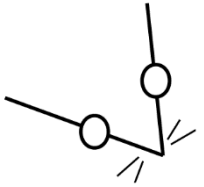


Analyzing, understanding and (hopefully) fixing degraded search performance

WOLFcon 2023

David Maus, State and University Library Hamburg

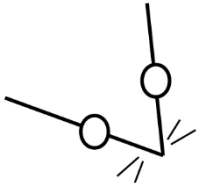
Online, 24.08.2023



About me

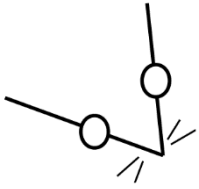
- Head of the Research & Development Department at State and University Library Hamburg
 - VuFind-based Discovery Systems
 - Repositories
 - Digital Humanities
- VuFind contributor
- Markup professional



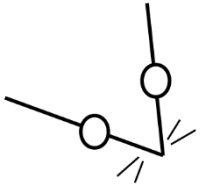


Agenda

- Katalog***plus*** in a Nutshell
- Introducing SolrProbe
- Findings using SolrProbe
- Conclusion

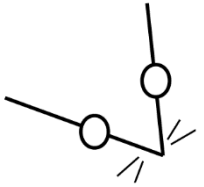


Katalog*plus*

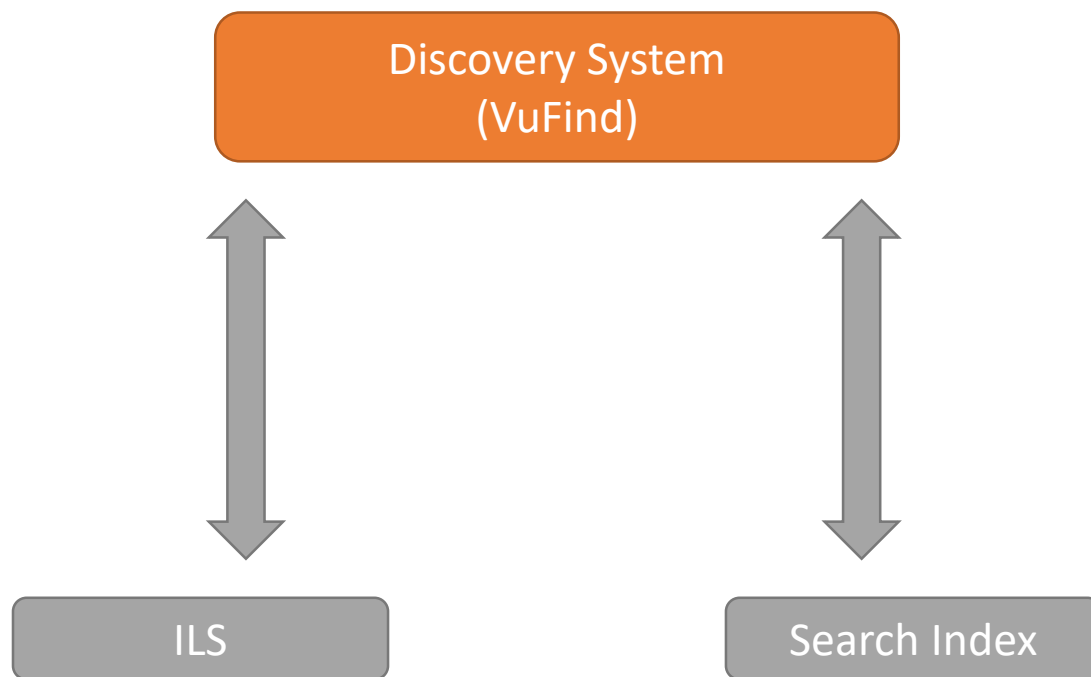


Katalog*plus* in a Nutshell

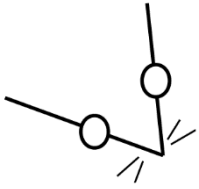
- Primary discovery platform for Hamburg University w/ additional patron-facing services
 - VuFind 8.1 with local modifications and modules
- Numbers
 - 8 millionen media units, ca. 100 millionen articles
 - 250.000 searches per month
 - 200.000 item investigations per month
- People
 - 10—15 people from different professions in different roles



Katalog*plus* Infrastructure

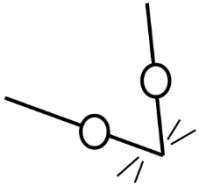


- Access to ILS and search index through a service provider
- Search index K10plus Zentral by Verbundzentrale of the Common Library Network (VZG of GBV)
- Fair use policy



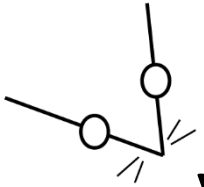
Und es war Sommer... (Peter Maffay)

- Summer 2022: Search index provider clarifies the fair use policy
 - Max. 175 requests per minute per discovery system
 - Throttling and hard errors (429 Too Many Requests)
 - Short notice
- Questions over questions
 - To which degree we affected by this policy change?
 - How many requests per minute do we actually send?
 - Which components of VuFind are sending the requests?
 - What is the average response time?
 - Who's to blame?



