Performance Optimizations for Wireless Wide-Area Networks

Rajiv Chakravorty

Suman Banerjee

Pablo Rodriguez

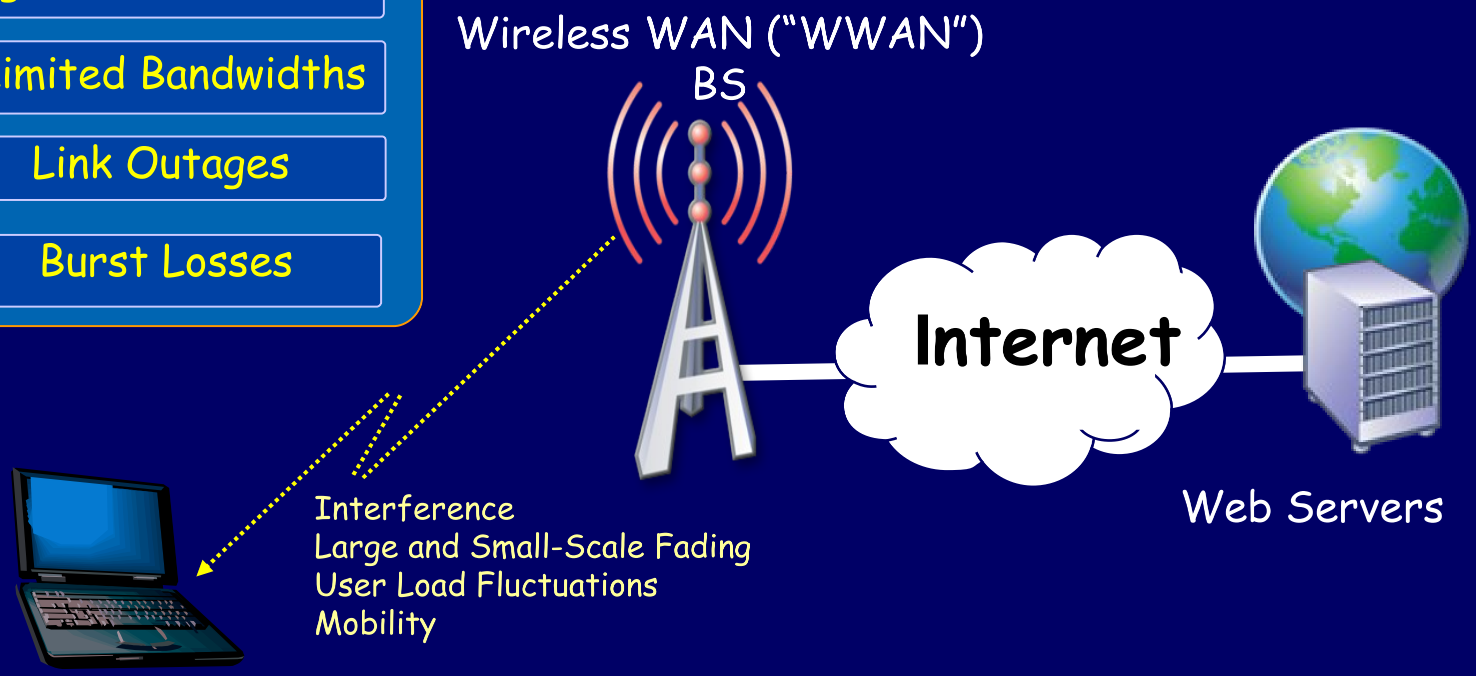
Julian Chesterfield

Ian Pratt

Motivation

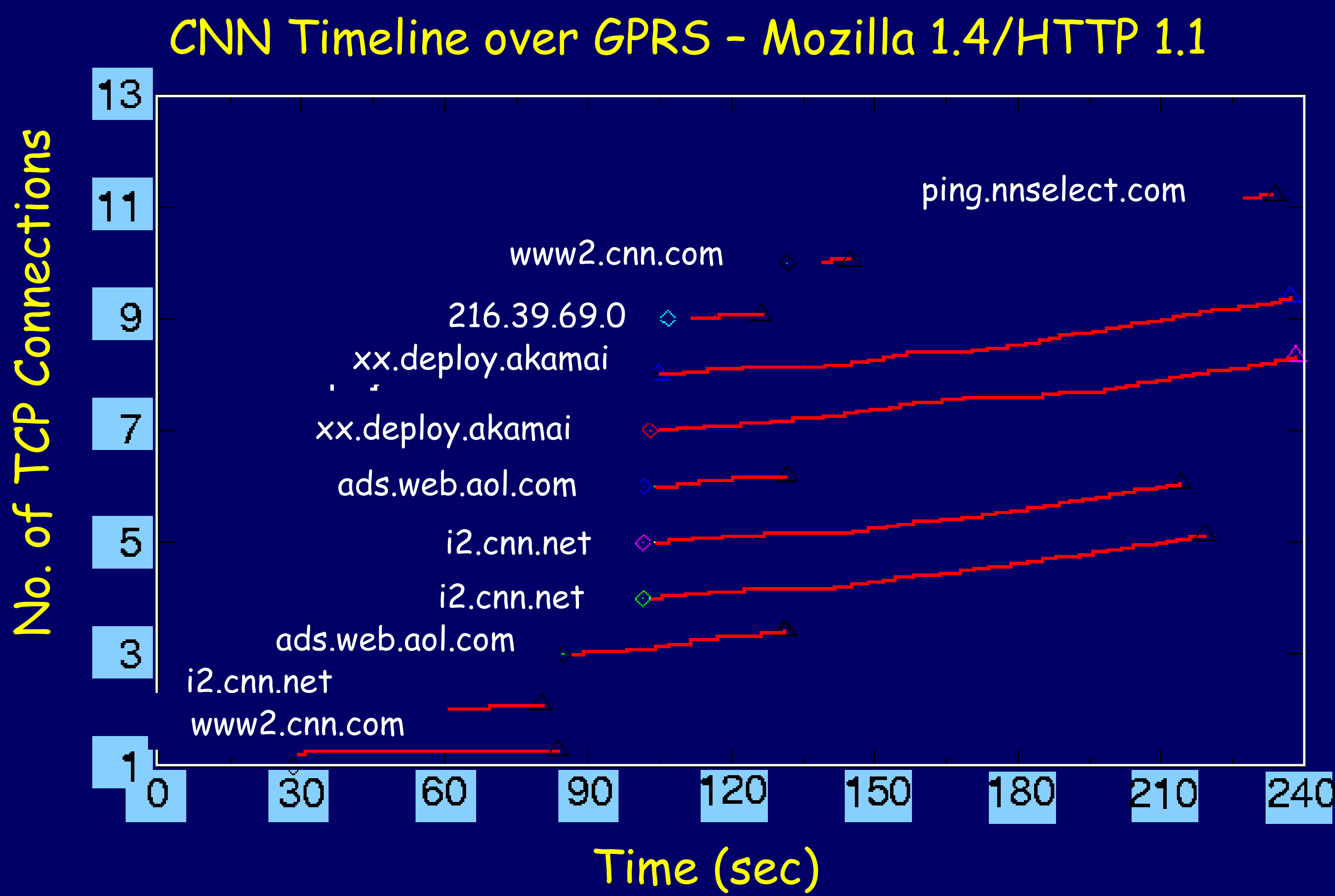
WWAN Links:

- High/Variable RTT
- Limited Bandwidths
- Link Outages
- Burst Losses



User Experience in "WWANs" significantly different from the relatively stable 802.11 WLAN

Why is the Web so Slow?



