University of Bern


# Exploring Security Issues in Open Source Software

Supervisor

*Dr. Mohammad Ghafari*

Student

*Noah Bühlmann*

June 9, 2020

# Motivation

- OSS has become impressively popular
- Recent work has mostly focused on security bug report prediciton
- Timely reaction to security issues is critical, otherwise:

EQUIFAX

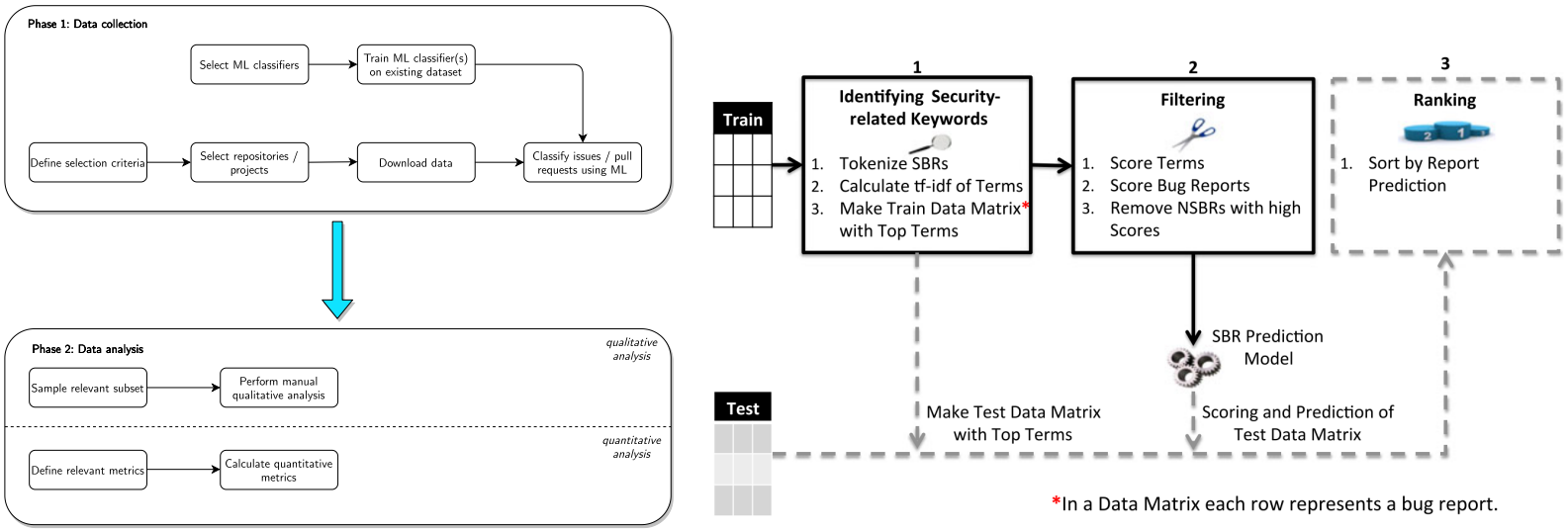
OpenSSL
Cryptography and SSL/TLS Toolkit

Struts

# Recap

Back in February, I thought I could use:

- ... the GHTorrent project.
- ... sophisticated ML classification.



# Large-scale analysis

- Define criteria for repository selection
- Write Python script for API querying
- Classify security issues using GitHub labels

  Labels

- Classify issue status (accepted, pending, rejected)

# Large-scale analysis

- Define criteria for repository selection
- Write Python script for API querying
- Classify security issues using GitHub labels


- Classify issue status (accepted, pending, rejected)

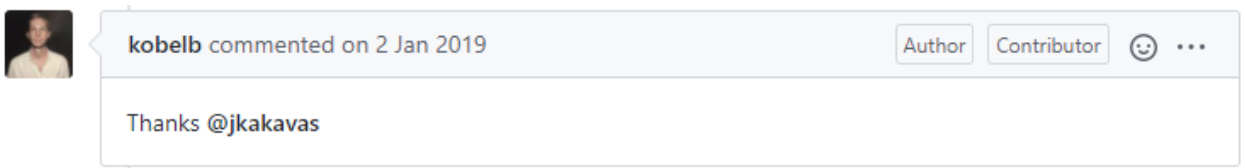
$\Longrightarrow$ Dataset with metadata from 182 repositories,
with $\approx$ 250'000 issues,
with $\approx$ 850'000 discussion comments,
with many quantitative and qualtitative features on repositories, issues and comments.

