Title: Software Composition Analysis Tools: SCRM Value Add or Lossy Noise Machines?

Proposed Length: 30 minutes

Presenter's Name: Robert Erbes, Micaela Gallegos

Affiliation: LLNL and INL

Associated Persons: Robert Erbes (INL), Hannah Kleinheider (INL), KD Villa (LLNL), and Micaela Gallegos (LLNL)

Software supply chain risk management (SCRM) depends upon accurate information regarding the software components that comprise any given software system. The collection of components included in a software package can be organized within a software bill of materials, or SBOM. SBOMs are ideally generated when the software components are put together, such as at compile time, but for many reasons that has not and is not always possible. For example, legacy or proprietary software packages often do not have SBOMs available to downstream consumers of that software. It's not just end users that are affected, manufacturers themselves also must deal with this problem.

To answer these questions, the market has seen the rise of several commercial software composition analysis (SCA) tools. These tools aim to peer into completed software systems, automatically identifying hidden software dependencies and looking up known vulnerabilities associated with those dependencies to enable end-users to enhance their cyber supply chain risk management processes. These tools are potentially a huge boon to end users of legacy and proprietary software – and a potential bane, depending on how accurate they are.

This research asks that question – how accurate are currently available binary SCA tools – and provides answers to several other questions: What does it mean to be "accurate"? What limitations do the tools have in identifying common edge cases that take place in modern software development? Can they help you avoid a devastating supply chain attack, or is it all just noise?

After researching SCA tools on the market, we identified three vendors that fit our use case and would provide analysis on compiled binaries. Using these tools, we submitted firmware for critical infrastructure devices for analysis and SBOM generation. The SBOM outputs were then cross referenced with SBOMs generated through manual analysis for comparison. In addition to the firmware samples, we also submitted edge case samples based off a popular open-source library that were specifically crafted to evaluate each tools' ability to accurately identify components. These samples were customized to be consistent with modifications we have seen in modern software development as well as a couple that are representative of supply chain attacks.

In the end, we found that accuracy in component identification varied across tools and edge case difficulty. To learn more, be sure to attend our talk!