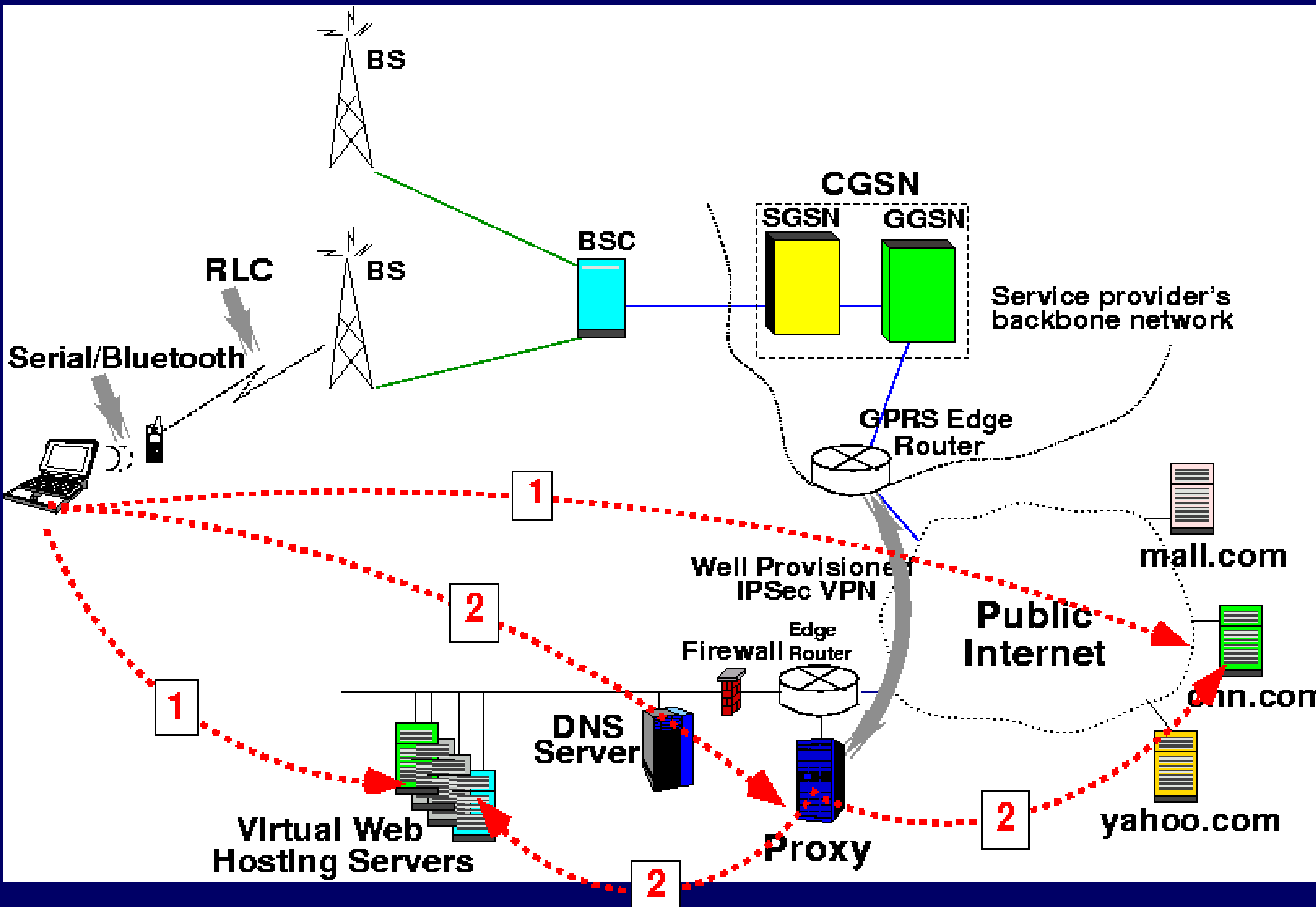
CNN Download takes well over 3 mins…


Ask Why
Mobile Web
sucks?

Three Main Contributions

- Benchmark— Web Browsers, Protocols and techniques over Wireless WANs
- Implement & Study— Range of Optimizations at different layers of the stack and their cross-layer impact on applications
- Introduce— A methodology for realistic and repeatable web experiments over WWAN

