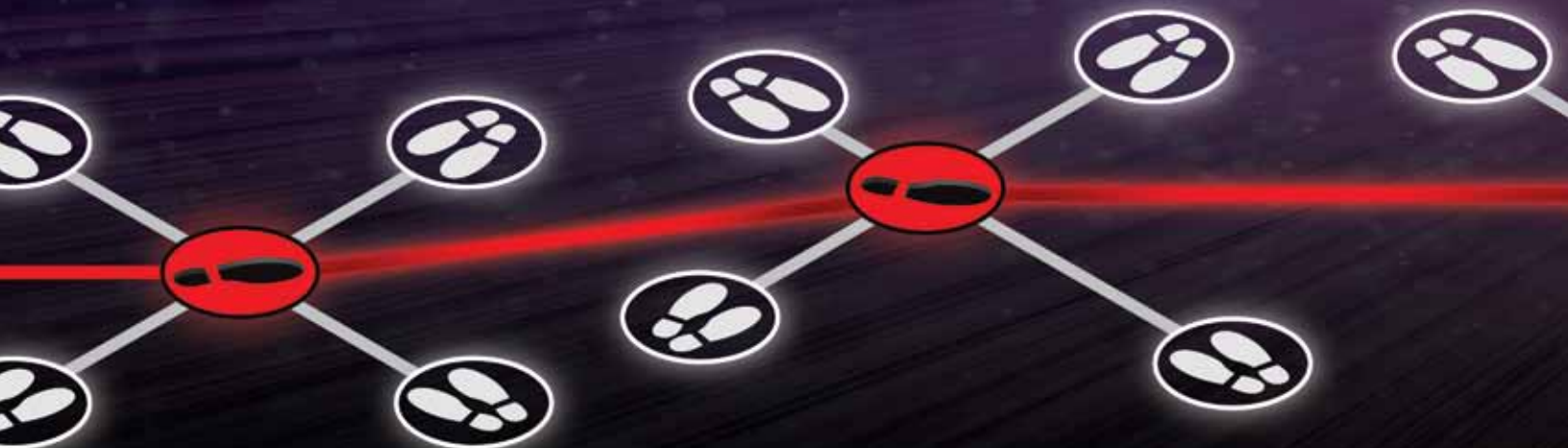


INTRUDER ALERT FOR THE CYBER WORLD

SOPHISTICATED HACKERS ARE GOING TO GET IN. THE TRICK IS TO FIND THEM ON THE



On January 7, 2013, Los Alamos National Laboratory was hacked by an unlikely adversary: itself.

“It’s a case of good guys acting like bad guys to test our defenses,” explains Josh Neil, a cyber security statistics expert at Los Alamos. “In this case, it was our own Engineering and Security Services group.” Neil leads a project that develops software to protect against cyber attacks. The software, called PathScan, has been up and running at Los Alamos for about a year, ready and able to catch intruders. Indeed, it proved capable of spotting the friendly attack and continues to staff its cyber sentry post, ever watchful for unfriendly ones.

Can it do the same for other protected networks? All evidence points to yes. It has already run or is currently running in a trial phase on other government and industry networks, with another trial run about to begin.

“I think we’ve got a ground-breaking technology for network defense,” Neil says. “This could really change the cyber security landscape for the better.”

Byte background

Conventional cyber security software works by scanning network activity for particular packets of information, or byte strings, that correspond to previously reported cyber crimes or other malware activity. Because this approach matches current network byte strings to well-identified malicious byte strings, it is extremely accurate at diagnosing known attacks. In this sense, it is similar to antivirus software running on an individual computer. It affords reliable protection against a long list of established viruses despite its vulnerability to something new. But that’s where the virus analogy ends.

“Sophisticated attacks generally are not virus-like,” says Neil. “Exponentially exploding, automated intrusions may sound worrisome, but in practice, they’re the easier ones to catch.”



Even if the particular form of a virus-like attack on an individual computer has never been encountered before, its explosive nature still tends to give it away on the network scale: it radiates outward from one computer to many, forming a pattern known as a star. Such an attack can be detected rapidly, with safeguards automatically engaged upon detection. For example, the affected computers might have their network connections disabled while alert messages are dispatched to network administrators. Attacks that can be handled in this way aren't the ones that most concern Neil.

