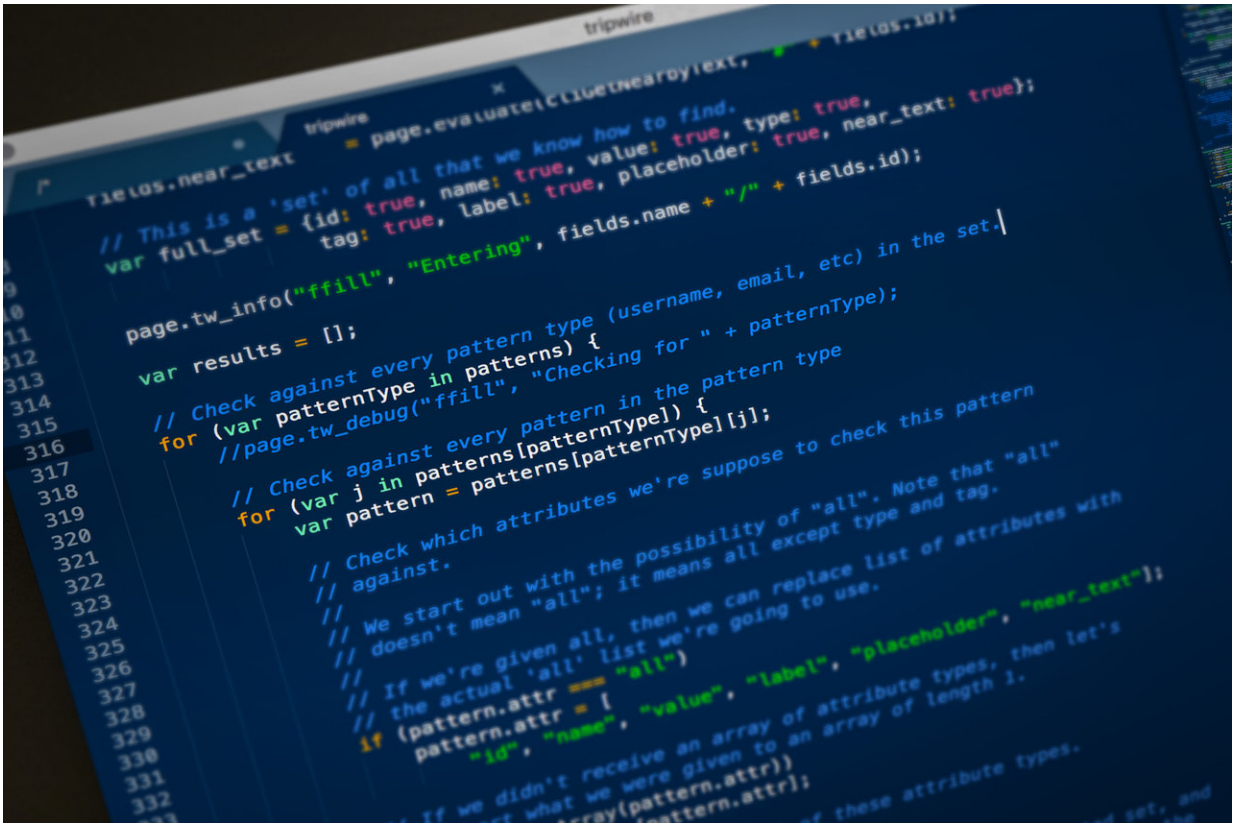
Computer scientists develop a simple tool to tell if websites suffered a data breach

December 12 2017



Some of the code engineers use to develop Tripwire.The entire code is available on GitHub. Credit: University of California San Diego

Computer scientists have built and successfully tested a tool designed to detect when websites are hacked by monitoring the activity of email

accounts associated with them. The researchers were surprised to find that almost 1 percent of the websites they tested had suffered a data breach during their 18-month study period, regardless of how big the companies' reach and audience are.

"No one is above this—companies or nation states— it's going to happen; it's just a question of when," said Alex C. Snoeren, the paper's senior author and a professor of computer science at the Jacobs School of Engineering at the University of California San Diego.

One percent might not seem like much. But given that there are over a billion sites on the Internet, this means tens of millions of websites could be breached every year, said Joe DeBlasio, one of Snoeren's Ph.D. students and the paper's first author.

Even scarier, the researchers found that popular sites were just as likely to be hacked as unpopular ones. This means that out of the top-1000 most visited sites on the Internet, ten are likely to be hacked every year.

"One percent of the really big shops getting owned is terrifying," DeBlasio said.

The team of researchers at UC San Diego presented the tool in November at ACM Internet Measurement Conference in London.

The concept behind the tool, called Tripwire, is relatively simple. DeBlasio created a bot that registers and creates accounts on a large number of websites—around 2,300 were included in their study. Each account is associated with a unique email address. The tool was designed to use the same password for the email account and the website account associated with that email. Researchers then waited to see if an outside party used the password to access the email account. This would indicate that the website's account information had been leaked.

To make sure that the breach was related to hacked websites and not the email provider or their own infrastructure, researchers set up a control group. It consisted of more than 100,000 email accounts they created with the same email provider used in the study. But computer scientists did not use the addresses to register on websites. None of these email accounts were accessed by hackers.

In the end, researchers determined 19 websites had been hacked, including a well-known American startup with more than 45 million active customers.

Once the accounts were breached, researchers got in touch with the sites' security teams to warn them of the breaches. They exchanged emails and phone calls. "I was heartened that the big sites we interacted with took us seriously," Snoeren said.

Yet none of the websites chose to disclose to their customers the breach the researchers had uncovered. "I was somewhat surprised no one acted on our results," Snoeren said.

The researchers decided not to name the companies in their study.

"The reality is that these companies didn't volunteer to be part of this study," Snoeren said. "By doing this, we've opened them up to huge financial and legal exposure. So we decided to put the onus on them to disclose."

Interestingly, very few of the breached accounts were used to send spam once they became vulnerable. Instead, the hackers usually just monitored email traffic. DeBlasio speculates that the hackers were monitoring emails to harvest valuable information, such as bank and credit card accounts.

Researchers went a step further. They created at least two accounts per website. One account had an "easy" password—strings of seven-character words with their first letter capitalized and followed by a single digit. These kinds of passwords are usually the first passwords that hackers will guess. The other account had a "hard" password—random 10-character strings of numbers and letters, both in lower and upper case, without special characters.

Seeing which of the two accounts got breached allowed researchers to make a good guess about how websites store passwords. If both the easy and hard passwords were hacked, the website likely just stores passwords in plain text, contrary to typically-followed best practice. If only the account using the easy password was breached, the sites likely used a more sophisticated method for password storage: an algorithm that turns passwords into a random string of data—with random information added to those strings.

The computer scientists had a few pieces of advice for Internet users: don't reuse passwords; use a password manager; and ask yourself how much you really need to disclose online.

"Websites ask for a lot of information," Snoeren said. "Why do they need to know your mother's real maiden name and the name of your dog?"

DeBlasio was less optimistic that these precautions would work.

"The truth of the matter is that your information is going to get out; and you're not going to know that it got out," he said.

Snoeren and colleagues are not planning to pursue further research on Tripwire.

"We hope to have impact through companies picking it up and using it themselves," he said. "Any major email provider can provide this service."

Provided by University of California - San Diego