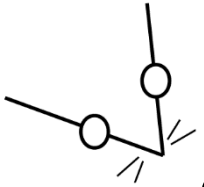
SolrProbe

...utilizing VuFind's event bus for fun and profit



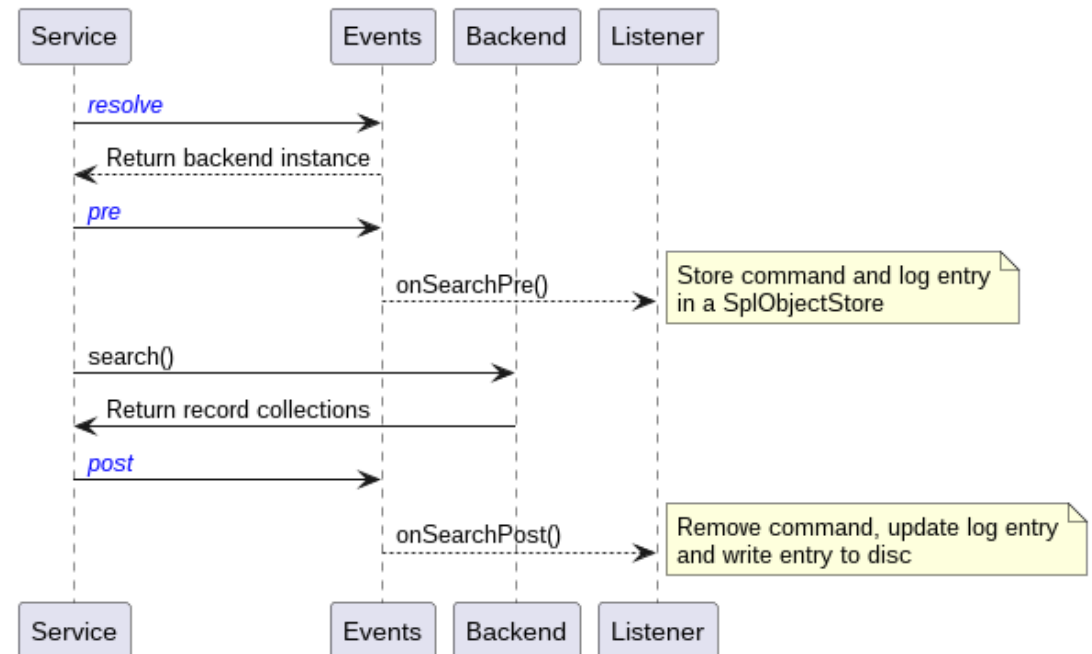
VuFindSearch Event Bus

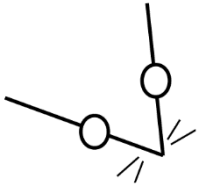
- VuFind uses an *event bus* to resolve search backend instances and to allow for the manipulation of search request and search response
 - VuFindSearch\Service::EVENT_RESOLVE
 - VuFindSearch\Service::EVENT_PRE
 - VuFindSearch\Service::EVENT_POST
 - VuFindSearch\Service::EVENT_ERROR



SolrProbe Listener

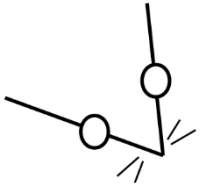
- Attach listener to the event bus for *pre*, *post*, and *error* events
- Store search command and log entry in a SplObjectStorage on receiving the *pre* event
- Remove command, update log entry and write to handler on *post* and *error* event





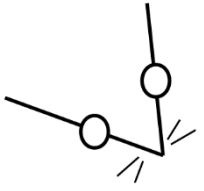
Logging

- Log entry
 - Response time
 - Command class
 - Backend identifier
 - Query time (self-reported by Solr)
 - Response status
 - User interaction (roughly: URL that caused the index query)
- Logfile
 - Simple textfile with timestamp and JSON-encoded data