# Manual analysis

- Questionnaire with focus on qualitative discussion analysis and pull requests
- Random sample with 333 security issues
- Answer questionnaire for 333 issues and 1335 comments
- Notes of other interesting observations

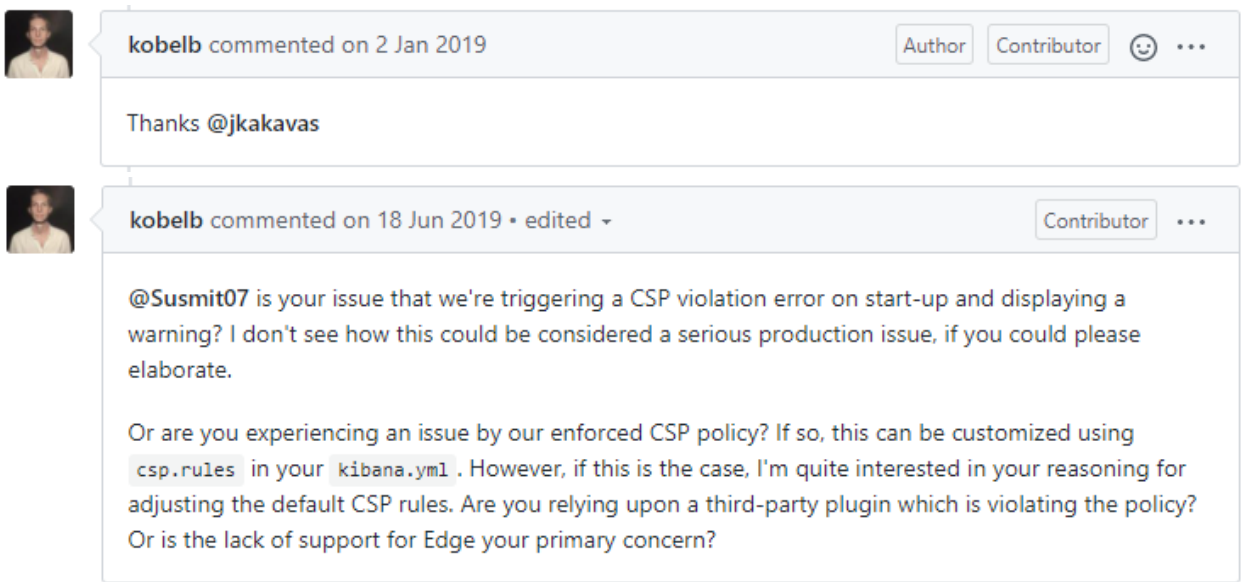
# Manual analysis

- Questionnaire with focus on qualitative discussion analysis and pull requests
- Random sample with 333 security issues
- Answer questionnaire for 333 issues and 1335 comments
- Notes of other interesting observations



# Manual analysis

- Questionnaire with focus on qualitative discussion analysis and pull requests
- Random sample with 333 security issues
- Answer questionnaire for 333 issues and 1335 comments
- Notes of other interesting observations



