UNIVERSITAS SCIENTIARUM SZEGEDIENSIS
UNIVERSITY OF SZEGED
Department of Software Engineering

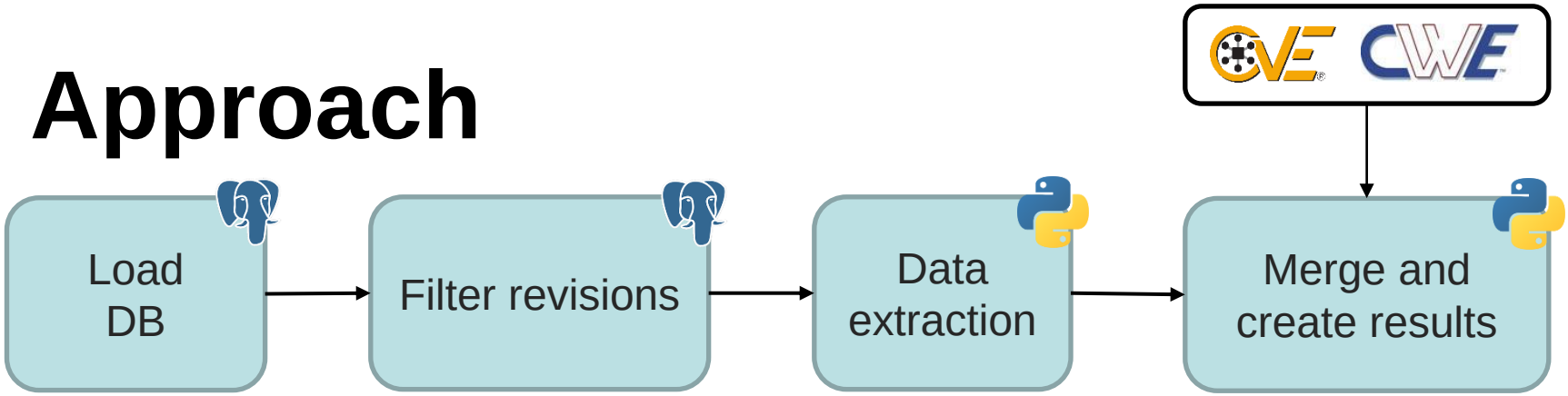
Exploring the Security Awareness of the Python and JavaScript Open Source Communities

Gábor Antal, Márton Keleti, Péter Hegedűs
{antal, keletim, hpeter}@inf.u-szeged.hu

Introduction

- Importance of software security
  - financial damage, compromise infrastructure, threaten countries
- Software Heritage Graph Dataset
  - A very large dataset containing the development history of publicly available software (fully-deduplicated Merkle DAG)
- Python and JavaScript
- Our focus:
  - What are the **typical security vulnerability types** the JavaScript and Python open-source communities mitigate and how do they relate to each other? (RQ1)
  - **How quickly** the JavaScript and Python open-source communities **mitigate** a newly published security vulnerability? (RQ2)

Approach



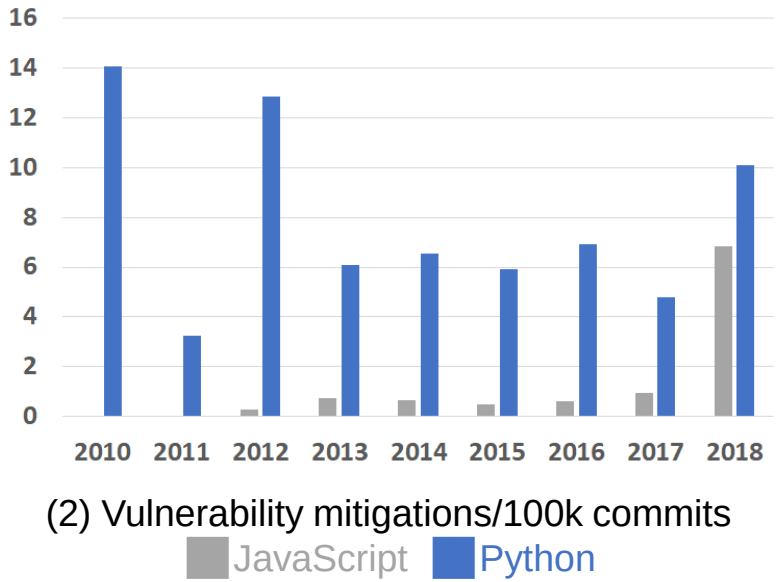
1. Load database
   - Difficulties because of the size of SWHGD
2. Filter revisions (SQL scripts)
3. Fine filtering and data extraction (Python)
4. Merge and create results
   - Download and load vulnerability data from MITRE (CVE/CWE)
   - Merge with the mined data
   - Create the results

Results

- 6,342 vulnerability mitigation commits
  - 3,458 for JavaScript
  - 2,884 for Python

