Searching and Analyzing HTTP Data with the WASE Framework

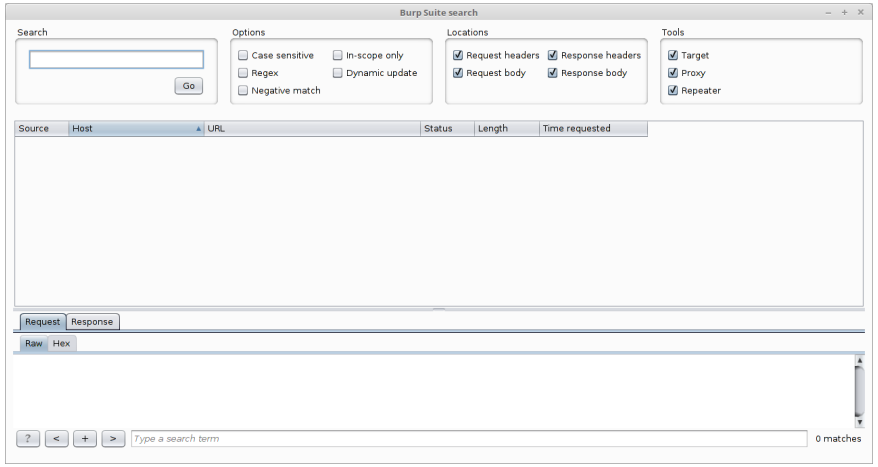
Thomas Patzke
German OWASP Day 2016

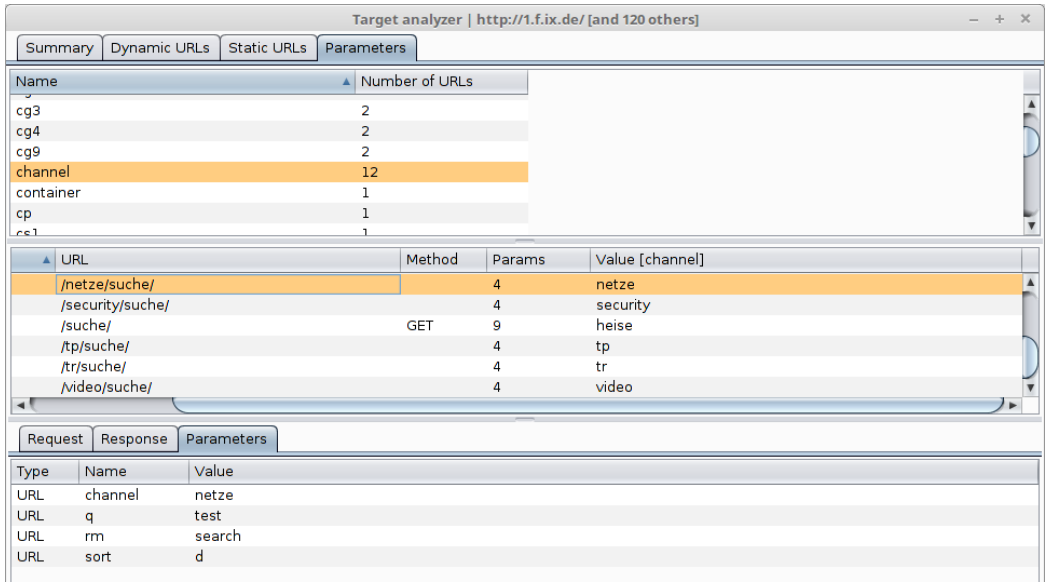
Content

1. Why?
2. Introduction to WASE and the toolchain
3. Usage Examples
   - Security Testing of web applications
   - Mass-Scanning the Alexa Top 1M
4. Future Development & Ideas

What is the Problem?

Unflexible!





Bad Performance!