"The ones you really have to worry about are those being driven by a human being in real time," he says.

In contrast to a virus-like attack that spreads rapidly from computer to computer and therefore induces a fairly obvious abnormality in the network usage pattern, a single user can try to hide in the daily bustle of network activity. The Los Alamos unclassified computer network, for example, contains about 20,000 computers generating more than 500 million communication events (from one computer to another) every day—presenting an enormous background in which a hacker may effectively disappear.

The main virtue of PathScan is its ability to find hidden hackers by recognizing the subtle, small-scale network abnormalities that result from their intrusions. It does this as its name implies, by scanning for the paths taken by hackers as they move around within the target network.

Your enemies closer

Hackers typically want to steal digital information, including personal and proprietary information. In some cases they may be individuals who intend to use the stolen information themselves or sell it for profit. In other cases they may be well-funded nation-states that plan to use the stolen information for national gain.

The first line of defense against such cyber theft is a firewall. A firewall analyzes data packets entering a protected network and rejects those that pose a security concern, such as external login attempts without proper authentication. As a result, a hacker can't access computers beyond the firewall directly and must instead obtain some kind of insider access. A brute-force approach might be to physically break into a

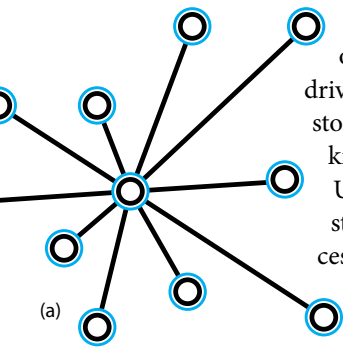
facility and stick a malware-containing USB drive into a computer. But a more common (and less risky) approach is phishing: sending an official-looking email to employees on the inside that encourages each employee to click on an external link or open an attached file. Either action delivers malware to the employee's computer. Because the employee accesses the link or the attachment deliberately—that is, the connection is initiated from within by an authorized user—the firewall may not prevent the subsequent malware download. This allows the external hacker to access the employee's computer.

The most prevalent way defend against phishing attacks (without disallowing all web activity originating from email clicks) is to train employees to recognize suspicious emails. But this defense is not perfect. "We can't rely on being able to stop every phishing attack from getting through while maintaining current network usability," says Neil, "and that means we have to be able to detect and stop an intruder inside the network."

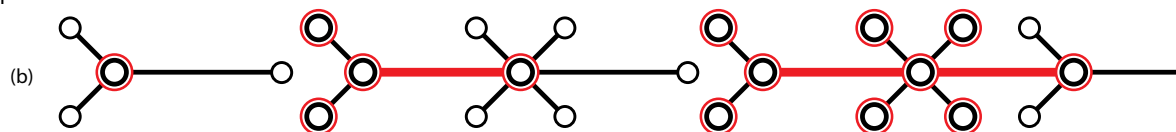
Crawling criminals

Fortunately, an employee who gets tricked by a phishing attack rarely has whatever the hacker is looking for right there on his or her computer. That means the hacker must hop from the hacked computer to other computers on the network in search of data worth stealing. This usually isn't quick or easy, because not every computer can access every other computer.

Once a hacker arrives at one computer, he or she has the ability to log in to all the other machines normally accessible from that computer because login credentials to those other machines are stored on its hard drive. (Passwords may not be directly stored, but encrypted versions of them, known as hashed passwords, are. Unfortunately, hashed passwords can still provide login access in a process known as "pass the hash," even if the hacker can't read the original passwords.)



Cyber attack patterns: (a) A star formation occurs when an attacker, whether human-driven or self-replicating like a computer virus, uses one hacked computer to reach out to many more, checking each for vulnerabilities, useful data, or login credentials to other computers. (b) A caterpillar is a human-controlled attack formation in which the hacker progresses from each computer to a handful of others. Moves that further the hacker's goals form the caterpillar's body, while backtracked dead-ends form the legs.



