Visualizing the Vulnerability Landscape of Major Scientific Cyberinfrastructure GitHub Ecosystems

by Ben Lazarine (Indiana U.), Dalya Manatova (Indiana U.), Sagar Samtani (Indiana U.), and Hongyi Zhu (UT San Antonio)

Overview:

- Ensuring the cybersecurity for scientific cyberinfrastructure (CI) is of growing importance.
- Scientific CIs have an increasing quantity of GitHub **repositories with source code to execute scientific procedures.** However, these repositories often possess **thousands of vulnerabilities.**
  - **Challenge:** This scale of vulnerabilities is often difficult to discover, prioritize, and remediate.
- Visualizations can be a promising approach to effectively identifying key **development** trends, **vulnerability** trends, repositories to prioritize, and individuals to offer selected security training.

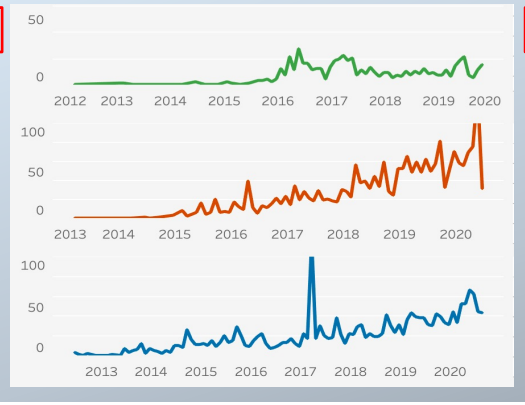
Solution:

We have developed a GitHub Vulnerability dashboard that includes 4,856 repositories and user data collected through the GitHub API. The dashboard that aims to provide CI developers with:

1. An **overview** of their repository trends for three scientific CIs [1a] and a word cloud of **key users** [1b] contributing the most vulnerabilities (anonymized to protect identities).
2. A complete **vulnerability assessment** overview [2] is filterable by multiple levels of granularity from CI to repository types to specific repositories is enabled by the dashboard.
3. A **detailed breakdown** of every scanned repository [3] filterable by both repository type and vulnerability type is provided.

[1] CI Development Trends on GitHub

a



b

[3] Repository Summaries

Repository	Language	Main Vulnerability	Bytes	Commits	Forks	Vulnerabilit..
	None	Insecure Input	230	103	0	3
	Ruby	Secret	18,683	3,170	0	15,019
	HTML	Insecure Input	112	28	0	3,956
	C	Insecure Functions	40,568	6,206	0	6
	JavaScript	Insecure Functions	39,695	1,275	0	6
	Java	Insecure Functions	130	151	0	809
	C	Insecure Input	81,477	13,707	0	6,299
	Python	Insecure Input	5,426	670	0	27
	C	Insecure Input	82,932	14,010	0	6,403
	C++	Insecure Functions	446,338	8,988	0	1,300
	C++	N/A	441,648	8,870	0	0
	C++	N/A	334,060	5,965	0	0
	None	File Permission	16,051	4,965	0	69
	None	Insecure Functions	13,613	840	0	7

[3] The Repository Summary tab allows users to drill down to language, most common vulnerability type, size, number of forks, and number of vulnerabilities for each repository.

[2] Vulnerability Assessment Overview

Category	Vulnerability	Description	Total
Secret	Secret	A potential password/key	1,820,616
	Password	Word password found	806,261
	Weak cryptography	Insufficient cryptographic method	182,632
	Filetype	File that may contain secrets	3,966
Insecure	File permission	File may have dang. Permissions	320,149
	Insecure function	Function can be vulnerable	264,725
	Insecure module	Module can be vulnerable	9,416
	Deprecated library	Library no longer supported	67
	Insecure conn.	Dangerous internet connections	667
Attack	Insecure input	Dangerous handling of user input	2,049,433
	SQL injection	Hardcoded SQL expressions	413
	XML attack	Dangerous XML library	3,028
	XSS vulnerability	Dangerous library usage	969

[2] **Top common vulnerabilities:** secrets, dangerous file permissions, insecure functions and dangerous handling of user input.