




Looking for clues
in the code
when everyone is
trying to **GET IN**

Critical Error
 Los Alamos encounters tens of thousands of cyber attacks each day.
Close

http://www.lanl.gov/discover/publications/...catching-code-02.php

to guard their own information; they are also studying these types of threats to improve protection strategies for everyone. Analysts at Los Alamos manually evaluate APT malware sets on a continuous basis. The wealth of expertise they have garnered has helped the Laboratory establish a world-class research program that now develops tools for automated malware detection and characterization.

A malware analyst has multiple jobs to do: recognize malicious code entering the network, determine what the code is intended to do, and if possible, identify the source of the attack. The entire process is called reverse engineering (RE), and it can take days or weeks to accomplish. One key issue is that although some threats are familiar, emerging threats often prove more difficult to characterize.

Christine Anderson-Cook is a Los Alamos statistician who has been analyzing APT malware for a number of years. Her team focuses on initial screening—trying to identify and classify threats as they are detected. She explains that traditional commercially available antivirus software will not suffice for APT attacks because it functions by looking for an exact match between the malware code and a known code in the antivirus software’s library of threats.

“In an APT, the code is constantly evolving because it is associated with an active attack by a team of hackers,” says Anderson-Cook. “So we need to use statistical analysis to determine a probability-based match, instead of an exact one.” For a complex evolving threat, this strategy leads to

better detection and characterization of the entire threat landscape, such as what types of attacks are coming in and how many are related to known attacks or to each other.

Act Fast... Critical Error
 The worldwide cyber security market is projected to be a whopping \$170 billion in 2020.
Close



http://www.lanl.gov/...catching-code-01.php

THERE IS A REASON IT’S CALLED A VIRUS. Computer viruses, like biological viruses, cause damage to their hosts, spread between hosts, and modify and replicate themselves. And like natural viruses, computer viruses carry out these tasks by following instructions found in their “genomic” source code.

The first digital viruses to spread autonomously from computer to computer culminated in harmless messages displayed on the screen of the victim machine. One of the first, developed in 1982 by a 15-year-old boy, simply displayed a short poem, and another declared a “Universal Message of Peace.” Unfortunately, today’s malicious software, or malware, has evolved tremendously and is used in a broad spectrum of cyber attacks that are far from innocuous. One end of the spectrum includes attackers who, often for financial gain, target individual devices by deleting files, degrading system performance, or stealing personal information. On the other end of spectrum, sophisticated large-scale attacks (by groups of hackers) on specific organizations—such as Sony Pictures, the U.S. Office of Personnel Management, and the Democratic National Committee—have shown the potential for extensive, lasting damage through stealing trade secrets or confidential information.

This latter case, a coordinated attack on a specific target by a dedicated group of hackers, is called an advanced persistent threat (APT). To protect against APTs, companies and government institutions alike are spending billions annually to protect their own valuable data. But at Los Alamos, computer scientists are not only working

“Sometimes the new code is only slightly different,” says Juston Moore, a Los Alamos data scientist. Moore explains that quality software—even malicious software—is expensive to create, so many malware developers simply recycle existing code, making only those changes necessary to circumvent antivirus software. In this case, finding the small differences is key to understanding what distinguishes the new threat. Is it simply a cosmetic change to obfuscate, or hide, the code, or could the small change be a significant new strategy on the part of the attacker?

On the other hand, what stays the same is also an important signature. “Coders actually have a style, or voice,” says Anderson-Cook, comparing them to songwriters or playwrights whose word choices or patterns of prose make their work recognizable. For that reason, the code that has been conserved might give insight into the source of the threat. This broad analysis of APTs as a whole, especially with an interest in attribution, distinguishes Los Alamos researchers from other anti-malware efforts where the focus is largely just on blocking malware rather than studying it.

Because reverse engineering an evolving threat is so complicated and time consuming, Anderson-Cook and her colleagues have spent the last few years working on two algorithms to automate the RE process and essentially triage the threats coming in so the engineers can save their valuable time for the most difficult ones. The first algorithm searches the raw code, which is basically a long list of instructions. At a high level, these involve simple directions to respond to commands by an attacker, read or modify files, or open applications. However, certain patterns of these mundane instructions can be revealing. For instance, common sequences or co-occurrences of instructions can reveal a connection between malware families, and clues hidden in the predominant instruction can reveal the intent of the code.

The second algorithm looks at the next level of complexity: the patterns in subroutines, which are groups of instructions that together accomplish a specific function. While the first algorithm is analogous to observing patterns

in the way a writer organizes words a sentence, the second algorithm examines the more complex idea of how the writer organizes a whole story. Anderson-Cook explains that by combining these two algorithms,

Critical Error
On average, 416 days pass before a company knows it has been hacked.



her team has developed a unique ability to quickly compare and classify new malware either as a member of a known family of previously identified codes or as a brand-new threat.

Moore’s team has also developed a statistically guided RE toolset called REDUCE that can expedite the analysis process significantly by evaluating multiple pieces of malware at a time and identifying reoccurring patterns. These reoccurring patterns can be used to improve manual analyses and develop predictive signatures, useful in the detection of new variants of APT malware. Moore’s most recent work involves uncovering similarities in obfuscated malware code, an especially difficult task.

But even with these successes, Anderson-Cook and Moore won’t be letting their guard down anytime soon. Keeping up with rapidly adapting and innovative adversaries requires anticipating new types of threats and more sophisticated versions of existing ones. Moore explains that malware analysis won’t prevent all cyber attacks and that the future of cyber security might instead rely heavily on behavior analysis of an

already-infected machine rather than just screening for malware as it arrives. Just as physicians without access to modern lab results must attempt to identify viruses from a patient’s symptoms, cyber analysts of the future may need to track down malicious intruders by evaluating the symptoms of the computer’s illness instead of catching the code at all. **LDRD**

—Rebecca McDonald

Special Offer!



Critical Error
It’s not just personal computers that are at risk; critical national infrastructure systems are run by computers.