**kobelb** commented on 2 Jan 2019 — Author — Contributor

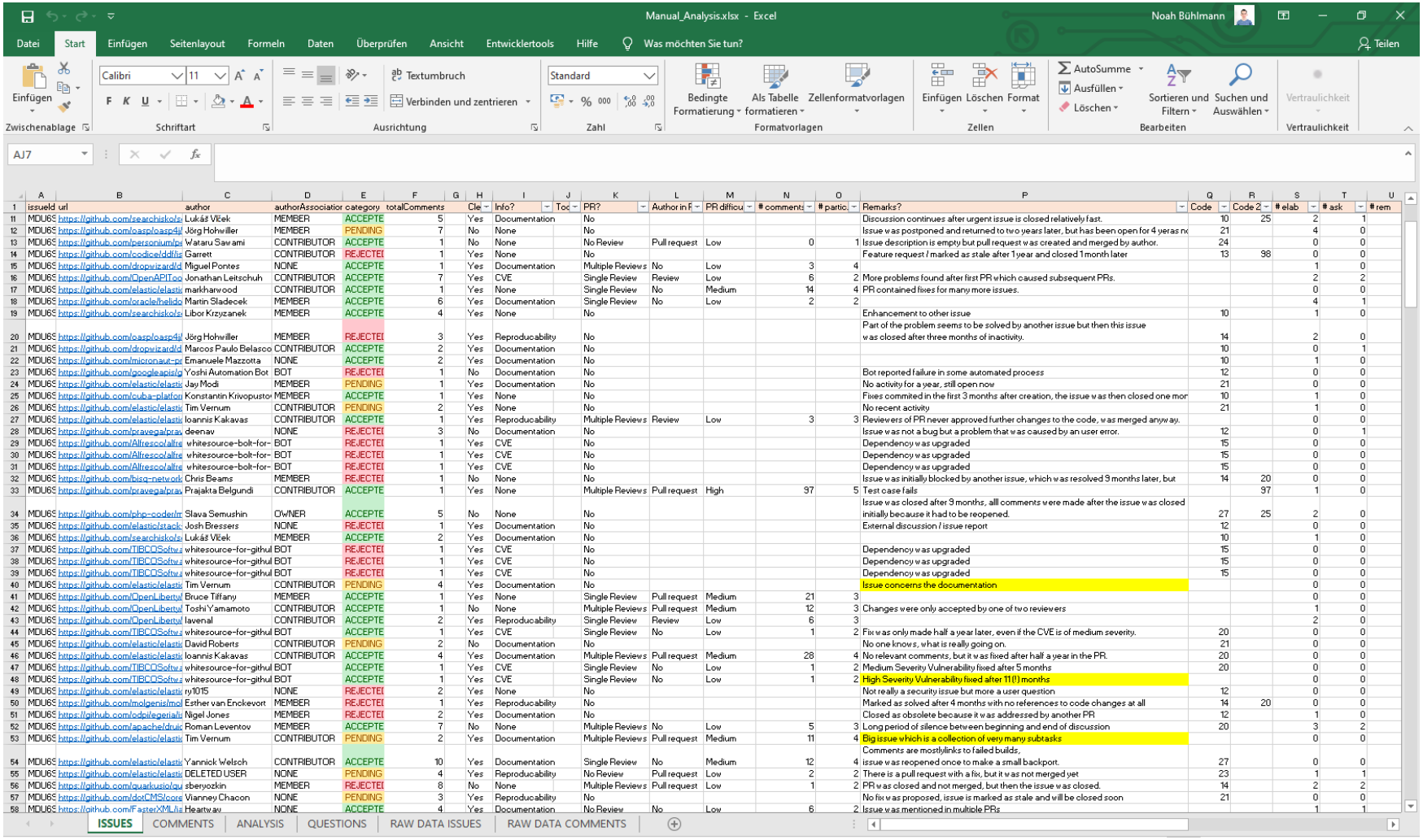
Thanks **@jkakavas**

**kobelb** commented on 18 Jun 2019 • edited — Contributor

**@Susmit07** is your issue that we're triggering a CSP violation error on start-up and displaying a warning? I don't see how this could be considered a serious production issue, if you could please elaborate.

Or are you experiencing an issue by our enforced CSP policy? If so, this can be customized using `csp.rules` in your `kibana.yml`. However, if this is the case, I'm quite interested in your reasoning for adjusting the default CSP rules. Are you relying upon a third-party plugin which is violating the policy? Or is the lack of support for Edge your primary concern?