...or in Words

Try to do one of the following with your tool of choice:

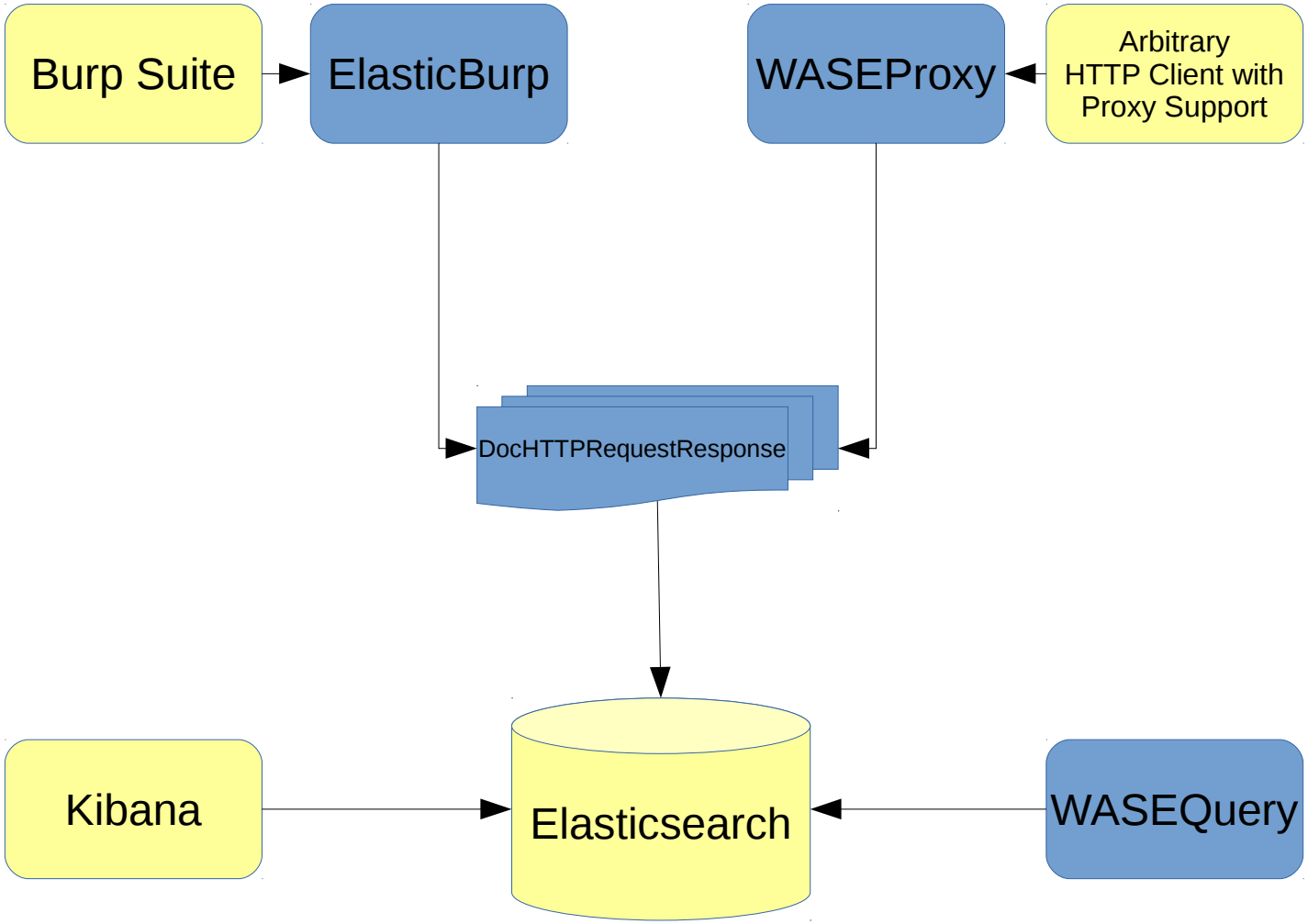
- Search all POST requests that don‘t contain a CSRF token
- List all values of a parameter or cookie that encountered while a web application security test
- List all values of a security header with its corresponding URL
- List all URLs where inferred content type is HTML while the server tells something different about its content type
- Show all HTML responses without a doctype definition
- Find all external script references
- Discover unsafe or nonse HTTP security header values

