



Toolforge webservices are in the final stages of [migrating to the toolforge.org domain](#) .
Please help us clean up older documentation referring to tools.wmflabs.org!

Incident documentation/20170607-WikiScrape

[< Incident documentation](#)

Contents [\[hide\]](#)

- [1 Summary](#)
- [2 Timeline](#)
- [3 Conclusions](#)
- [4 Actionables](#)

Summary

A high request rate from a single bot (same UA/IP) of some specific queries on the API caused high concurrency for such queries on the database for enwiki. As a consequence, those requests were served very slowly. The slowness then propagated to the non-API mediawiki servers due to lagging on the API databases, and those caused an increased load on both the API and standard appserver cluster, see:

<https://grafana.wikimedia.org/dashboard/db/prometheus-apache-hhvm-dc-stats?panelId=17&fullscreen&orgId=1&from=1496806612437&to=1496818165547>

Timeline

- 04:44 A first, smaller spike of 5xx happens on text caches
- 05:08 Icinga sends the first alarm to IRC <icinga-wm> PROBLEM - Text HTTP 5xx reqs/min on graphite1001 is CRITICAL: CRITICAL: 11.11% of data above the critical threshold [1000.0]
- 05:28 _joe_ notices the flapping alerts and starts investigating
- 05:29 time correlation with a l10-update is suggested by mutante, but is determined to be a red herring
- 05:31 Investigation on the varnish layer starts. _joe_ notices there are no icinga alarms for mailbox lag and that most 5xx are evenly distributed on all caches
- 05:39 Varnish is ruled out as source of the issue
- 05:43 Investigation on logstash and oxygen is inconclusive.
- 05:53 The problem seems to happen for APIs only, mostly concentrated on enwiki. HHVM and apache logs at a first glance do not indicate any issue at the appserver level
- 05:56 A sudden shower of ProxyFetch failures from pybal is noticed for all appserver pools (not just API):

```
< _joe > Jun 7 05:55:45 lvs1003 pybal[6265]: [api-https_443 ProxyFetch] WARN:
mw1204.eqiad.wmnet (enabled/up/pooled): Fetch failed, 5.891 s since pybal is fetching
Special:BlankPage on both clusters, a problem with s1 is suggested. marostegui is investigating
```

- 08:02 _joe_ finds evidence of slowness in the API apache logs, in short bursts of a few seconds around the time of the pybal alerts
- 08:01 All service send out recoveries, and the issue doesn't seem to be repeating on the load balancers
- 08:20 After more digging in the logs, _joe_ notices a bot, coming from a single ip at Michigan State University with UA wikiScrape/0.0.0 is doing tons of requests in short bursts to the API for enwiki.
- 08:25 ema confirms that UA/ip combination made huge bursts of requests to the API between 04:44 and 05:56 resulting in 503s. Peak requests were at 1200 req/s
- 09:40 jynus narrows the issue down to the fact that `ApiQueryContributions::execute`, `Revision::getParentLengths`, normally taking 1 second to execute (already slow), at the time was taking >10 seconds to execute on S1 because of the extra load

Conclusions

It seems clear that this partial outage was caused by db resources starvation due to a heavy hitter on the APIs. The only way to solve this would be to have a good rate-limiting solution on the API.

[Main page](#)
[Recent changes](#)
[Server admin log \(Prod\)](#)
[Server admin log \(RelEng\)](#)
[Deployments](#)
[SRE/Operations Help](#)
[Incident status](#)

[Cloud VPS & Toolforge](#)

[Cloud VPS documentation](#)

[Toolforge documentation](#)

[Request Cloud VPS project](#)

[Server admin log \(Cloud VPS\)](#)

[Tools](#)

[What links here](#)

[Related changes](#)

[Special pages](#)

[Permanent link](#)

[Page information](#)

[Cite this page](#)



[Print/export](#)

[Create a book](#)

[Download as PDF](#)

[Printable version](#)

Actionables

-  **Done** Ban the UA/IP in case of need - prepare a patch
- Implement better rate-limiting for the API
- Investigate and implement varnish-level rate-limiting <https://phabricator.wikimedia.org/T163233>

Category: [Incident documentation](#)

This page was last edited on 7 June 2017, at 16:06.

Text is available under the [Creative Commons Attribution-ShareAlike License](#); additional terms may apply. See [Terms of Use](#) for details.

[Privacy policy](#) [About](#)
[Wikitech](#)

[Disclaimers](#) [Code of Conduct](#) [Developers](#) [Statistics](#) [Cookie statement](#) [Mobile view](#)