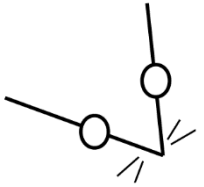


Analysis

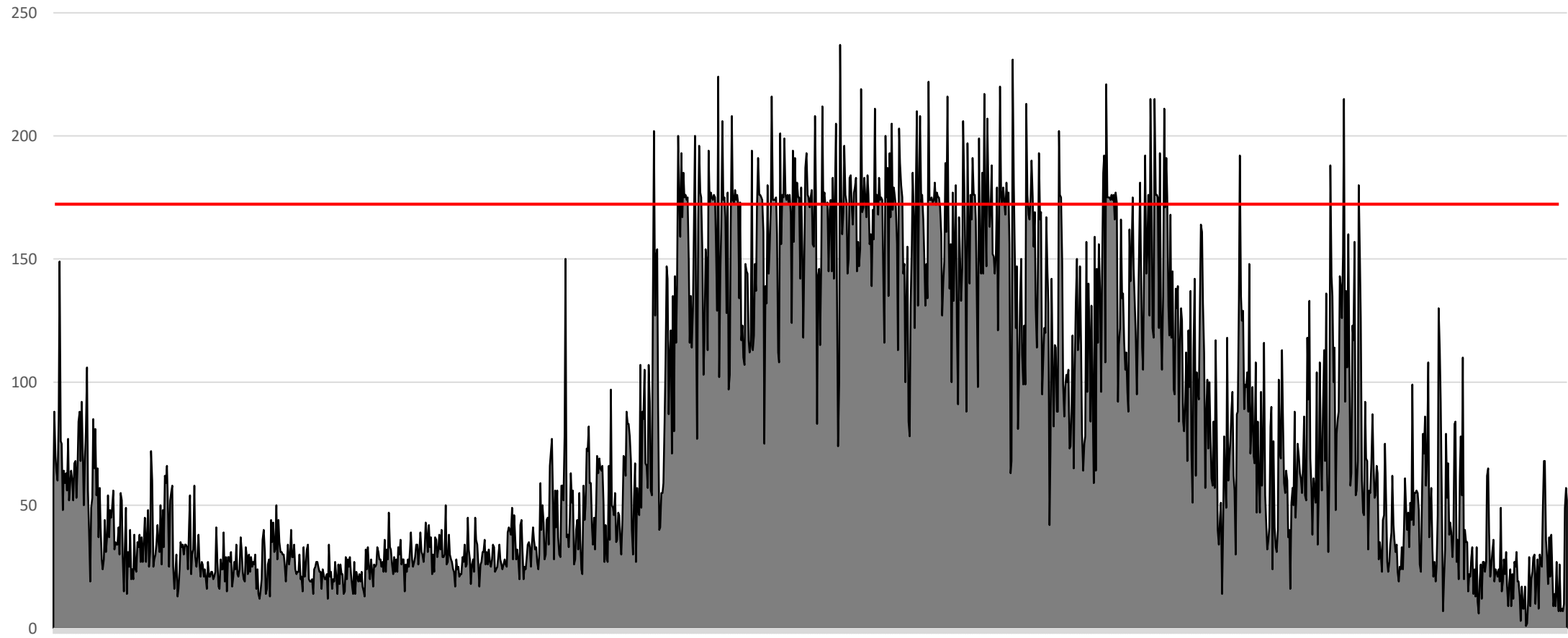
- Hodgepodge of ad-hoc scripts and programs
 - Excel
 - Gnuplot
 - RRD Tool
- Metrics
 - Active sessions (per hour, per minute)
 - Search commands (per hour, per minute)
 - User interactions
 - Response time

