Keeping web-browsing data safe from hackers

June 9 2022, by Adam Zewe

MIT researchers analyzed a powerful cyberattack, known as a website-fingerprinting attack, and then developed strategies that dramatically reduce the attacker's chances of success. Pictured, from left to right: graduate student Jules Drean, Mengjia Yan, the Homer A. Burnell Career Development Assistant Professor of Electrical Engineering and Computer Science, and Jack Cook '22. Credit: Jose-Luis Olivares, MIT

Malicious agents can use machine learning to launch powerful attacks that steal information in ways that are tough to prevent and often even more difficult to study.

Attackers can capture data that "leaks" between software programs running on the same computer. They then use machine-learning algorithms to decode those signals, which enables them to obtain passwords or other private information. These are called "side-channel attacks" because information is acquired through a channel not meant for communication.

Researchers at MIT have shown that machine-learning-assisted side-channel attacks are both extremely robust and poorly understood. The use of machine-learning algorithms, which are often impossible to fully comprehend due to their complexity, is a particular challenge. In a new paper, the team studied a documented attack that was thought to work by capturing signals leaked when a computer accesses memory. They found that the mechanisms behind this attack were misidentified, which would prevent researchers from crafting effective defenses.

To study the attack, they removed all memory accesses and noticed the attack became even more powerful. Then they searched for sources of information leakage and found that the attack actually monitors events that interrupt a computer's other processes. They show that an adversary can use this machine-learning-assisted attack to exploit a security flaw and determine the website a user is browsing with almost perfect accuracy.

With this knowledge in hand, they developed two strategies that can thwart this attack.

"The focus of this work is really on the analysis to find the root cause of the problem. As researchers, we should really try to delve deeper and do

more analysis work, rather than just blindly using black-box machine-learning tactics to demonstrate one attack after another. The lesson we learned is that these machine-learning-assisted attacks can be extremely misleading," says senior author Mengjia Yan, the Homer A. Burnell Career Development Assistant Professor of Electrical Engineering and Computer Science (EECS) and a member of the Computer Science and Artificial Intelligence Laboratory (CSAIL).

The lead author of the paper is Jack Cook '22, a recent graduate in computer science. Co-authors include CSAIL graduate student Jules Drean and Jonathan Behrens Ph.D. '22. The research will be presented at the International Symposium on Computer Architecture.

A side-channel surprise

Cook launched the project while taking Yan's advanced seminar course. For a class assignment, he tried to replicate a machine-learning-assisted side-channel attack from the literature. Past work had concluded that this attack counts how many times the computer accesses memory as it loads a website and then uses machine learning to identify the website. This is known as a website-fingerprinting attack.

He showed that prior work relied on a flawed machine-learning-based analysis to incorrectly pinpoint the source of the attack. Machine learning can't prove causality in these types of attacks, Cook says.

"All I did was remove the memory access and the attack still worked just as well, or even better. So, then I wondered, what actually opens up the side channel?"

This led to a research project in which Cook and his collaborators embarked on a careful analysis of the attack. They designed an almost identical attack, but without memory accesses, and studied it in detail.

They found that the attack actually records a computer's timer values at fixed intervals and uses that information to infer what website is being accessed. Essentially, the attack measures how busy the computer is over time.

A fluctuation in the timer value means the computer is processing a different amount of information in that interval. This is due to system interrupts. A system interrupt occurs when the computer's processes are interrupted by requests from hardware devices; the computer must pause what it is doing to handle the new request.

When a website is loading, it sends instructions to a web browser to run scripts, render graphics, load videos, etc. Each of these can trigger many system interrupts.

An attacker monitoring the timer can use machine learning to infer high-level information from these system interrupts to determine what website a user is visiting. This is possible because interrupt activity generated by one website, like CNN.com, is very similar each time it loads, but very different from other websites, like Wikipedia.com, Cook explains.

"One of the really scary things about this attack is that we wrote it in JavaScript, so you don't have to download or install any code. All you have to do is open a website. Someone could embed this into a website and then theoretically be able to snoop on other activity on your computer," he says.

The attack is extremely successful. For instance, when a computer is running Chrome on the macOS operating system, the attack was able to identify websites with 94% accuracy. All commercial browsers and operating systems they tested resulted in an attack with more than 91% accuracy.

There are many factors that can affect a computer's timer, so determining what led to an attack with such high accuracy was akin to finding a needle in a haystack, Cook says. They ran many controlled experiments, removing one variable at a time, until they realized the signal must be coming for system interrupts, which often can't be processed separately from the attacker's code.

Fighting back

Once the researchers understood the attack, they crafted security strategies to prevent it.

First, they created a browser extension that generates frequent interrupts, like pinging random websites to create bursts of activity. The added noise makes it much more difficult for the attacker to decode signals. This dropped the attack's accuracy from 96% to 62%, but it slowed the computer's performance.

For their second countermeasure, they modified the timer to return values that are close to, but not the actual time. This makes it much harder for an attacker to measure the computer's activity over an interval, Cook explains. This mitigation cut the attack's accuracy from 96% down to just 1%.

"I was surprised by how such a small mitigation like adding randomness to the timer could be so effective. This mitigation strategy could really be put in use today. It doesn't affect how you use most websites," he says.

Building off this work, the researchers plan to develop a systematic analysis framework for machine-learning-assisted side-channel attacks. This could help the researchers get to the root cause of more attacks, Yan says. They also want to see how they can use machine learning to

discover other types of vulnerabilities.

"This paper presents a new interrupt-based side channel attack and demonstrates that it can be effectively used for website fingerprinting attacks, while previously, such attacks were believed to be possible due to cache side channels," says Yanjing Li, assistant professor in the Department of Computer Science at the University of Chicago, who was not involved with this research. "I liked this paper immediately after I first read it, not only because the new attack is interesting and successfully challenges existing notions, but also because it points out a key limitation of ML-assisted side-channel attacks—blindly relying on machine-learning models without careful analysis cannot provide any understanding on the actual causes/sources of an attack, and can even be misleading. This is very insightful and I believe will inspire many future works in this direction."

More information: There's Always a Bigger Fish: A Clarifying Analysis of a Machine-Learning-Assisted Side-Channel Attack. people.csail.mit.edu/mengjia/d … r_Fish_ISCA_2022.pdf

This story is republished courtesy of MIT News (web.mit.edu/newsoffice/), a popular site that covers news about MIT research, innovation and teaching.

Provided by Massachusetts Institute of Technology