research!rsc Thoughts and links about programming, by Russ Cox

Timeline of the xz open source attack

Posted on Monday, April 1, 2024.
Updated Wednesday, April 3, 2024.

Over a period of over two years, an attacker using the name "Jia Tan" worked as a diligent, effective contributor to the xz compression library, eventually being granted commit access and maintainership. Using that access, they installed a very subtle, carefully hidden backdoor into liblzma, a part of xz that also happens to be a dependency of OpenSSH sshd on Debian, Ubuntu, and Fedora, and other systemd-based Linux systems that patched sshd to link libsystemd. (Note that this does not include systems like Arch Linux, Gentoo, and NixOS, which do not patch sshd.) That backdoor watches for the attacker sending hidden commands at the start of an SSH session, giving the attacker the ability to run an arbitrary command on the target system without logging in: unauthenticated, targeted remote code execution.

The attack was publicly disclosed on March 29, 2024 and appears to be the first serious known supply chain attack on widely used open source software. It marks a watershed moment in open source supply chain security, for better or worse.

This post is a detailed timeline that I have constructed of the social engineering aspect of the attack, which appears to date back to late 2021. (See also my analysis of the attack script.)

Corrections or additions welcome on Bluesky, Mastodon, or email.

Prologue

2005–2008: Lasse Collin, with help from others, designs the .xz file format using the LZMA compression algorithm, which compresses files to about 70% of what gzip did [1]. Over time this format becomes widely used for compressing tar files, Linux kernel images, and many other uses.

Jia Tan arrives on scene, with supporting cast

2021-10-29: Jia Tan sends first, innocuous patch to the xz-devel mailing list, adding ".editorconfig" file.

2021-11-29: Jia Tan sends second innocuous patch to the xz-devel mailing list, fixing an apparent reproducible build problem. More patches that seem (even in retrospect) to be fine follow.

2022-02-07: Lasse Collin merges first commit with "jiat0218@gmail.com" as author in git metadata ("liblzma: Add NULL checks to LZMA and LZMA2 properties encoders").

2022-04-19: Jia Tan sends yet another innocuous patch to the xz-devel mailing list.

2022-04-22: "Jigar Kumar" sends first of a few emails complaining about Jia Tan's patch not landing. ("Patches spend years on this mailing list. There is no reason to think anything is coming soon.") At this point, Lasse Collin has already landed four of Jia Tan's patches, marked by "Thanks to Jia Tan" in the commit message.

2022-05-19: "Dennis Ens" sends mail to xz-devel asking if XZ for Java is maintained.

2022-05-19: Lasse Collin replies apologizing for slowness and adds "Jia Tan has helped me off-list with XZ Utils and he might have a bigger role in the future at least with XZ Utils. It's clear that my resources are too limited (thus the many emails waiting for replies) so something has to change in the long term."

2022-05-27: Jigar Kumar sends pressure email to patch thread. "Over 1 month and no closer to being merged. Not a surprise."

2022-06-07: Jigar Kumar sends pressure email to Java thread. "Progress will not happen until there is new maintainer. XZ for C has sparse commit log too. Dennis you are better off waiting until new maintainer happens or fork yourself. Submitting patches here has no purpose these days. The current maintainer lost interest or doesn't care to maintain anymore. It is sad to see for a repo like this."

2022-06-08: Lasse Collin pushes back. "I haven't lost interest but my ability to care has been fairly limited mostly due to longterm mental health issues but also due to some other things. Recently I've worked off-list a bit with Jia Tan on XZ Utils and perhaps he will have a bigger role in the future, we'll see. It's also good to keep in mind that this is an unpaid hobby project."

2022-06-10: Lasse Collin merges first commit with "Jia Tan" as author in git metadata ("Tests: Created tests for hardware functions"). Note also that there was one earlier commit on 2022-02-07 that had the full name set only to jiat75.

2022-06-14: Lasse Collin merges only commit with "jiat75@gmail.com" as author. This could have been a temporary git misconfiguration on Jia Tan's side forgetting their fake email address.

2022-06-14: Jugar Kumar sends pressure email. "With your current rate, I very doubt to see 5.4.0 release this year. The only progress since april has been small changes to test code. You ignore the many patches bit rotting away on this mailing list. Right now you choke your repo. Why wait until 5.4.0 to change maintainer? Why delay what your repo needs?"

2022-06-21: Dennis Ens sends pressure email. "I am sorry about your mental health issues, but its important to be aware of your own limits. I get that this is a hobby project for all contributors, but the community desires more. Why not pass on maintainership for XZ for C so you can give XZ for Java more attention? Or pass on XZ for Java to someone else to focus on XZ for C? Trying to maintain both means that neither are maintained well."

2022-06-22: Jigar Kumar sends pressure email to C patch thread. "Is there any progress on this? Jia I see you have recent commits. Why can't you commit this yourself?"