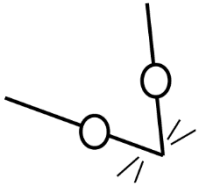


Findings

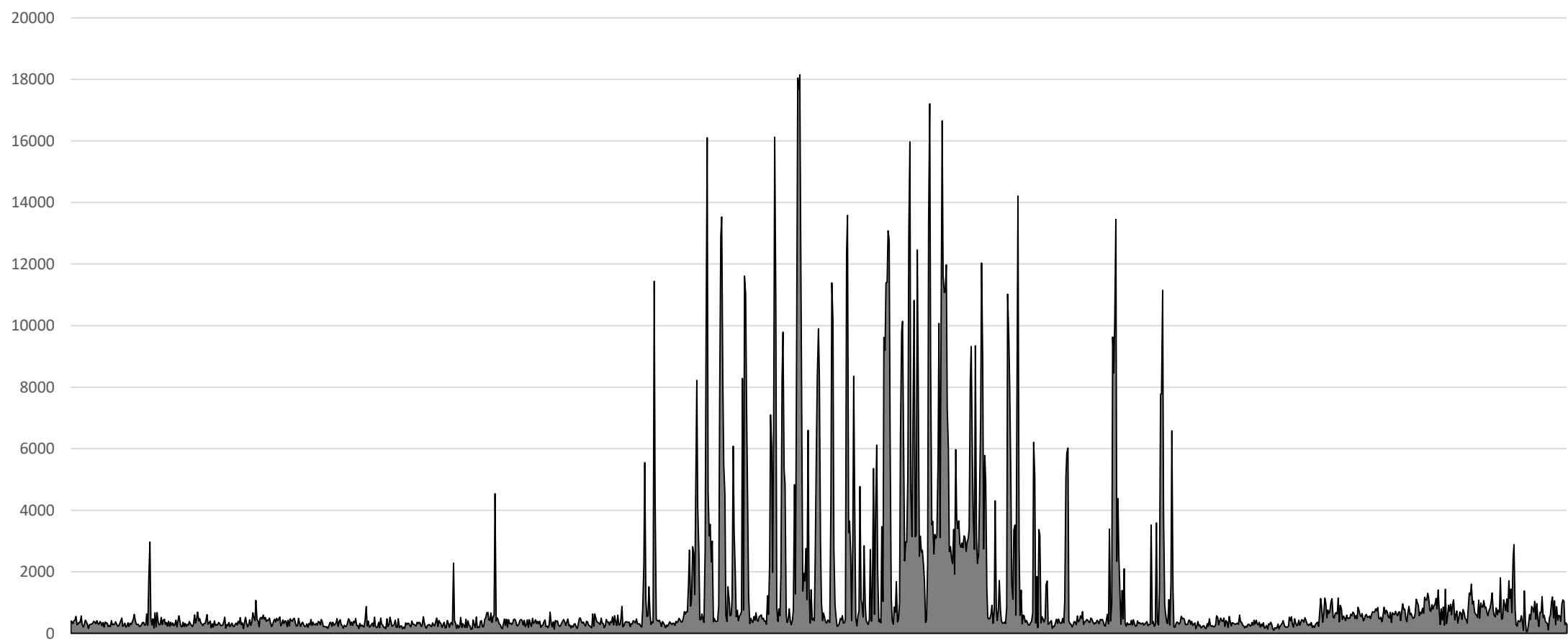


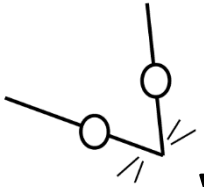
Commands per Minute (08.02.2023)





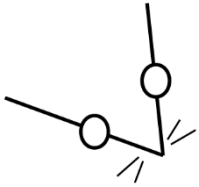
Average Request Time in ms (08.02.2023)



