What is Log4j? A cybersecurity expert explains the latest internet vulnerability, how bad it is and what's at stake

December 23 2021, by Santiago Torres-Arias

Credit: Pixabay/CC0 Public Domain

Log4Shell, an internet vulnerability that affects millions of computers, involves an obscure but nearly ubiquitous piece of software, Log4j. The

software is used to record all manner of activities that go on under the hood in a wide range of computer systems.

Jen Easterly, director of the U.S. Cybersecurity & Infrastructure Security Agency, called Log4Shell the most serious vulnerability she's seen in her career. There have already been hundreds of thousands, perhaps millions, of attempts to exploit the vulnerability.

So what is this humble piece of internet infrastructure, how can hackers exploit it and what kind of mayhem could ensue?

What does Log4j do?

Log4j records events—errors and routine system operations—and communicates diagnostic messages about them to system administrators and users. It's open-source software provided by the Apache Software Foundation.

A common example of Log4j at work is when you type in or click on a bad web link and get a 404 error message. The web server running the domain of the web link you tried to get to tells you that there's no such webpage. It also records that event in a log for the server's system administrators using Log4j.

Similar diagnostic messages are used throughout software applications. For example, in the online game Minecraft, Log4j is used by the server to log activity like total memory used and user commands typed into the console.

How does Log4Shell work?

Log4Shell works by abusing a feature in Log4j that allows users to

specify custom code for formatting a log message. This feature allows Log4j to, for example, log not only the username associated with each attempt to log in to the server but also the person's real name, if a separate server holds a directory linking user names and real names. To do so, the Log4j server has to communicate with the server holding the real names.

Unfortunately, this kind of code can be used for more than just formatting log messages. Log4j allows third-party servers to submit software code that can perform all kinds of actions on the targeted computer. This opens the door for nefarious activities such as stealing sensitive information, taking control of the targeted system and slipping malicious content to other users communicating with the affected server.

It is relatively simple to exploit Log4Shell. I was able to reproduce the problem in my copy of Ghidra, a reverse-engineering framework for security researchers, in just a couple of minutes. There is a very low bar for using this exploit, which means a wider range of people with malicious intent can use it.

Log4j is everywhere

One of the major concerns about Log4Shell is Log4j's position in the software ecosystem. Logging is a fundamental feature of most software, which makes Log4j very widespread. In addition to popular games like Minecraft, it's used in cloud services like Apple iCloud and Amazon Web Services, as well as a wide range of programs from software development tools to security tools.

This means hackers have a large menu of targets to choose from: home users, service providers, source code developers and even security researchers. So while big companies like Amazon can quickly patch their web services to prevent hackers from exploiting them, there are many

more organizations that will take longer to patch their systems, and some that might not even know they need to.

The damage that can be done

Hackers are scanning through the internet to find vulnerable servers and setting up machines that can deliver malicious payloads. To carry out an attack, they query services (for example, web servers) and try to trigger a log message (for example, a 404 error). The query includes maliciously crafted text, which Log4j processes as instructions.

These instructions can create a reverse shell, which allows the attacking server to remotely control the targeted server, or they can make the target server part of a botnet. Botnets use multiple hijacked computers to carry out coordinated actions on behalf of the hackers.

A large number of hackers are already trying to abuse Log4Shell. These range from ransomware gangs locking down minecraft servers to hacker groups trying to mine bitcoin and hackers associated with China and North Korea trying to gain access to sensitive information from their geopolitical rivals. The Belgian ministry of defense reported that its computers were being attacked using Log4Shell.

Although the vulnerability first came to widespread attention on Dec. 10, 2021, people are still identifying new ways to cause harm through this mechanism.

Stopping the bleeding

It is hard to know whether Log4j is being used in any given software system because it is often bundled as part of other software. This requires system administrators to inventory their software to identify its presence. If some people don't even know they have a problem, it's that

much harder to eradicate the vulnerability.

Another consequence of Log4j's diverse uses is there is no one-size-fits-all solution to patching it. Depending on how Log4j was incorporated in a given system, the fix will require different approaches. It could require a wholesale system update, as done for some Cisco routers, or updating to a new version of software, as done in Minecraft, or removing the vulnerable code manually for those who can't update the software.

Log4Shell is part of the software supply chain. Like physical objects people purchase, software travels through different organizations and software packages before it ends up in a final product. When something goes wrong, rather than going through a recall process, software is generally "patched," meaning fixed in place.

However, given that Log4j is present in various ways in software products, propagating a fix requires coordination from Log4j developers, developers of software that use Log4j, software distributors, system operators and users. Usually, this introduces a delay between the fix being available in Log4j code and people's computers actually closing the door on the vulnerability.

Some estimates for time-to-repair in software generally range from weeks to months. However, if past behavior is indicative of future performance, it is likely the Log4j vulnerability will crop up for years to come.

As a user, you are probably wondering what can you do about all this. Unfortunately, it is hard to know whether a software product you are using includes Log4j and whether it is using vulnerable versions of the software. However, you can help by heeding the common refrain from computer security experts: Make sure all of your software is up to date.