sonatype

Attacks from Wonderland

XZ, Model Poisoning, and Reflections on Trust

Dr. Stephen Magill
VP, Product Innovation

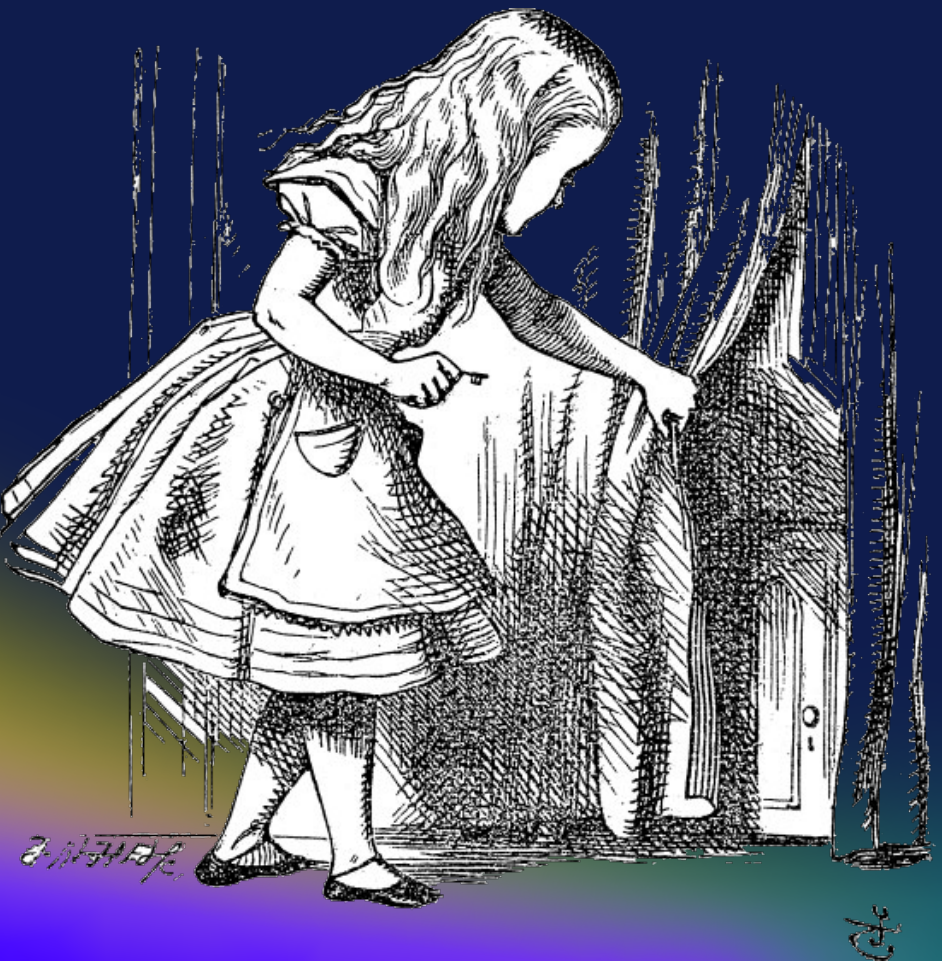
The XZ Project

- Contains the liblzma compression library
- This compression algorithm is used in .xz files
- Used for Ubuntu, Debian, and Fedora packages
- (Indirect) dependency of OpenSSH

- Subject to an attack that could have compromised most IT infrastructure

sonatype

March 29, 2024: Andres Freund discovers the attack



AndresFreundTec

@AndresFreundTec@mastodon.social

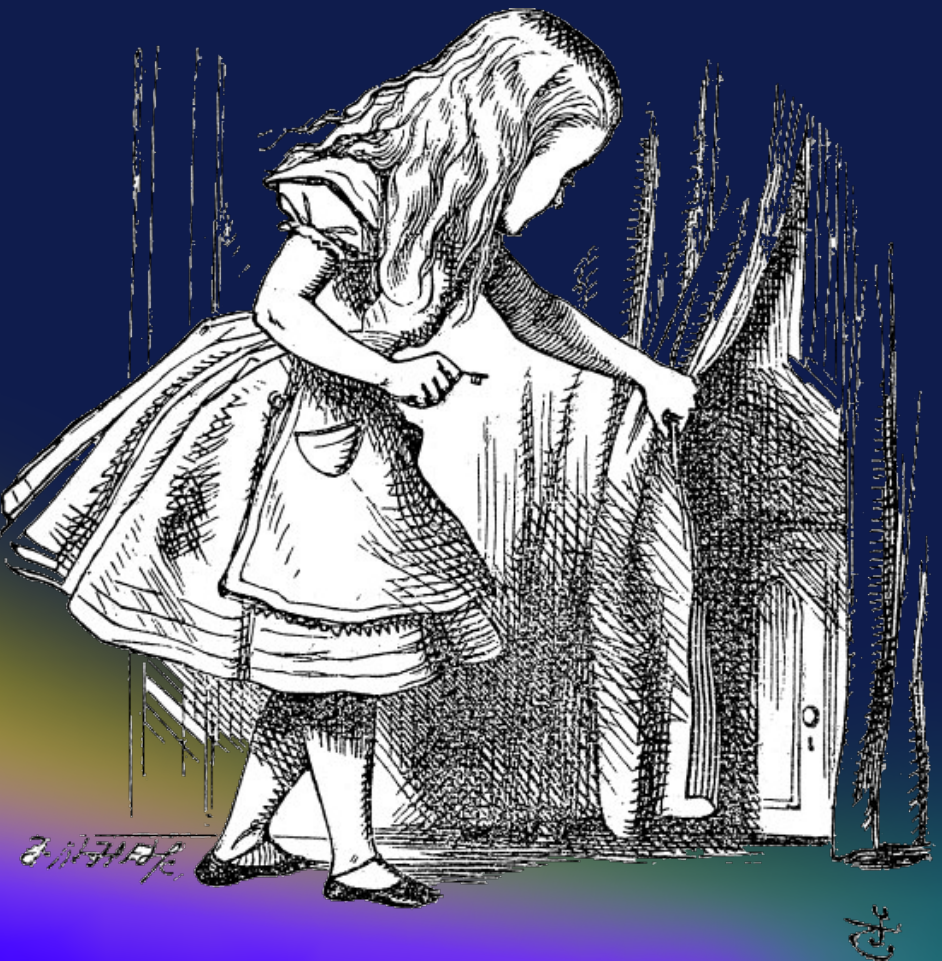
I was doing some micro-benchmarking at the time, needed to quiesce the system to reduce noise. Saw sshd processes were using a surprising amount of CPU, despite immediately failing because of wrong usernames etc. Profiled sshd, showing lots of cpu time in liblzma, with perf unable to attribute it to a symbol. Got suspicious. Recalled that I had seen an odd valgrind complaint in automated testing of postgres, a few weeks earlier, after package updates.