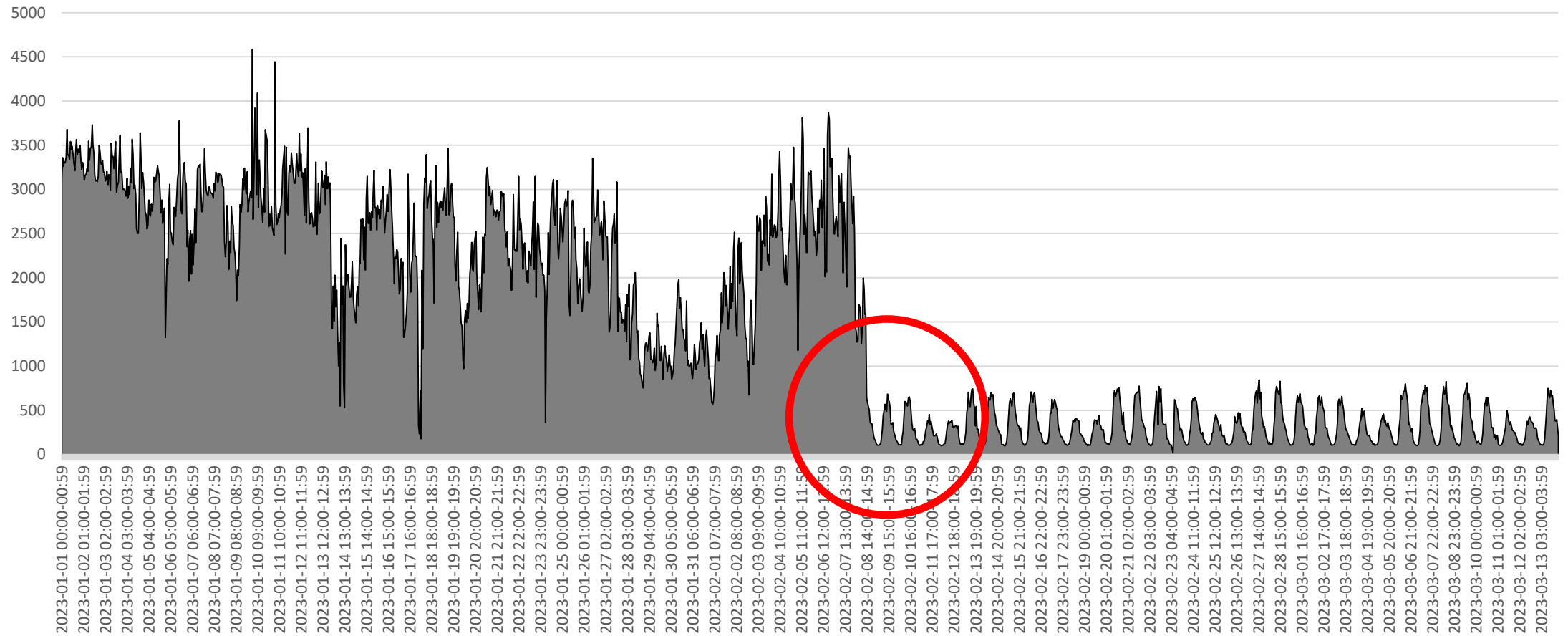


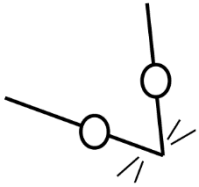
What is the cause of all the requests?

- Hypothesis 1: Bots and crawlers
- Hypothesis 2: Features
- Hypothesis 3: Suboptimal programming