2022-06-29: Lasse Collin replies: "As I have hinted in earlier emails, Jia Tan may have a bigger role in the project in the future. He has been helping a lot off-list and is practically a co-maintainer already. :-) I know that not much has happened in the git repository yet but things happen in small steps. In any case some change in maintainership is already in progress at least for XZ Utils."

Jia Tan becomes maintainer

At this point Lasse seems to have started working even more closely with Jia Tan. Brian Krebs observes that many of these email addresses never appeared elsewhere on the internet, even in data breaches (nor again in xz-devel). It seems likely that they were fakes created to push Lasse to give Jia more control. It worked. Over the next few months, Jia started replying to threads on xz-devel authoritatively about the upcoming 5.4.0 release.

2022-09-27: Jia Tan gives release summary for 5.4.0. ("The 5.4.0 release that will contain the multi threaded decoder is planned for December. The list of open issues related to 5..4.0 [sic] in general that I am tracking are...")

2022-10-28: Jia Tan added to the Tukaani organization on GitHub. Being an organization member does not imply any special access, but it is a necessary step before granting maintainer access.

2022-11-30: Lasse Collin changes bug report email from his personal address to an alias that goes to him and Jia Tan, notes in README that "the project maintainers Lasse Collin and Jia Tan can be reached via xz@tukaani.org".

2022-12-30: Jia Tan merges a batch of commits directly into the xz repo ("CMake: Update .gitignore for CMake artifacts from in source build"). At this point we know they have commit access. Interestingly, a few commits later in the same batch is the only commit with a different full name: "Jia Cheong Tan".

2023-01-11: Lasse Collin tags and builds his final release, v5.4.1.

2023-03-18: Jia Tan tags and builds their first release, v5.4.2.

2023-03-20: Jia Tan updates Google oss-fuzz configuration to send bugs to them.

2023-06-22: Hans Jansen sends a pair of patches, merged by Lasse Collin, that use the "GNU indirect function" feature to select a fast CRC function at startup time. The final commit is reworked by Lasse Collin and merged by Jia Tan. This change is important because it provides a hook by which the backdoor code can modify the global function tables before they are remapped read-only. While this change could be an innocent performance optimization by itself, Hans Jansen returns in 2024 to promote the backdoored xz and otherwise does not exist on the internet.

2023-07-07: Jia Tan disables ifunc support during oss-fuzz builds, claiming ifunc is incompatible with address sanitizer. This may well be innocuous on its own, although it is also more groundwork for using ifunc later.

2024-01-19: Jia Tan moves web site to GitHub pages, giving them control over the XZ Utils web page. Lasse Collin presumably created the DNS records for the xz.tukaani.org subdomain that points to GitHub pages. After the attack was discovered, Lasse Collin deleted this DNS record to move back to tukaani.org, which he controls.

Attack begins

2024-02-23: Jia Tan merges hidden backdoor binary code well hidden inside some binary test input files. The README already said (from long before Jia Tan showed up) "This directory contains bunch of files to test handling of .xz, .lzma (LZMA_Alone), and .lz (lzip) files in decoder implementations. Many of the files have been created by hand with a hex editor, thus there is no better "source code" than the files themselves." Having these kinds of test files is very common for this kind of library. Jia Tan took advantage of this to add a few files that wouldn't be carefully reviewed.

2024-02-24: Jia Tan tags and builds v5.6.0 and publishes an xz-5.6.0.tar.gz distribution with an extra, malicious build-to-host.m4 that adds the backdoor when building a deb/rpm package. This m4 file is not present in the source repository, but many other legitimate ones are added during package as well, so it's not suspicious by itself. But the script has been modified from the usual copy to add the backdoor. See my xz attack shell script walkthrough post for more.

2024-02-24: Gentoo starts seeing crashes in 5.6.0. This seems to be an actual ifunc bug, rather than a bug in the hidden backdoor, since this is the first xz with Hans Jansen's ifunc changes, and Gentoo does not patch sshd to use libsystemd, so it doesn't have the backdoor.

2024-02-26: Debian adds xz-utils 5.6.0-0.1 to unstable.

2024-02-27: Jia Tan starts emailing Richard W.M. Jones to update Fedora 40 (privately confirmed by Rich Jones).

2024-02-28: Debian adds xz-utils 5.6.0-0.2 to unstable.

2024-02-28: Jia Tan breaks landlock detection in configure script by adding a subtle typo in the C program used to check for landlock support. The configure script tries to build and run the C program to check for landlock support, but since the C program has a syntax error, it will never build and run, and the script will always decide there is no landlock support. Lasse Collin is listed as the committer; he may have missed the subtle typo, or the author may be forged. Probably the former, since Jia Tan did not bother to forge committer on his many other changes. This patch seems to be setting up for something besides the sshd change, since landlock support is part of the xz command and not liblzma. Exactly what is unclear.