Really required a lot of coincidences.

Mar 29, 2024, 18:32 818 retoots

sonatype

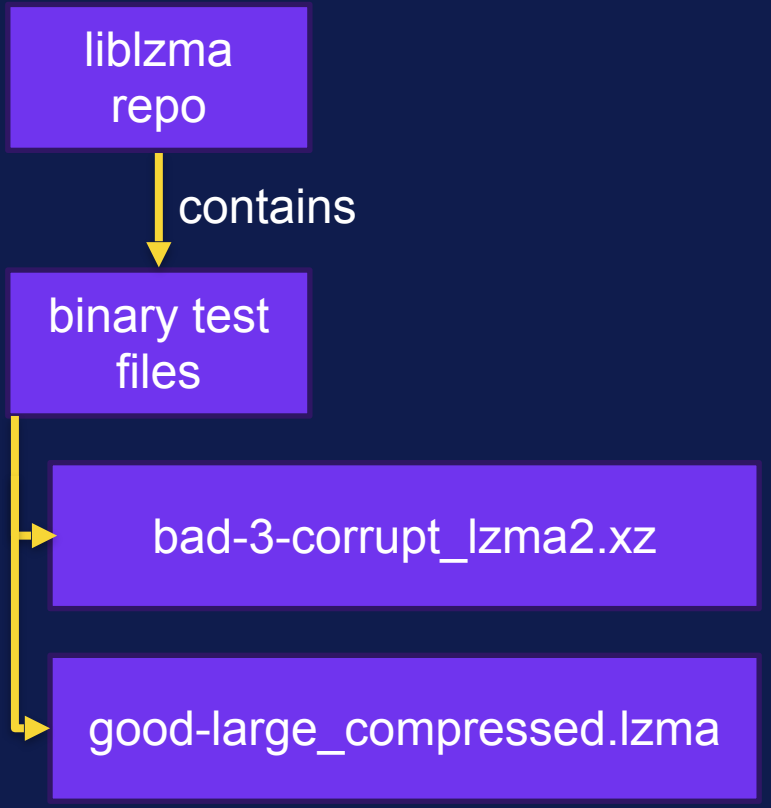
March 29, 2024: Andres Freund discovers the attack



- Happens to be running Debian Testing
- Doing performance analysis, so notices even small amounts of CPU load
- Remembers a random valgrind complaint
- Digs into the issue rather than just killing OpenSSH or adjusting benchmarking approach
- Knows enough to follow the code all the way to the exploit

