Technology | DOI:10.1145/1839676.1839683 Gary Anthes

Security in the Cloud

Cloud computing offers many advantages, but also involves security risks. Fortunately, researchers are devising some ingenious solutions.

"Computing may some day be organized as a public utility, just as the telephone system is a public utility," Massachusetts Institute of Technology (MIT) computer science pioneer John McCarthy noted in 1961.

We aren't quite there yet, but cloud computing brings us close. Clouds are all the rage today, promising convenience, elasticity, transparency, and economy. But with the many benefits come thorny issues of security and privacy.

The history of computing since the 1960s can be viewed as a continuous move toward ever greater specialization and distribution of computing resources. First we had mainframes, and security was fairly simple. Then we added minicomputers and desktop and laptop computers and client-server models, and it got more complicated. These computing paradigms gave way in turn to *n*-tier and grid computing and to various types of virtualization.

As hardware infrastructures grew more complicated and fragmented, so did the distribution of software and data. There seemed no end to the ways that users could split up their computing resources, and no end to the security problems that arose as a result. Part of the problem has been one of moving targets—just as one computing paradigm seemed solid, a new, more attractive one beckoned.

In a sense, cloud computing simplifies security issues for users by outsourcing them to another party, one that is presumed to be highly skilled at dealing with them. Cloud users may think they don't have to worry about the security of their software and data anymore, because they're in expert hands.

But such complacency is a mistake, say researchers at Hewlett-Packard (HP) Laboratories in Bristol, U.K. They are prototyping Cells as a Service, by which they hope to automate security management in the cloud. A cell, managed as a single administrative domain using common security policies, contains a bundle of virtual machines, storage volumes, and networks running across multiple physical machines. Around the cells HP inserts various sensors, detectors, and mitigators that look for viruses, intrusions, and other suspicious behavior. Virtualization enables these agents to be very close to the action without being part of it or observed by it, according to HP.

Cloud computing simplifies security issues for users by outsourcing them to companies such as Microsoft, which recently opened a $550 million data center in Chicago.

"People often think of virtualization as adding to security problems, but it is fundamentally the answer to a lot of those problems," says Martin Sadler, director of HP's Systems Security Lab. "You can do all sorts of things you can't do when these things are physical machines." For example, the sensors can watch CPU activity, I/O patterns, and memory usage and, based on models of past behavior, recognize suspicious activity. They can also assess the probability of certain events happening and take action accordingly. They might, for instance, throttle back the CPU, stop all I/O to a virtual machine (VM), or take a clone of the VM and move it elsewhere for evaluation. Agents could be deployed by cloud users, cloud service providers, or third parties such as a virus protection company, Sadler says.

But these agents introduce their own management challenges. There might be as many as 30 agents, interacting in various ways and with varying drains on system resources. HP Labs is developing analytic tools that can generate playbooks that script system behavior. These templates, tailorable by users, employ cost/benefit analyses and reflect what is most important to users and what cost they are willing to bear for various types of protection.

Virtual Machine Introspection

IBM Research is pursuing a similar approach called "virtual machine introspection." It puts security inside a protected VM running on the same

physical machine as the guest VMs running in the cloud. The security VM employs a number of protective methods, including the whitelisting and blacklisting of guest kernel functions. It can determine the operating system and version of the guest VM and can start monitoring a VM without any beginning assumption of its running state or integrity.

Instead of running 50 virus scanners on a machine with 50 guest VMs, virtual machine introspection uses just one, which is much more efficient, says Matthias Schunter, a researcher at IBM Research's Zurich lab. "Another big advantage is the VM can't do anything against the virus scan since it's not aware it's being scanned," he says.

Another variation, called "lie detection," puts a tiny piece of software inside the VM to look at the list of running processes as seen by the user. Introspection software outside the VM can reliably determine all the processes actually running on the VM; if there is any difference between the two lists, some malware, such as a rootkit, is suspected of running on the VM.

