'Trojan Source' bug a novel way to attack program encodings

November 3 2021, by Bob Yirka

Credit: CC0 Public Domain

A pair of security experts at TrojanSource have found a novel way to attack computer source code—one that fools a compiler (and human reviewer) into thinking code is safe. Nicholas Boucher and Ross

Anderson, both with the University of Cambridge, have posted a paper on the TrojanSource web page detailing the vulnerability and ways that it might be fixed.

As Boucher and Anderson describe it, the vulnerability involves adversarial attacks being committed by nefarious types using Unicode control characters to reorder characters in source code that appears to programmers to be legitimate. More specifically, the vulnerability involves the use of a 'Bidi' algorithm, in Unicode (an international encoding standard that can be used in different languages) where characters can be placed both left to right and right to left—because some languages, such as Hebrew and Arabic are written and read right to left.

The vulnerability exists because the algorithms that process such code do not take into consideration that some of the characters that are being read left to right, can have a different meaning or purpose if they are read right to left. Because virtually all of the most popular programming languages in use today—C, C+, Java, Python, Go, Rust and JavaScript—allow Unicode, that means that virtually all programs are potentially at risk.

As an example, Boucher and Anderson show that a line of code such as:

/* begin admins only */ if (isAdmin) {

Could be changed to:

/* if (isAdmin) { begin admins only */

The first line is a harmless comment inserted by a programmer, the second is code that could be used to conduct a desired outcome by a hacker. The researchers suggest the vulnerability represents a serious

threat to software supply chains—if such vulnerabilities were exploited, they could impact downstream software by allowing them to inherit the same vulnerability.

Because the vulnerability exists for such a wide variety of programming languages, its disclosure was first coordinated with officials charged with maintaining the rules for such languages giving them time to add changes to compilers and interpreters to account for and mitigate such a threat.

More information: Report: www.trojansource.codes/trojan-source.pdf

TrojanSource: www.trojansource.codes/

Citation: 'Trojan Source' bug a novel way to attack program encodings (2021, November 3) retrieved 19 April 2024 from https://techxplore.com/news/2021-11-trojan-source-bug-encodings.html