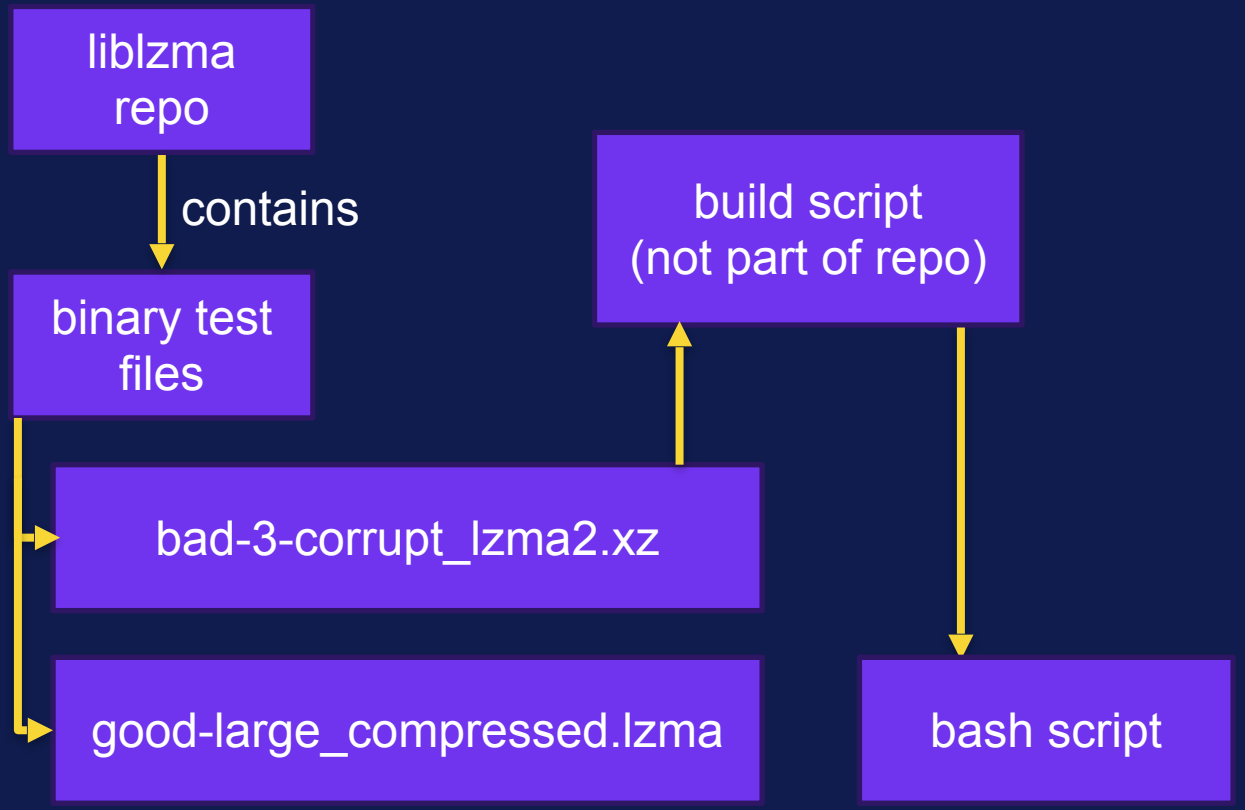
sonatype

The Attack



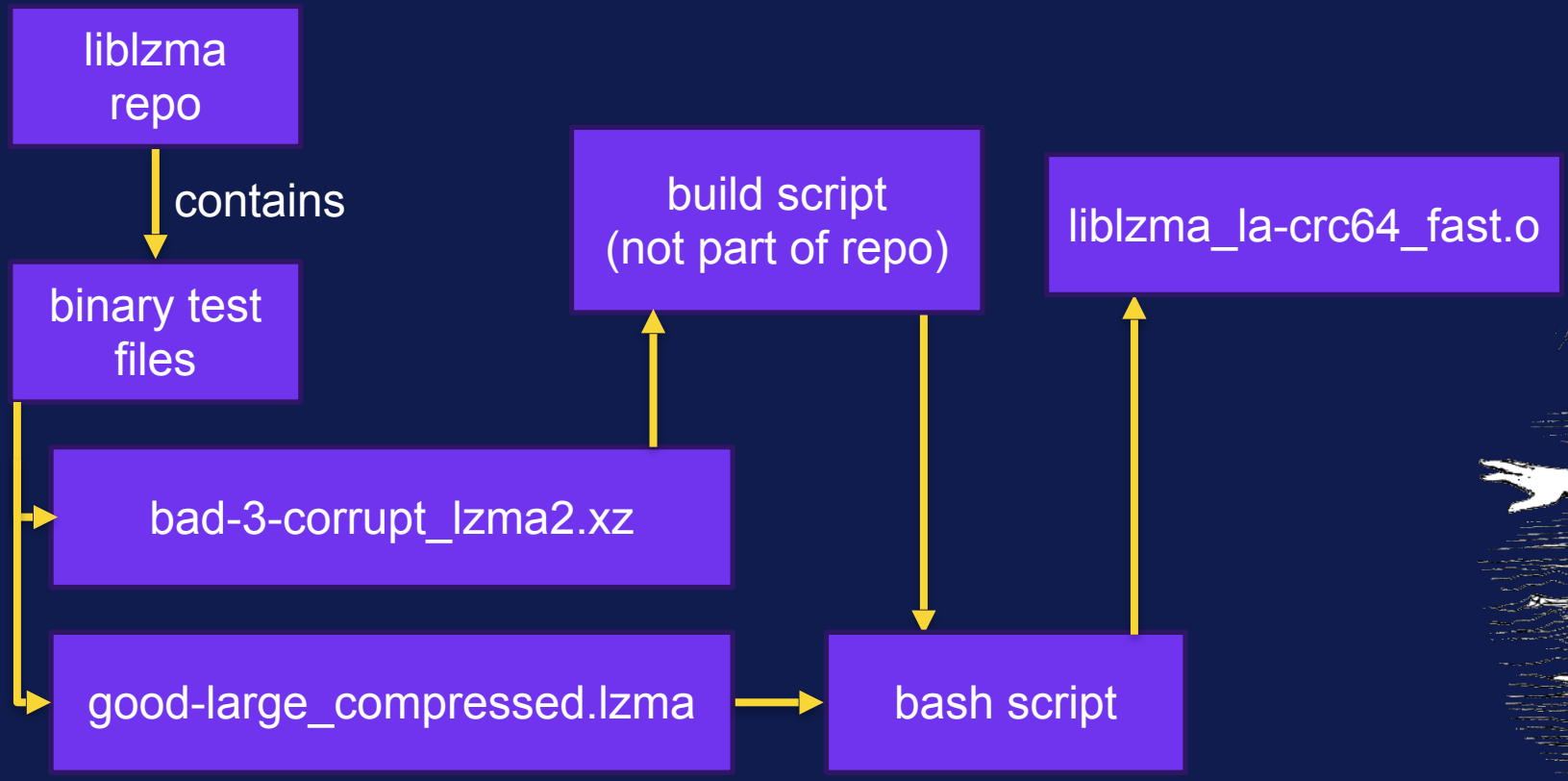
sonatype

The Attack



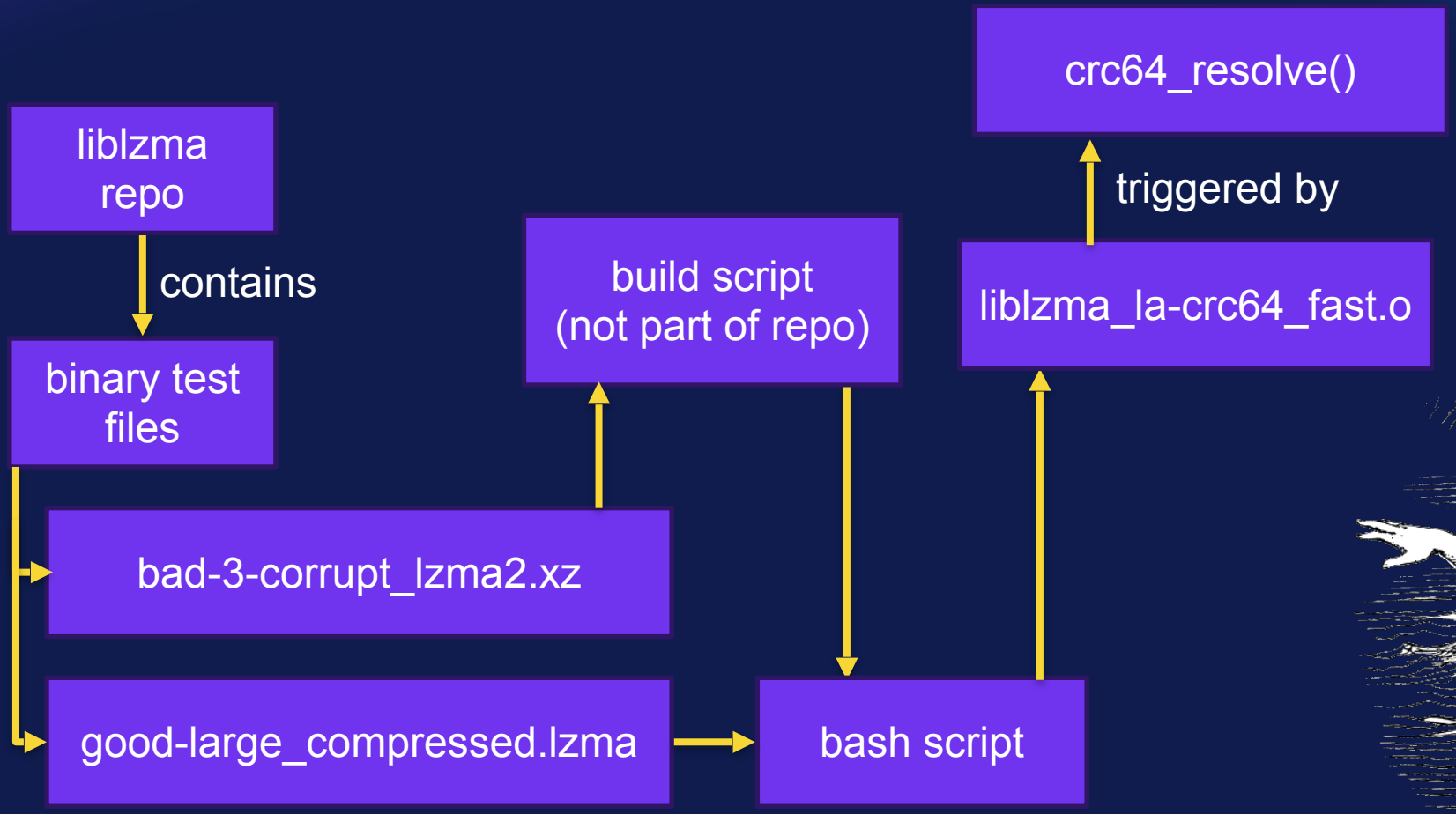
sonatype

The Attack



