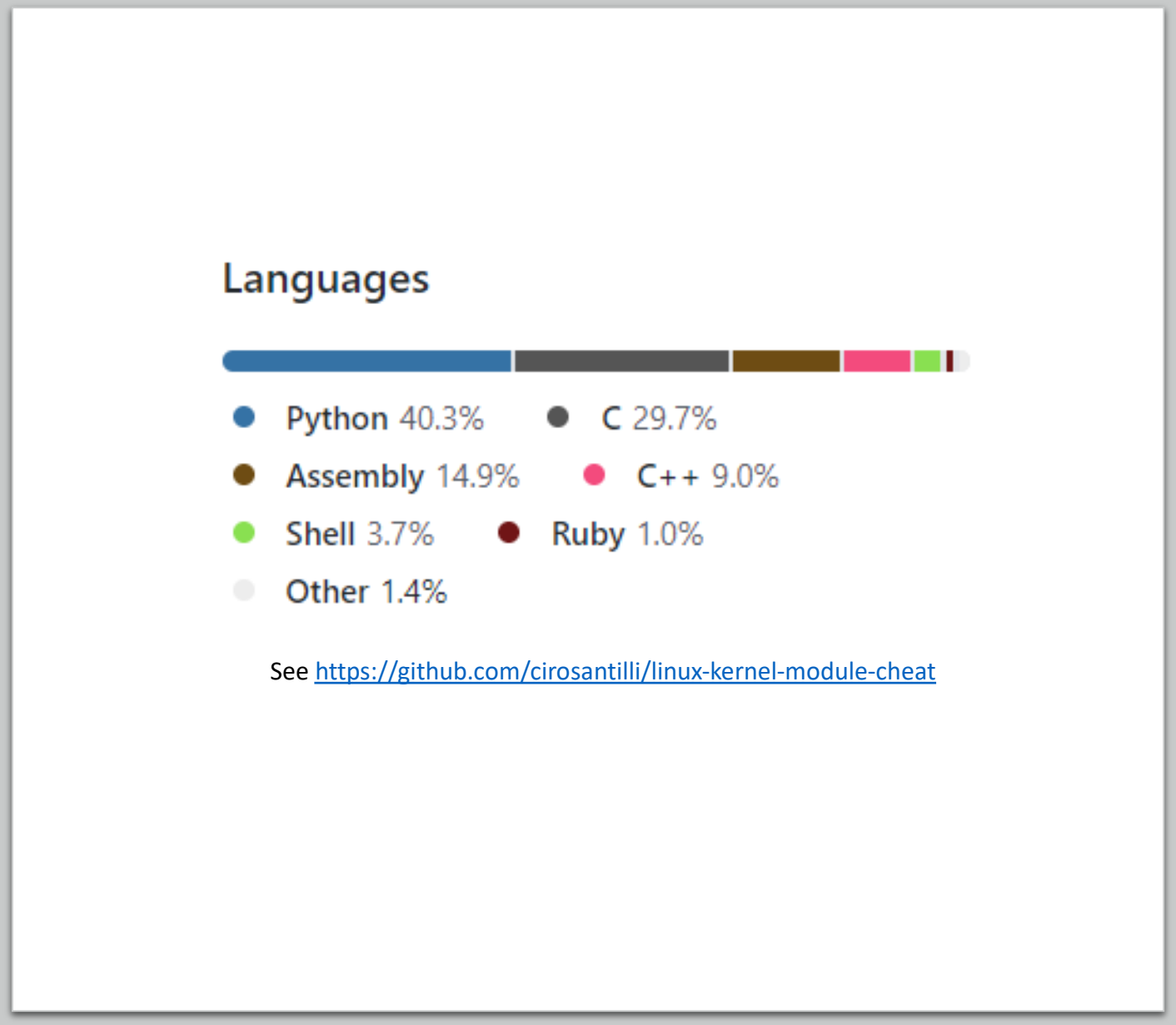
Impact of Language Choice on the Vulnerability of Multi-Language Software

Wen Li
Yashovardhan Sharma

Alissa Cielecki
Arcadia University
Dr. Haipeng Cai

The Problem:

- Majority of modern applications written in **multiple** programming languages
- Lack of information on vulnerabilities introduced by multi-language development
- Software vulnerabilities are a **major** cybersecurity threat