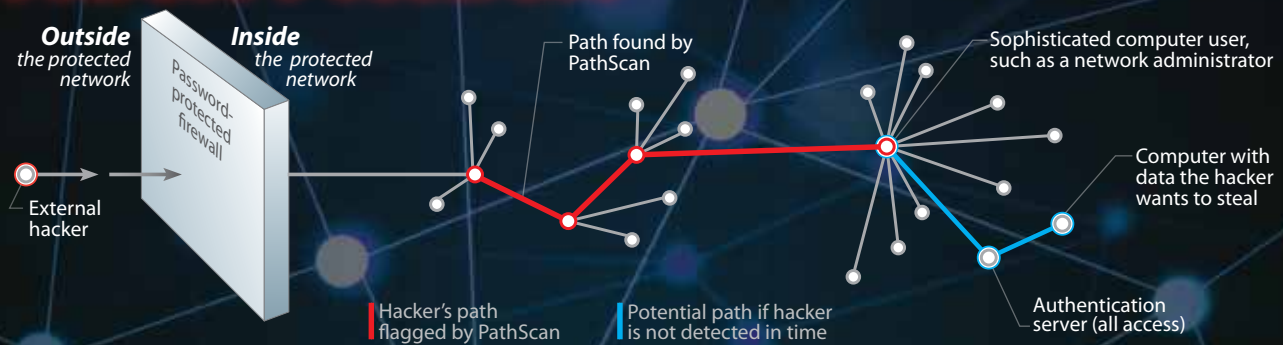
Typically, the credentials stored on one computer deliver access to many other machines, including email servers, network drives, shared printers, and even other office computers. These are not always useful to the hacker, however, partly because the login credentials obtained from one computer usually convey few access privileges to the files on another computer—effectively "look but don't touch"—and partly because the most sensitive machines require progressively higher-level credentials.

As a result, a hacker must hop from one machine to another in search of more privileged login credentials, advancing (and sometimes backtracking) across the network. This creates a tell-tale pattern of motion, hopping from one computer to several others, settling on the most promising one, and repeating the process from there. The pattern, when drawn on paper, appears as a line of "correct" hops with a bunch of "incorrect" hops extending off to the sides. The pattern is named for its resemblance to a caterpillar, with the body formed from the line of correct hops and the legs formed from all the others. PathScan searches for these caterpillars.

The process of navigating from one computer to the next is slow: the hacker has to examine each machine to see what information and login credentials it stores. And most computers on the network have little or no useful information to steal and no elevated login credentials beyond what the hacker already used to get that far. However, some users are more sophisticated in terms of their network use, such as system administrators who oversee all the computers in some large group within the organization. If an intruder successfully enters one of these sophisticated-user machines, he or she will find a set of login credentials with complete administrative access privileges to more valuable target servers. With these privileges, the hacker can view or copy anything and even install permanent programs like backdoors, which allow the hacker to secretly return to a machine with full administrative access at any time in the future.

There is one machine in particular that the intruder really wants to access, known as an authentication server. It holds and recognizes the login credentials for users all over the organization and is involved every time any computer on the network authenticates to any other. This machine is the hacker's ultimate all-access pass; if a hacker invades the authentication server with a valid administrator credential, every other computer becomes available, including those containing the data the hacker wants to steal. Therefore, the trick is to detect the caterpillar's motion and shut it down

Hack Attack



A successful cyber attack, usually attempted for the purpose of stealing proprietary data, requires several steps:

(1) The attacker must access a computer inside the protected network by establishing a connection across the organization's firewall. This requires password authentication, and if the hacker can't obtain valid login credentials, he or she may make a phishing attempt instead. The attacker contacts an employee of the organization by email and tricks the employee into clicking on something malicious—either an attached file or a link to a compromised website—which then allows access to that employee's computer.

(2) Since the information the attacker is after is unlikely to reside on the first computer reached, and because the original firewall crossing might have alerted computer security personnel to the intrusion, the attacker must hop to another computer using login credentials found on the original hacked computer. In this way, the intruder moves from one computer to another, continually searching for higher-level login credentials that convey elevated access privileges on protected systems and servers.

(3) Any combination of three consecutive hops considered out of the ordinary by the Los Alamos PathScan software (when compared to normal network activity) alerts computer security personnel to investigate. They determine the extent of the attack and how to respond. It may be necessary, for example, to disconnect part or all of the organization's network from the outside world, wipe certain computers clean, or replace certain logins. (The red line illustrates a detectable intruder path.)