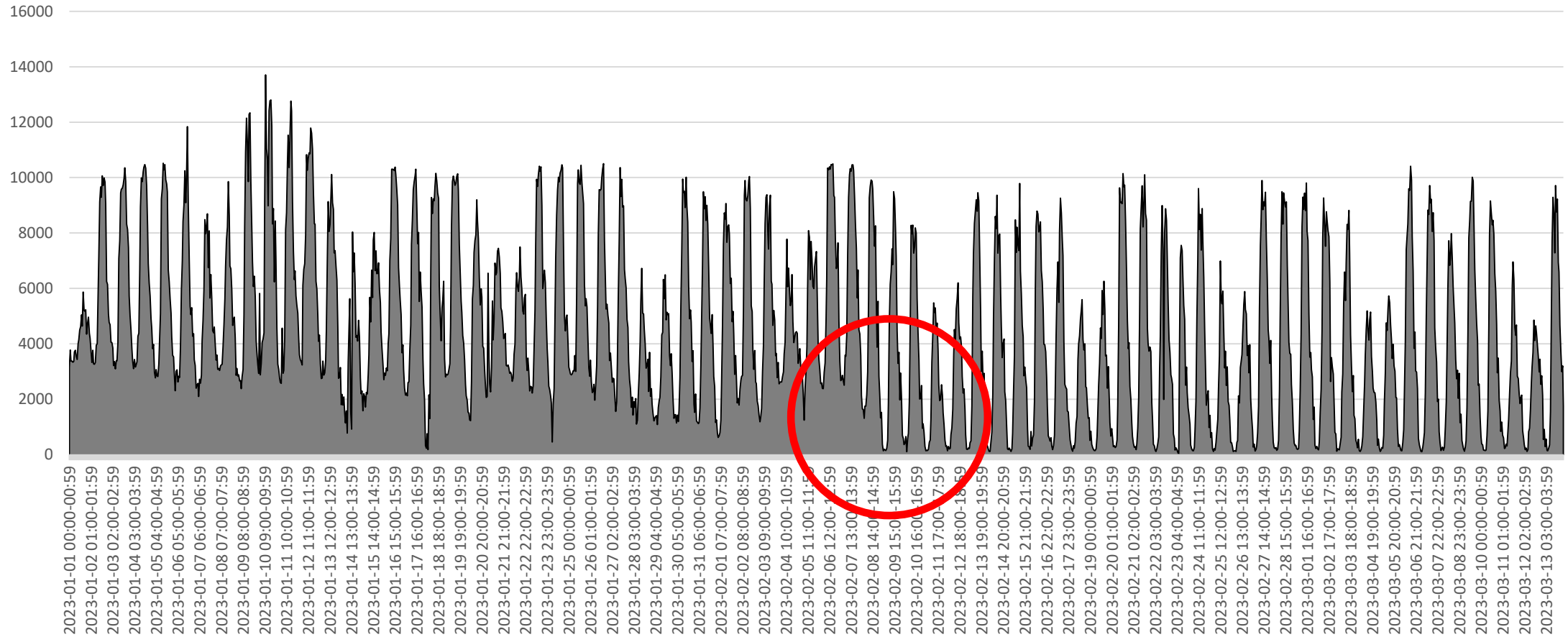


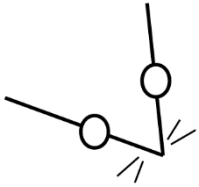
Sessions per Hour (01.01.23 – 14.03.23)





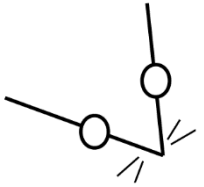
Commands per Hour (01.01.23 – 14.03.23)



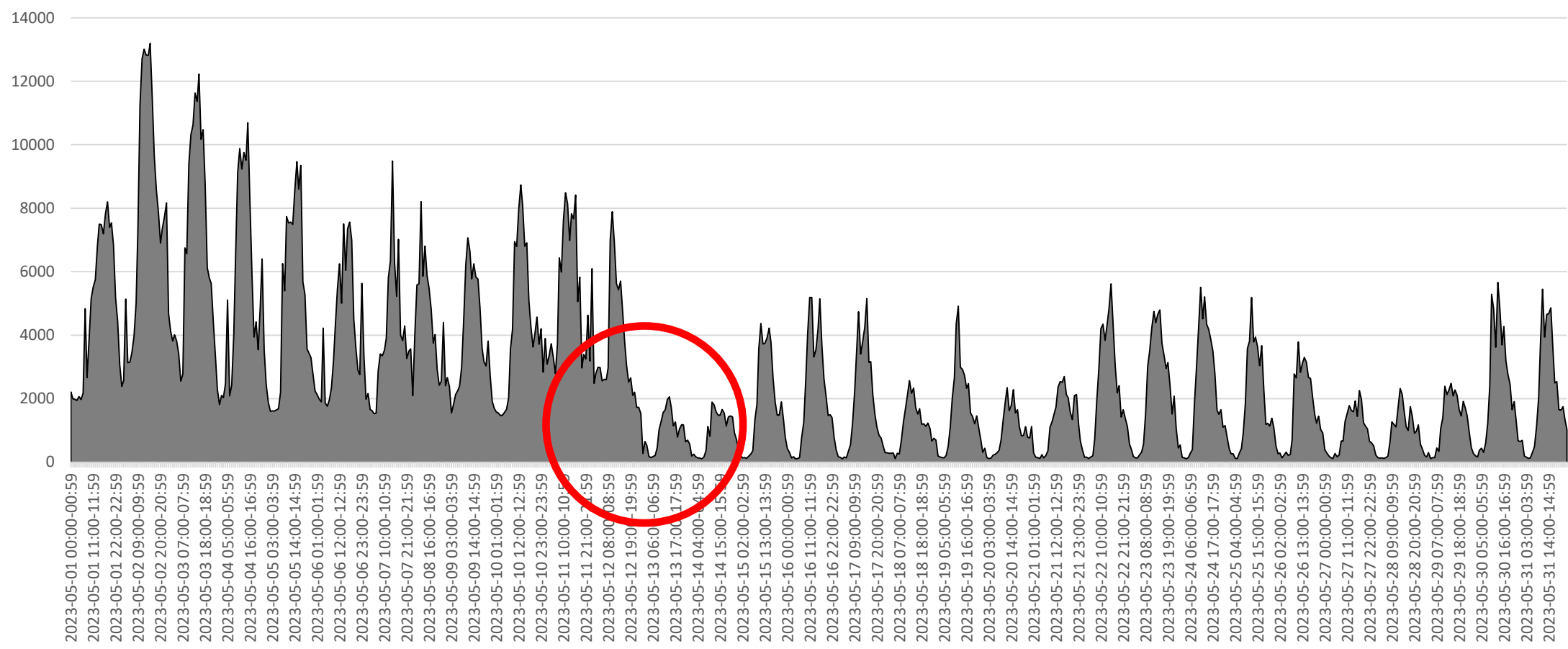


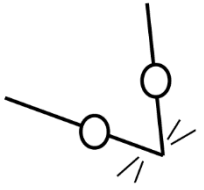
User Interactions (01.01.23 – 31.03.23)

