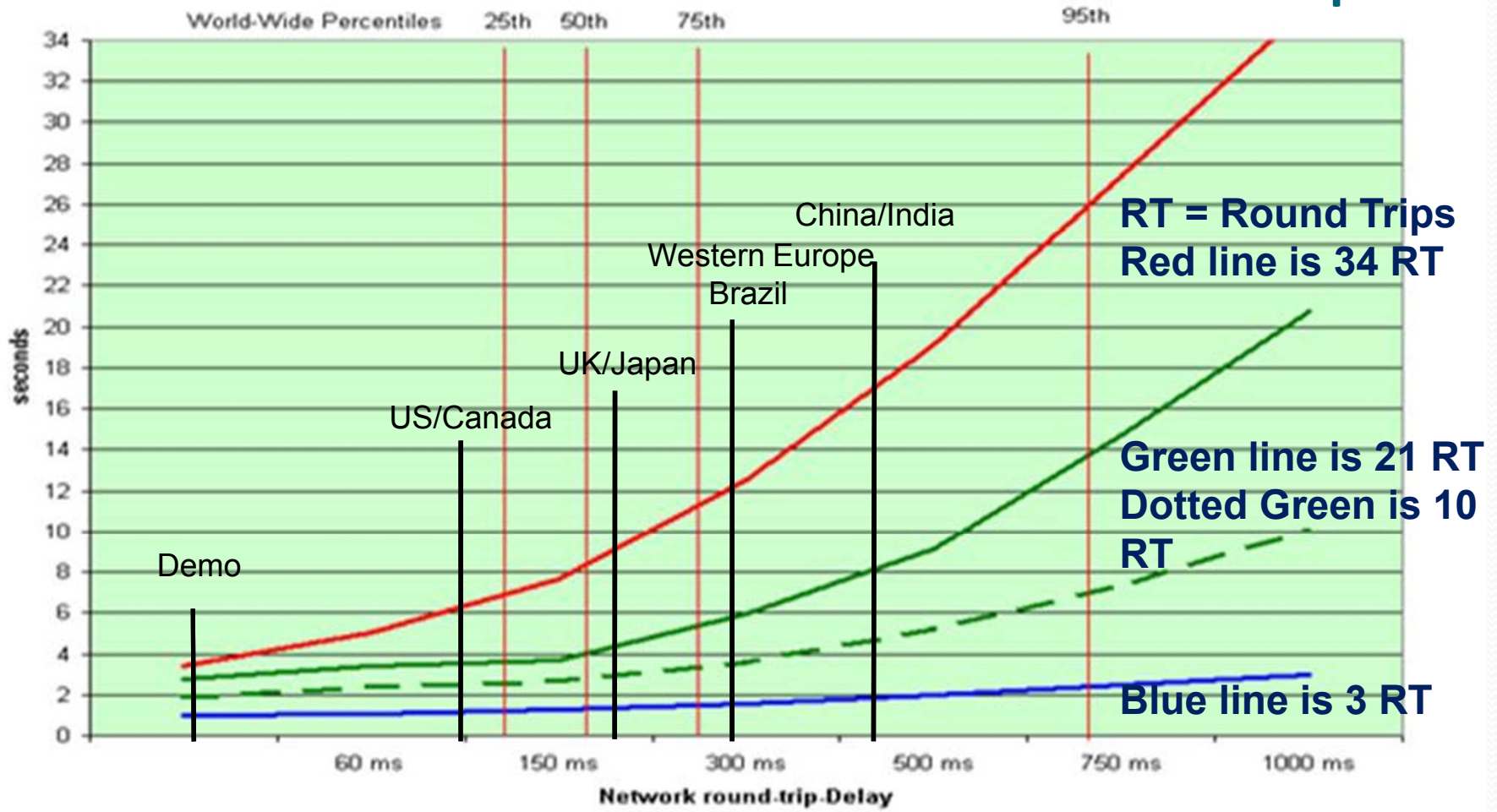


## Notes

- Minimum bandwidth to be seen on this map is 14 Gigs
- Does not report bandwidth within a country



