The New York Times

Finding a Needle in a Digital Haystack

By TIM WHITE
JAN. 22, 2014

PLEASANTON, Calif. — LAST year the private sector spent $67.2 billion on cybersecurity services. Nevertheless, according to a recent investigation by Verizon, 60 percent of successful hacks were not detected until months after the attacks began. In the wake of recent high-profile hacker attacks against Target, Neiman Marcus and other retailers, the obvious question is: Why hasn't all that money done any good?

It's not for lack of trying. Much of the money is well spent, paying for armies of technical engineers and state-of-the-art security applications.

The problem is not the resources, or the personnel, or the data. It's that many organizations simply don't know how to arrange the data to identify suspicious patterns and weaknesses, at least not fast enough. There's too much data, and not enough perspective.

What we need, then, is not necessarily more money or information, but a better way of knowing what it means — of interpreting the data to discover an unknown attack as it happens or, even better, anticipate the next attack.

The amount of security-related data that the typical enterprise manages today is staggering. A Fortune 500-size company could have over an exabyte (a billion gigabytes) of data scattered across thousands of servers and hundreds of thousands of software applications. All told, a detailed investigation of a complex, stealthy attack could involve trillions of discrete data points.

Most of those data points, even security-related points, are noise. Warnings are generated each time an easily blocked virus tries to gain access, or an employee enters the wrong password. Across a large company, that could mean thousands of warnings a day, all of which must be investigated. This noise makes the job of rapidly identifying serious attacks an overwhelming task.

The solution lies in finding a way to examine the data so that analysts can quickly identify suspicious patterns. Instead of programs to generate more data, we need different tools to understand them. And it turns out that the best tools are right in our heads: our eyes.

The fast-growing field of graph analytics involves the creation of a visual representation of "objects" within an organization: people, devices, computers, applications. The relationships are displayed as lines linking objects. (My company, YarcData, develops graph-analytic platforms for a variety of applications.)

With this type of analysis, objects and their relationships uncover connections that are hidden in the numbers but obvious to the eye. It allows a user to see, immediately, the object's broader context. A failed attack on one computer might not raise a flag on its own — but it could if an analyst saw, simultaneously, that the target was connected to other critical targets, or that the attack was a modification of a recent failed attack.

Our brains conduct such contextual assessments every day, albeit on a much simpler, smaller scale. Think about when one hears a strange noise. Within a split second our mind connects the noise with its context. And we can ask creative questions, drawing on seemingly unrelated correlations that a computer program could never be robust enough to handle. Was it late at night? Is the family dog in that room? If it was in the middle of the day and the dog was in that room, we would react differently than if it was a strange noise late at night, with the dog asleep at our feet and no one else in the house.

Consider another example: My phone has just mysteriously connected to a website I have never visited, or intended to visit. Our minds ask the same sorts of contextual questions: Am I using the phone? What time of day is it? How many websites have I visited in the last five minutes? The last 30 minutes? Could a new app be connecting to sites to bump ad revenues?

All this information is already present in a company's data network. Graph analytics facilitates presentation in a visual fashion to allow our eyes, and brains, to do the top-level processing, enabling rapid understanding of all these events, then identifying the difference that would explain whether it was a malicious application or just a harmless bit of spam. Someone, somewhere, has to decide on a fast, accurate response.

Why hasn't this been done already? The industry has tried. But the speed and scope in computing power were not available until recently. And of course, with so much invested in traditional cybersecurity methods, many companies have been loath to try something new.

That must change soon. Target has pledged to make "significant changes," and it's likely that name-brand retailers around the globe will be undertaking security audits and taking measures to harden themselves against attacks and protect customer data.

Winning at cybersecurity today isn't necessarily about collecting more data. It is about unleashing the information in the data that's already there — and doing so not in weeks, days and hours, but in minutes and seconds.

Tim White is the global head of government and intelligence for YarcData, a data analytics firm.