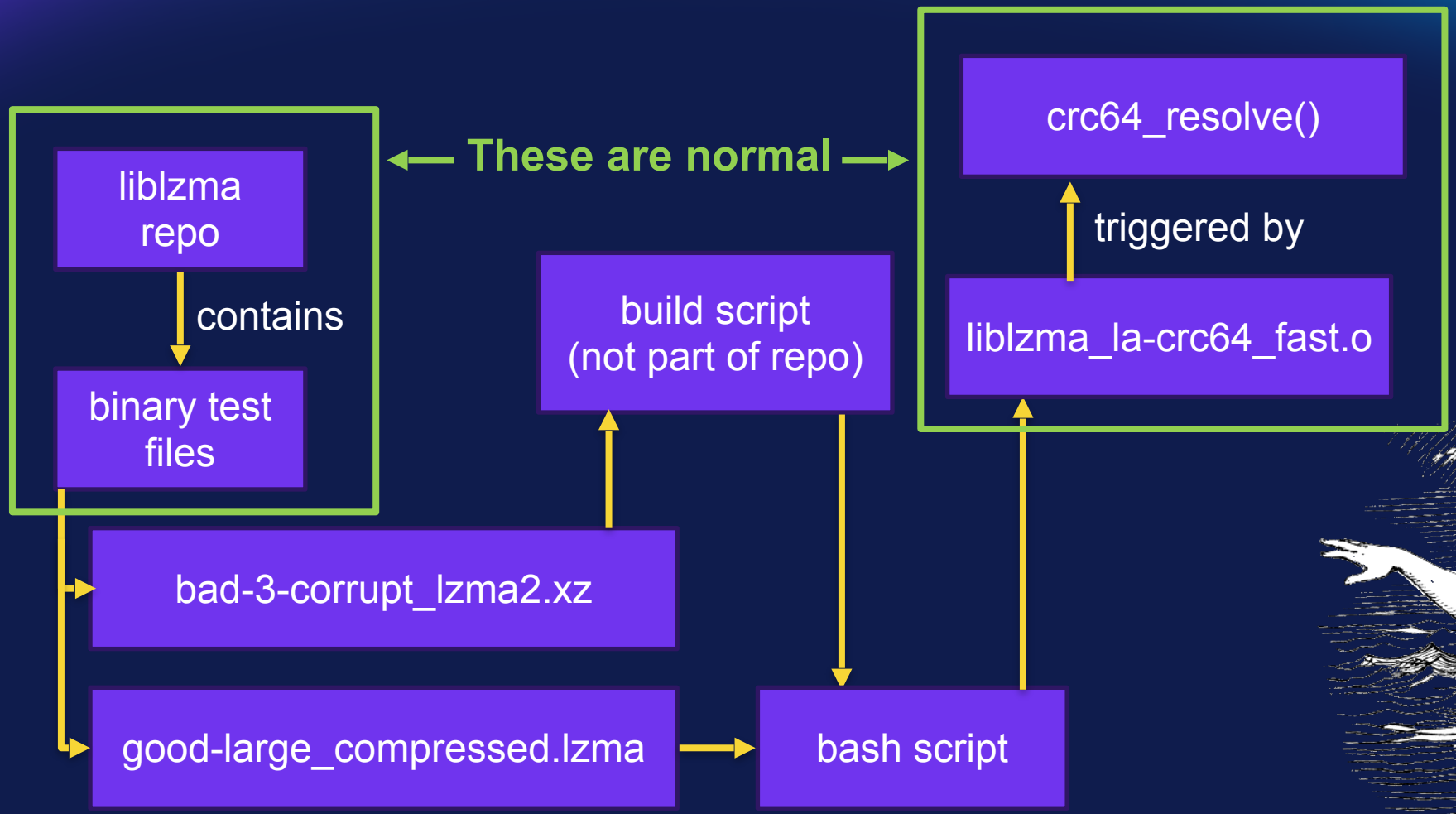
sonatype

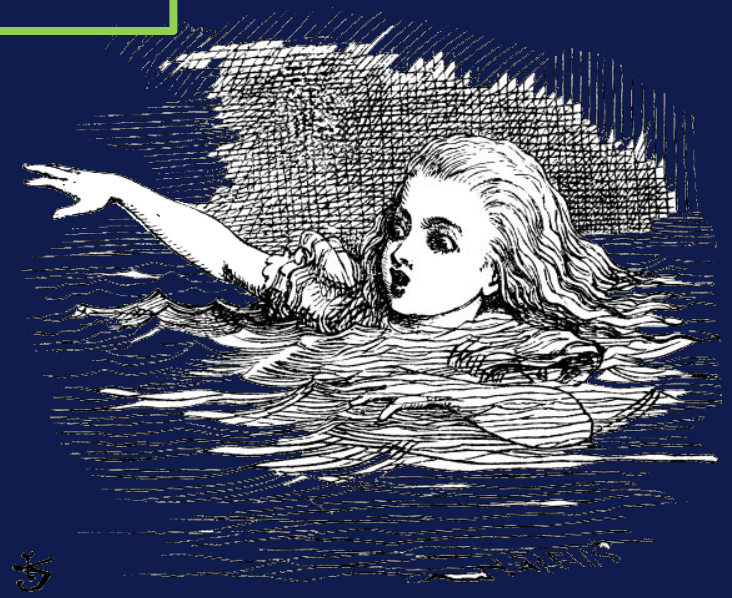
The Attack



sonatype

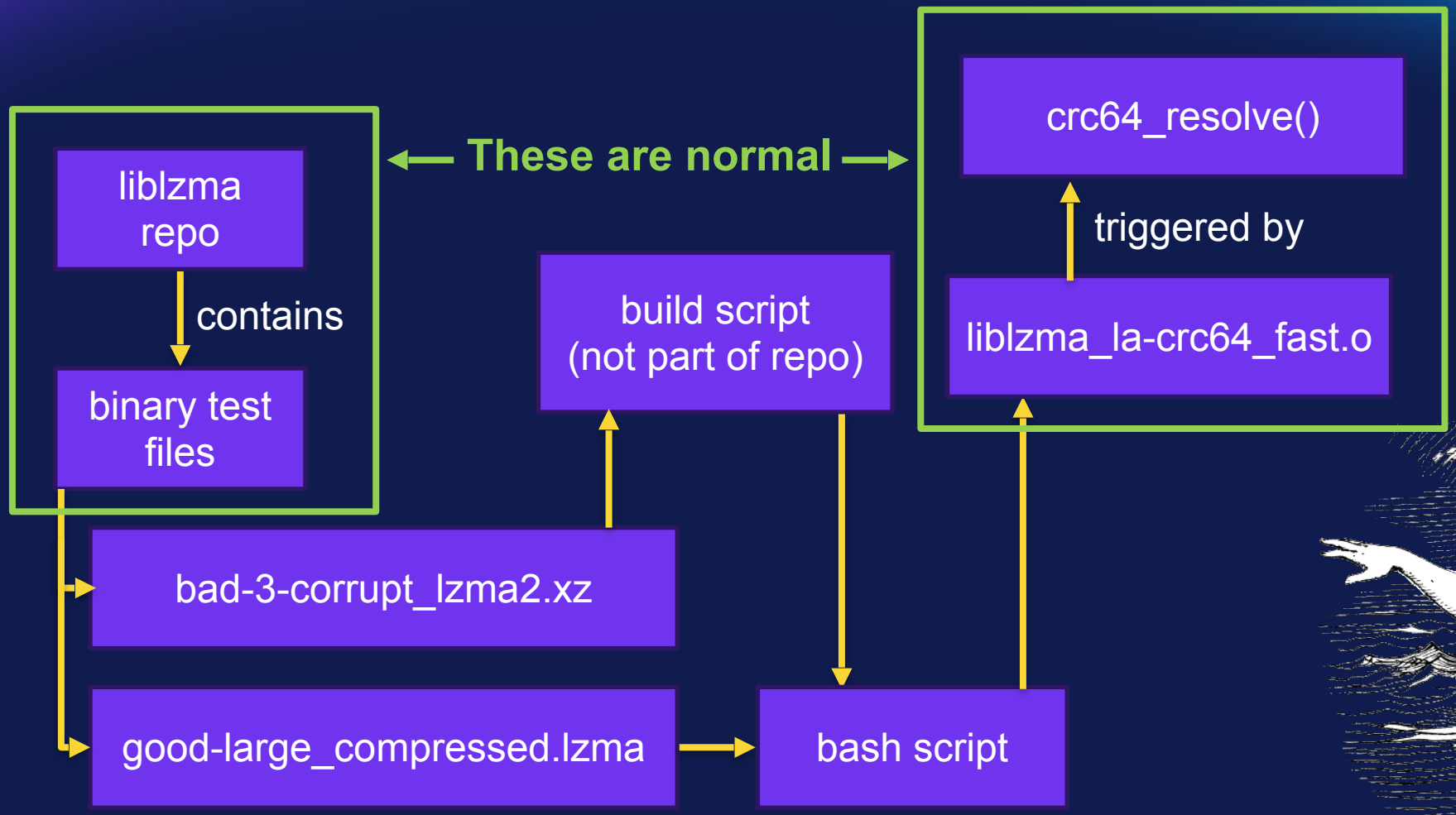
The Attack



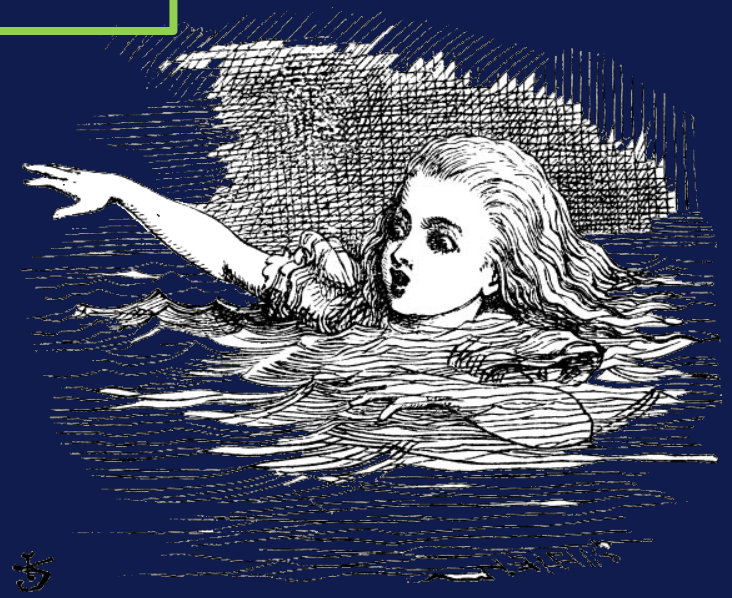


sonatype

