HTTP Header Encoding Results

Hervé Ruellan

Goal

- Obtain information for designing a good HTTP header representation
- Statistics on HTTP headers
  - Number of header per message, header frequencies...
- Compression measurements
  - Encoding mechanisms: indexing, delta-encoding...
  - With or without DEFLATE
  - Short to long-lived connection

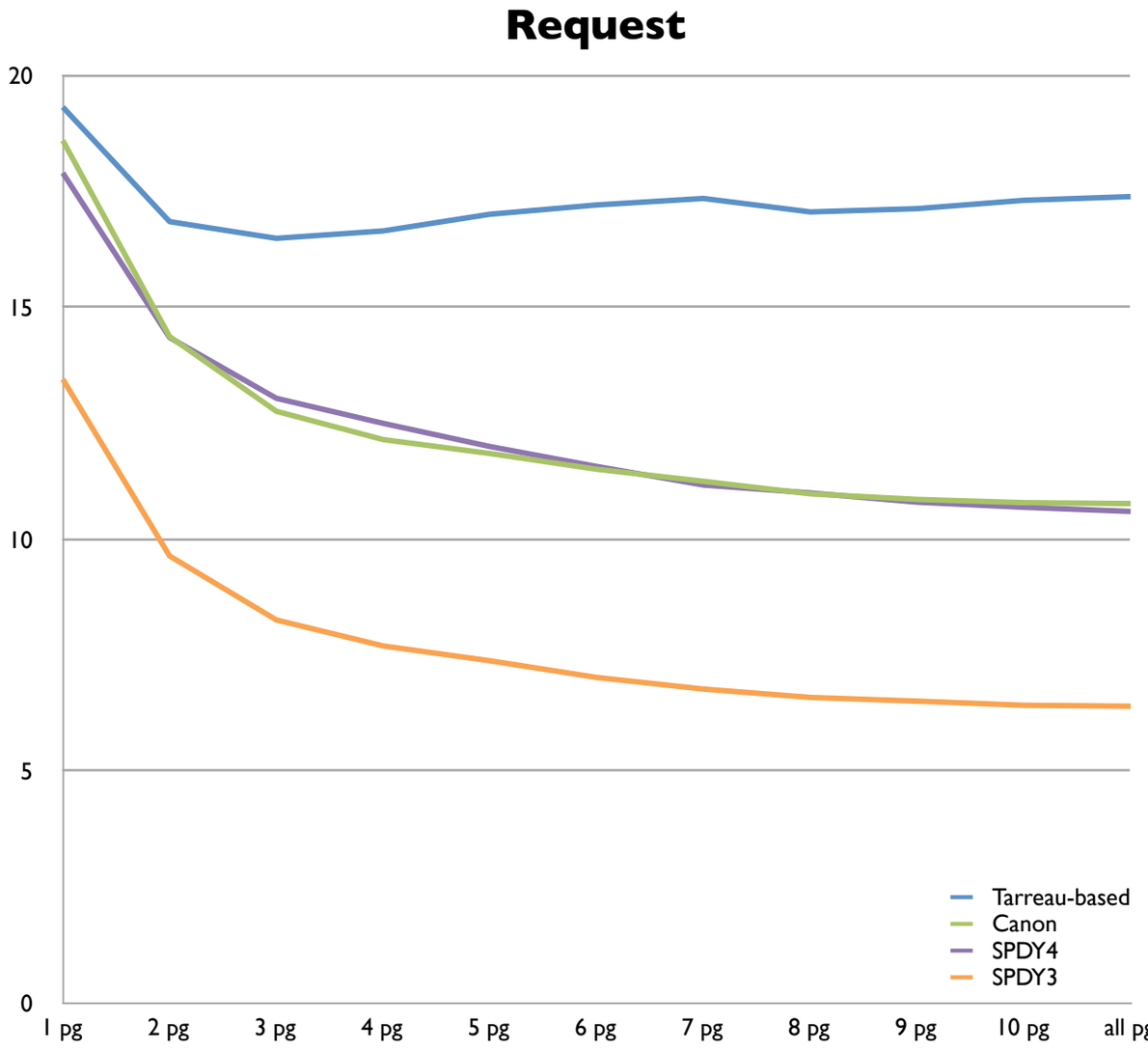
Testing Corpus

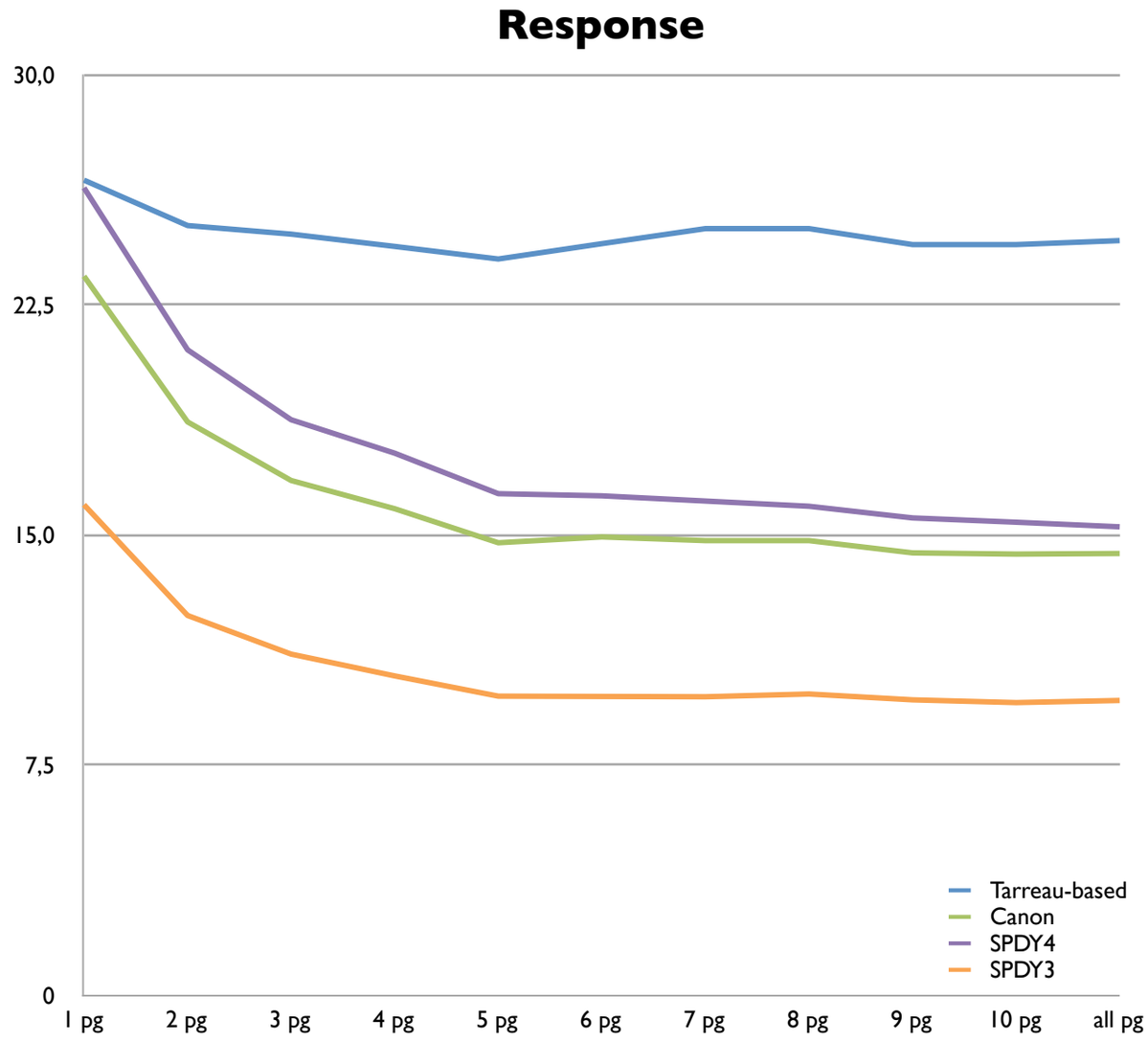
- 20 Sites from Alexa's top
  - French sites to avoid automatic redirections
- For each site:
  - Main page
  - 10 pages selected randomly from main page links
- Browser used: Chrome
- Note: need to validate representativeness of this corpus

Results

- HTTP Header statistics
  - See Appendices
- Compression results
  - Not a codec benchmark
  - Insights on different encoding techniques
- Need to consolidate testing corpus
  - More browsers, international sites, more types of HTTP usage