# Manual analysis



Manual_Analysis.xlsx - Excel

| issueId | url | author | authorAssociation | category | totalComments | Cle | Info? | Toc | PR? | Author in F | PR difficu | # comments | # partic | Remarks? | Code | Code 2 | # elab | # ask | # rem |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| MDUS | https://github.com/searchisko/s | Lukáš Vlček | MEMBER | ACCEPTE | 5 | Yes | Documentation | | No | | | | | Discussion continues after urgent issue is closed relatively fast. | 10 | 25 | 2 | 1 | |
| MDUS | https://github.com/oasp/oasp4j | Jörg Hohwiller | MEMBER | PENDING | 7 | No | None | | No | | | | | Issue was postponed and returned to two years later, but has been open for 4 yeras n | 21 | | 4 | 0 | |
| MDUS | https://github.com/perzonium/pe | Wataru Sawami | CONTRIBUTOR | ACCEPTE | 1 | No | None | | No Review | Pull request | Low | 0 | 1 | Issue description is empty but pull request was created and merged by author. | 24 | | 0 | 0 | |
| MDUS | https://github.com/codice/ddf/is | Garrett | CONTRIBUTOR | REJECTEI | 1 | Yes | None | | No | | | | | Feature request / marked as stale after 1 year and closed 1 month later | 13 | 98 | 0 | 0 | |
| MDUS | https://github.com/dropwizard/d | Miguel Pontes | NONE | ACCEPTE | 1 | Yes | Documentation | | Multiple Reviews | No | Low | 3 | 4 | | | | 1 | 0 | |
| MDUS | https://github.com/OpenAPITod | Jonathan Leitschuh | CONTRIBUTOR | ACCEPTE | 7 | Yes | CVE | | Single Review | Review | Low | 6 | 2 | More problems found after first PR which caused subsequent PRs. | | | 2 | 2 | |
| MDUS | https://github.com/elastic/elastic | markharwood | CONTRIBUTOR | ACCEPTE | 1 | Yes | None | | Single Review | No | Medium | 14 | 4 | PR contained fixes for many more issues. | | | 0 | 0 | |
| MDUS | https://github.com/oracle/helido | Martin Sladecek | MEMBER | ACCEPTE | 6 | Yes | Documentation | | Single Review | No | Low | 2 | 2 | | | | 4 | 1 | |
| MDUS | https://github.com/searchisko/s | Libor Krzyzanek | MEMBER | ACCEPTE | 4 | Yes | None | | No | | | | | Enhancement to other issue | 10 | | 1 | 0 | |
| MDUS | https://github.com/oasp/oasp4j | Jörg Hohwiller | MEMBER | REJECTEI | 3 | Yes | Reproducability | | No | | | | | Part of the problem seems to be solved by another issue but then this issue was closed after three months of inactivity. | 14 | | 2 | 0 | |
| MDUS | https://github.com/dropwizard/d | Marcos Paulo Belasco | CONTRIBUTOR | ACCEPTE | 2 | Yes | Documentation | | No | | | | | | 10 | | 0 | 1 | |
| MDUS | https://github.com/micronaut-pr | Emanuele Mazzotta | NONE | ACCEPTE | 2 | Yes | Documentation | | No | | | | | | 10 | | 1 | 0 | |
| MDUS | https://github.com/googleapis/g | Yoshi Automation Bot | BOT | REJECTEI | 1 | No | Documentation | | No | | | | | Bot reported failure in some automated process | 12 | | 0 | 0 | |
| MDUS | https://github.com/elastic/elastic | Jay Modi | MEMBER | PENDING | 1 | Yes | Documentation | | No | | | | | No activity for a year, still open now | 21 | | 0 | 0 | |
| MDUS | https://github.com/cuba-platfor | Konstantin Krivopusto | MEMBER | ACCEPTE | 1 | Yes | None | | No | | | | | Fixes commited in the first 3 months after creation, the issue was then closed one mor | 10 | | 1 | 0 | |
| MDUS | https://github.com/elastic/elastic | Tim Vernum | CONTRIBUTOR | PENDING | 2 | Yes | None | | No | | | | | No recent activity | 21 | | 1 | 0 | |
| MDUS | https://github.com/elastic/elastic | Ioannis Kakavas | CONTRIBUTOR | ACCEPTE | 1 | Yes | Reproducability | | Multiple Reviews | Review | Low | 3 | 3 | Reviewers of PR never approved further changes to the code, was merged anyway. | | | 0 | 0 | |
| MDUS | https://github.com/pravega/prav | deenav | NONE | REJECTEI | 3 | No | Documentation | | No | | | | | Issue was not a bug but a problem that was caused by an user error. | 12 | | 0 | 1 | |
| MDUS | https://github.com/Alfresco/alfre | whitesource-bolt-for- | BOT | REJECTEI | 1 | Yes | CVE | | No | | | | | Dependency was upgraded | 15 | | 0 | 0 | |
| MDUS | https://github.com/Alfresco/alfre | whitesource-bolt-for- | BOT | REJECTEI | 1 | Yes | CVE | | No | | | | | Dependency was upgraded | 15 | | 0 | 0 | |
| MDUS | https://github.com/Alfresco/alfre | whitesource-bolt-for- | BOT | REJECTEI | 1 | Yes | CVE | | No | | | | | Dependency was upgraded | 15 | | 0 | 0 | |
| MDUS | https://github.com/bisq-network | Chris Beams | MEMBER | REJECTEI | 1 | No | None | | No | | | | | Issue was initially blocked by another issue, which was resolved 9 months later, but | 14 | 20 | 0 | 0 | |
| MDUS | https://github.com/pravega/prav | Prajakta Belgundi | CONTRIBUTOR | ACCEPTE | 1 | Yes | None | | Multiple Reviews | Pull request | High | 97 | 5 | Test case fails | | 97 | 1 | 0 | |
| MDUS | https://github.com/php-coder/m | Slava Semushin | OWNER | ACCEPTE | 5 | No | None | | No | | | | | Issue was closed after 9 months, alll comments were made after the issue was closed initially because it had to be reopened. | 27 | 25 | 2 | 0 | |
