EXTERNAL REPOSITORY RELIANCE

Thesis committee: Dr. ing. Sebastian Proksch,
Dr. ing. Casper Bach Poulsen & Shujun Huang

What is the Impact of External Repositories on Packages in Maven Central?

What is the Problem?

98% of all software uses open-source code

Artifact repositories like Maven Central are often used to retrieve open-source libraries to include them in applications

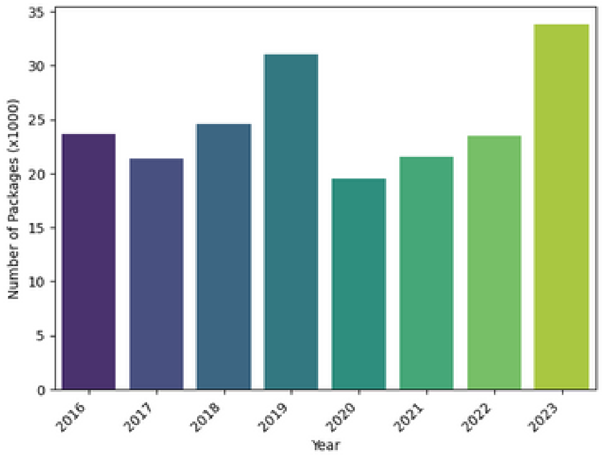
Usage of external repositories poses risks:
Deprecated/malicious software
Mirror Package Override Attacks
Legal Complications
Increased complexity of project
Unintended behavior

How has it been Investigated ?

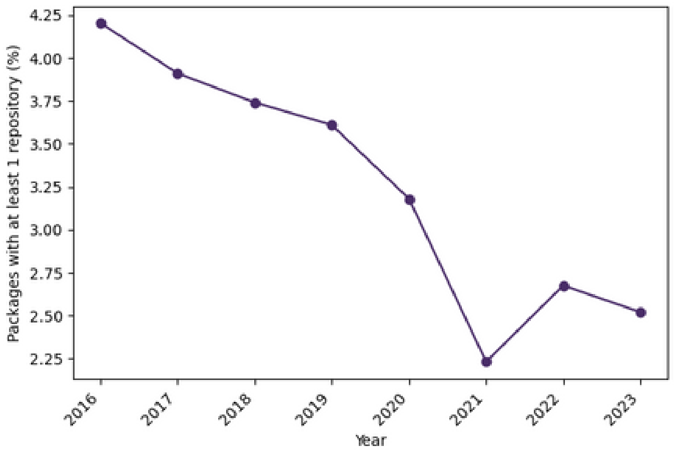
Leveraged Maven indexes to collect packages from every 50th index from 400 till 800

Random sampling to include one version per package to avoid data skewing by popular packages

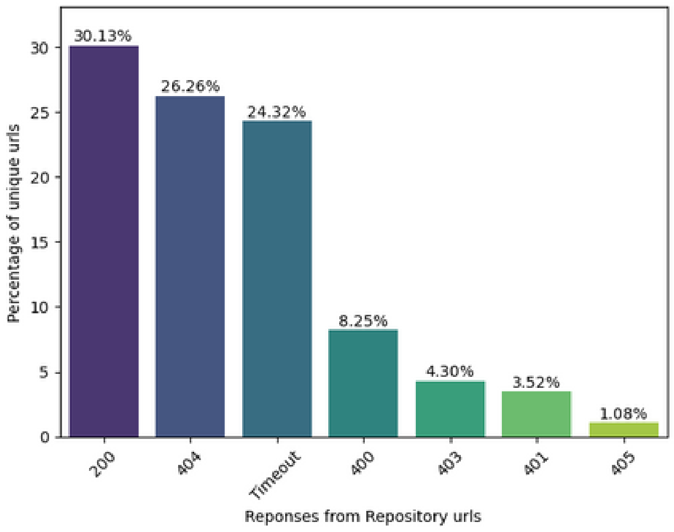
Total packages collected: 934,266
Total sample size: 199,188



Usage of repositories is declining



69.58% unreachable repositories.



What has been found?

3.29% of packages have one or more repositories

External repositories have no significant effect on error rate

19.58% of errors of packages with repositories were caused by one of their repositories

id and url collisions can cause problems and unintended behavior like retrieving packages from the wrong repositories

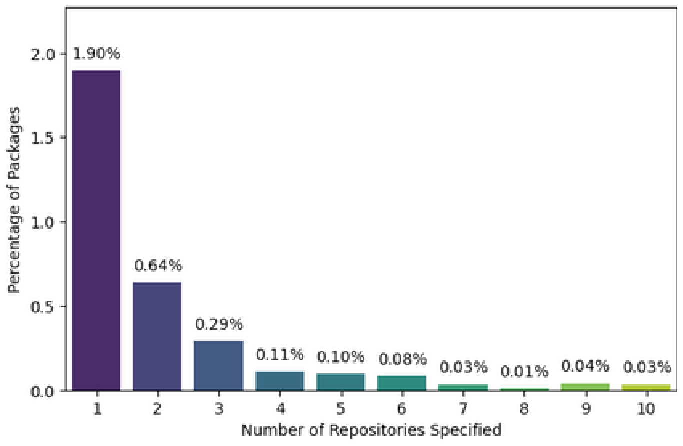
19.31% of unique ids have collisions

20.75% of unique urls have collisions

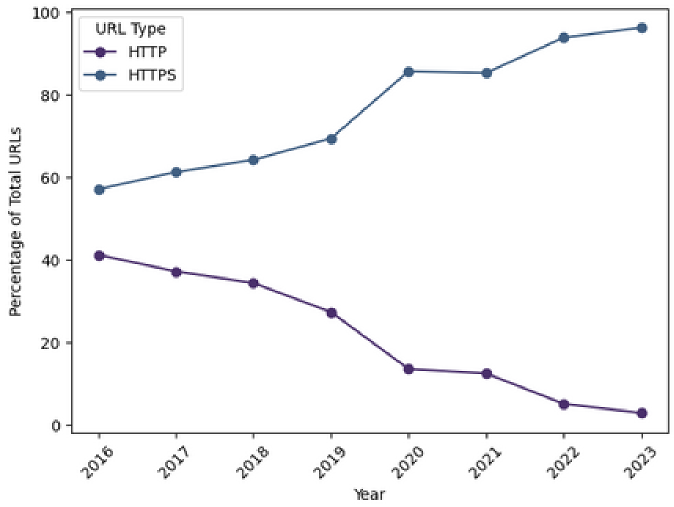
central id used 555 times with 31 different urls

7.8 unique urls per id

One external repository is the most common



HTTP usage declined to almost zero



Conclusion

Packages in Maven Central are at risks of unintended behavior, unknowingly importing malicious code an the other problems we identified.

Url and id collisions indicate that more guidelines around the naming is necessary

Threats to Validity

Generalizability to other ecosystems (e.g. npm, PyPI)
Accuracy of data collection
Accuracy of error collection
Reachability of urls

Future Work

Unified repository id naming convention
Effect of disabled directory listings

Recommendations

Evaluate...
...the need for external repositories
...the security and alternatives of the used external repositories
...the existing POM and settings files for url and id collisions
...whether repositories still resolve and HTTPS is used

JELLE SANDIFORT

☎ +31 6 83 15 69 69

✉ j.w.sandifort@student.tudelft.nl

TUDelft