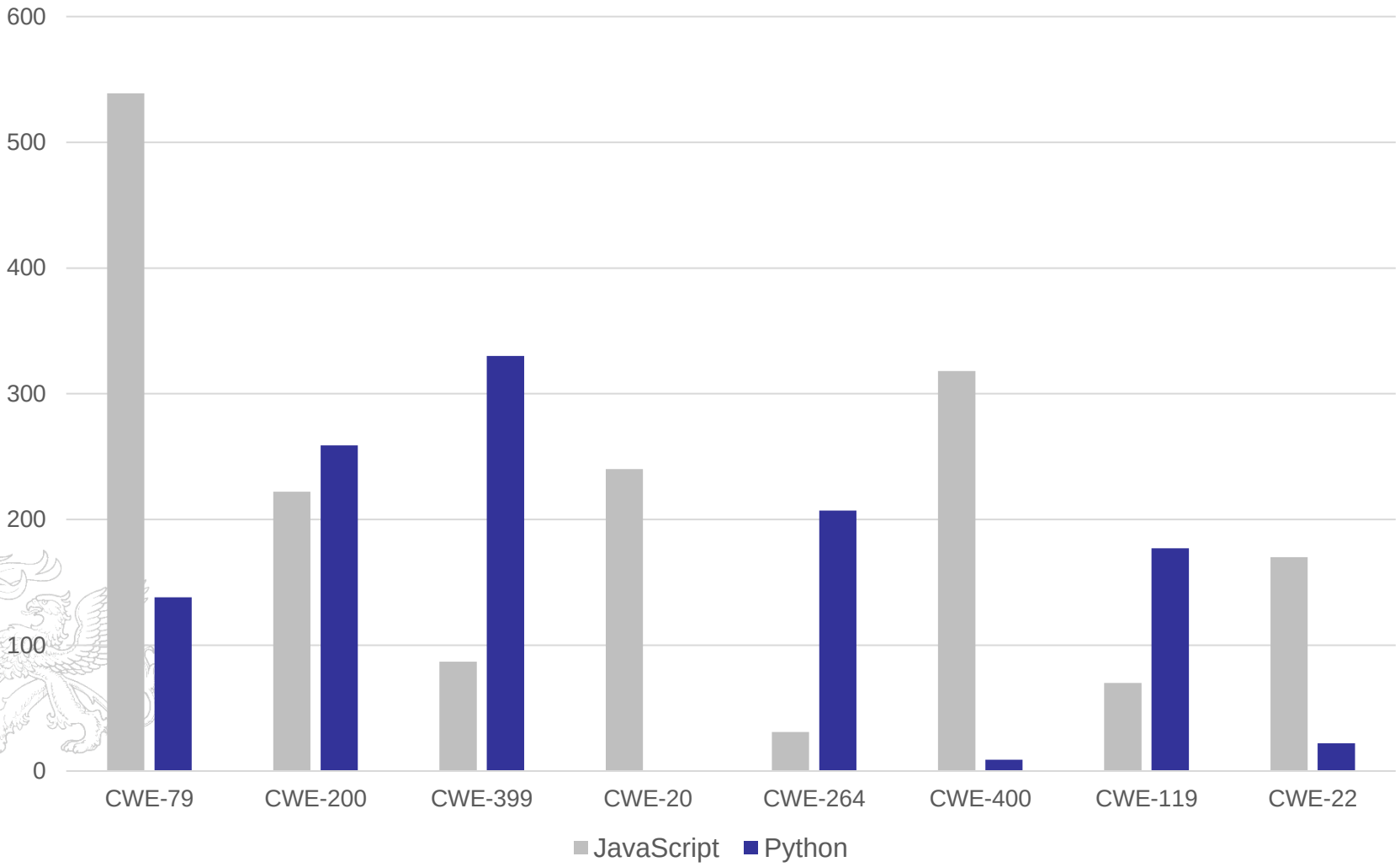
Year	Vuln. JS	Vuln. PY	Total JS	Total PY
2010	0	225	102,525	1,597,160
2011	0	67	675,492	2,068,155
2012	6	343	2,078,887	2,663,836
2013	41	209	5,705,696	3,436,804
2014	84	291	12,692,836	4,440,660
2015	111	328	23,794,463	5,537,294
2016	239	453	38,990,699	6,527,350
2017	393	329	40,883,417	6,835,803
2018	2584	639	37,729,971	6,315,866

(1) Commit statistics per year



(2) Vulnerability mitigations/100k commits

Results – RQ1


600
500
400
300
200
100
0
CWE-79
CWE-200
CWE-399
CWE-20
CWE-264
CWE-400
CWE-119
CWE-22
JavaScript
Python