| MDUS | https://github.com/elastic/stack | Josh Bressers | NONE | REJECTEI | 1 | Yes | Documentation | | No | | | | | External discussion / issue report | 12 | | 0 | 0 | |
| MDUS | https://github.com/searchisko/s | Lukáš Vlček | MEMBER | ACCEPTE | 2 | Yes | Documentation | | No | | | | | | 10 | | 1 | 0 | |
| MDUS | https://github.com/TIBCOSoftwa | whitesource-for-githul | BOT | REJECTEI | 1 | Yes | CVE | | No | | | | | Dependency was upgraded | 15 | | 0 | 0 | |
| MDUS | https://github.com/TIBCOSoftwa | whitesource-for-githul | BOT | REJECTEI | 1 | Yes | CVE | | No | | | | | Dependency was upgraded | 15 | | 0 | 0 | |
| MDUS | https://github.com/TIBCOSoftwa | whitesource-for-githul | BOT | REJECTEI | 1 | Yes | CVE | | No | | | | | Dependency was upgraded | 15 | | 0 | 0 | |
| MDUS | https://github.com/elastic/elastic | Tim Vernum | CONTRIBUTOR | PENDING | 4 | Yes | Documentation | | No | | | | | Issue concerns the documentation | | | 0 | 0 | |
| MDUS | https://github.com/OpenLiberty/ | Bruce Tiffany | MEMBER | ACCEPTE | 1 | Yes | None | | Single Review | Pull request | Medium | 21 | 3 | | | | 0 | 0 | |
| MDUS | https://github.com/OpenLiberty/ | Toshi Yamamoto | CONTRIBUTOR | ACCEPTE | 1 | No | None | | Multiple Reviews | Pull request | Medium | 12 | 3 | Changes were only accepted by one of two reviewers | | | 1 | 0 | |
| MDUS | https://github.com/OpenLiberty/ | lavenal | CONTRIBUTOR | ACCEPTE | 2 | Yes | Reproducability | | Single Review | Review | Low | 6 | 3 | | | | 2 | 0 | |
| MDUS | https://github.com/TIBCOSoftwa | whitesource-for-githul | BOT | ACCEPTE | 1 | Yes | CVE | | Single Review | No | Low | 1 | 2 | Fix was only made half a year later, even if the CVE is of medium severity. | 20 | | 0 | 0 | |
| MDUS | https://github.com/elastic/elastic | David Roberts | CONTRIBUTOR | PENDING | 2 | No | Documentation | | No | | | | | No one knows, what is really going on. | 21 | | 0 | 0 | |
| MDUS | https://github.com/elastic/elastic | Ioannis Kakavas | CONTRIBUTOR | ACCEPTE | 4 | Yes | Documentation | | Multiple Reviews | Pull request | Medium | 28 | 4 | No relevant comments, but it was fixed after half a year in the PR. | 20 | | 0 | 0 | |
| MDUS | https://github.com/TIBCOSoftwa | whitesource-for-githul | BOT | ACCEPTE | 1 | Yes | CVE | | Single Review | No | Low | 1 | 2 | Medium Severity Vulnerability fixed after 5 months | 20 | | 0 | 0 | |
| MDUS | https://github.com/TIBCOSoftwa | whitesource-for-githul | BOT | ACCEPTE | 1 | Yes | CVE | | Single Review | No | Low | 1 | 2 | High Severity Vulnerability fixed after 11 (!) months | | | 0 | 0 | |
| MDUS | https://github.com/elastic/elastic | ry1015 | NONE | REJECTEI | 2 | Yes | None | | No | | | | | Not really a security issue but more a user question | 12 | | 0 | 0 | |
| MDUS | https://github.com/molgenis/mol | Esther van Enckevort | MEMBER | REJECTEI | 1 | Yes | Reproducability | | No | | | | | Marked as solved after 4 months with no references to code changes at all | 14 | 20 | 0 | 0 | |
| MDUS | https://github.com/odpi/egeriali | Nigel Jones | MEMBER | REJECTEI | 2 | Yes | Documentation | | No | | | | | Closed as obsolete because it was addressed by another PR | 12 | | 1 | 0 | |
| MDUS | https://github.com/apache/druic | Roman Leventov | MEMBER | ACCEPTE | 7 | No | None | | Multiple Reviews | No | Low | 5 | 3 | Long period of silence between beginning and end of discussion | 20 | | 3 | 2 | |
| MDUS | https://github.com/elastic/elastic | Tim Vernum | CONTRIBUTOR | PENDING | 2 | Yes | Documentation | | Multiple Reviews | Pull request | Medium | 11 | 4 | Big issue which is a collection of very many subtasks | | | 0 | 0 | |
| MDUS | https://github.com/elastic/elastic | Yannick Welsch | CONTRIBUTOR | ACCEPTE | 10 | Yes | Documentation | | Single Review | No | Medium | 12 | 4 | Comments are mostlylinks to failed builds, issue was reopened once to make a small backport. | 27 | | 0 | 0 | |
| MDUS | https://github.com/elastic/elastic | DELETED USER | NONE | PENDING | 4 | Yes | Reproducability | | No Review | Pull request | Low | 2 | 2 | There is a pull request with a fix, but it was not merged yet | 23 | | 1 | 1 | |
| MDUS | https://github.com/quarkusio/qu | sberyozkin | MEMBER | REJECTEI | 8 | No | None | | Multiple Reviews | Pull request | Low | 1 | 2 | PR was closed and not merged, but then the issue was closed. | 14 | | 2 | 2 | |
| MDUS | https://github.com/dotCMS/core | Vianney Chacon | NONE | PENDING | 3 | Yes | Reproducability | | No | | | | | No fix was proposed, issue is marked as stale and will be closed soon | 21 | | 0 | 0 | |
| MDUS | https://github.com/FasterXML/ja | Heartwau | NONE | ACCEPTE | 4 | Yes | Documentation | | No Review | No | Low | 6 | 2 | Issue was mentioned in multiple PRs | | | 1 | 1 | |