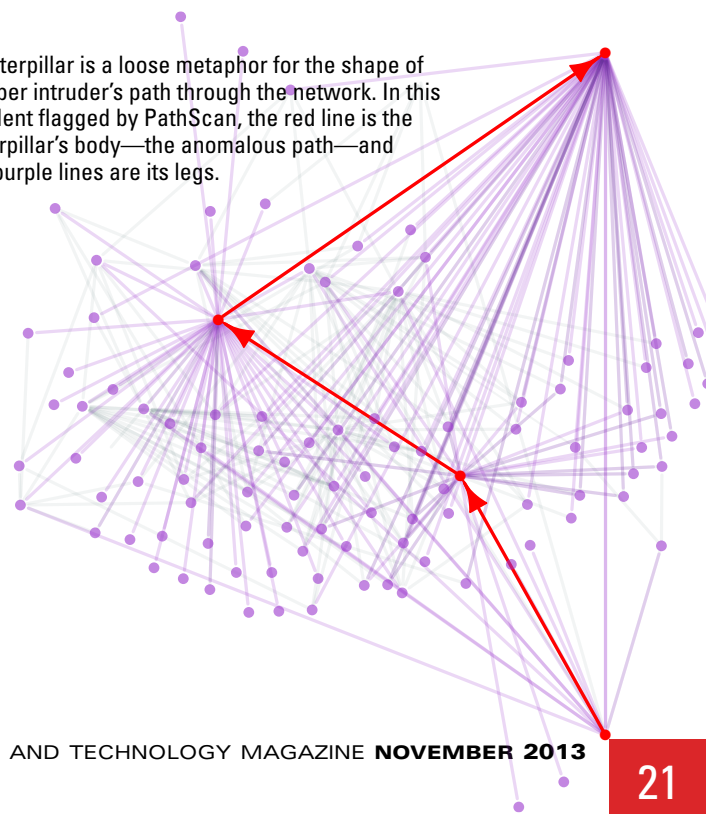
(4) With each hop, the attacker tries to move toward a computer with greater access, such as a computer used by a network administrator or, better yet, the organization's authentication server, which stores login credentials to all other computers on the network. If the attacker is not detected before reaching these machines, then the computer containing the data he or she intends to steal will become accessible. (The blue line indicates this access sequence if the attack is not stopped earlier.)

before it reaches the authentication server. Otherwise, if that server is hacked, the entire network will have to be taken offline while every login in the whole organization is changed and every computer in the organization is analyzed for tampering and theft. It may even be necessary to wipe many computers completely, which would represent a substantial cost in terms of downtime and lost productivity, even if the data theft was ultimately unsuccessful.

Guarding the edges

Neil and his PathScan colleague Curtis Storlie are statisticians, and PathScan is primarily a probability and statistics analysis tool. It maintains a baseline statistical pattern for normal network behavior and identifies network communications that deviate from that pattern. What is the prob-

A caterpillar is a loose metaphor for the shape of a cyber intruder's path through the network. In this incident flagged by PathScan, the red line is the caterpillar's body—the anomalous path—and the purple lines are its legs.



ability that computer A contacts computer B at a particular time on a Wednesday afternoon? What if it's the third time A has contacted B in the last hour? What if A just contacted C 10 minutes earlier, and D 25 minutes before that? And what if B contacts E after being contacted by A? Are these events normal or suspicious?

