CORE INFRASTRUCTURE INITIATIVE | LINUX FOUNDATION

Census Project Summary

A Linux Foundation publication

www.coreinfrastructure.org
www.linuxfoundation.org

The Heartbleed vulnerability in the open source software (OSS) program OpenSSL was a serious one with widespread impact.

In hindsight, experts say Heartbleed could have been detected before it was deployed. Certainly, Heartbleed highlighted that vulnerabilities in some widely used and depended-upon software programs can have serious ramifications.

It also clearly showed that while some software projects have many participants, perform in-depth security analyses, and produce high-quality software with strong security, other projects rely on small teams and limited budget and time to do the tasks necessary to ensure security. Adding more frustration for all is the fact that it's incredibly difficult to quickly and accurately identify which critical projects fall into the second bucket. Even the most savvy security experts and developers struggle with measuring the security of software. Just the process of identifying which projects require additional scrutiny and might be in need of more support and funds is complicated and often overlooked. Case in point is OpenSSL, which had relatively few developers and many bug reports that languished without response for long periods of time prior to Heartbleed. Unfortunately there's no perfect set of metrics to guarantee that software is secure or not.

To solve these problems, The Linux Foundation and industry leaders like Amazon Web Services, Facebook, Google, IBM, Microsoft, and others created the Core Infrastructure Initiative, a multi-million dollar effort to collaboratively identify and fund critical open source projects in need of assistance.

CII is launching an open source program called The Census Project, a new program that analyzes popular open source projects to establish which ones are critical to Internet infrastructure and also most in need of increased support and funding from a security point of view. CII hopes the industry will find The Census Project an efficient barometer for assessing software security.

The Census Project represents CII's current view of the open source ecosystem and which projects are at risk, and therefore strong candidates to receive CII funding. It does not assess the security of the projects themselves. CII members expect The Census Project to accelerate the process by which projects in need receive more resources. We look forward to feedback on the effort in order to improve the census itself and subsequently the software that we all depend on for our privacy and security.

Who Oversees The Census Project

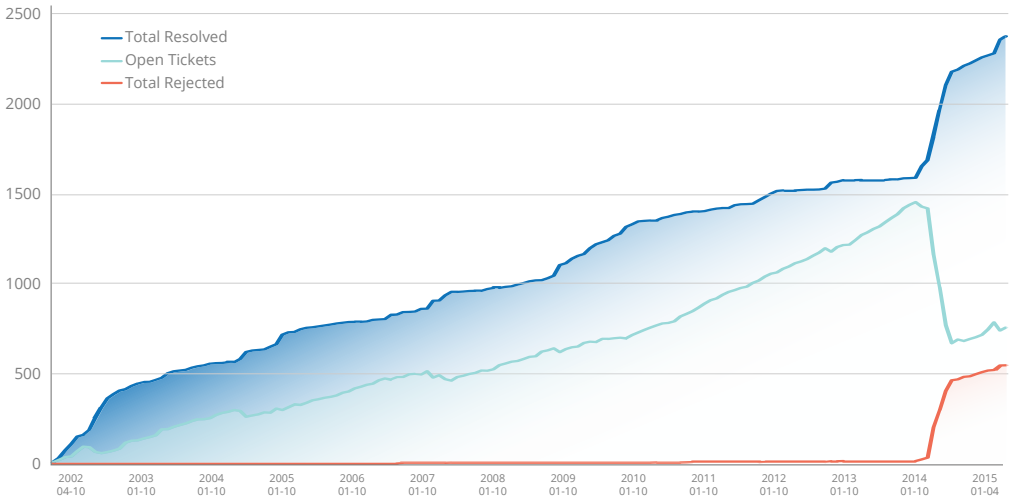
Coordination for The Census Project is overseen by the Institute for Defense Analyses (IDA), a nonprofit organization that operates three federally funded research and development centers and exists to promote national security, preserve the public welfare, and advance scientific learning by analyzing, evaluating, and reporting on matters of interest to the United States Government. IDA's work on The Census Project is funded by CII and by the U.S. Department of Homeland Security Homeland Open Security Technology (DHS HOST) program at Georgia Tech Research Institute (GTRI).