ISSUES | COMMENTS | ANALYSIS | QUESTIONS | RAW DATA ISSUES | RAW DATA COMMENTS

# Results

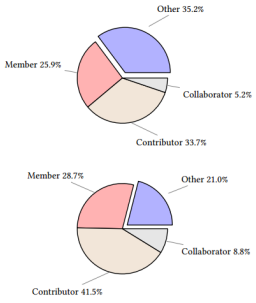



**Figure 9: Reporters of non-security issues (top) and security issues (bottom)**

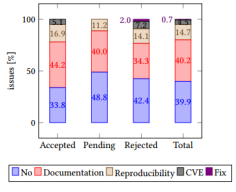



**Figure 14: Further information via hyperlinks in issue report**

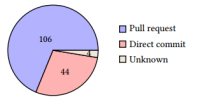



**Figure 20: Implementation of accepted security issues**

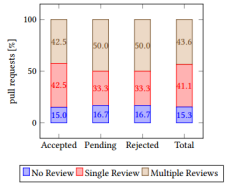
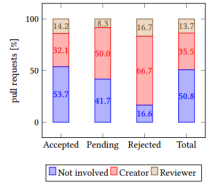



**Figure 21: Review situation of pull requests**



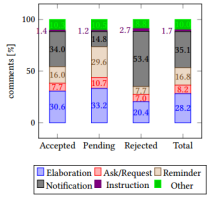
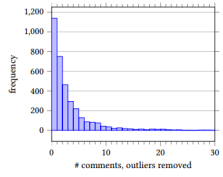




**Figure 16: Mean composition of security issue discussions**



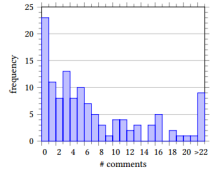




**Figure 23: Number of comments in pull request discussions**

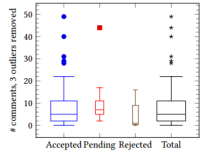



**Figure 24: Number of comments in pull request discussions**

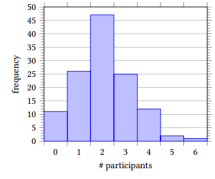
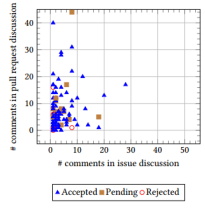



**Figure 25: Number of participants in pull request discussions**



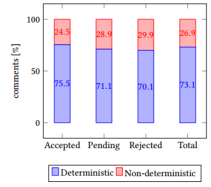




**Figure 17: Deterministic vs. non-deterministic elaboration**

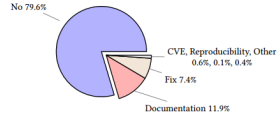



**Figure 18: Hyperlinks in issue discussions**

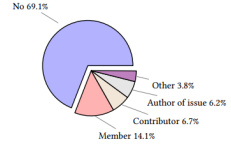



**Figure 19: Mentioning in issue discussions**

# Results

- Security issues, compared to non-security issues...