User Interaction	Index Requests
Result list „Books & More“	19%
Article availability (AJAX)	19%
Number of hits in other tab (AJAX)	16%
Location facet (AJAX)	14%
Detail view „Books & More“	12%
Depending works (AJAX)	10%
Other	10%



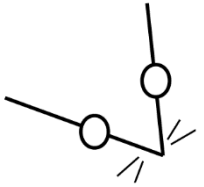
Commands per Hour (01.05.23 – 31.05.23)



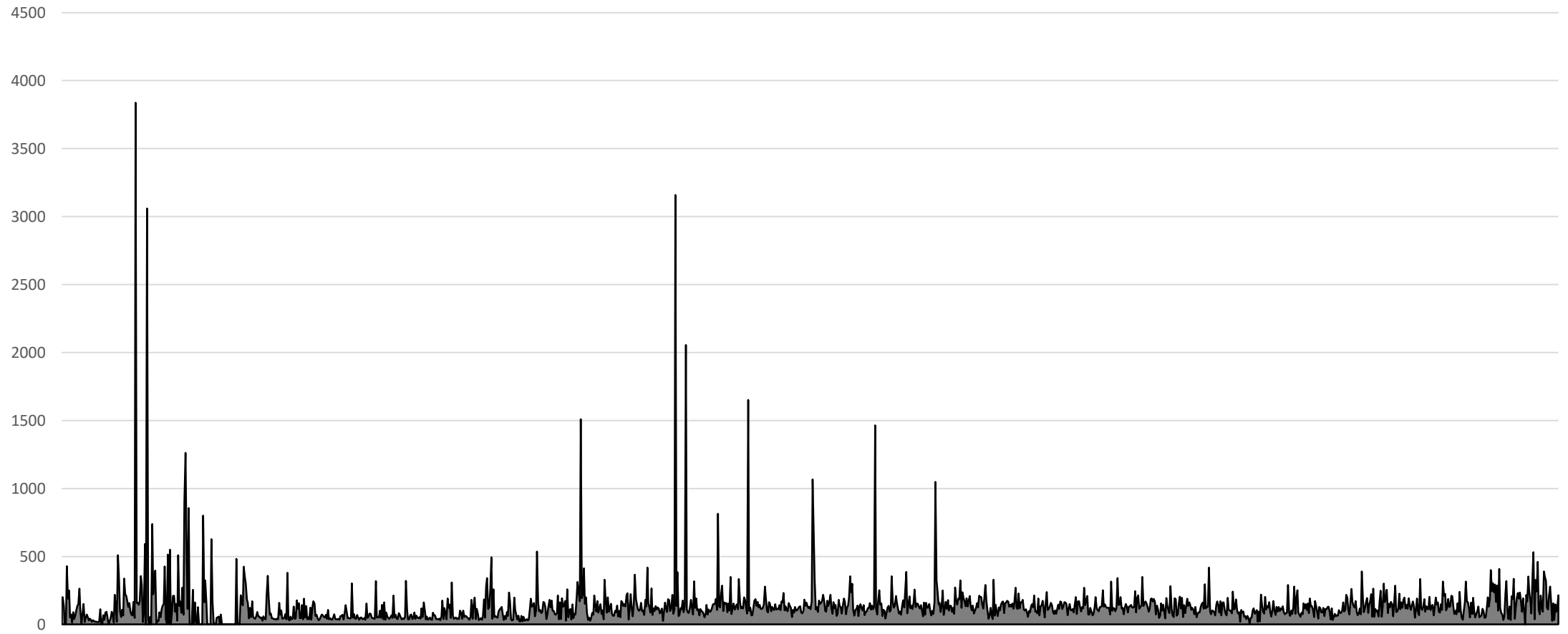


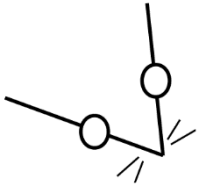
Addendum: SearchCache

- VuFind 8+ supports caching of search requests
- Activated for Katalog*plus* using the memcached backend
 - Memory 1GB
 - TTL 4h
 - Cache ratio ca. 20%