Cambridge Infrastructure and Testbed

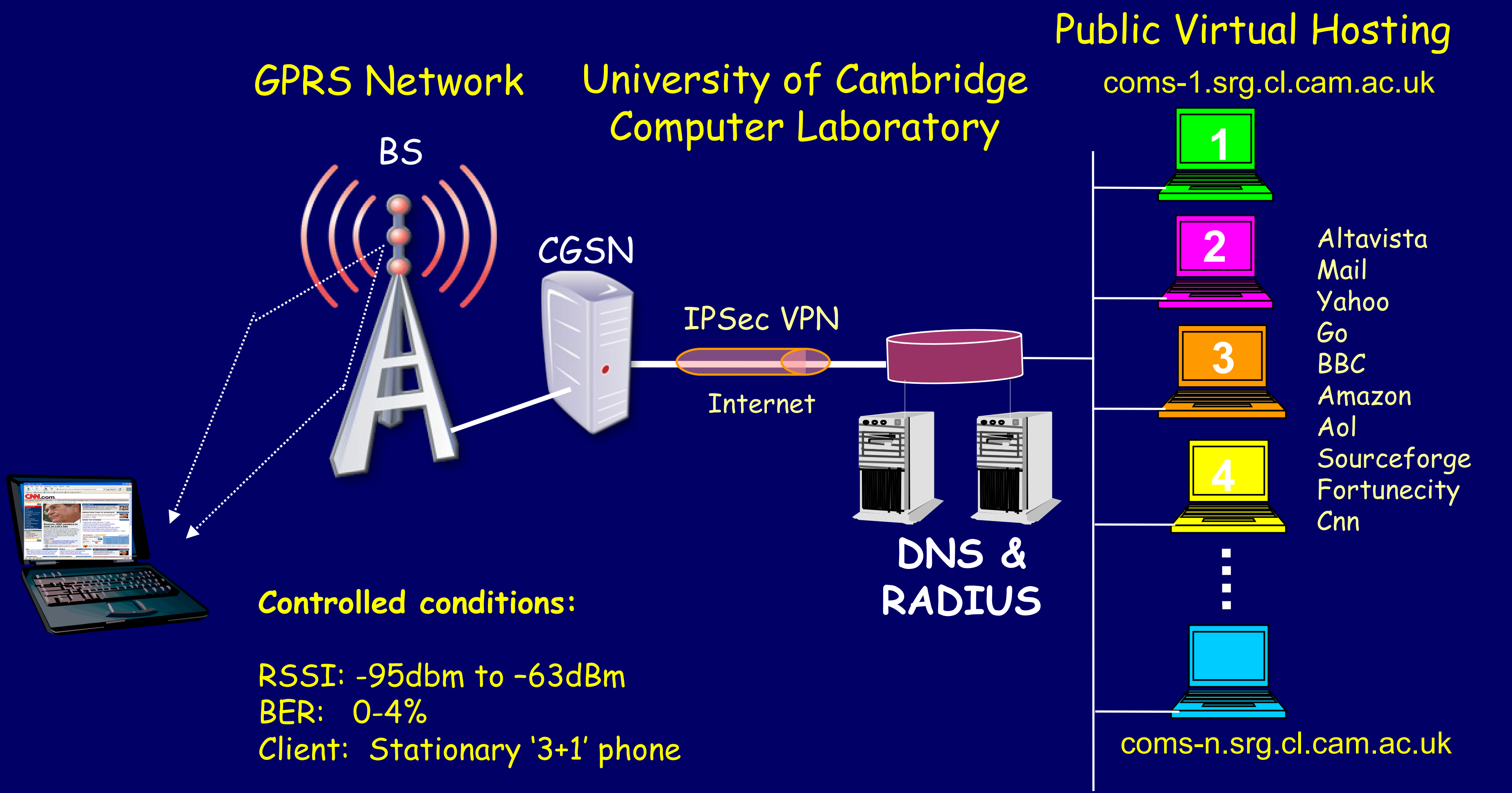


Experimental Methodology

- We use Virtual web-hosting
- Contents of some popular Web-sites change very frequently (e.g. CNN changes in minutes)
- We replicate the key components of some popular Internet Web Sites in our Lab
  (Replicate both Volume and Structure)

Virtual Web-hosting allows Web experiments to be Repeatable and Reproducible

Cambridge Open Mobile Virtual Hosting Infrastructure



Controlled conditions:

RSSI: -95dbm to –63dBm
BER: 0-4%
Client: Stationary '3+1' phone

Content Selection

- ✓ Web-sites Ranked in → 100hot.com
- ✓ Choice based on Content "diversity"

Web Page	No. of Servers	Object Count	Sum (KB)	Avg. (KB)
Mail	4	11	37KB	3.3KB
Yahoo	6	16	61KB	**3.8KB**
Amazon	3	42	92KB	2.2KB
CNN	6	67	187KB	**2.8KB**

"Diversity" = Number of Servers, Object count/size, Content types, volume and their distribution

Performance Benchmarks

HTTP 1.1

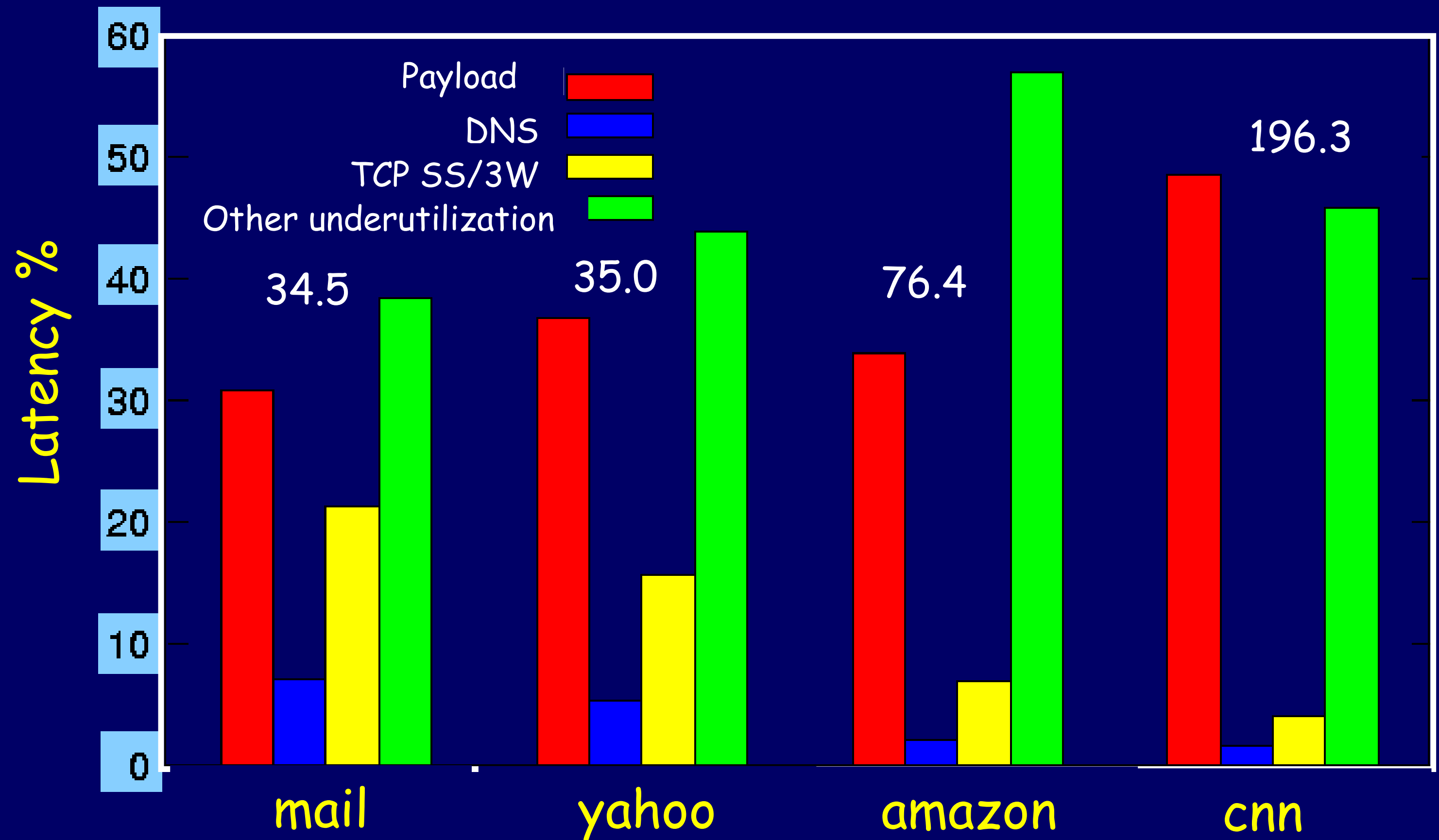
Website	Sum (KB)	HTTP T'put (Kbps)	%Dgr.
MAIL	37KB	8.5	-79%
YAHOO	69KB	13.8	-66%
AMAZON	92KB	9.6	-75%
CNN	187KB	7.6	-81%

TCP

File	TCP T'put (Kbps)	%Dgr.
5KB	18.1	-54%
50KB	29.7	-25%
100KB	30.5	-23%
200KB	29.9	-25%

When TCP tuned to work relatively well, why is the performance of HTTP 1.1 worse?