The Attack



When loaded, hooks
RSA_public_decrypt
in OpenSSH
and bypasses auth



sonatype

Twenty Years Earlier…

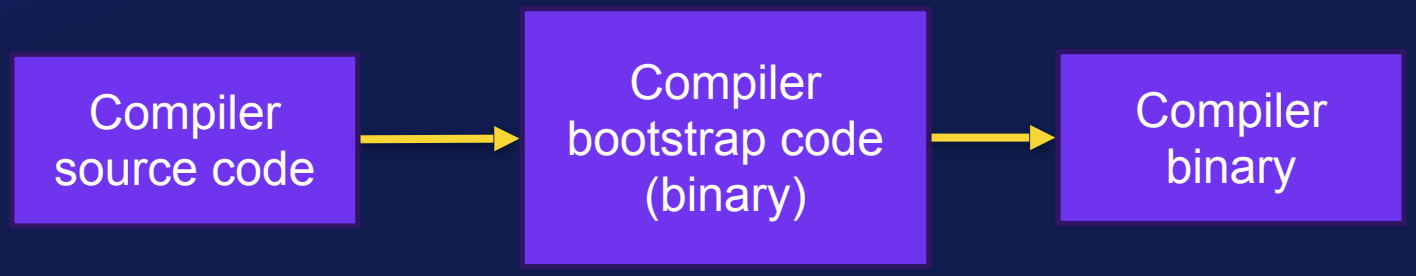
TURING AWARD LECTURE

Reflections on Trusting Trust

