This bot hunts software bugs for the Pentagon

reader comments

25 with 21 posters participating

Share this story

Late last year, David Haynes, a security engineer at the Internet infrastructure company Cloudflare, found himself gazing at a strange image. “It was pure gibberish,” he says. “A whole bunch of gray and black pixels, made by a machine.” He declined to share the image, saying it would be a security risk.

Haynes’ caution was understandable. The image was created by a tool called Mayhem that probes software to find unknown security flaws, made by a startup spun out of Carnegie Mellon University called ForAllSecure. Haynes had been testing it on Cloudflare software that resizes images to speed up websites and fed it several sample photos. Mayhem mutated them into glitchy, cursed images that crashed the photo-processing software by triggering an unnoticed bug, a weakness that could have caused headaches for customers paying Cloudflare to keep their websites running smoothly.

Cloudflare has since made Mayhem a standard part of its security tools. The US Air Force, Navy, and Army have used it, too. Last month, the Pentagon awarded ForAllSecure a $45 million contract to widen use of Mayhem across the US military. The department has plenty of bugs to find. A 2018 government report found that nearly all weapons systems the

Department of Defense tested between 2012 and 2017 had serious software vulnerabilities.

Mayhem isn’t sophisticated enough to fully replace the work of human bug finders, who use knowledge of software design, code-reading skills, creativity, and intuition to find flaws. But ForAllSecure co-founder and CEO David Brumley says the tool can help human experts get more done. The world’s software has more security holes than experts have time to find, and more flaws ship every minute. “Security isn’t about being either secure or insecure—it’s about how fast you can move,” says Brumley.

Mayhem originated in an unusual 2016 hacking contest in a Las Vegas casino ballroom. Hundreds of people showed up to watch the Cyber Grand Challenge, hosted by the Pentagon’s research agency DARPA. But there was nary a human on stage, just seven gaudily lit computer servers. Each hosted a bot that tried to find and exploit bugs in the other servers, while also finding and patching its own flaws. After eight hours, Mayhem, made by a team from Brumley’s Carnegie Mellon security lab, won the $2 million top prize. Its magenta-lit server landed in the Smithsonian.

Brumley, who is still a Carnegie Mellon professor, says the experience convinced him that his lab’s creation could be useful in the real world. He put aside the offensive capabilities of his team’s bot, reasoning defense was more important, and set about commercializing it. “The Cyber Grand Challenge showed that fully autonomous security is possible,” he says. “Computers can do a reasonably good job.”

US contract

The governments of China and Israel thought so, too. Both offered contracts, but ForAllSecure signed up with Uncle Sam. It got a contract with the Defense Innovation Unit, a Pentagon group that tries to fast-track new technology into the US military.

ForAllSecure was challenged to prove Mayhem's mettle by looking for flaws in the control software of a commercial passenger plane with a military variant used by US forces. In minutes, the auto-hacker found a vulnerability that was subsequently verified and fixed by the aircraft's manufacturer.

Other bugs found by Mayhem include one discovered earlier this year in the OpenWRT software used in millions of networking devices. Last fall, two interns at the company scored a payout from Netflix's bug-bounty program after they used Mayhem to find a flaw in software that lets people send video from their phone to a TV.

Brumley says interest from automotive and aerospace companies is particularly strong. Cars and planes rely increasingly on software, which needs to function reliably for years and is updated rarely, if at all.

Mayhem works only on programs for Linux-based operating systems and finds bugs in two ways, one scattershot, the other more targeted.

The first is a technique called fuzzing, which involves bombarding the target software with randomly generated input, such as commands or photos, and watching to see if any trigger exploitable crashes. The second, called symbolic execution, involves creating a simplified mathematical representation of the target software. That dumbed-down double can be analyzed to identify potential weak spots in the real target.

Fuzzing has become more widely used in computer security in recent years. Last year, Google released a fuzzing tool it says has found more than 16,000 bugs in its Chrome browser. But Haynes of Cloudflare says the technique is still not commonly used in industry because fuzzing tools usually require too much careful adaptation for each target program. ForAllSecure has crafted Mayhem to be more adaptable, he says, allowing Cloudflare to use fuzzing more routinely. Symbolic execution can find more complex bugs and has previously been used mostly in research labs, Haynes says.

Humans still necessary

Ruoyu Wang, a professor at Arizona State University, hopes Mayhem is just the start of a more automated future for computer security, but he says that will require bug-finding bots to collaborate more with humans.

Mayhem shows that automation can do useful work, Wang says, but existing auto bug finders can’t be much help with complex Internet services or software packages. The best software is nowhere near smart enough to understand the intent and functioning of programs as people do. Mayhem’s ability to try many different things more quickly than any human is no substitute. “Many of the hard problems in automatically finding vulnerabilities are nowhere close to being solved,” says Wang.

Wang was part of a team called Mechanical Phish that placed third in the 2016 DARPA tournament that gave Mayhem its start. He now works on a new research program from the agency called CHESS, trying to make more powerful bug-finding software that taps humans for help with things machines can’t grok. “Right now the state-of-the-art automation doesn’t know when it’s hitting a barrier,” Wang says. “It should realize that and consult a human.” Today Mayhem looks for bugs on its own, but its descendants may be team players.

This article first appeared on wired.com.

https://arstechnica.com/?p=1681690