Bonus points: try it without a coffee break :)

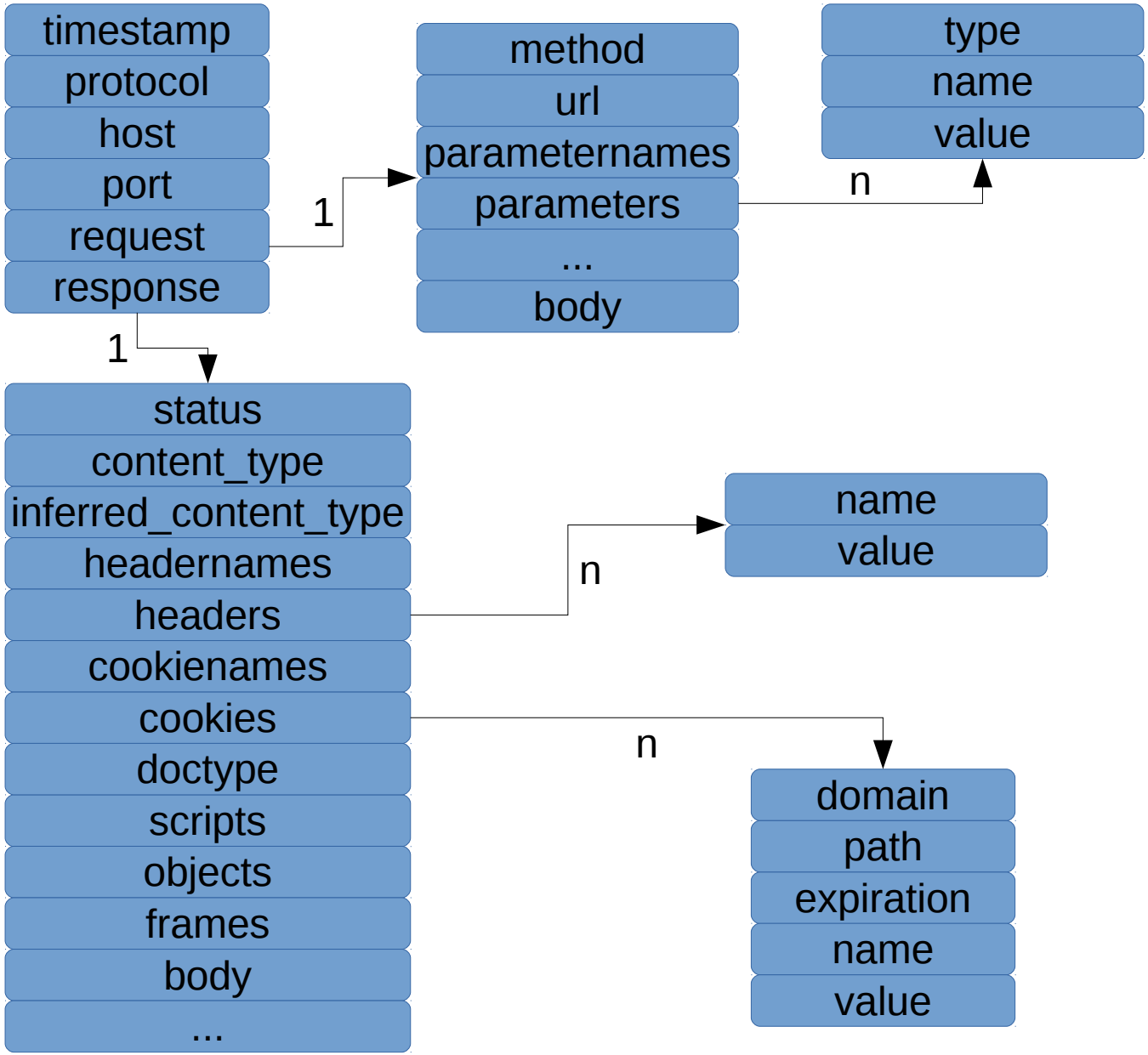
Elasticsearch, Kibana, WASE

- Elasticsearch: a search and analytics engine for textual data
- Kibana: web frontend for Elasticsearch
- WASE: **W**eb **A**udit **S**earch **E**ngine
  - Definition of a data structure for HTTP requests/responses for Elasticsearch
  - Toolchain: ElasticBurp, WASEProxy, WASEQuery