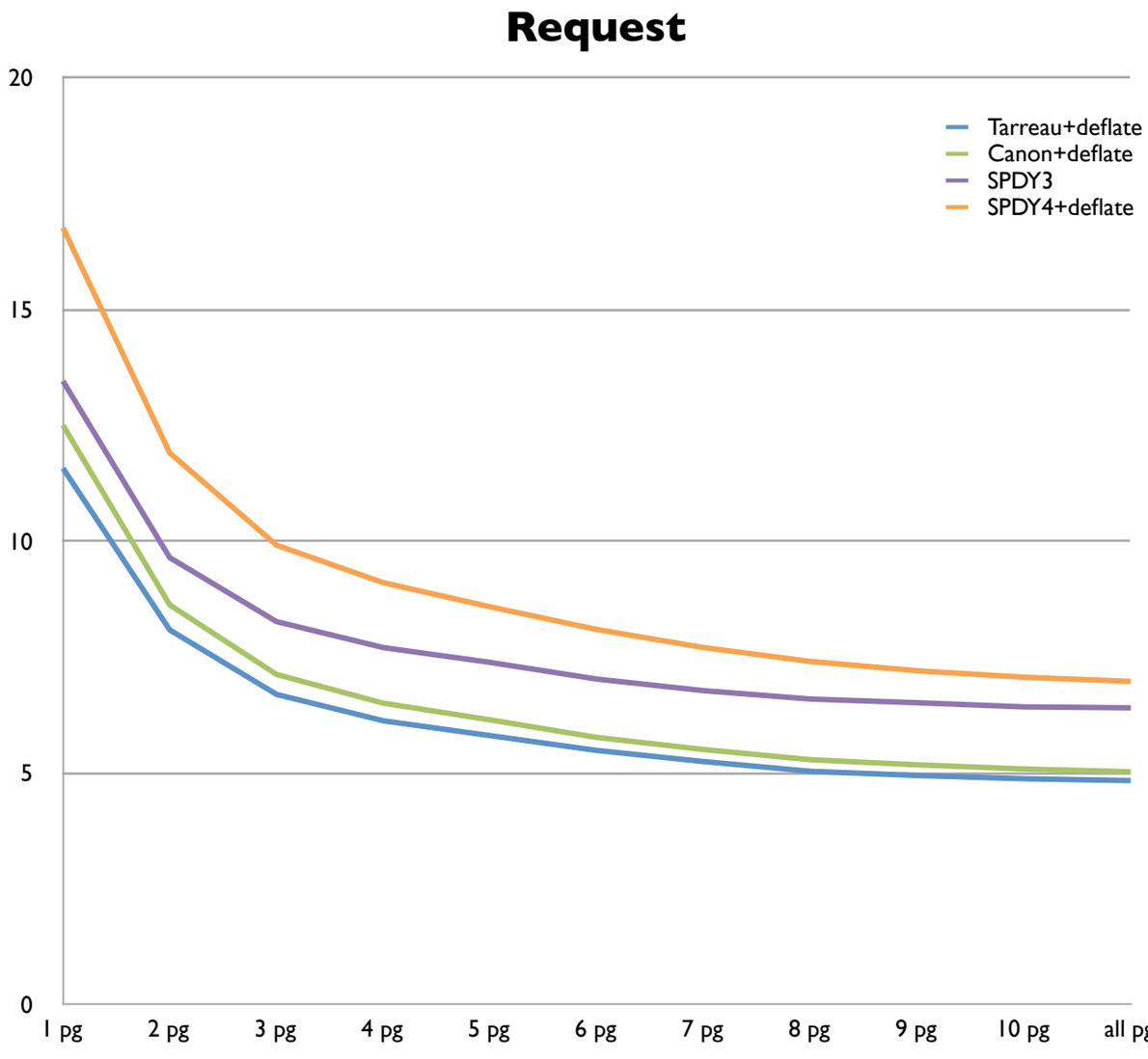
Compression Results
No DEFLATE

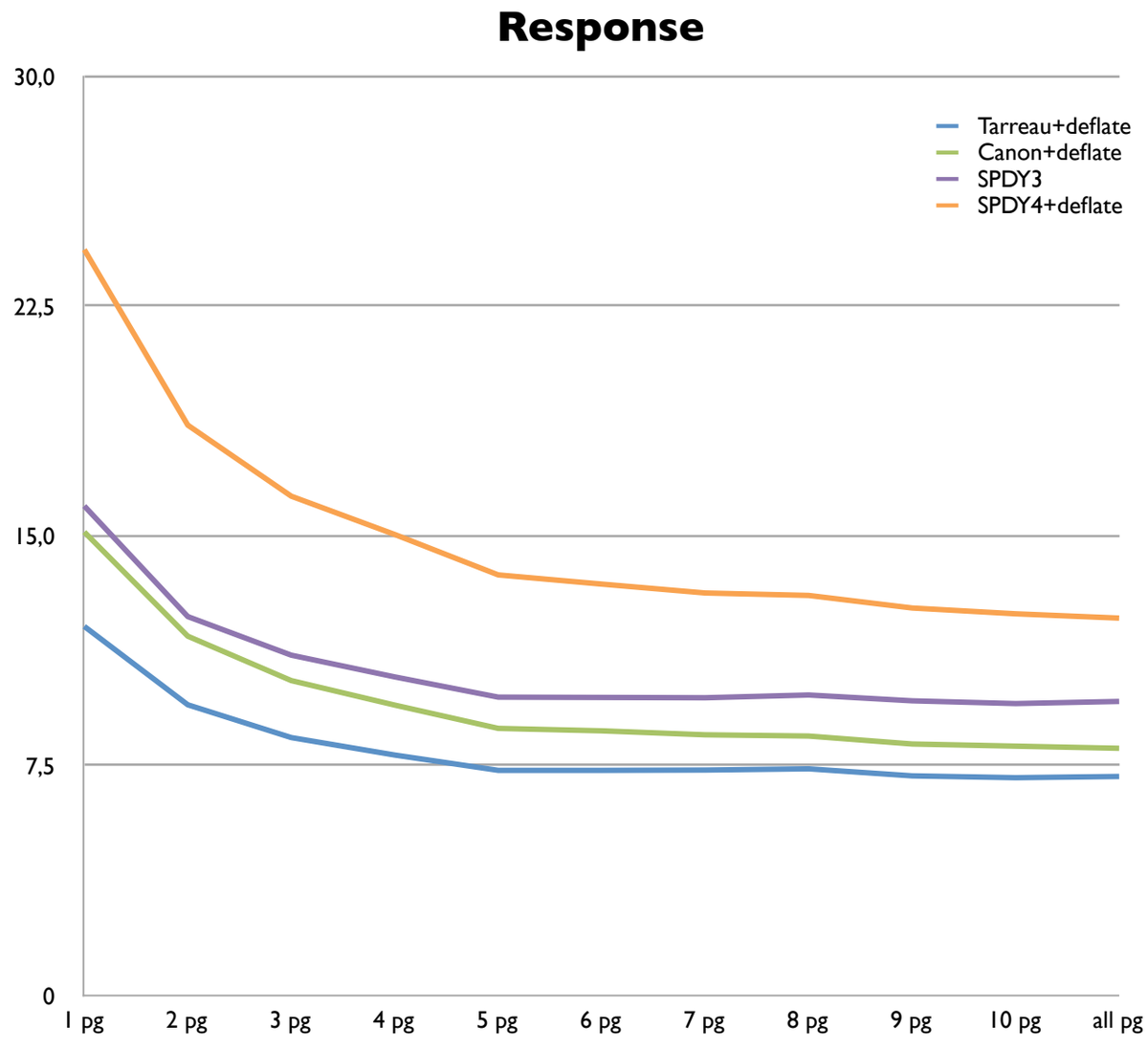




- Good compression can be obtained without DEFLATE

Compression Results With DEFLATE

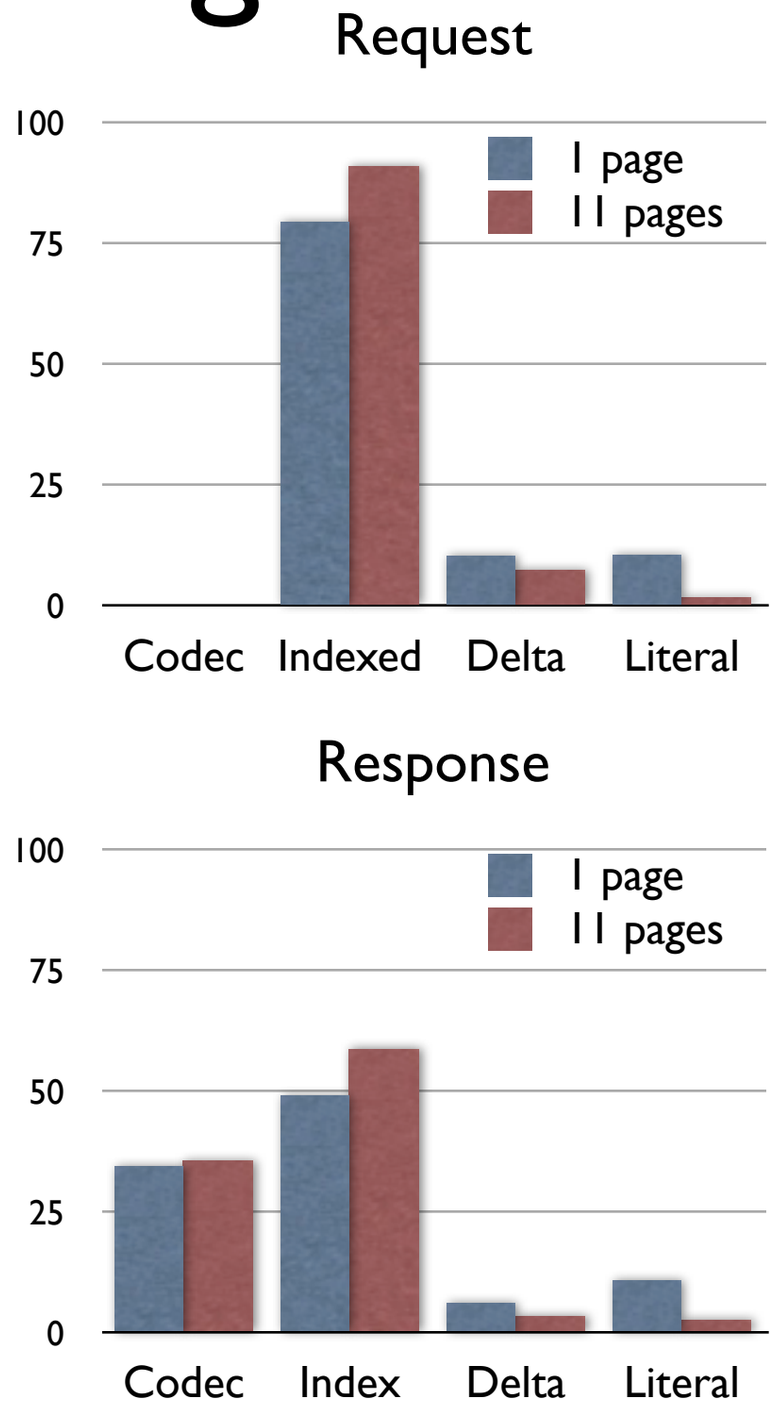




- Indexing makes DEFLATE more efficient
- Processing speed in same order of magnitude

Some Efficient Tools for Header Value Encoding

- Indexing
  - Buffer several values for a given name
- Specific value codecs
  - Integer and Date
  - Mainly for responses
- Delta-encoding
  - Common beginning shared with indexed value
  - Useful for URLs, with several indexed values



Summary

- Studied simple encoding mechanisms
  - Indexing, delta-encoding, type-specific codecs
- Obtained results
  - Provide good compression
  - Combine well with DEFLATE

Questions?

Appendices

Codec Used

- Binary
