Researchers develop automatic security tests for complex systems

March 15 2017

These tests produce millions of valid program inputs within minutes. In this manner the researchers can automatically extract the required information from the program they are examining. They will present further details at the Cebit computer fair in Hannover in Hall 6, Stand C47.

Andreas Zeller, professor of Software Engineering at the Saarland University and CISPA researcher, is working on uncovering security vulnerabilities before they are exploited by cyber-criminals. "Modern test generators can generate inputs for the program in question at a high speed," explains Zeller. "But for that to work it is essential to know how the input is structured, because the program immediately disallows invalid inputs. This is precisely what our researchers are working on, namely deciphering exactly how these program inputs need to be constructed."

By looking at a given program and its range of inputs, Zeller and his doctoral students Matthias Hoeschele and Alexander Kampmann are able to automatically extract a so-called "context-free grammar": This is a description of all valid inputs for one specific program, quite like the German grammar is a description of correct sentences in the German language. The CISPA researchers also named the matching software system they developed for this central approach. The prototype is called "Autogram", for "automatic" and "grammar", and first results were already presented in September 2016, at the Automated Software Engineering conference in Singapore.

"With the grammar that Autogram generates, we can produce millions of valid inputs in minutes, allowing us to test a program more comprehensively," Zeller explains. The sheer amount of inputs considerably reduces the likelihood of overlooking security gaps, according to Zeller.

In order to extract the grammar for one specific program, Autogram observes how the program handles a given input. Different parts of the entry are processed in different parts of the program, which allows the Autogram system to collect the relevant information - data on the structure of valid inputs and their relation to the program code. The extracted grammars themselves are in fact very readable for humans, since they use specific identifiers from the program code. "At present, we are testing our prototype by letting it analyze a wide range of input formats, such as JSON or table data. We use about one thousand valid inputs as a foundation," says Alexander Kampmann. Prospectively, these inputs will be omitted, though, so that in a next step the grammar could be gleaned from the program directly.

Based on the extracted grammar, the researchers can create new test entries that analyze the program systematically. How this can be done efficiently is being further researched in their project "tribble", which is also being presented at Cebit. "Tribble" uses the grammars as provided by Autogram and then systematically compiles all valid input variables and code snippets. The IT security researchers around Zeller already have a wide range of experience with grammar-based testing. In 2012, they presented their test generator LANGFUZZ, which comprehensively analyzed the Firefox web browser, using a hand-made grammar at the time. LANGFUZZ has been in daily use with Firefox developers for four years, and with its help, so far more than 4,000 errors and security gaps have been identified and corrected.

So now the researchers from Saarbruecken are extending their range,

from Firefox to virtually any program and input format. "The long term goal is fully automated security testing, applicable for all - from the smallest Internet of Things gadget to full-grown servers," says Zeller.

More information: www.st.cs.uni-saarland.de/models/autogram/

Provided by Saarland University