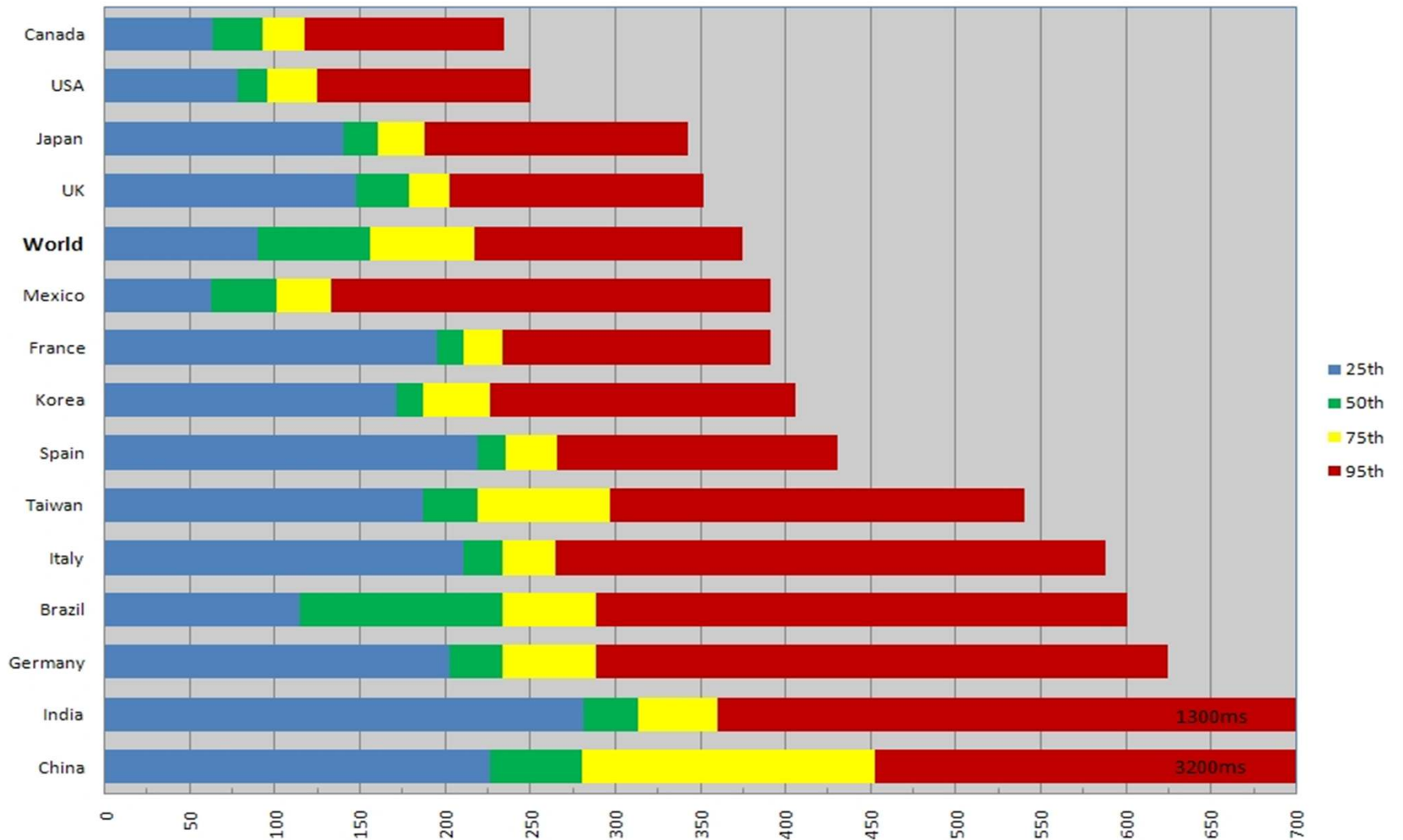
# Latency and the impact to page load times based on number of round trips



Source data for timings is 75<sup>th</sup> percentile for country in question from: <http://msncore/performance/netsmart/Netstats.asp>  
Microsoft Confidential

# Network Round trip delays...

Network Round Trip Delay from SVC (WH, BB only) Feb 2007



# Additional info

- The PingER Project for RTTs <http://www-iepm.slac.stanford.edu/pinger/>
- General stats on internet usage <http://www.internetworldstats.com/stats.htm>
- Internet submarine cables reference <http://www.telegeography.com/maps/index.php>

# Conclusion

- Measurements
  - Scenarios
  - Methodology
  - Passive – Data Collection / Reports
  - Active – Alerts / Triggers
  - Threshold / criteria
  - Triggers
- Improvements
  - Rigor
  - Data driven



# Microsoft India is hiring!

Contact: [Mukesh.Jain@microsoft.com](mailto:Mukesh.Jain@microsoft.com)

# Questions?

**Mukesh Jain**  
**Mukesh.Jain@microsoft.com**

# Why is Performance important?

- More online activities
- Competitors are faster
- Users have more options
- Attracting and retaining users
- We cannot overcome speed of light
- Release it and then fix it – no longer an option – we may lose Mind-Share and we may not get second chance