KEN THOMPSON

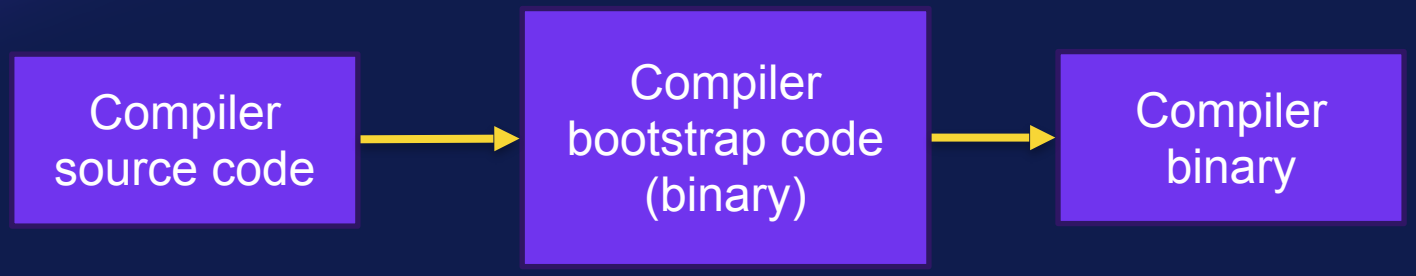
- Received Turing Award for creating UNIX
- His acceptance speech … is widely considered a seminal computer security work in its own right. (Wikipedia)