# Results

- Security issues, compared to non-security issues...
  - are very infrequent

# Results

- Security issues, compared to non-security issues...
  - are very infrequent
  - appear later in a project

# Results

- Security issues, compared to non-security issues...
  - are very infrequent
  - appear later in a project
  - are more often reported by core members of a project

# Results

- Security issues, compared to non-security issues...
  - are very infrequent
  - appear later in a project
  - are more often reported by core members of a project
  - are reported by a small circle of persons

# Results

- Security issues, compared to non-security issues...
  - are very infrequent
  - appear later in a project
  - are more often reported by core members of a project
  - are reported by a small circle of persons
  - receive a faster first reaction

# Results

- Security issues, compared to non-security issues...
  - are very infrequent
  - appear later in a project
  - are more often reported by core members of a project
  - are reported by a small circle of persons
  - receive a faster first reaction
  - have a slower proceeding discussion

# Results

- Security issues, compared to non-security issues...
  - are very infrequent
  - appear later in a project
  - are more often reported by core members of a project
  - are reported by a small circle of persons
  - receive a faster first reaction
  - have a slower proceeding discussion
  - have fewer comments in their discussions

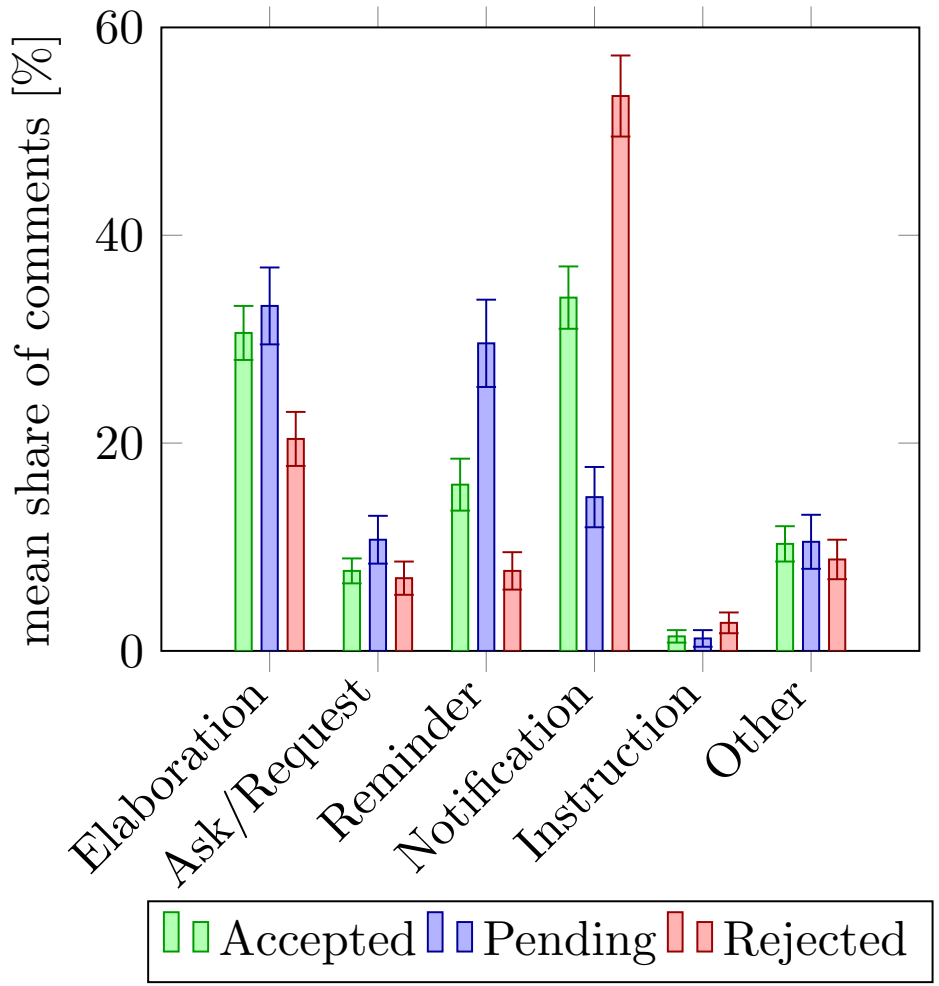
# Results

- Security issues, compared to non-security issues...
  - are very infrequent
  - appear later in a project
  - are more often reported by core members of a project
  - are reported by a small circle of persons
  - receive a faster first reaction
  - have a slower proceeding discussion
  - have fewer comments in their discussions
  - are resolved slower

# Results

- Security issues, compared to non-security issues...
  - are very infrequent
  - appear later in a project
  - are more often reported by core members of a project
  - are reported by a small circle of persons
  - receive a faster first reaction
  - have a slower proceeding discussion
  - have fewer comments in their discussions
  - are resolved slower
- Low number of comments and participants in security issues and pull requests

# Results


mean share of comments [%]
60
40
20
0
Elaboration
Ask/Request
Reminder
Notification
Instruction
Other
Accepted
Pending
Rejected

# Results

- Accepted security issues
  - are more often clearly explained
  - more often include additional documentation or reproducibility information
  - have a higher proportion of thematically relevant comments.

  compared to rejected security issues.