Factoring (Under) Performance



Payload and default HTTP 1.1 behavior impacts web downloads over WWANs

Multi-layer Optimizations

Application Layer

- HTTP Pipelining
- Content Compression
- Caching/Delta Encoding

Session Layer

- Optimizing Browser Conn.
- DNS/URL-Rewriting
- Server-side `Parse-n-Push'

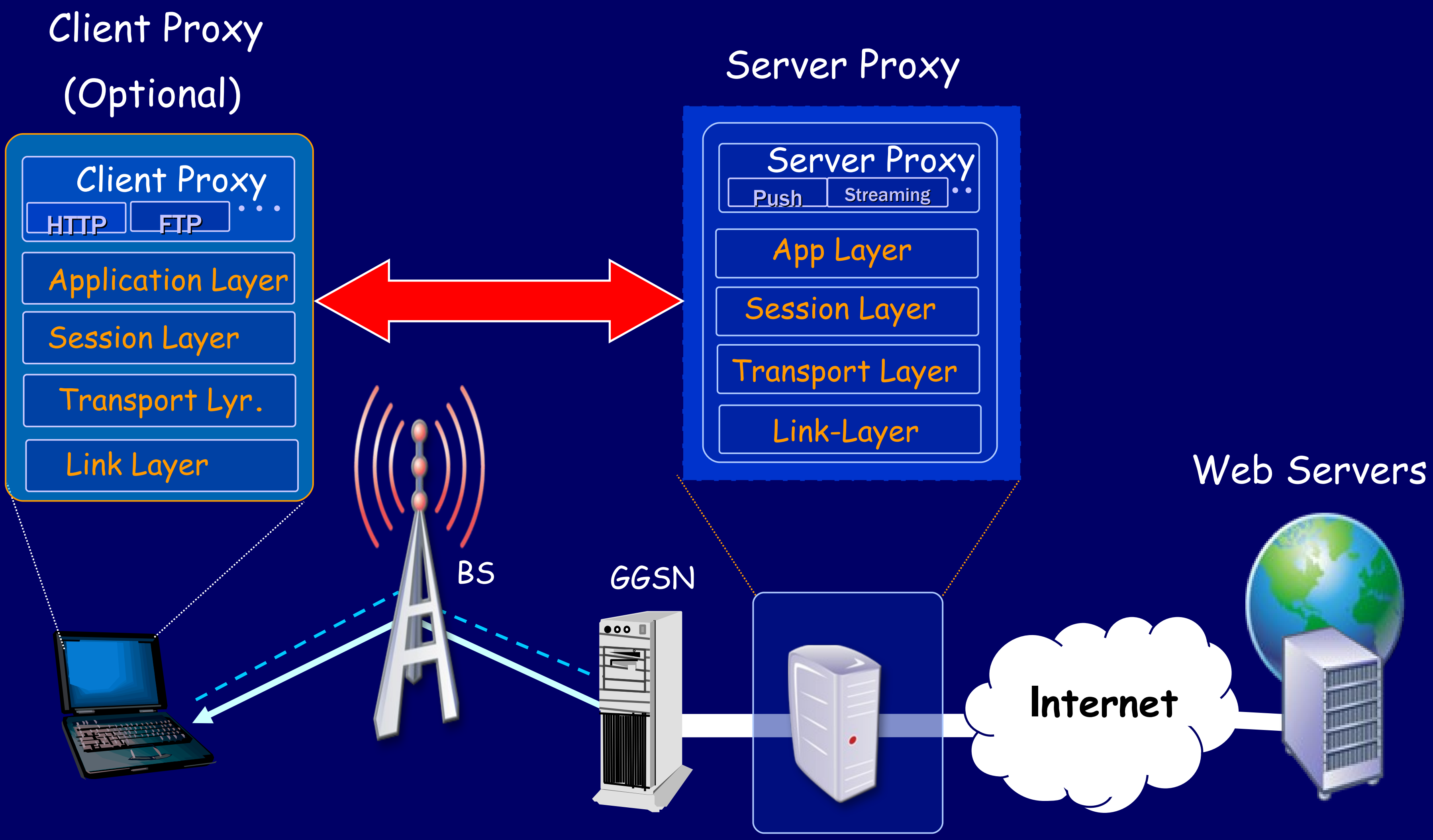
Transport Layer

- Link-adapted TCP Variant
- Custom Transport Protocol

Link Layer

- Dynamic FECs
  (trace-driven simulations)

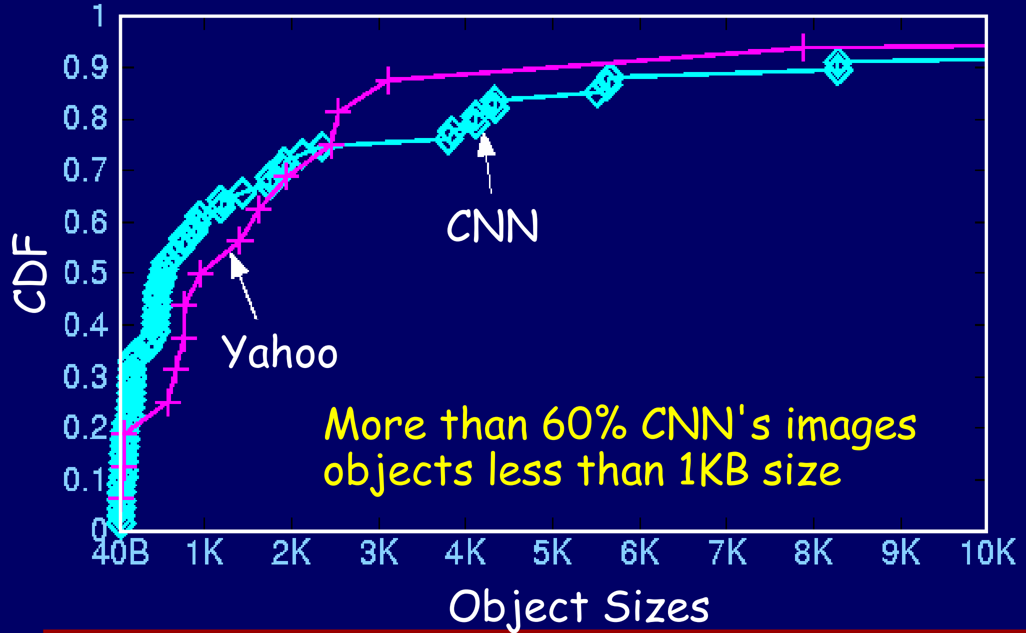
Proxy-based Optimizations



Dual-Proxy Mode

Compression Gains

Website	Mail	Yahoo	Amazon	Cnn
Compression	59%	59%	57%	52%
% Improvement	18%	41%	18%	12%

