

Web Performance Acceleration by Caching Rendering Results

-NTT Network Technology Labs
Yuusuke Nakano, Noriaki Kamiyama, and Kohei Shiimoto
-Cybermedia Center, Osaka University
Go Hasegawa
-Graduate School of Information and Science Technology, Osaka University
Masayuki Murata and Hideo Miyahara

Copyright © 2015 NTT Corp. All Rights Reserved.

Background

Revenue for service providers depends on web performance

Translating Web Performance to Dollars and Cents
Speed is a feature, and it is not simply speed for speed's sake. Well-publicized studies from Google, Microsoft, and Amazon all show that web performance translates directly to dollars and cents—e.g., **a 2,000 ms delay on Bing search pages decreased per-user revenue by 4.5%**

Similarly, an Aberdeen study of over 160 organizations determined that an extra **one-second delay in page load times led to 7% loss in conversions, 11% fewer page views, and a 16% decrease in customer satisfaction!**

Ilya Grigorik, High Performance Browser Networking, O'Reilly, 2013

Copyright © 2015 NTT Corp. All Rights Reserved. 2

Purpose of research

Reduce the longest portion of download time for files making up a web page
Blocked time is the longest (caused if an object has dependencies on other objects)
Purpose: accelerate web performance by reducing blocked time

Breakdown of download time

		Averages of each portion of download time for each host (ms)				
		Host	California	Argentina	France	Australia
Blocked	Blocked	255.36	552.92	418.45	377.77	
	DNS	0.68	0.30	0.43	0.85	
Connect	Connect	13.86	25.01	81.32	21.18	
	Send	0.02	0.01	0.01	0.02	
Wait	Wait	195.10	392.28	235.04	292.07	
	Receive	53.10	78.61	46.43	72.43	

Copyright © 2015 NTT Corp. All Rights Reserved. 3

Proposed method

Cache rendering results in network

- Rendering and cache function renders and caches in network
- Special browser gets rendering results from function and shows page

Copyright © 2015 NTT Corp. All Rights Reserved. 4

Problem in proposed method

Different objects between webpages with same URL
→ **EVERY CACHED RENDERING RESULT CAN NOT BE USED FOR EVERY USER**

Different objects in webpages with same URL

Ways to solve

- Cache and reuse rendering results of static objects
- Download dynamic objects by special browser

→ REUSE PRE-RENDERED IMAGES OF STATIC OBJECTS
→ REDUCE THE NUMBER OF DOWNLOADS & RENDERING OBJECTS
→ REDUCE THE NUMBER OF DEPENDENCIES
→ REDUCE BLOCKED TIME

Copyright © 2015 NTT Corp. All Rights Reserved. 5

Details of proposed method

Steps of method

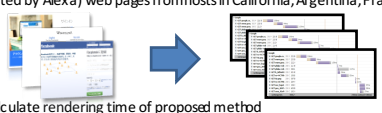
- The rendering and cache function caches the rendering result of static objects in a web page by extracting differences between web pages with the same URL when it receives multiple requests for the URL
- The function sends the cached rendering result of the static portion to the special browser in accordance with a request from the special browser
- The special browser shows the final rendering result in its screen by getting the dynamic portion from web servers and inserts it into the rendering result of the static portion

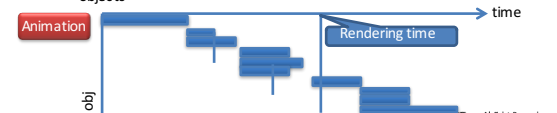
Copyright © 2015 NTT Corp. All Rights Reserved. 6

Evaluation method

Calculate rendering time of proposed method

Calculation method

1. Measure download time and start time of downloading of objects for 999 popular (rated by Alexa) web pages from hosts in California, Argentina, France and Australia.
 
2. Calculate rendering time of proposed method
 - I. Extract dynamic objects by finding different size objects downloaded by the 4 hosts that have the same URL
 - II. Calculate rendering time for dynamic objects considering dependencies among objects



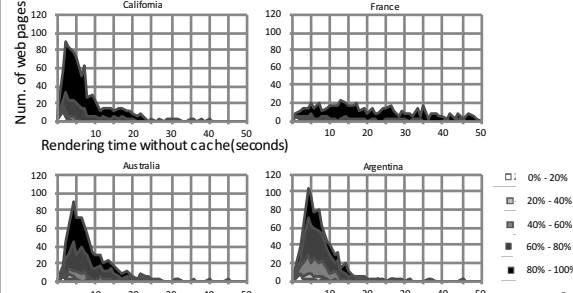
Copyright © 2015 NTT corp. All rights reserved. 7

Evaluation result

Reduced rendering time in Australia and Argentina where RTT is long

Rendering time histogram of every second for each host

Each area: percentage of rendering time with cache amongst rendering time without cache



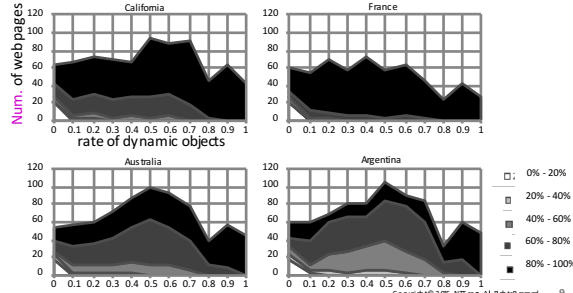
Copyright © 2015 NTT corp. All rights reserved. 8

Evaluation result

Reduced rendering time of web pages whose dynamic object rates are less than 80%

Dynamic object rate histogram of every 0.1 for each host

Each area: percentage of rendering time with cache amongst rendering time without cache



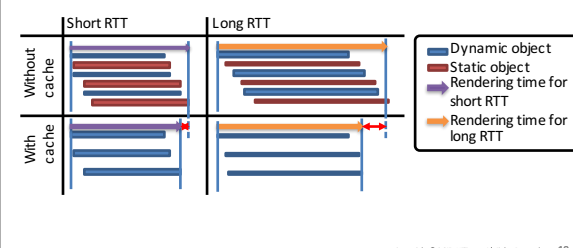
Copyright © 2015 NTT corp. All rights reserved. 9

Discussion

Cause of large reduction of long RTT hosts

Long RTT

- Long download time for each object
- Many dependencies
- Large reduction



Copyright © 2015 NTT corp. All rights reserved. 10

Conclusion and future work

- Method for accelerating web rendering time by caching rendering results
- Calculate rendering time of proposed method
 - Large reduction for long RTT hosts
 - Large reduction for webpages whose dynamic object rates are less than 80%
- Future work
 - Implementation and evaluation

Copyright © 2015 NTT corp. All rights reserved. 11