WASE Framework



DocHTTPRequestResponse



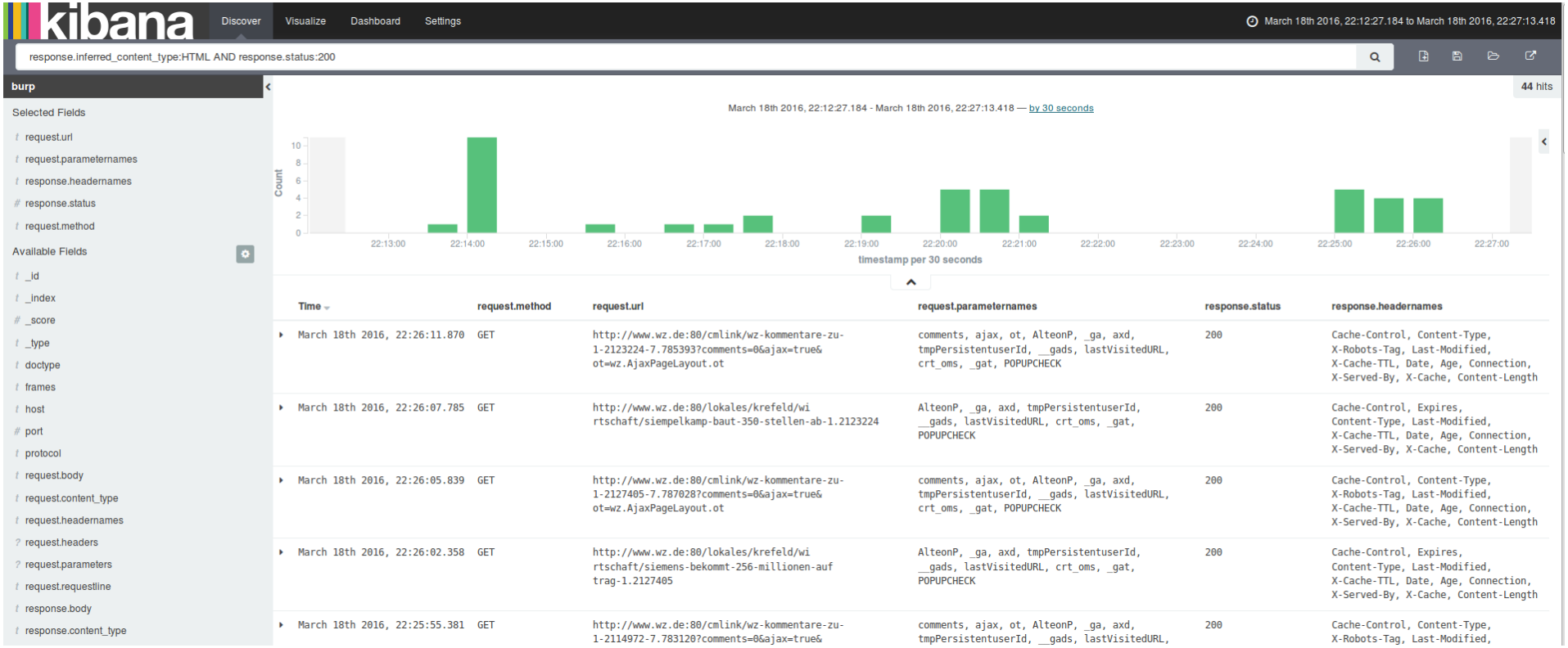
Usage Examples

- Complex searches and analytics in web application security tests
- Mass Scans of web sites
- Malware analysis (someones master thesis)