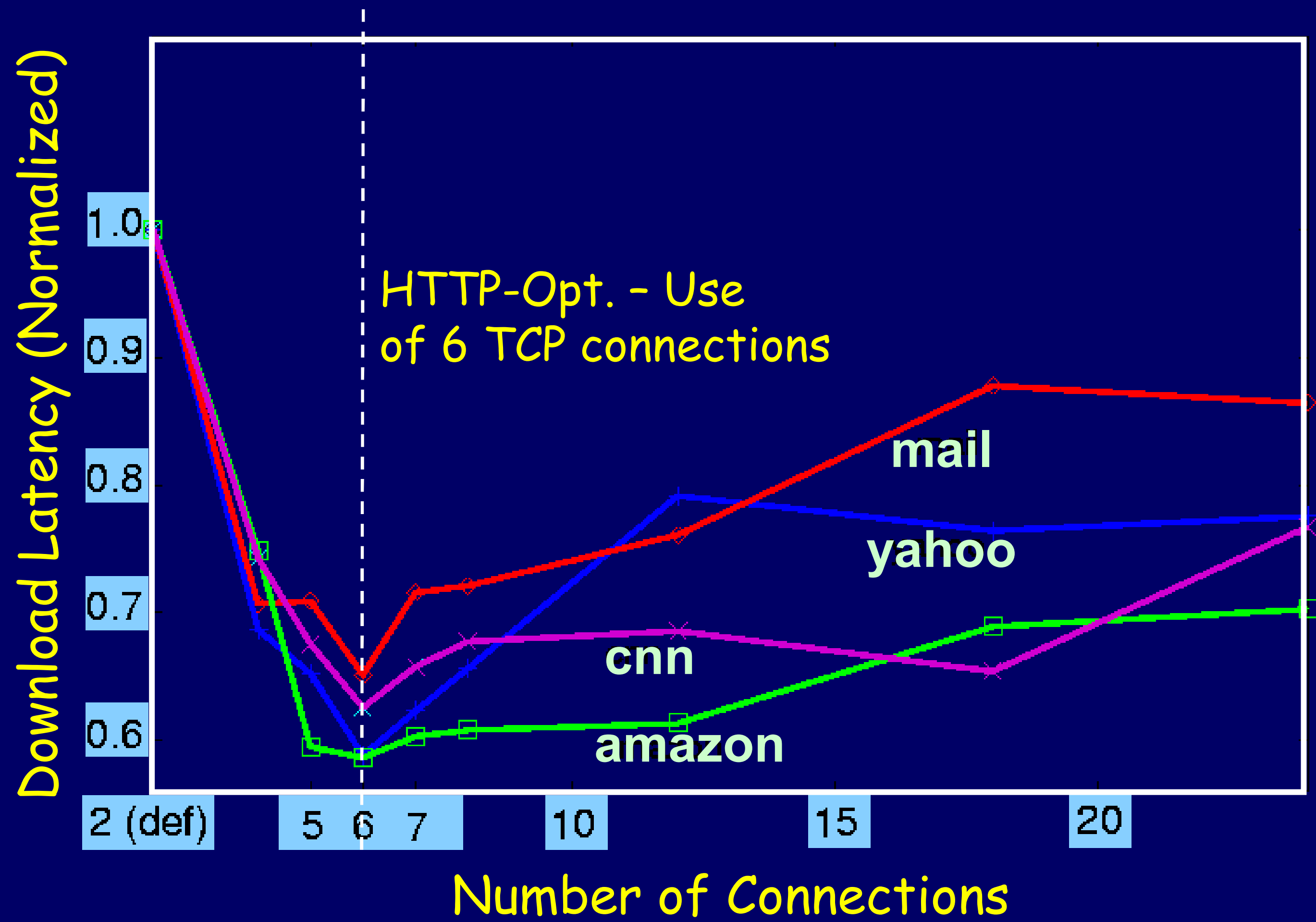


- Yahoo provides the best improvements through compression whereas CNN the least.

- Yahoo offers the best compression per image object size.

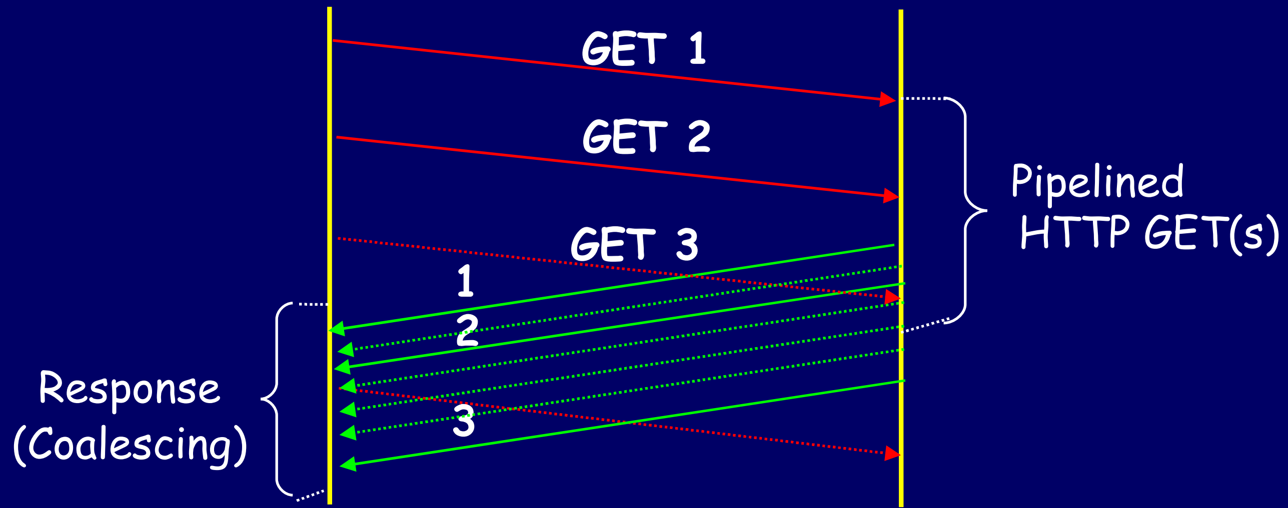
Compression Gains depends largely on the content characteristics of websites