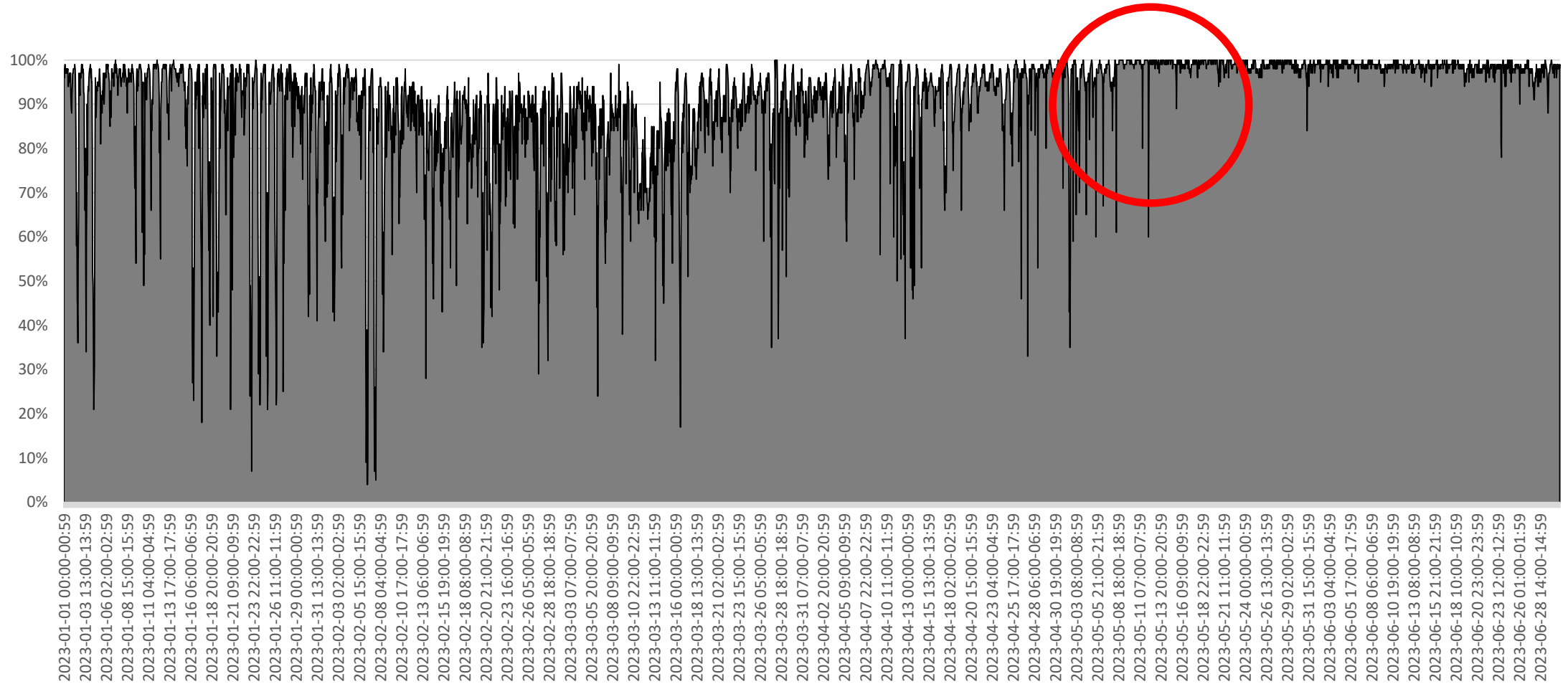


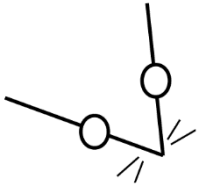
Average Request Time in ms (09.08.2023)





Commands w/ Response Time < 1000ms



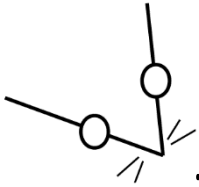


Conclusion



Summary

- We collect performance related data since appr. July 2022 and thus can measure the impact of our improvements.
- We used a combination of excluding bots & crawlers, caching, and disabling/refactoring features to improve the overall performance of our main discovery system.
- We found that there is „room for improvement“ in the VuFind core application, especially with regards to caching records.



Thank you!

- Composer-installable PHP component, GPLv3
<https://github.com/subhh/vufind-solrprobe>

David Maus

Email: david.maus@sub.uni-hamburg.de

Fediverse: @dmaus@dmaus.name