Ke, Jessica - Intern

From: Paul Anderson <paul@grammatech.com>
Sent: Tuesday, June 15, 2021 4:50 PM
To: SBOM_RFC
Subject: Response to SBOM RFC - binary SCA tools

Paul Anderson
VP of Engineering
GrammaTech, Inc.
paul@grammatech.com

I am submitting this comment in response to the Request for Comments on minimum elements for SBOMs.

I urge that any recommendation for minimum elements of an SBOM be considerate of the special properties of SBOMs that are produced by Software Component Analysis (SCA) tools, particularly SCA tools that analyze binary code. The RFC describes the special risks associated with binary software in section 3c.

SCA tools that analyze binary code face unique challenges. Given a binary executable or library, with no knowledge of how that binary was produced, the SCA tool attempts to identify the original source-code components that have been compiled into that binary. This is akin to unbaking a cake and recovering the original ingredients. The compiler and other parts of the toolchain used to generate the binary are irreversible transformers; it is impossible to reliably reconstruct the original source code from the binary. Consequently, the SCA tool must base its conclusions on fragmentary evidence gleaned from the binary or its metadata. State-of-the-art tools use machine learning techniques to sift through pieces of evidence to yield conclusions about the origin. Consequently the conclusions of the SCA tool may be imprecise, and potentially accompanied by a "confidence" measure. For example, a tool may claim "I conclude with 86% confidence that this binary contains code from libXyz versions 1.2-1.4". The tool may also report the evidence used to arrive at that conclusion.

In such a situation, it is meaningless to assign a cryptographic hash to the component. Hashes apply to chunks of information with known boundaries (such as an entire file), but any such boundaries that existed in the original source code are long gone by the time the binary is produced. We recommend that hashes be an optional part of the SBOM.

Similarly, the exact version may not be recoverable. Two contiguous versions of a library may exhibit only very small differences, and so there may be no evidence available to reliably distinguish them. We recommend that version numbers allow for some amount of variation.

Thanks for your consideration, and I wish you the best of luck in this important endeavor.

-Paul

--
Paul Anderson, VP of Engineering, GrammaTech, Inc.
531 Esty St., Ithaca, NY 14850
Tel: +1 607 273-7340 x118; https://www.grammatech.com