How Compilers Work

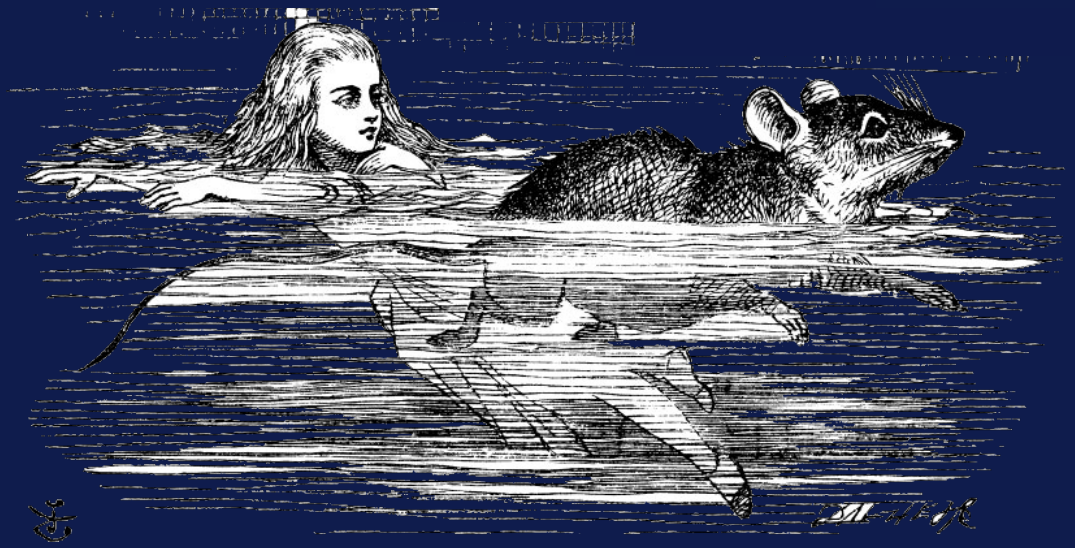


sonatype

How Compilers Work

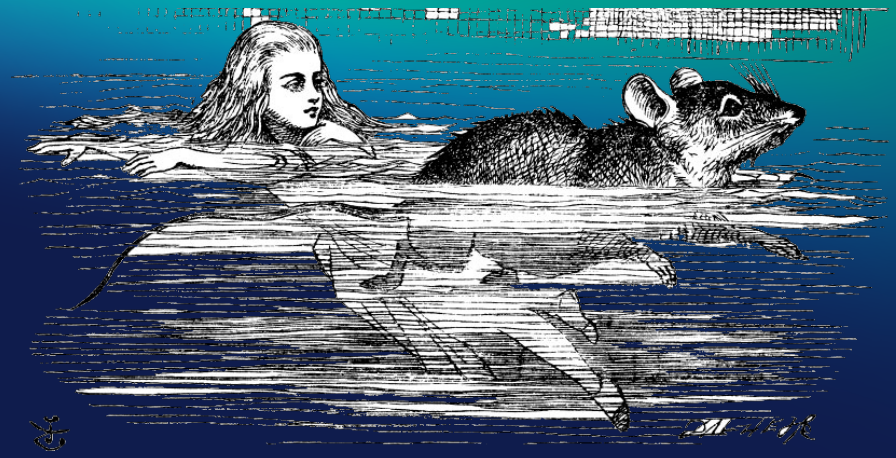


Hide a backdoor
In here

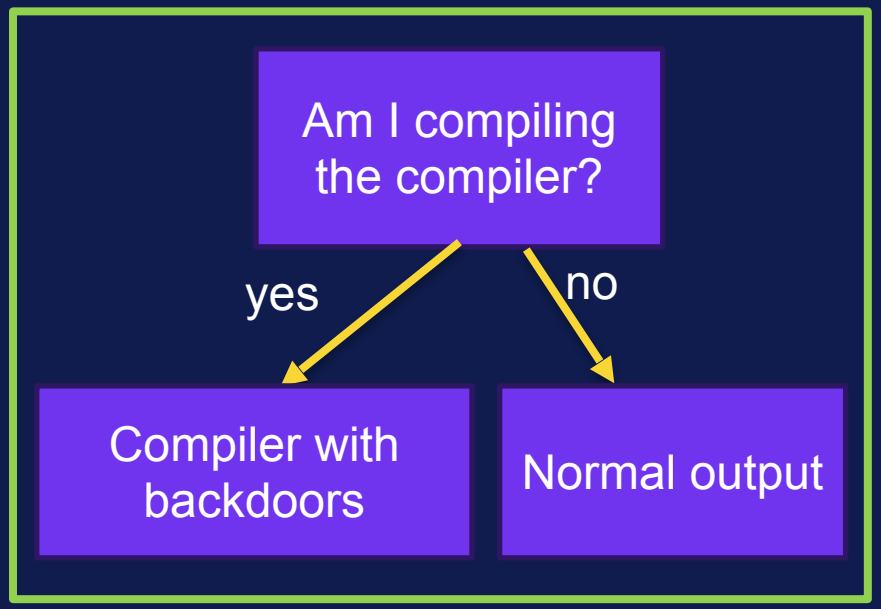


sonatype

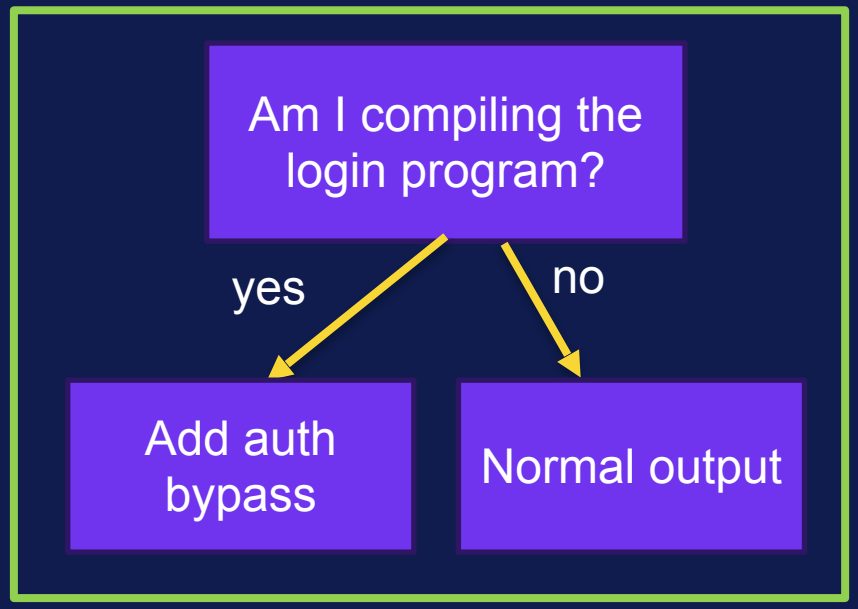
Trusting Trust Backdoor



Compiler Backdoor



Login Backdoor



sonatype

Comparison

“Trusting Trust”	xz attack
Backdoor gets inserted during compilation	Backdoor gets inserted during build / pkg
Backdoor is hidden in compiler binary	Backdoor is hidden in binary test file
Backdoor bypasses login prompt	Backdoor bypasses ssh authentication
Operates by modifying another program during compilation	Operates by modifying another program during loading

Twenty Years Earlier…

“In demonstrating the possibility of this kind of attack, I picked on the C compiler. I could have picked on any program-handling program such as an assembler, a **loader**, or even hardware microcode. As the level of program gets lower, these bugs will be harder and harder to detect. A well-installed microcode bug will be almost impossible to detect.”

--Ken Thompson

Comparison

“Trusting Trust”	AI code gen attack
Backdoor gets inserted during compilation	Backdoor gets inserted during training
Backdoor is hidden in compiler binary	Backdoor is hidden in model weights
Backdoor bypasses login prompt	Backdoor introduces incorrect code
Operates by modifying another program during compilation	Operates by modifying code suggestions

Comparison

“Trusting Trust”	General AI attack
Backdoor gets inserted during compilation	Backdoor gets inserted during training
Backdoor is hidden in compiler binary	Backdoor is hidden in model weights
Backdoor bypasses login prompt	Backdoor introduces bias / lies / errors
Operates by modifying another program during compilation	Operates by modifying decisions other automated systems make

The Timeline

- Oct 2021: Jia Tan starts contributing

- …

- Dec 2022: Jia Tan becomes a maintainer
- Feb 2024: Backdoor code inserted
- March 2024: Backdoor detected