  - Roughly Willy Tarreau's proposal (June 2012)
  - DEFLATE windowBits set to 13
- CRF
  - Indexing (with controlled memory usage) + Delta-encoding
  - DEFLATE windowBits set to 12
- SPDY3
  - DEFLATE windowBits set to 15
- SPDY4
  - Roberto Peon's Python prototype
  - DEFLATE windowBits set to 13

Header Encoding Statistics

	Request		Responses	
	1 page	11 pages	1 page	11 pages
Messages	540	7687	540	7687
Headers	4948	64701	7052	93107
Average	9.16	8.42	13.06	12.11
Peak	9	9	12-13	13-14
Min	3	0	0	0
Max	11	14	23	23

Note: resource's path not included here

Optimal Name Encoding

	Requests		Responses	
	1 page	11 pages	1 page	11 pages
Messages	540	7687	540	7687
Headers	4948	64701	7052	93107
Initial dictionary	12	12	26	26
Maximal final dictionary	13	17	30	35
Indexed	4943	64661	7003	92979
Literal	5	40	49	128

Optimal Value Encoding

	Requests		Responses	
	1 page	11 pages	1 page	11 pages
Messages	540	7687	540	7687
Headers	4948	64701	7052	93107
Date codec	0	64	1395	19127
Int codec	0	6	1029	13850
Indexed	4414	62839	3450	54611
Delta	19	750	425	3110
Literal	514	1024	753	2409

Header Occurrences

	Requests		Responses	
	1 page	11 pages	1 page	11 pages
100 %	2	0	0	0
95 %	7	2	7	0
75 %	0	6	2	9
50 %	0	0	3	2
10 %	1	2	8	7
0 %	5	12	29	54

Most Frequent Headers

	Requests			Responses	
	1 page	11 pages		1 page	11 pages
accept	99 %	93 %	age	96 %	91 %
accept-charset	99 %	93 %	content-length	94 %	90 %
accept-encoding	99 %	93 %	content-type	98 %	92 %
accept-language	99 %	93 %	date	99 %	92 %
cache-control	99 %	40 %	last-modified	85 %	86 %
host	99 %	93 %	server	99 %	92 %
proxy-connection	96 %	90 %	via	96 %	91 %
referer	100 %	95 %	x-cache	96 %	90 %
user-agent	100 %	95 %	x-cache-lookup	96 %	90 %