Looking from both within the VM and without, the lie detector can also compare the lists of files on disk, the views of open sockets, the lists of loaded kernel modules, and so on. "Each of these lie tests improves the chances of detecting potential malware, but none of them can prove that no malware exists," says IBM researcher Klaus Julisch.

In a third application, a virtual intrusion detection system runs inside the physical machine to monitor traffic among the guest VMs. The virtual networks hidden inside a physical machine are not visible to conventional detectors because the detectors usually reside in a separate machine, Schunter says.

Indeed, snooping between VMs inside a machine was shown to be a real possibility by researchers last year. Computer scientists Thomas Ristenpart, Hovav Shacham, and Stefan Savage at the University of California, San Diego and Eran Tromer at MIT proved it was possible for an adversary to get his or her VM co-located with a target's VM on a cloud's physical machine 40% of the time. In a paper, "Hey, You, Get Off of My Cloud," they showed how the adversary could launch a side-channel attack based on the VM's sharing of physical resources such as CPU data caches. The researchers also outlined a number of mitigation steps, but concluded the only practical and foolproof protection is for cloud users to require that their VMs run on dedicated machines, which is potentially a costly solution.

"People often think of virtualization as adding to security problems, but it is fundamentally the answer to a lot of those problems," says Martin Sadler, director of HP's Systems Security Lab.

Difficulties With Encryption

Encryption is sometimes seen as the ultimate security measure, but it also presents difficulties in the cloud. At present, processing encrypted data means downloading it and decrypting it for local use and then possibly uploading the results, which is a cumbersome and costly process.

The ability to process encrypted data in place has been a dream of cryptographers for years, but it is now demonstrating some progress. Last year, Craig Gentry, first at Stanford University and then at IBM Research, proved it is possible to perform certain operations on data without first decrypting it. The technique, called "fully homomorphic encryption," was hailed as a conceptual breakthrough, but is so computationally demanding that practical applications are years away, experts say.

Meanwhile, the more limited ability to search encrypted data is closer to reality. In "Cryptographic Cloud Stor-

Society

Pew Report on Mobile Apps

Although a greater number of adults are turning to mobile phones to text and access the Internet, age and gender differences exist, according to a report by Pew Research Center's Internet & American Life Project and The Nielsen Company.

The report, titled *The Rise of Apps Culture*, found that 35% of U.S. adults have software applications or apps on their phones, yet only 24% of adults use those apps. Overall, today's apps culture—essentially born a couple of years ago with the introduction of Apple's iPhone—is predominantly male, younger, and more affluent.

Eighteen to 29-year-olds comprise only 23% of the U.S. adult population but constitute 44% of the apps-using population. By contrast, 41% of the adult population is age 50 and older but this group makes up just 14% of apps users. Younger adopters also use apps, including games and social media, more frequently.

Gender differences were also apparent. Women are more likely to rely on social networking apps such as Facebook and Twitter while men are inclined to use productivity and financial apps.

Nevertheless, adoption is growing rapidly. The Nielsen Company found that the average number of apps on a smartphone has swelled from 22 in December 2009 to 27 today. Not surprisingly, iPhone owners top the list with an average of 40 apps, while Android users claim 25 and BlackBerry owners 14.

The next few years will likely usher in dramatic changes. "Every metric we capture shows a widening embrace of all kinds of apps by a widening population, states Roger Entner, coauthor of the report and senior vice president at Nielsen. "It's ... not too early to say that this is an important new part of the technology world."

—Samuel Greengard

age," a paper published earlier this year, researchers Seny Kamara and Kristin Lauter of Microsoft Research described a virtual private storage service that aims to provide the security of a private cloud and the cost savings of a public cloud. Data in the cloud remains encrypted, and hence protected from the cloud provider, court subpoenas, and the like. Users index their data, then upload the data and the index, which are both encrypted, to the cloud. As needed, users can generate tokens and credentials that control who has access to what data.

Given a token for a keyword, an authorized user can retrieve pointers to the encrypted files that contain the keyword, and then search for and download the desired data in encrypted form. Unauthorized observers can't know anything useful about the files or the keywords.