Searches

- All POSTs without CSRF-Token:
  `request.method:POST -request.parameternames.raw:"csrftoken"`
- 2xx Responses recognised as HTML without <! DOCTYPE ...>:
  `response.status:[200 TO 299] AND response.inferred_content_type:html -doctype`
- HTML Responses not declared as such ones:
  `response.inferred_content_type:html -response.content_type:html`
- Responses without XFO:
  `NOT response.headernames:"X-Frame-Options"`

Searching with Kibana



Searching with WASEQuery

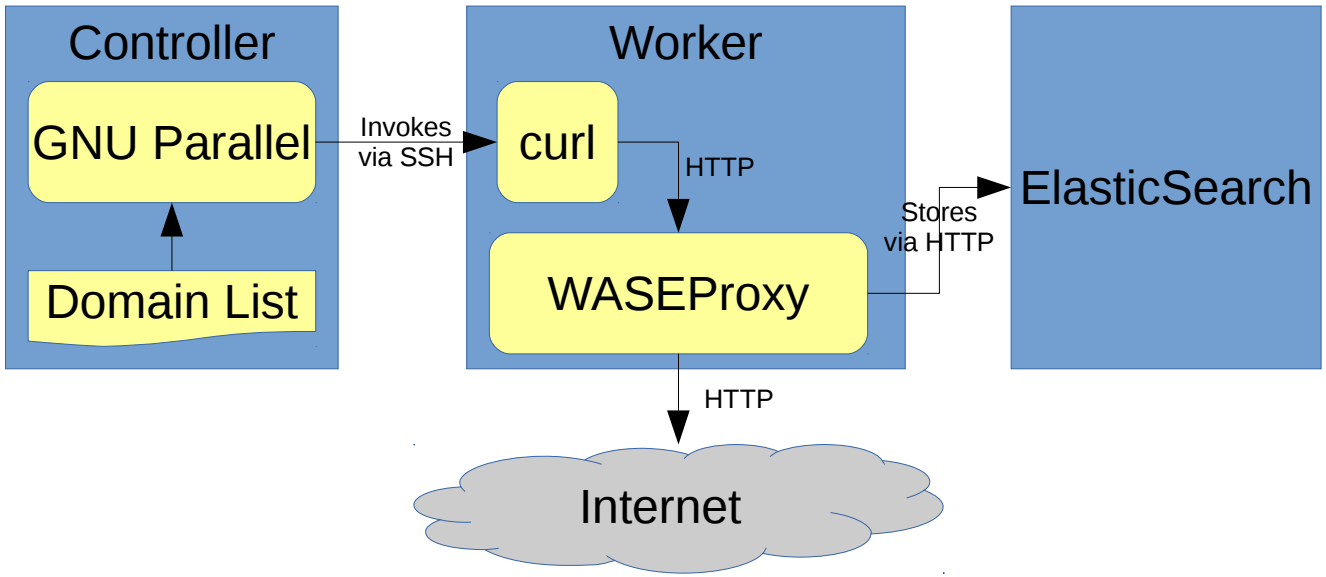
- Kibana doesn‘t likes nested data structures and doesn‘t exposes many ElasticSearch features
- WASEQuery: collection of few useful queries