Tuning Browser Performance



Optimal Connection Setting in Browsers improves performance by 25 - 45%

HTTP Pipelining

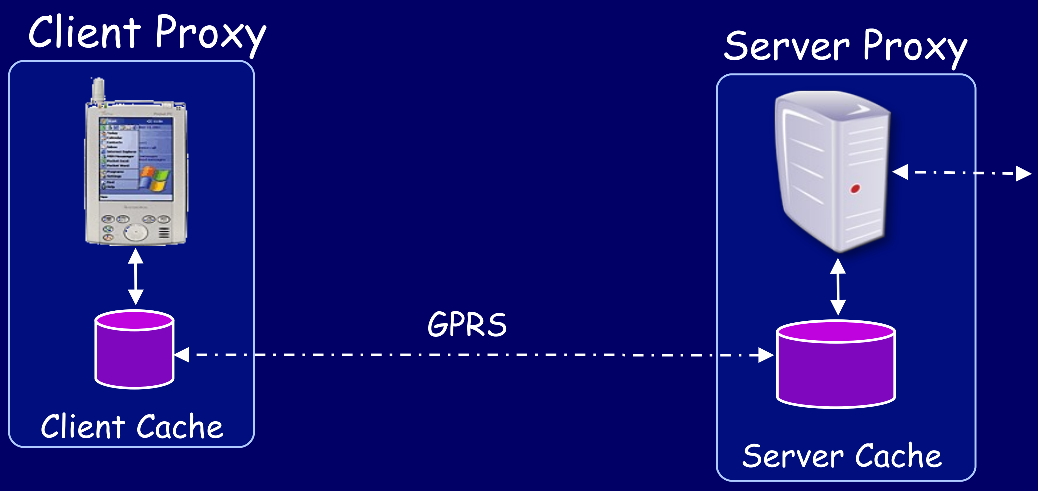


Website	Mail	Yahoo	Amazon	Cnn
HTTP-opt.	38%	31%	43%	37%
HTTP-pipe	56%	35%	49%	55%

HTTP Pipelining improves utilization with 5 - 20% additional gain over HTTP-opt.

CHK-based Caching

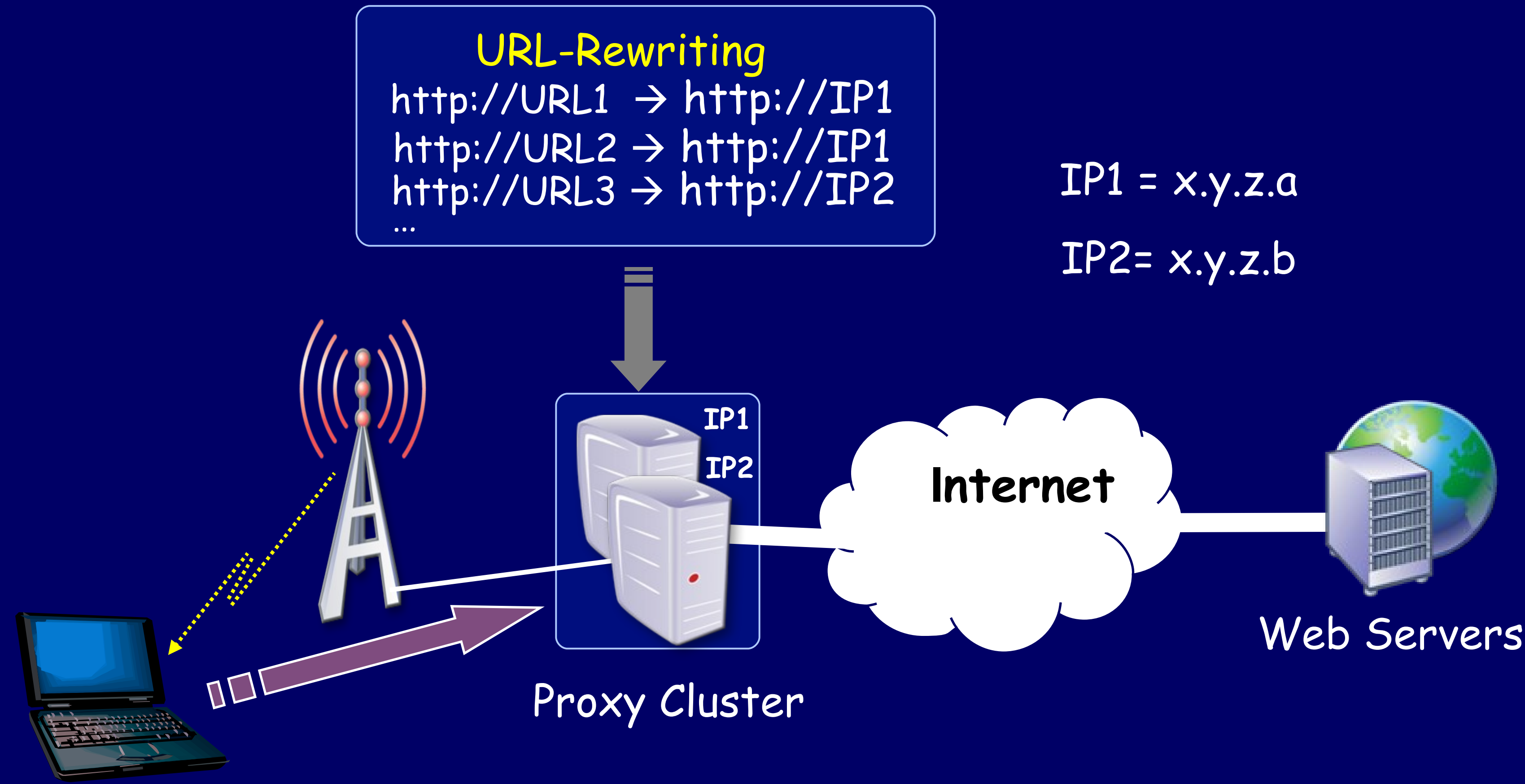
Improves client cache hit rates, reduces redundant data transfers and optimizes bandwidth requirements



- ✓ Cache objects with SHA-1 fingerprint (CHK)
- ✓ URL-to-CHK mapping (offers "alias" protection)
- ✓ Deltas - Send "updates" as the difference

Caching and delta-encoding improves perf. By 5 - 9% depending on the web-site

URL-Rewriting



URL/DNS-Rewriting eliminates DNS lookups to provide 5 - 9% additional gain

Transport Layer Solutions

Link-Adapted TCP (TCP-WWAN)

- ✓ No TCP slow-start
- ✓ Prevent Spurious Timeouts
- ✓ Enhanced Recovery
- ✓ Avoid Excess Queuing

Custom Transport (UDP-GPRS)