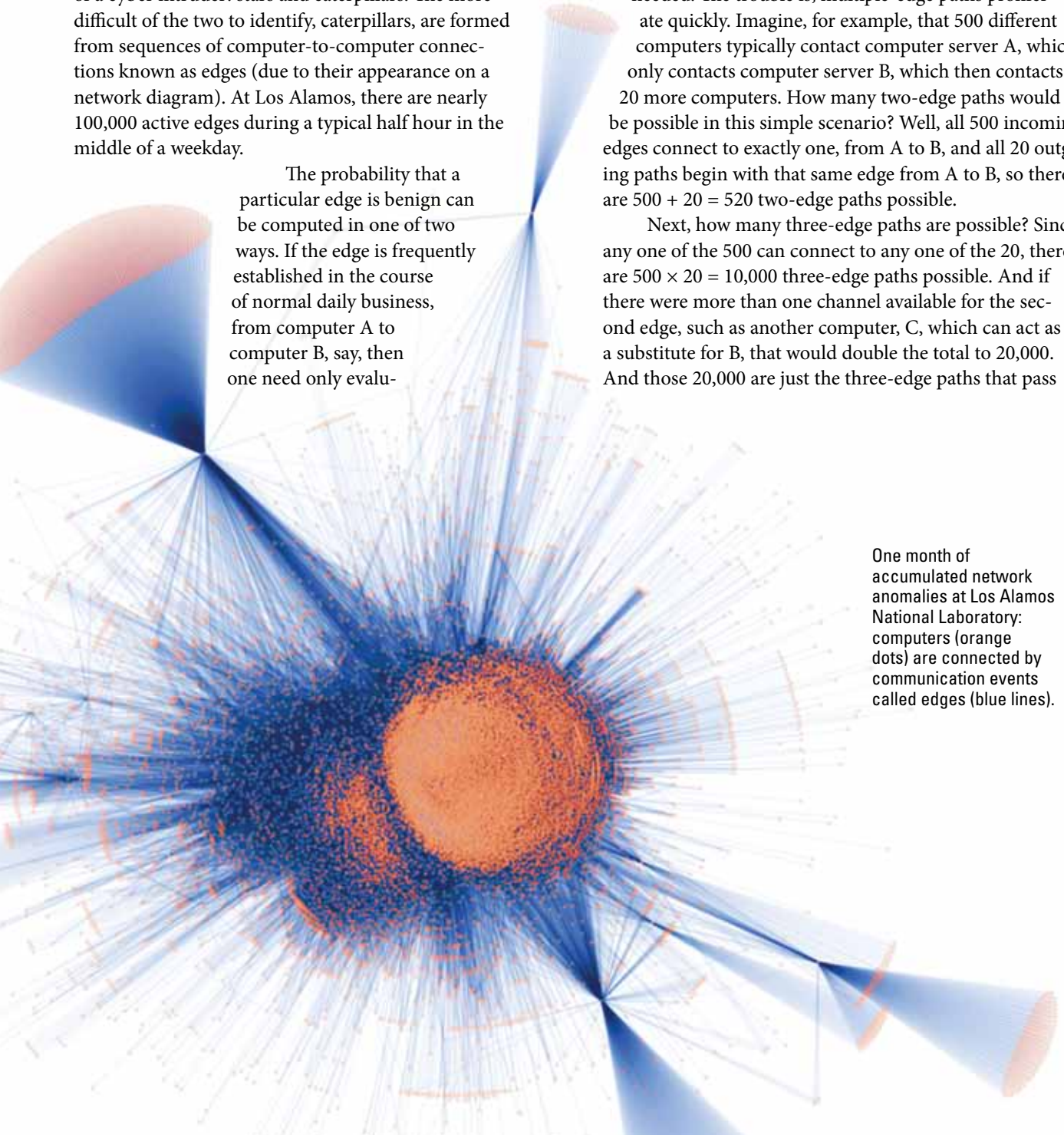
Also on the team are system architect Curtis Hash, large-scale data-management specialist Alexander Brugh, and cyber security expert Mike Fisk. Together, the team has designed the software to sift through an enormous volume of network data in real time, looking for the major footprints of a cyber intruder: stars and caterpillars. The more difficult of the two to identify, caterpillars, are formed from sequences of computer-to-computer connections known as edges (due to their appearance on a network diagram). At Los Alamos, there are nearly 100,000 active edges during a typical half hour in the middle of a weekday.

The probability that a particular edge is benign can be computed in one of two ways. If the edge is frequently established in the course of normal daily business, from computer A to computer B, say, then one need only evalu-

ate the frequency. If that connection is typically established 50 times in a given time period, then how unusual would 75 be? A question like this can be answered probabilistically. Alternatively, if the edge is completely new because A has never contacted B before, then the probability that it's benign is modeled as a logistic regression, blending three distinct rates: the rate at which new edges appear on the network overall, the rate at which A initiates them, and the rate at which B receives them.