Who Oversees The Census Project

Available on GitHub and CII's website, The Census Project automatically gathers important metrics automatically, with a focus on less active projects. IDA and CII experts estimate a program's exposure to attack using an algorithm to evaluate the data collected, which generates a list of projects that require more scrutiny. The algorithm also considers factors such as recent activity and if a project web site exists, to assign a risk index number ranging from 0-16. Metrics are based on data that can easily and quickly be acquired, such as:

- Common Vulnerabilities and Exposures (CVEs): how many common vulnerabilities and exposures filed and open;
- Contributors: contributor count in the past 12 months;
- Popularity: if the package is in the top 90 percent of most popular Debian packages;
- Network Exposure: if the package is directly exposed to the network (whether client or server), if it is used to process data provided by a network, if it typically runs as root (either via suid or directly), or controls access to such, and therefore is a risk for local privilege escalation.
- Application Data Only: if the Debian database reports that it is "Application Data" or "Standalone Data" rather than an application.

Data is sourced from Black Duck Open Hub, a free online community resource for discovering, evaluating, tracking and comparing open source code and projects, where possible, as well as from the Debian package repository.



A scoring system heuristically combines these automatically gathered metrics with the exposure estimate to narrow in on projects that require more scrutiny to better determine the security and health of the projects. A risk index, based on a weighted point system, assigns a number ranging from 0-16 for each project. It looks at factors such as:

- Popularity: If the package is in the top 90% of most popular Debian packages, it receives a point.
- Main Language: If the project's main language is C or C++, add 2 points.
- Project website: 1 point if there is no identified project websites.
- CVE vulnerability reports: 3 points if 4+ , 2 points for 2–3, 1 point for 1. The CVE count is direct from Debian.
- twelve_month_contributor_count: 5 points for 0 contributors, 4 points for 1–3 contributors, 2 points if the number is unknown (blank).
- Exposure values: 2 points if directly exposed to the network (as a server or client), 1 point if it is often used to process data provided by a network, and 1 point if it could be used for could be used for local privilege escalation.
- Three points are subtracted from the score if the package contains only data.

David A. Wheeler, an open source and security research expert at IDA, CII and others involved in The Census Project also carefully review the results to define a subset of projects. Final results of this cumulative process are available online with the ability to sort projects by risk score, CVE count, contributor count and popularity.

Projects with a higher ranking are especially in need of reinforcements and funding. A high score means that the project may not be getting the attention that it deserves and that it merits further investigation. Using this process, the Risk Index pinpointed several projects CII already funds, including OpenSSL (Risk Score=8), OpenSSH (Risk Score=8), NTP (Risk Score=8), GnuPG (Risk Score=7), which is great validation for the organization's early work. The highest risk score given during the latest Census run was eleven.

The Census Project is an open source, iterative project written in Python by Samir Khakimov at IDA under the MIT license. It will evolve and improve thanks to ongoing community contributions and suggestions. Analyzing crash report data, bug fix data and the use of static analysis tools are a few of the areas where community involvement would be extremely beneficial. Developers and security experts are invited to help improve all facets of The Census Project, from suggesting projects to add and improving data sources to tweaking the metrics and algorithm to make it better for all.

For more information, the IDA white paper titled "Open Source Software Projects Needing Security Investments" summarizes past research and approaches used to calculate risk as well as Wheeler's newest Census Project findings and methodology.

Additional Resources

The Census Project: How To Join

The Census Project GitHub Repository

The Census Project report: “Open Source Software Projects Needing Security Investments.”

The Linux Foundation promotes, protects and standardizes Linux by providing unified resources and services needed for open source to successfully compete with closed platforms.

To learn more about The Linux Foundation or our other initiatives please visit us at www.linuxfoundation.org