2024-02-29: On GitHub, @teknoraver sends pull request to stop linking liblzma into libsystemd. It appears that this would have defeated the attack. Kevin Beaumont speculates that knowing this was on the way may have accelerated the attacker's schedule. @teknoraver commented on HN that the liblzma PR was one in a series of dependency slimming changes for libsystemd; there were two mentions of it in late January.

2024-03-04: RedHat distributions start seeing Valgrind errors in liblzma's `_get_cpuid` (the entry to the backdoor). The race is on to fix this before the Linux distributions dig too deeply.

2024-03-05: The libsystemd PR is merged to remove liblzma. Another race is on, to get liblzma backdoor'ed before the distros break the approach entirely.

2024-03-05: Debian adds xz-utils 5.6.0-0.2 to testing.

2024-03-05: Jia Tan commits two ifunc bug fixes. These seem to be real fixes for the actual ifunc bug. One commit links to the Gentoo bug and also typos an upstream GCC bug.

2024-03-08: Jia Tan commits purported Valgrind fix. This is a misdirection, but an effective one.

2024-03-09: Jia Tan commits updated backdoor files. This is the actual Valgrind fix, changing the two test files containing the attack code.

"The original files were generated with random local to my machine. To better reproduce these files in the future, a constant seed was used to recreate these files."

2024-03-09: Jia Tan tags and build v5.6.1 and publishes xz 5.6.1 distribution, containing a new backdoor. To date I have not seen any analysis of how the old and new backdoors differ.

2024-03-20: Lasse Collin sends LKML a patch set replacing his personal email with both himself and Jia Tan as maintainers of the xz compression code in the kernel. There is no indication that Lasse Collin was acting nefariously here, just cleaning up references to himself as sole maintainer. Of course, Jia Tan may have prompted this, and being able to send xz patches to the Linux kernel would have been a nice point of leverage for Jia Tan's future work. We're not at trusting trust levels yet, but it would be one step closer.

2024-03-25: Hans Jansen is back (!), filing a Debian bug to get xz-utils updated to 5.6.1. Like in the 2022 pressure campaign, more name###@mailhost addresses that don't otherwise exist on the internet show up to advocate for it.

2024-03-27: Debian updates to 5.6.1.

2024-03-28: Jia Tan files an Ubuntu bug to get xz-utils updated to 5.6.1 from Debian.

Attack detected

2024-03-28: Andres Freund discovers bug, privately notifies Debian and distros@openwall. RedHat assigns CVE-2024-3094.

2024-03-28: Debian rolls back 5.6.1, introducing 5.6.1+really5.4.5-1.

2024-03-28: Arch Linux changes 5.6.1 to build from Git.

2024-03-29: Andres Freund posts backdoor warning to public oss-security@openwall list, saying he found it "over the last weeks".

2024-03-29: RedHat announces that the backdoored xz shipped in Fedora Rawhide and Fedora Linux 40 beta.

2024-03-30: Debian shuts down builds to rebuild their build machines using Debian stable (in case the malware xz escaped their sandbox?).

2024-03-30: Haiku OS moves to GitHub source repo snapshots.

Further Reading

- Evan Boehs, Everything I know about the XZ backdoor (2024-03-29).
- Filippo Valsorda, Bluesky re backdoor operation (2024-03-30).
- Michał Zalewski, Techies vs spies: the xz backdoor debate (2024-03-30).
- Michał Zalewski, OSS backdoors: the folly of the easy fix (2024-03-31).
- Connor Tumbleson, Watching xz unfold from afar (2024-03-31).
- nugxperience, Twitter re awk and rc4 (2024-03-29)
- birchb0y, Twitter re time of day of commit vs level of evil (2024-03-29)
- Dan Feidt, 'xz utils' Software Backdoor Uncovered in Years-Long Hacking Plot (2024-03-30)
- smx-smz, [WIP] XZ Backdoor Analysis and symbol mapping
- Dan Goodin, What we know about the xz Utils backdoor that almost infected the world (2024-04-01)
- Akamai Security Intelligence Group, XZ Utils Backdoor — Everything You Need to Know, and What You Can Do (2024-04-01)
- Kevin Beaumont, Inside the failed attempt to backdoor SSH globally — that got caught by chance (2024-03-31)
- amlweems, xzbot: notes, honeypot, and exploit demo for the xz backdoor (2024-04-01)
- Rhea Karty and Simon Henniger, XZ Backdoor: Times, damned times, and scams (2024-03-30)
- Andy Greenberg and Matt Burgess, The Mystery of 'Jia Tan,' the XZ Backdoor Mastermind (2024-04-03)
- Risky Business #743 -- A chat about the xz backdoor with the guy who found it (2024-04-03)