sonatype

The Timeline

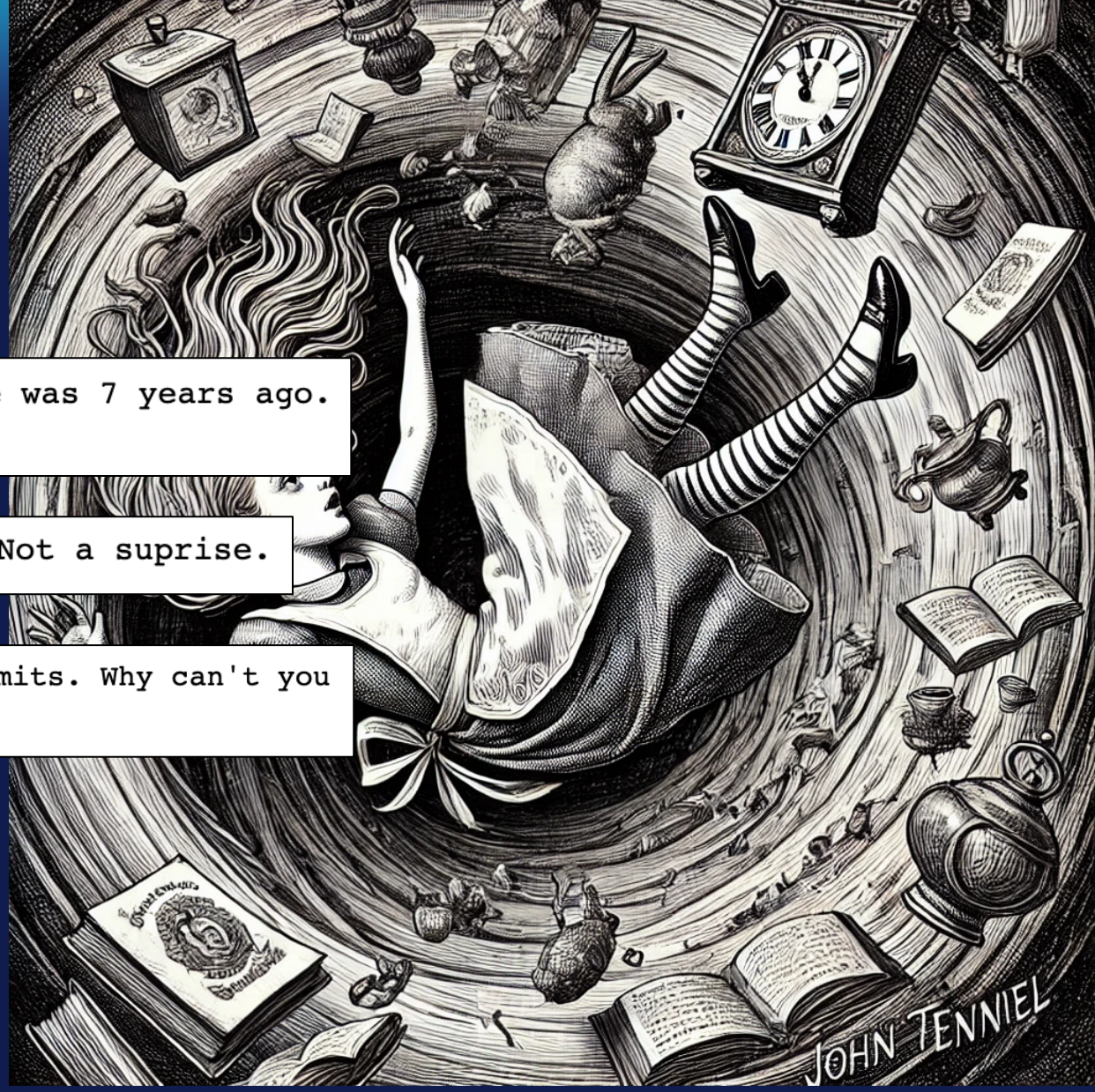
- Oct 2021: Jia Tan starts contributing

```
Patches spend years on this mailing list. 5.2.0 release was 7 years ago.
There is no reason to think anything is coming soon.
```

```
Over 1 month and no closer to being merged. Not a suprise.
```

```
Is there any progress on this? Jia I see you have recent commits. Why can't you
commit this yourself?
```

- Dec 2022: Jia Tan becomes a maintainer
- Feb 2024: Backdoor code inserted
- March 2024: Backdoor detected

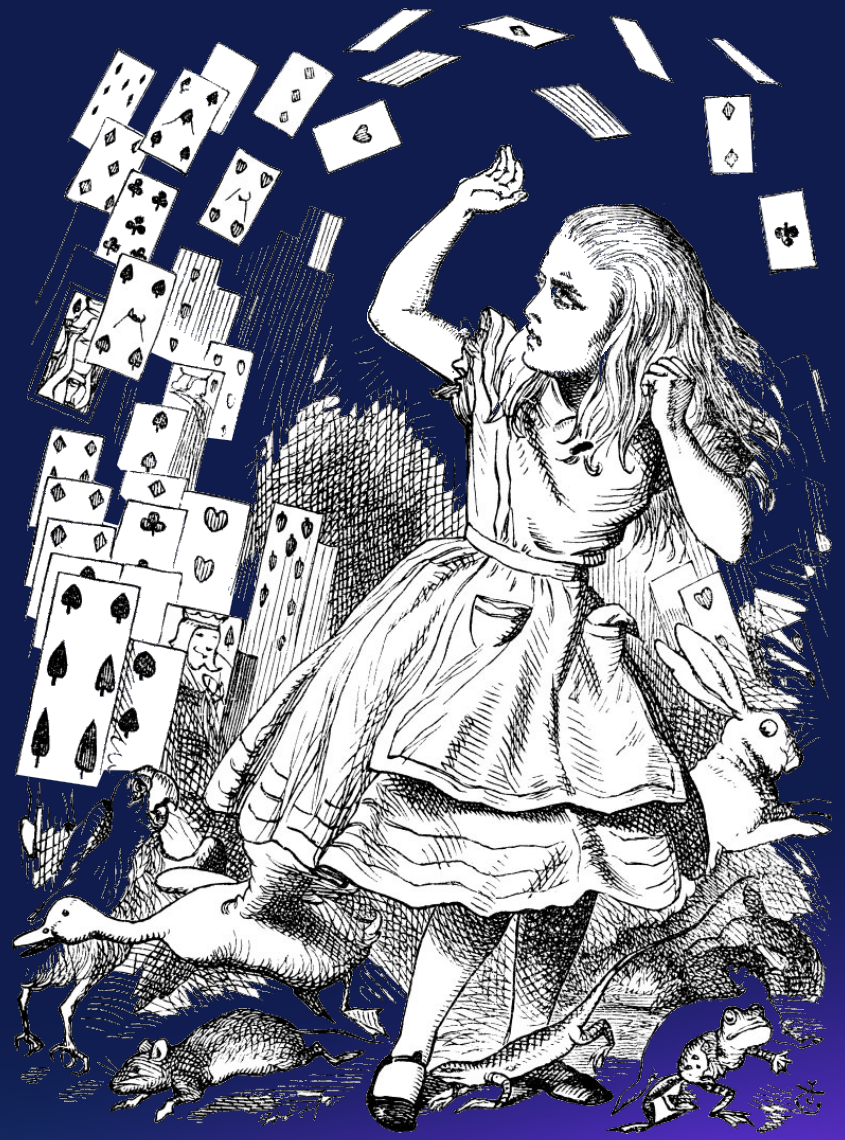


A Parallel Event

- OpenJS targeted by similar attack
- Maintainers of Node.js, jQuery, Lodash, Electron, Webpack
- Larger community, more maintainers
- Pressure to add maintainer is met with skepticism

Resisting These Attacks

- Transparency is key
  - Which repos include binaries?
  - Which projects are well-funded?
  - Which versions am I using?
  - What are the specific hashes?
  - What models am I using?
  - How were they derived?
- Technologies that can help
  - Additional project metadata
  - SBOMs
  - AIBOMs



sonatype