Example: List of all CSPs that contain the word *unsafe*

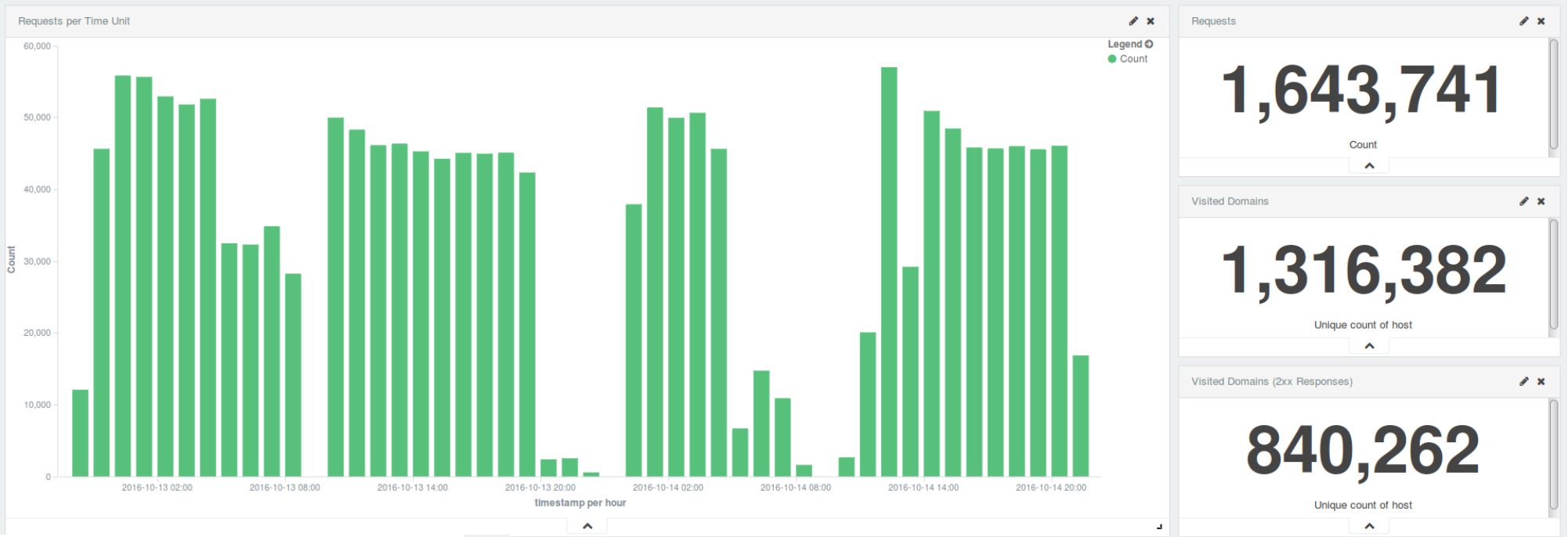
```
[thomas:~/Devel/WASE] master(+0/-0)* 2s ± ./WASEQuery.py headervalues --values '*unsafe*' content-security-policy | cat -n
     1  default-src * data: blob:;script-src *.facebook.com *.fbcdn.net *.facebook.net *.google-analytics.com *.virtualearth.net *.google.com 127.0.0.1:* *.spoti
c-b-a.akamaihd.net *.atlassolutions.com blob: data:;style-src * 'unsafe-inline' data:;connect-src *.facebook.com *.fbcdn.net *.facebook.net *.spotilocal.com:* *.
ssolutions.com attachment.fbsbx.com ws://localhost:* blob:;
     2  default-src * data: blob:;script-src *.facebook.com *.fbcdn.net *.facebook.net *.google-analytics.com *.virtualearth.net *.google.com 127.0.0.1:* *.spoti
c-b-a.akamaihd.net *.atlassolutions.com blob: data:;style-src * 'unsafe-inline' data:;connect-src *.facebook.com *.fbcdn.net *.facebook.net *.spotilocal.com:* *.
ssolutions.com attachment.fbsbx.com ws://localhost:* blob: chrome-extension://boadgeojelhgndaghljhdicfkmllpafd;
     3  report-uri /_/ConsentHttp/cspreport;script-src 'unsafe-inline' 'self' 'unsafe-eval' https://apis.google.com https://ssl.gstatic.com https://www.google.co
     4  script-src 'self' https://addons.mozilla.org https://www.paypalobjects.com https://www.google.com/recaptcha/ https://www.gstatic.com/recaptcha/ https://s
mg-src 'self' data: blob: https://www.paypal.com https://ssl.google-analytics.com https://addons.cdn.mozilla.net https://static.addons.mozilla.net https://sentry
'unsafe-inline' https://addons.cdn.mozilla.net; frame-src 'self' https://ic.paypal.com https://paypal.com https://www.google.com/recaptcha/ https://www.paypal.co
-src 'self' https://addons.cdn.mozilla.net; report-uri /__cspreport__
     5  script-src https://connect.facebook.net https://cm.g.doubleclick.net https://ssl.google-analytics.com https://graph.facebook.com https://twitter.com 'uns
er.com https://publish.twitter.com https://ton.twitter.com https://syndication.twitter.com https://www.google.com https://t.tellapart.com https://platform.twitte
'; frame-ancestors 'self'; font-src https://twitter.com https://*.twimg.com data: https://ton.twitter.com https://fonts.gstatic.com https://maxcdn.bootstrapcdn.c
s://*.twimg.com https://ton.twitter.com blob: 'self'; connect-src https://graph.facebook.com https://*.giphy.com https://*.twimg.com https://api.twitter.com http
tps://upload.twitter.com https://api.mapbox.com 'self'; style-src https://fonts.googleapis.com https://twitter.com https://*.twimg.com https://translate.googleap
ttps://maxcdn.bootstrapcdn.com https://netdna.bootstrapcdn.com 'self'; object-src https://twitter.com https://pbs.twimg.com; default-src 'self'; frame-src https:
r.vimeo.com https://pay.twitter.com https://www.facebook.com https://ton.twitter.com https://syndication.twitter.com https://vine.co twitter: https://www.youtube
.ak.facebook.com 'self' https://donate.twitter.com; img-src https://graph.facebook.com https://*.giphy.com https://twitter.com https://*.twimg.com data: https://
ook.com https://ton.twitter.com https://*.fbcdn.net https://syndication.twitter.com https://media.riffsy.com https://www.google.com https://stats.g.doubleclick.n
report-uri https://twitter.com/i/csp_report?a=NVQWGYLXFVZXO2LGOQ%3D%3D%3D%3D%3D%3D&ro=false;
```