The experimental Microsoft service also offers users "proof of storage," a protocol by which a server can prove to a client that it did not tamper with its encrypted data. The client encodes the data before uploading it and can verify the data's integrity at will.

Not all cloud security risks arise from technology, says Radu Sion, a computer science professor at Stony Brook University. There is scant legal or regulatory framework, and few precedents, to deal with issues of liability among the parties in cloud arrangements, he notes. "What happens when your data is on a server in China but you outsourced to a cloud service in New York?" asks Sion. "Or what if you have the legal resources to fight a subpoena for your data, but they subpoena your cloud provider instead? *You* will be under scrutiny for moving to the cloud by your shareholders and everyone else."

In "Cryptographic Cloud Storage," Microsoft researchers Seny Kamara and Kristin Lauter describe a virtual private storage service that provides the security of a private cloud and the cost savings of a public cloud.

Nevertheless, Sion says all but the most sophisticated enterprises will be safer putting their computing resources in the expert hands of one of the major cloud providers. "Companies like Google and Amazon and Microsoft have hundreds of people devoted to security," he says. "How many do you have?"

Further Reading

Christodorescu, M., Sailer, R., Schales, D., Sgandurra, D., and Zamboni, D.
Cloud security is not (just) virtualization security, *Proceedings of the 2009 ACM Workshop on Cloud Computing Security*, Chicago, IL, Nov. 13, 2009.

Gentry, C.
Fully homomorphic encryption using ideal lattices, *Proceedings of the 41st Annual ACM Symposium on Theory of Computing*, Bethesda, MD, May 31–June 2, 2009.

Kamara, S. and Lauter, K.
Cryptographic cloud storage, *Proceedings of Financial Cryptography: Workshop on Real-Life Cryptographic Protocols and Standardization*, Tenerife, Canary Islands, Spain, January 25–28, 2010.

Ristanpart, T., Tromer, E., Sacham, H., and Savage, S.
Hey, you, get off of my cloud: exploring information leakage in third-party compute clouds, *Proceedings of the 16th ACM Conference on Computer and Communications Security*, Chicago, IL, Nov. 9–13, 2009.

Shi, E., Bethencourt, J., Chan, T-H., Song, D., and Perrig, A.
Multi-dimensional range query over encrypted data, *Computer Science Technical Report CMU-CS-06-135R*, Carnegie Mellon University, March 2007.

Gary Anthes is a technology writer and editor based in Arlington, VA.

Distributed Computing

Math at Web Speed

"Many hands make light work," goes the old adage. Now there's data to prove it.

In recent weeks, both Yahoo! and Google have announced the results of separate mathematical experiments that demonstrate the computational power of large clusters of networked PCs.

At Yahoo!, a team led by researcher Tsz-Wo Sze broke the world record for calculating the digits of pi, crunching the famously irrational number to the two-quadrillionth bit by stitching together more than 1,000 computers to complete the calculation over a 23-day period. The researchers estimate that a typical computer would have taken at least 500 years to carry out the same operation.

Another group of researchers recently took advantage of Google's distributed computing infrastructure to tackle another famously thorny computational challenge: Rubik's Cube. The team developed an algorithm capable of solving any Rubik's Cube configuration in 20 moves or less, resolving a conundrum that has puzzled mathematicians for three decades. The computers simulated all 43 quintillion possible combinations of the cube in just a few weeks, a task the researchers estimate would have taken a single computer 35 years.

Google has yet to release the details of its technical solution, but it probably bears some resemblance to the approach used at Yahoo!, where the team used Apache Hadoop, open-source software originally developed at Google (and later developed extensively by Yahoo!) that allows developers to stitch together thousands of computers over the network into a powerful cloud computer.

"We believe that our Hadoop clusters are already more powerful than many other supercomputers," says Sze, who conceived of the project as part of an internal Yahoo! contest to demonstrate the capabilities of Hadoop.

In both cases, the mathematical problems proved particularly well-suited to distributed computing because the calculations can be parceled out over the network into much smaller operations, capable of running on a standard-issue PC. Making light work indeed.

—Alex Wright