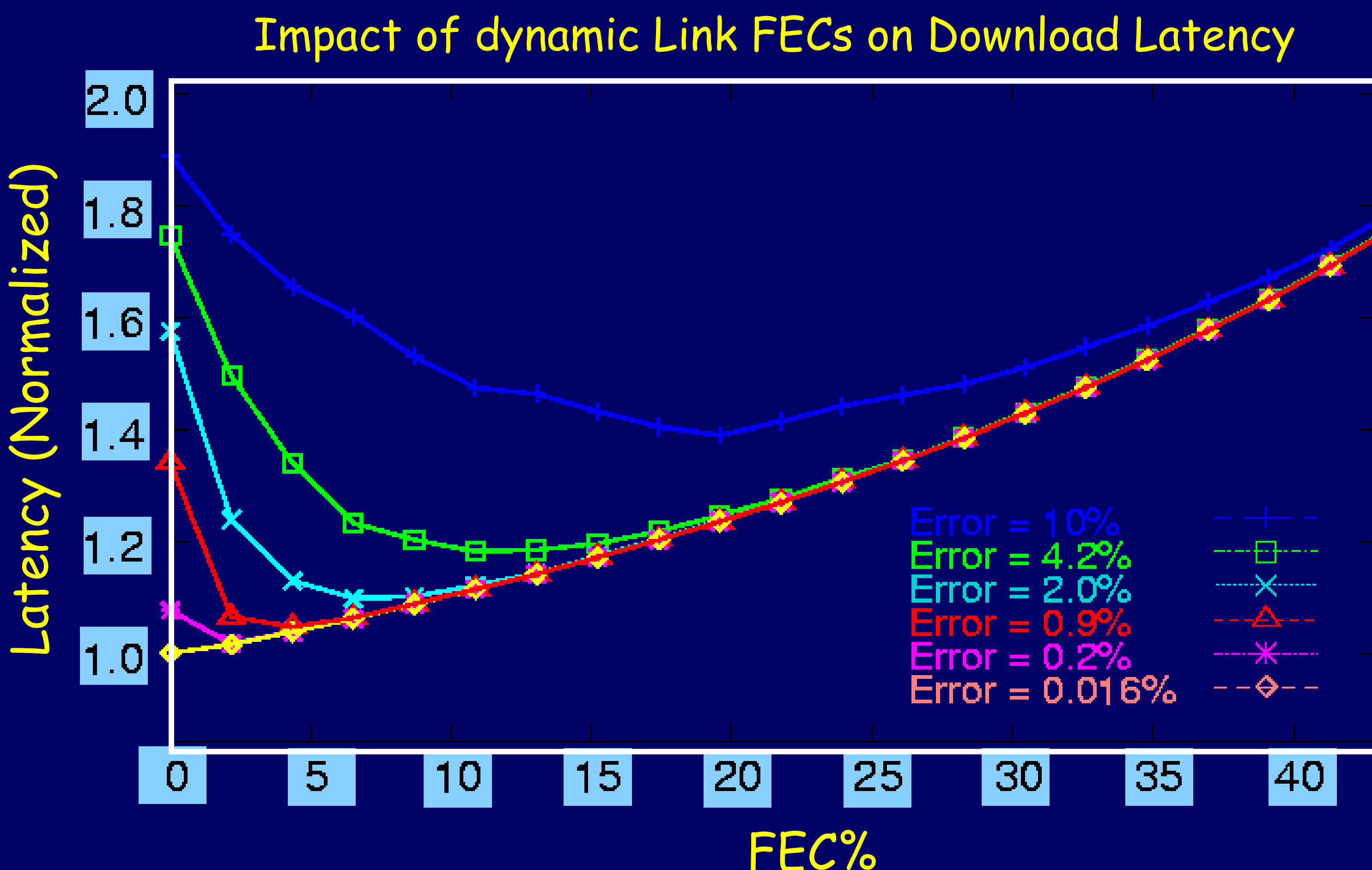
- ✓ No slow-start
- ✓ No TCP Transaction Cost
- ✓ Credit-based Flow Control
- ✓ Messages-based Protocol
- ✓ NACK-based Selective Repeat for recovery

Optimized Transport solutions provide further improvements (5-14%) in performance

Link Layer Optimizations

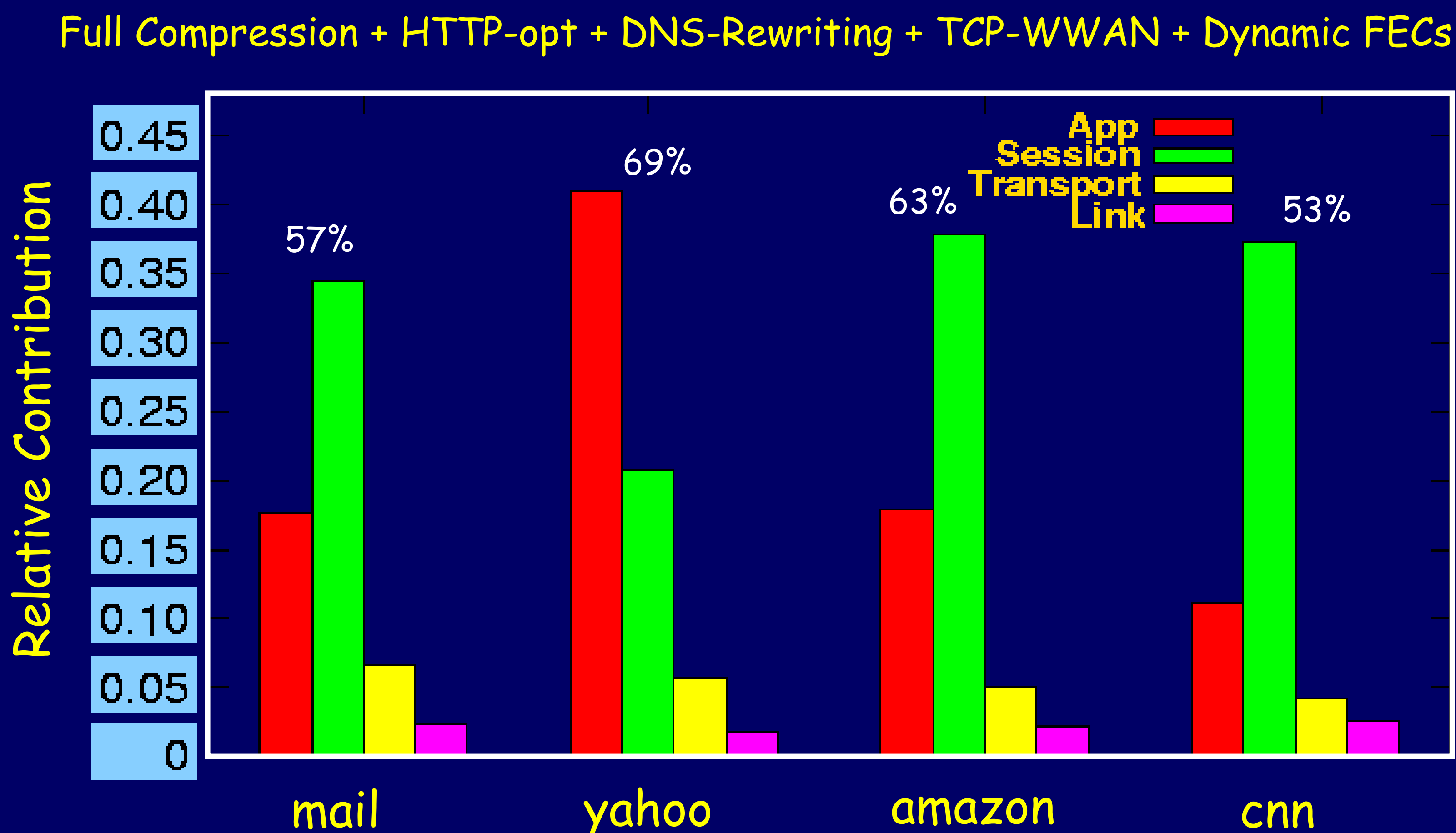
- GPRS supports four FEC schemes (CS1-CS4)
- Most GPRS networks support static CS-2
- Trace-driven Evaluation to examine how Dynamic FECs benefit Application Performance
- We use link-layer traces where we can infer slots received in error within RLC blocks