But individual, anomalous edges are usually insufficient to indicate an intruder; rather, multiple edges are needed. The trouble is, multiple-edge paths proliferate quickly. Imagine, for example, that 500 different computers typically contact computer server A, which only contacts computer server B, which then contacts 20 more computers. How many two-edge paths would be possible in this simple scenario? Well, all 500 incoming edges connect to exactly one, from A to B, and all 20 outgoing paths begin with that same edge from A to B, so there are $500 + 20 = 520$ two-edge paths possible.

Next, how many three-edge paths are possible? Since any one of the 500 can connect to any one of the 20, there are $500 \times 20 = 10,000$ three-edge paths possible. And if there were more than one channel available for the second edge, such as another computer, C, which can act as a substitute for B, that would double the total to 20,000. And those 20,000 are just the three-edge paths that pass



One month of accumulated network anomalies at Los Alamos National Laboratory: computers (orange dots) are connected by communication events called edges (blue lines).

through computer A after the first edge. On the Los Alamos unclassified network, mid-day on a weekday, a typical 30-minute window contains about 300 million three-edge paths.

Based on their extensive cyber experience, Neil and his team made the decision to stop there rather than include four- and five-edge paths. Their reasoning? Not only does the increase in path length beyond three make the number of paths more computationally expensive, but it also requires that intruders make more moves to get caught and therefore misses those who achieve their objective in fewer moves.

With so many three-edge paths to examine, PathScan needs to be very discriminating about which ones it considers suspicious. It does this by computing a probability for each path it observes on the network: the probability that the path in question should emerge at that given time, assuming that no cyber attack is actually underway. That is to say, under normal business conditions, what is the probability of a particular path occurring? If the probability is too slim, PathScan generates an alarm, and cyber security personnel investigate the potential intrusion.

How slim is too slim? That's open to debate. If the threshold probability is set too low, then very few anomalous paths will be reported, and it may be possible for an intruder to slip by. If it is set too high, that will create additional workload for analysts chasing down false alarms. Additionally, the optimal sensitivity setting may differ from one organization to the next, with a national security laboratory like Los Alamos choosing a relatively more conservative threshold, calibrated against historical network data collected during the past 10 years. Within that data, PathScan identified several sophisticated attacks. Such sophisticated attacks do not come along often, and PathScan has demonstrated exceptional reliability at isolating these exceedingly rare events within the mountains of data processed.

Path forward

Pilot programs for PathScan have been or are being conducted at a number of organizations other than Los Alamos, including the U.S. military and the oil and gas industry. And through a partnership with the Department of Homeland Security's Transition to Practice Program, another pilot program is set to begin at the Department of Veterans Affairs

in the coming months. Upon the successful conclusion of these pilot programs, the team intends to deploy PathScan more widely. But that won't be the end of the story. Cyber criminals present an ever-evolving threat, and cyber security efforts have to keep up. PathScan must continue to evolve as well, and Los Alamos remains the best place for that to happen.

"Los Alamos is a leader in network data collection," says Neil "I couldn't do this anywhere else. The data we track—bytes and packets sent and received, types of communications occurring—usually doesn't get comprehensively reported, even at very wealthy companies. The time has come for other enterprises to collect and analyze their network data more effectively, as we do."

To stay ahead of cyber crime, PathScan is learning to search for additional sorts of anomalous events captured by these network statistics. Imagine, for example, that the exact same number of bytes is sent from computer A to computer B, then C, then D. Even if the path ABCD doesn't register as anomalous, the constant byte count would. "If that transmission were legitimate, it would have gone straight from A to D," Neil says. His team is already testing prototype enhancements of this sort.

The next step will be training PathScan to figure out on its own what to look for, based on real-time network usage. Although anomalous three-edge paths are a strong indicator of malicious activity, under certain network conditions, other indicators might do even better. The team wants to create models of new and ingenious cyber attack methodologies and teach PathScan how to adapt and reprioritize its search objectives accordingly, on the fly. Neil is convinced this can be done and hopes that cyber security can outpace cyber crime as a result.

"Lately, you hear about successful cyber attacks on high-profile company networks, and you get the impression we're losing all these individual battles," he says. "But I think we can defend our networks and the intellectual gold they contain much better than we do now—and get some real cyber defense wins." **LDRD**

—Craig Tyler

The Los Alamos PathScan team, clockwise from bottom, is Josh Neil, Curtis Storlie, Curtis Hash, Mike Fisk, and Alexander Brugh.