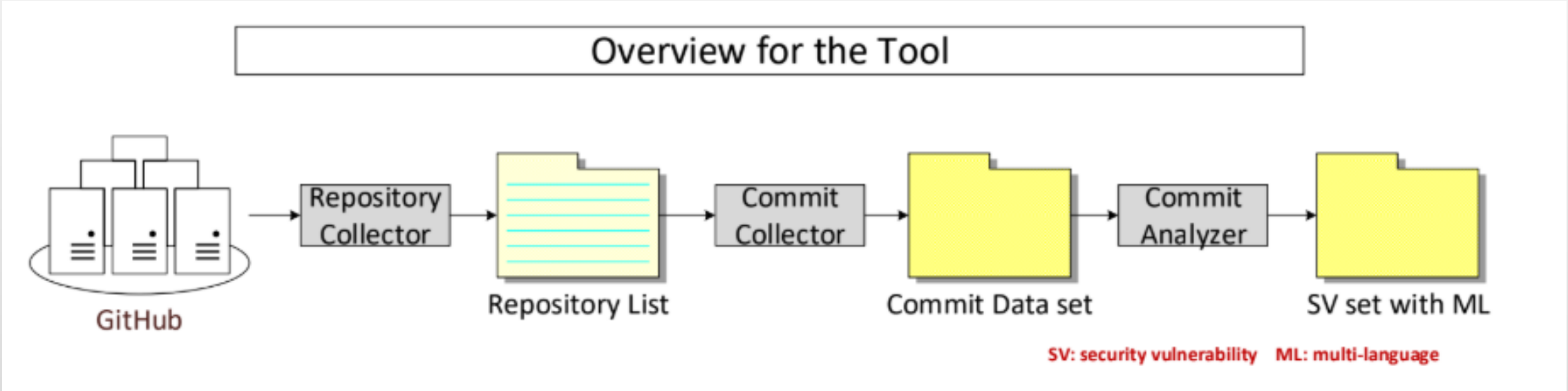


See https://github.com/cirosantilli/linux-kernel-module-cheat

Approach: Garner an Empirical Understanding

- Dataset: top 7,000 multi-language projects from Github
- Study the projects in context of our RQs:
  - RQ1: Are some multi-language selections more vulnerable than others?
  - RQ2: Are some language combinations associated with certain types of vulnerabilities?
  - RQ3: Is the vulnerability of language combinations related to functionality domains?
  - RQ4: What evolutionary perspective can we gain through such analysis?

Study Design



Collect projects and project metadata

Collect all commits, or **versions**, for each project

Analyze commits to find security vulnerabilities

Data Collection

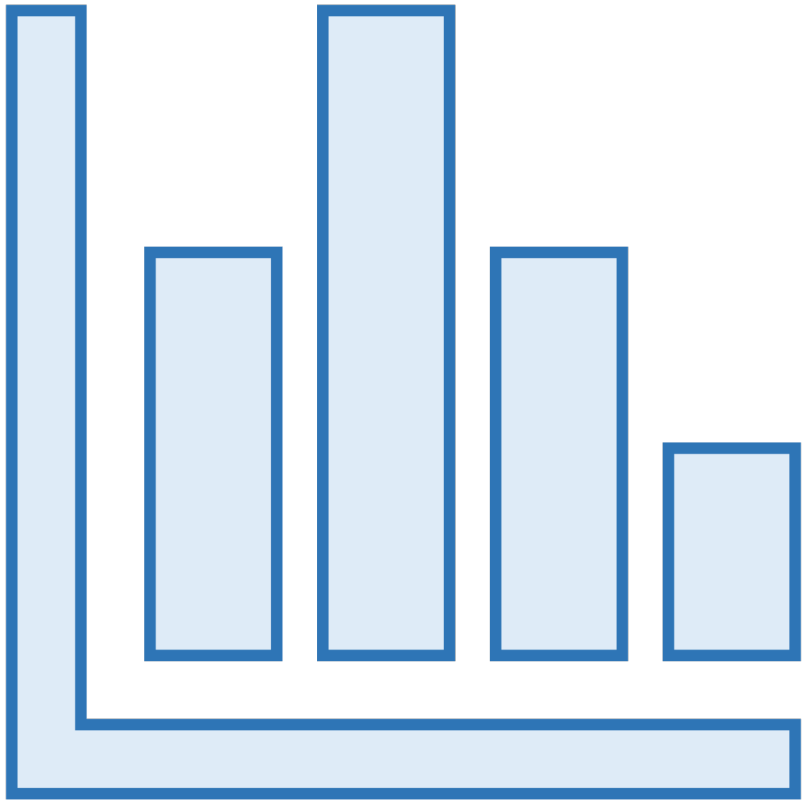
Code for data collection tools hosted at https://github.com/acielecki/WSU-REU/