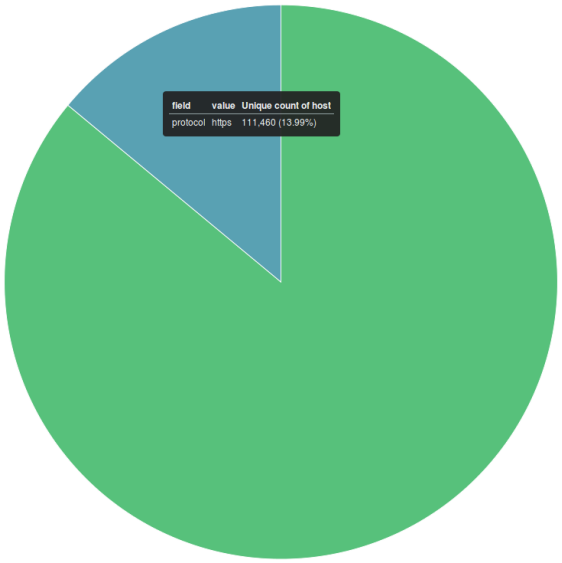
Mass-Scanning the Alexa Top 1m

- Scanning from AWS EC2 instances
  - 1 x t2.micro as scan controller (misused a bit as worker)
  - 4 x m4.large spot instances as scan workers
  - 4 x t2.micro.elasticsearch
  - 2h x 3 x r3.xlarge.elasticsearch for final analysis (required much RAM for some complex queries)
  - 1 day for scanning complete 1m list
  - ~1,50€ scanning costs, ~3€ analysis
- Tools:
  - GNU Parallel
  - curl
  - WASEProxy
- No response bodys!
- 35,6 GB ES Indexes
- 15.311.855 ES Docs



Results





filters	Unique count of host
X-Frame-Options	66,753
X-Content-Type-Options	63,977
X-XSS-Protection	50,453
Strict-Transport-Security	28,597
Content-Security-Policy	3,450
Public-Key-Pins	359
Number of Domains	777,704