# Results

- Accepted security issues
  - are more often clearly explained
  - more often include additional documentation or reproducibility information
  - have a higher proportion of thematically relevant comments.

  compared to rejected security issues.
- Security issues are resolved faster if a fix is proposed in the comments or a CVE report is present.
- Security issues are resolved faster if someone is assigned to them.

# Results

- Accepted security issues
  - are more often clearly explained
  - more often include additional documentation or reproducibility information
  - have a higher proportion of thematically relevant comments.

  compared to rejected security issues.
- Security issues are resolved faster if a fix is proposed in the comments or a CVE report is present.
- Security issues are resolved faster if someone is assigned to them.
- Security issues are resolved using pull requests in 3 of 4 cases.
- Pull requests get reviewed in nearly 85% of the cases.

# Results

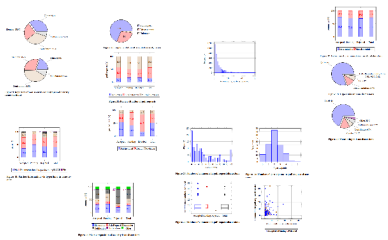
- Accepted security issues
  - are more often clearly explained
  - more often include additional documentation or reproducibility information
  - have a higher proportion of thematically relevant comments.

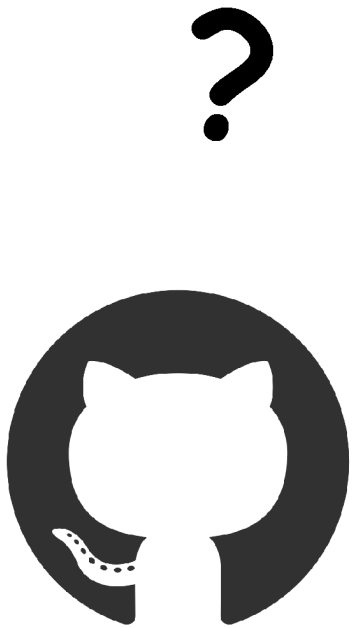
  compared to rejected security issues.
- Security issues are resolved faster if a fix is proposed in the comments or a CVE report is present.
- Security issues are resolved faster if someone is assigned to them.
- Security issues are resolved using pull requests in 3 of 4 cases.
- Pull requests get reviewed in nearly 85% of the cases.
- The acceptance of security issues does not significantly differ between the reporters gender (male 72.0%/female 70.4%).

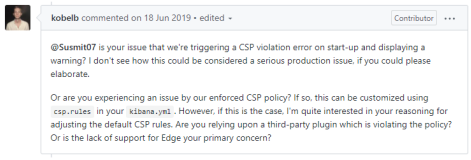
# Conclusion

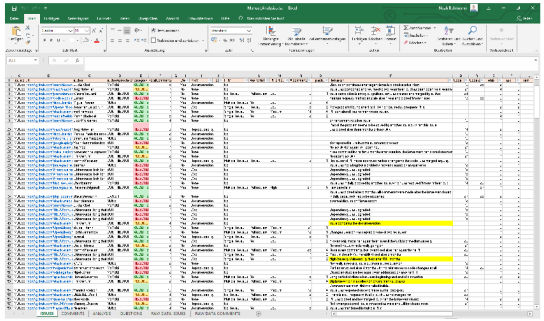
- Many possibilities for future work and improvement:
  - Increase representativeness (more issue trackers, other programming languages)
  - Better security issue classification using ML
  - More features and metadata (number of words, source code snippets, 'difficulty' of fix)
  - NLP on content of issue reports and comments

# Thank you!






kobelb commented on 18 Jun 2019 • edited
Contributor
@Susmit07 is your issue that we're triggering a CSP violation error on start-up and displaying a warning? I don't see how this could be considered a serious production issue, if you could please elaborate.
Or are you experiencing an issue by our enforced CSP policy? If so, this can be customized using csp.rules in your kibana.yml. However, if this is the case, I'm quite interested in your reasoning for adjusting the default CSP rules. Are you relying upon a third-party plugin which is violating the policy? Or is the lack of support for Edge your primary concern?



# Appendix: Sources

- Images: Wikimedia - Creative Commons
- GitHub: elastic, kibana, https://github.com/elastic/kibana/issues/30468
- Peters, Fayola, et al. “Text filtering and ranking for security bug report prediction.” IEEE Transactions on Software Engineering (2017).