Dynamic Link-layer FEC



For a given channel condition there is an optimal value of FEC that minimizes latency

Summary of Optimizations (1)



App and Session Layer Optimizations Dominate Performance Benefits (48-61%)

Summary of Optimizations (2)

	Transparent Proxy Opt.				Dual-Proxy Opt.	
Website (Virtual)	HTTP-Opt. (No reconf-I)		HTTP-Pipelining (No reconf.-II)		Client-reconf. (dual-proxy)	
	Lat.(s)	%Impr.	Lat.(s)	%Impr.	Lat.(s)	%Impr.
Mail	15.7	54%	13.2	62%	12.4	64%
Yahoo	11.6	67%	11.4	68%	9.9	72%
Amazon	30.8	60%	27.4	64%	24.3	68%
CNN	96.2	51%	65.1	67%	59.3	69%

Dual-Proxy Solution provides additional (5-18%) performance benefits

Main Observations

- Severe Mismatch in the performance of Default HTTP and TCP in WWANs
- Standard web-browsers fail to utilize the meagre resources of WWAN links
- Significant benefits can be realized from session and application layer optimizations
- Proxy-based solutions are most effective in improving performance for mobile end-users

Implications for 3G Links

Should we expect similar benefits for 3G?

Network	Web CNN Page (Avg. T'put)	FTP 200KB (Avg. T'put)	%Dgr.
GPRS	7.6 Kbps	29.9 Kbps	-75%
3G-1X	35 Kbps	91 Kbps	-61%
3G-UMTS	62 Kbps	178 Kbps	-65%

TCP and HTTP mismatch seen even in CDMA 3G-1X and UMTS 3G

Wired dial-ups Links

Are WWAN a special case of
low-bandwidth high-latency links e.g. dial-ups?

Network	Web CNN Page (Avg.T'put)	FTP 200KB (Avg.T'put)	%Dgr.
GPRS	7.6 Kbps	29.9 Kbps	-75%
Dial-up V.90	38.5 Kbps	45.8 Kbps	-19%

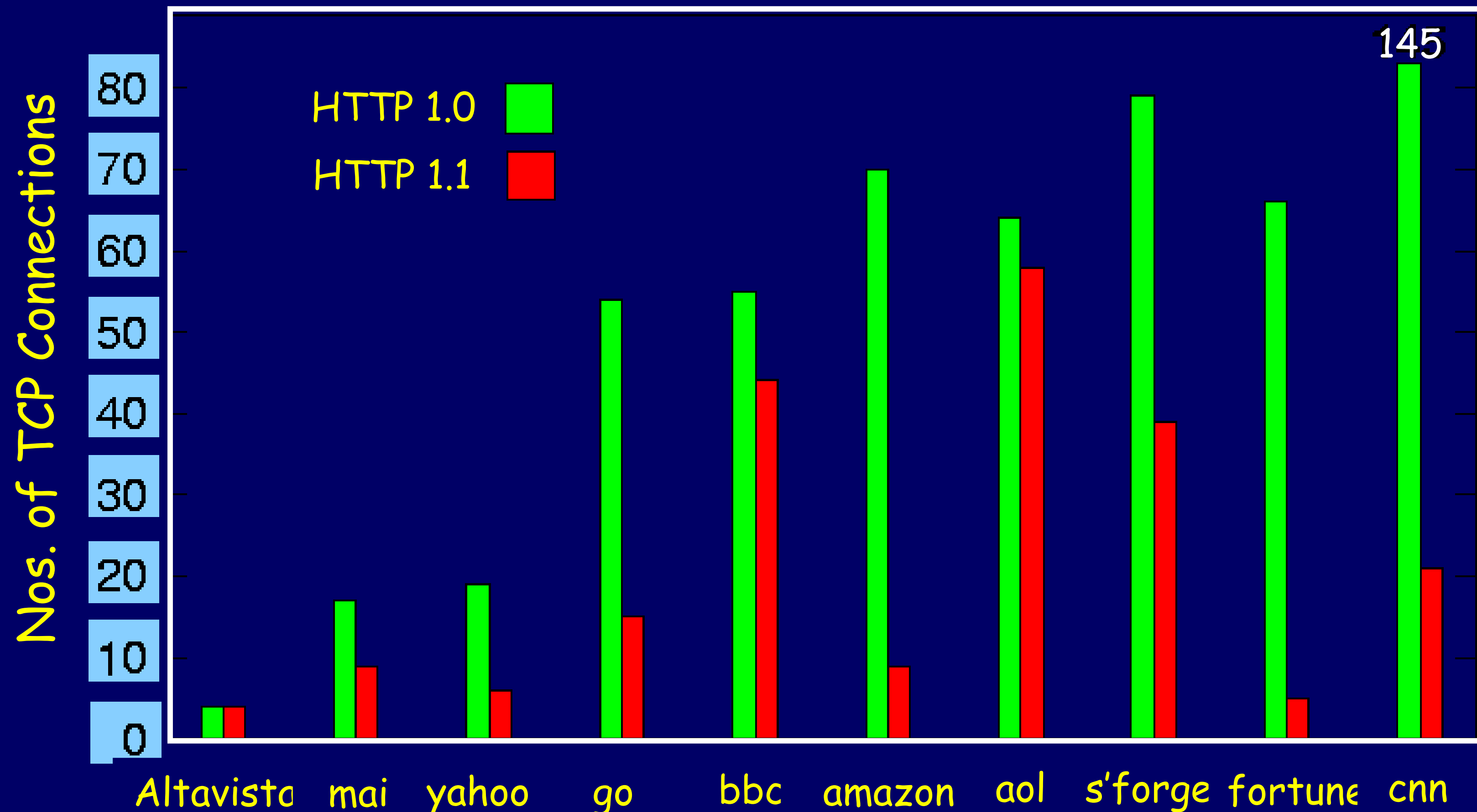
No noticeable mismatch in wired dial-ups

Virtual Web-hosting,
Tools, Source Code, traces:

http://www.cl.cam.ac.uk/users/rc277/wwan.html

Q?

Impact of Web Server FIN'ing



Benefits of HTTP 1.1 not realized due to explicit Web Server FIN'ing