Results: Popular Embedded Objects

Top 20 response.objects.raw	Unique count of host
core.RE	204
http://www.xatech.com/web_gear/chat/chat.swf	125
//www.youtube.com/get_player	104
index.swf	77
	69
banner.swf	67
http://swf.yowindow.com/yowidget3.swf	56
https://res.egtmgs.com/release/core/EmptySwf.swf	48
images/banner.swf	46
images/logo.swf	45
http://releases.flowplayer.org/swf/flowplayer-3.2.1.swf	36
/flash/mjupl4li.swf?123	35
//video.limelight.com/player/loader.swf	34
main.swf	34
source/plugin/study_nge/images/clock.swf	32
data:application/x-silverlight-2,	31
FLVPlayer_Progressive.swf	30
header.swf	30
images/top.swf	30
top.swf	29

filters	Unique count of host
Scanned Domains	777,704
with JS References	709,235
with JS refs but no external	109,716

Results: DOCTYPE Declarations

Top 20 response.doctype.raw	Unique count of host
DOCTYPE html	433,509
DOCTYPE html PUBLIC "-//W3C//DTD XHTML 1.0 Transitional//EN" "http://www.w3.org/TR/xhtml1/DTD/xhtml1-transitional.dtd"	123,066
doctype html	63,853
DOCTYPE html PUBLIC "-//W3C//DTD XHTML 1.0 Strict//EN" "http://www.w3.org/TR/xhtml1/DTD/xhtml1-strict.dtd"	33,757
DOCTYPE HTML	29,659
DOCTYPE HTML PUBLIC "-//W3C//DTD HTML 4.01 Transitional//EN" "http://www.w3.org/TR/html4/loose.dtd"	11,113
DOCTYPE HTML PUBLIC "-//W3C//DTD HTML 4.01 Transitional//EN"	7,701
doctype html	6,853
DOCTYPE html PUBLIC "-//W3C//DTD XHTML 1.1//EN" "http://www.w3.org/TR/xhtml11/DTD/xhtml11.dtd"	4,992
DOCTYPE HTML PUBLIC "-//W3C//DTD HTML 4.0 Transitional//EN"	3,832
DOCTYPE HTML PUBLIC "-//W3C//DTD HTML 4.01//EN" "http://www.w3.org/TR/html4/strict.dtd"	3,777
DOCTYPE html PUBLIC "-//W3C//DTD XHTML+RDFa 1.0//EN" "http://www.w3.org/MarkUp/DTD/xhtml-rdfa-1.dtd"	3,244
DOCTYPE html PUBLIC "-//W3C//DTD HTML 4.01 Transitional//EN" "http://www.w3.org/TR/html4/loose.dtd"	1,892
DOCTYPE HTML PUBLIC "-//W3C//DTD HTML 4.01 Transitional//EN" "http://www.w3.org/TR/html4/loose.dtd"	1,456
DOCTYPE html	1,364
DOCTYPE html PUBLIC "-//W3C//DTD XHTML 1.0 Transitional//EN" "http://www.w3.org/TR/xhtml1/DTD/xhtml1-transitional.dtd"	1,336
DOCTYPE html PUBLIC "-//W3C//DTD XHTML+RDFa 1.0//EN" "http://www.w3.org/MarkUp/DTD/xhtml-rdfa-1.dtd"	1,152
DOCTYPE html PUBLIC "-//W3C//DTD HTML 4.01//EN" "http://www.w3.org/TR/html4/strict.dtd"	1,112
DOCTYPE HTML PUBLIC "-//W3C//DTD HTML 3.2 Final//EN"	1,053
DOCTYPE html PUBLIC "-//W3C//DTD HTML 4.01 Transitional//EN"	1,041

What‘s Next – Future Development

- Documentation and Automation of Mass-Scan Setup
- Extraction of further attributes
- Further Input Frontends:
  - PCAP
  - Raw Text Files
  - OWASP ZAP
- Development of a fancy query language EQUEL
- Search interface in Burp Extension

That‘s it!

Get it on GitHub:
https://github.com/thomaspatzke/WASE
...Pull Requests are Welcome! :)

Live Demo: http://wase-demo.patzke.org

Questions?
Mail: thomas@patzke.org
Twitter: @blubbfiction