Results – RQ1

Typical Security Issue Types

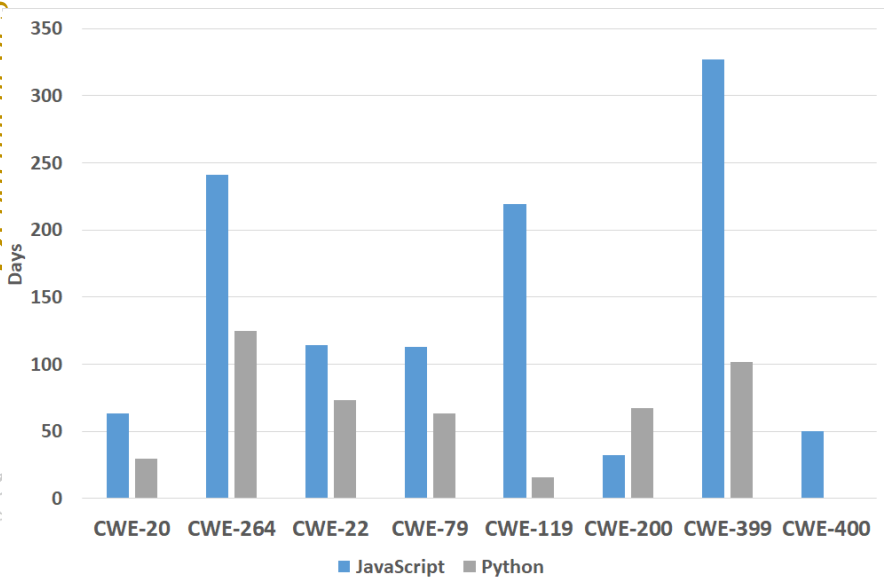
- Python
  - Information Exposure (CWE-200)
  - Resource Management Errors (CWE-399)
  - Permissions, Privileges, and Access Controls (CWE-264)
  - Improper Restriction of Operations within the Bounds of a Memory Buffer (CWE-119)
- JavaScript
  - Information Exposure (CWE-200)
  - Cross-site Scripting (CWE-79)
  - Path Traversal (CWE-22)
  - Improper Input Validation (CWE-20)
  - Uncontrolled Resource Consumption (CWE-400)

Results – RQ2

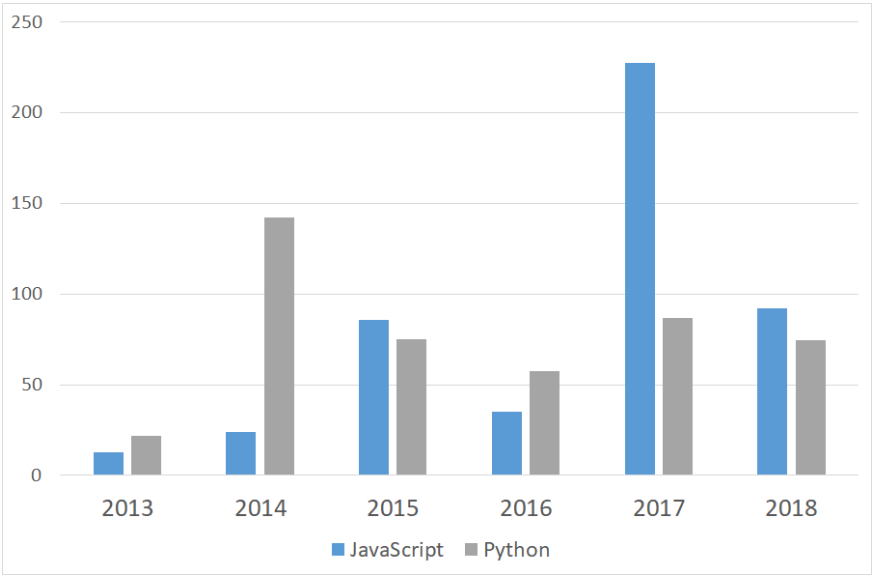
Reaction Times to Security Issues

- None of the communities react fast to vulnerabilities
  - Average: ~100 days
- Eight most prevalent CWE categories
  - Python community
    - Usually react faster than JavaScript communities
    - Average: 50 days or less
  - JavaScript community
    - React fast to information exposure, 32.5 days on average
    - Vulnerabilities falling into the CWE categories characteristic to Python are mitigated after 200 days or more
- Mitigation time is decreasing
  - Security issues are gaining attention

Results – RQ2



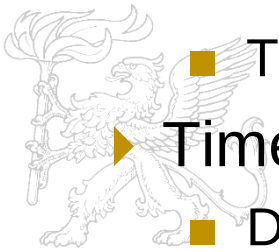
(1) Average number of days between mitigation commit date and CVE publish date grouped by CWE



(2) Average number of days between mitigation commit date and CVE publish date grouped by years

Conclusion

- Security is getting more important
  - More vulnerabilities are being found year by year
  - Luckily, more and more fixes are being released
- The percentage of vulnerability mitigation commits compared to the total number of commits in projects show a growing tendency
- Different languages have different typical vulnerability categories
  - There can be overlaps
- Time needed to fix a vulnerability is decreasing
  - Developers are fixing security issues faster



UNIVERSITAS SCIENTIARUM SZEGEDIENSIS
UNIVERSITY OF SZEGED
Department of Software Engineering

Thank you for your attention!

Péter Hegedűs

hpeter@inf.u-szeged.hu

University of Szeged and

MTA-SZTE Research Group on Artificial Intelligence

Szeged, Hungary