Repository List

id	size	forks	open_issues	subscribers_count	stargazers_count	language_dictionary	owner_type	pushed_at	topics	description
#	122706	23609	289	8383	308356	{'css': 71539, 'html': 17627, 'javascript': 9927	Organization	2020-01-28T	['careers', 'certificat	the https and and www
#	27961	23497	435	6005	156213	{'css': 11301, 'html': 17839, 'javascript': 2832	Organization	2020-01-28T	['framework', 'front	vue js is a incrementally
#	150292	27430	574	6641	142959	{'c': 5225, 'c++': 44278, 'coffeescript': 16554,	Organization	2020-01-28T	['declarative', 'front	a and flexible javascript
#	152366	67974	433	7190	138322	{'css': 424038, 'html': 242563, 'javascript': 62	Organization	2020-01-28T	['bootstrap', 'css-fra	the most popular and ja
#	6163	18584	812	2585	102566	{'makefile': 59, 'perl': 3811, 'python': 2685, 'r	Organization	2020-01-28T	['productivity', 'term	a delightful community
#	3029	17883	115	3508	92096	{'javascript': 67844}	Organization	2020-01-28T	['arrow-functions', '	javascript style guide
#	41467	21750	5	4032	89726	{'javascript': 2401}	Organization	2020-01-28T	['visualization']	bring data to life with ca
#	2780119	30006	323	6852	85813	{'assembly': 9571217, 'awk': 38903, 'c': 7937	User	2020-01-28T	[]	linux kernel source tree

Commit Dataset

sha	author	date	message	commits	parents
fc9a03138c959f0d24301c6401f800eb6e7d3af9	Phil Hagelberg	2020-02-09T17:17:37Z	Merge pull request #36 from gonewest818/no-litteri...	https://api.github.com/repos/technomancy/better-de...	e051c5278f71af8414f8c6101695197dd4365cca bd4ee4a66...
bd4ee4a66198d725cfa2849675501bc7a39f2546	Neil Okamoto	2020-02-05T18:31:46Z	package-lint: should declare (emacs "25.1") as a r...	https://api.github.com/repos/technomancy/better-de...	45f6f46da18294422bbfe6414f38115532e8e821
45f6f46da18294422bbfe6414f38115532e8e821	Neil Okamoto	2020-02-05T16:20:12Z	avoid overriding `backup-directory-alist` if alrea...	https://api.github.com/repos/technomancy/better-de...	4f2f9227e56bd6ad6ec97d20b9d9745f17cce9ab
4f2f9227e56bd6ad6ec97d20b9d9745f17cce9ab	Neil Okamoto	2020-02-05T16:07:30Z	no need to set `save-place-file` because default i...	https://api.github.com/repos/technomancy/better-de...	e051c5278f71af8414f8c6101695197dd4365cca
e051c5278f71af8414f8c6101695197dd4365cca	Phil Hagelberg	2020-01-01T22:57:56Z	Merge pull request #34 from fabio-porcedda/savepla...	https://api.github.com/repos/technomancy/better-de...	01b8f00235c37bddf3e1a949e943dc0c9980863 c72e7d544...
c72e7d5448cd0d4fbca4af455455f7b2d1335c7f	Fabio Porcedda	2019-12-30T13:12:05Z	Updated Save Place activation to support GNU Emacs...	https://api.github.com/repos/technomancy/better-de...	01b8f00235c37bddf3e1a949e943dc0c9980863



Evaluation & Results

- Study currently in progress
- Target results:
  - To see a correlation between language combinations and vulnerabilities (or lack thereof)
  - Answers to our four RQs
- Evaluation:
  - Quality of information/data acquired
  - Recommendations we can make based on findings

Conclusions

Potential Impact

To scientific community: gaining new knowledge

To society: new knowledge -> new tools to address software security -> more secure digital lives

Next Steps

Expanding the dataset

Refining how we determine vulnerabilities and security fixes

Thank You

Acknowledgement

This material is based upon work supported by the National Science Foundation Research Experiences for Undergraduates Program under Grant No. 1757632.