# Analyze Root Cause

- Find Root Cause of the problems
  - Analyze the data from Measure phase
  - Identify vital few variables (x)
  - Perform correlation and regression analysis
  - Data stratification
  - Use 5 Why techniques
  - Hypothesis testing
  - Sources of variation
  - Use Cause-Effect diagram
  - Plot data on graphs (trend, releases)
  - Special Cause / Common Cause

*# of files on the site is 10, some of these files can be combined, 2 files are not compressed. Majority of the users who abandon the site are from UK (Latency) and Dial-up users from US (slow connection). The problem started to happen from March 01 (when release 2.2 went live)*



# Improve Performance

- Improvement strategy & Plan
- Improvement solution selection
  - Generate ideas – Involve diverse team
  - Identify and rank solution alternatives
  - Pilot solutions (<http://experiments>) and select final solution
- Test and implement final solution
- Communication plan
  - Track improvement, monitor trend
  - Share success/failure stories → Best Practices

# Control – Sustain Performance

- Process Monitoring/Control Plan
  - Mistake Proofing (Poka-Yoke)
  - Control Chart
  - Response Plan
- Quality of Service (QoS) Program
- Document standard process/procedure
- Train resources
- Share Learning
- Mindset
  - If somebody else can find defect in your work, why can't you yourself find it

# Demo

- HttpWatch
- WanSim
- Yslow (Firebug)
- Bing Toolbar\*
- PLTWeb\*
- WebRunner\*
- WebHancer\* → Bing toolbar
- Keynote
- Gomez



# Performance/Load/Stress Testing

- Performance testing
  - User-Scenarios testing (typical case & best case)
  - Establish Baseline & perform trend analysis
  - Detect performance issues
  - Tools: WebRunnerPLT, WANSim, HttpWatch, etc.
- Load testing (volume/longevity/endurance)
  - Expected MAX # of concurrent users
  - Volume of data
  - Very Long active sessions
- Stress testing (negative testing)
  - What happens with the load exceeds significantly OR system goes thru serious resource constraints/failures
  - Does the system gracefully recovers from failure?

# Define Perf User Scenarios

- Understand the User (Voice of the Customer)
  - First time visitor / Guest user / authenticated user
  - Returning users with cache/no cache
  - User on the same site/session
  - User from other MS sites/domains (w/ passport)
  - User Demographics (Geo, Home/Office, machine config, consumer/info worker/social, connection speed)
  - Typical user transactions, Back/Forward/Refresh usage
- Perf goal
  - Regional competitor performance
  - What is acceptable performance
- Do not use “it should be fast”, try “*JP broadband users should be able to get the page in 4 seconds, when they visit for the first time (PLT<sub>1</sub>), 2 seconds for PLT<sub>2</sub>*”
- **If it appears slow – it is slow**, irrespective of what the data says

# Measure Performance

- Measurement Process – Data Collection Plan
  - Testing, Monitoring, Measurement system analysis, sampling
- Identify Key measures/drivers of performance
  - $Y = F(X_1, X_2, X_3, X_4, \dots)$
  - Ishikawa (Fishbone) diagram – Cause & Effect diagram
- Internal / External Benchmarking
- Baseline Current performance & Impact
  - Identify and measure current performance and its impact on customer, collect more data if required

*On 300kbps connection & 300ms Round Trip delay, it takes 6 seconds to load the page for the PLT1 case, 20% of our user abandon the page before it loads.*

*Web Page X have 2 HTML files, 3 .js files, 3.css, 5 images, the web App opens 2 parallel tcp ports to download them.*

# Types of Performance Measurements

- Client UI Response Time
- Server Response Time
- Load/Stress
- Bytes over wire/Throughput
- Availability
- Latency (anywhere in the world)
- Browser Processing & Rendering time
- User Machine - Resource utilization
- **Perceived Performance**

*If the user feels the product is slow, your product **is** slow – no matter what our data says*

# Measure: Key variables for Performance

- Factors attributing to Web page performance (Page load time)
  - # of Files, Static/Dynamic content
  - Bytes download / Compression
  - DNS Lookup time / Redirects
  - Peak hours / Load & Stress
  - User spread / Global?
  - Data Centers / CDN
  - Multiple versions of the web app
  - Web Page Architecture (parallel/sequential download)
  - # of parallel TCP connections
  - Expiry dates / Keep-Alive
  - PLT<sub>1</sub> / PLT<sub>2</sub> (Caching?)
  - JS Processing/CSS
  - Strict HTML
  - Typical User config (machine, connection speed, Geo, etc.)



# Measure

- Other Measurements
  - # of unique users
  - # of page views
  - TCP connect time
  - # of Errors
  - % of people on PLT<sub>1</sub> / PLT<sub>2</sub>
  - Click-Thru-Rate (CTR)
  - Abandon rate
    - Incomplete
    - Closed
    - Click-Away