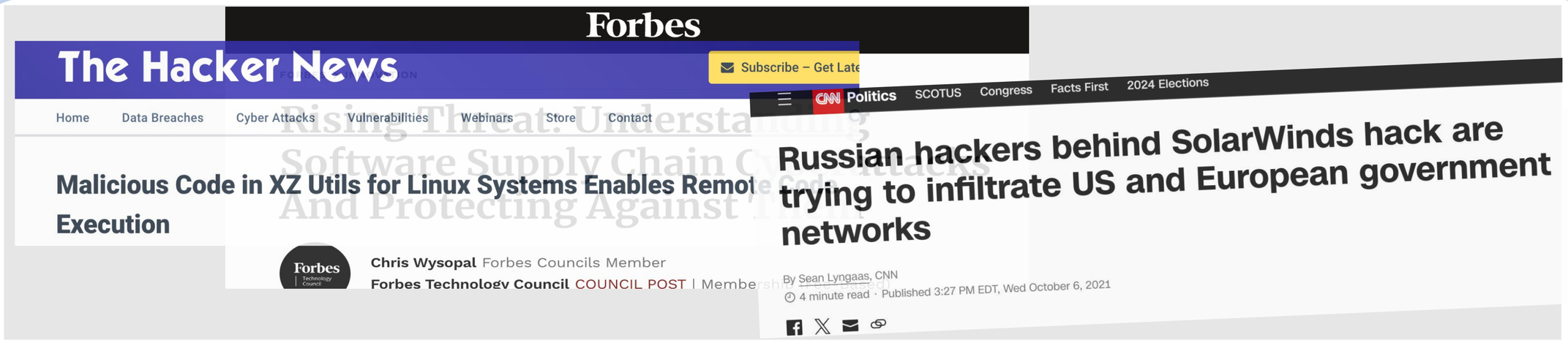
Veracity Technology Spearhead

Enabling end-to-end veracity within value exchange ecosystems

Improving Software Supply Chain Security: A Dataset for Binary Equivalence

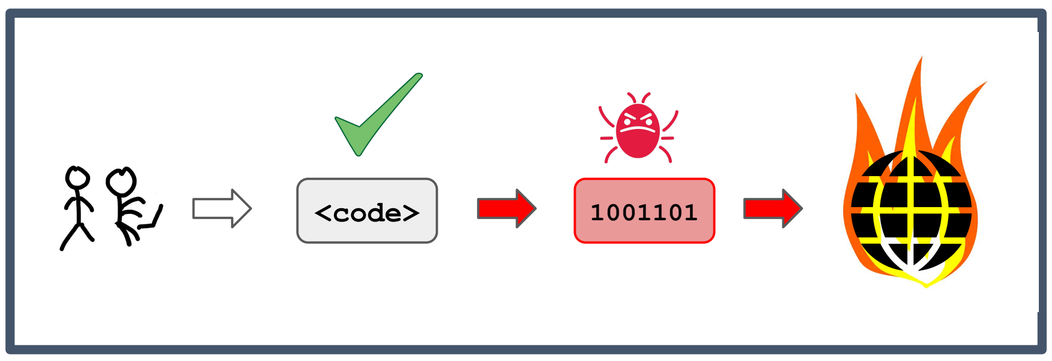
Jens Dietrich, Tim White – Victoria University Of Wellington, Behnaz Hassanshahi – Oracle Labs Australia

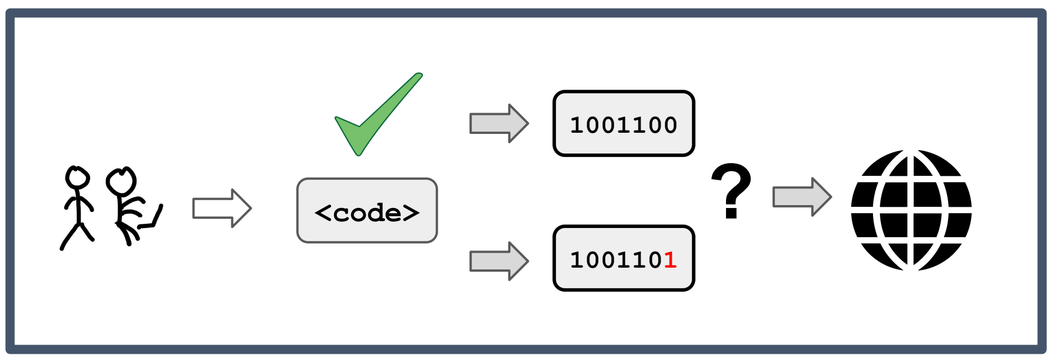


The Problem

Hackers can create vulnerabilities by compromising the highly automated processes used to build software from source code and distribute it. This approach has been employed to create some of the most severe and sophisticated cyber attacks (*Solarwinds*) in recent years.

A common approach to detect compromised programs is to reproduce builds in a secure environment, such as Google's *Assured Open Source*. The question arises whether two given builds are equivalent, and how comparisons can be used to detect vulnerabilities in compromised builds.





Our Approach: Data-Driven Equivalence

Binary equivalence is a difficult problem: trivial bitwise comparison has little value, and proving (functional) equivalence is impossible ("undecidable"). Pragmatic solutions need datasets for evaluation and training.

We have created such a dataset for Java bytecode from real world programs and a docker-based setup to mutate the compiler being used:

1. 300k+ pairs of binaries that are not identical but equivalent
2. 11k+ pairs of binaries that are similar but not equivalent due to confirmed structural ("*API*") changes between versions
3. 142k+ pairs of binaries that are similar but not equivalent due to injected semantic changes (aka *"mutations"*)
4. 202 pairs of binaries that are similar but not equivalent due to vulnerability patches

An evaluation of common approaches to equivalence based on common decompilers, disassemblers, locality-sensitive hashes and state-of-the-art academic tools against this dataset has revealed significant shortcomings of all of those methods.

Where from Here

We will continue working with an industry partner (Oracle Labs Brisbane) to develop novel approaches to binary equivalence, and tools to detect vulnerabilities in open source software based on this approach.

We want to find the next Solarwinds!

For more information please email **jens.dietrich@vuw.ac.nz**

Oracle Labs

National Science Challenges SCIENCE FOR TECHNOLOGICAL INNOVATION Kia kotahi mai – Te Ao Pūtaiao me Te Ao Hangarau

Victoria University of Wellington
University of Auckland University of Otago
University of Canterbury University of Waikato