CYBERSECURITY LAW & STRATEGY

JUNE 2018

This is Not Your Father's Cloud

By Adam Cohen

Part One of this two-part article is aimed at demystifying the hesitations behind cloud security and analyzing the fast-growing transformation to a range of newer technical approaches with important consequences for legal practice.

Business lawyers interested in providing *informed* professional judgment, should focus on the *facts* about the cloud services to which their clients are migrating *en masse* — not legal theory based on an abstract concept of cloud computing. Services provided by tech titans Amazon (AWS) and Microsoft (Azure) are replacing traditional IT environments at an astonishing rate. Their pricing and delivery models are enabling widespread distribution of cutting-edge technologies and vast computing, storage and network resources. Lawyers need a solid understanding of these services and how their unstoppable global proliferation with clients everywhere impacts legal practice.

The urgency and importance of understanding these services should be obvious. Lawyers, bound to a holy duty of confidentiality, routinely exchange information with clients over enterprise IT systems. Litigators, as part of an adversarial process called E-Discovery, identify, preserve, collect, search, process, analyze, review and produce digital artifacts from client IT. Corporate attorneys depend on the client's financial, accounting and other data to form legal opinions. These days, critical client information is likely to reside in AWS or Azure.

New Cloud Order: AWS and Azure Lead the Market

Enterprise customers of AWS, Azure and competing platforms from other galactic tech giants such as Google Cloud, enter a magical world of super-technology buzzwords — *serverless*, *containerized*, *orchestrated* and *hyper-converged*, *software-defined* with "*infrastructure as code*." Lawyers may be disoriented by the jargon, unfamiliar contracts, seemingly infinite IT power and unique pricing structures, all of which raise legal questions about security, e-discovery, and regulatory compliance, among many others.

Instead of pouring money into hardware and software for data centers as tradition dictates, clients of AWS and Azure adopt a model where payment for IT resources is metered to use. Such resources include foundations of IT, such as storage or processing power ("compute"), delivered through a variety of services, each offering options, with corresponding cost, performance and security implications. Providers like AWS and Azure offer capabilities like artificial intelligence (*e.g.*, Alexa), device farms for testing (*e.g.*, for mobile app

developers) and streaming analytics on real-time data (*e.g.*, video game developers can adjust online gaming environments as play unfolds).

Within the massive cloud platforms, the tech titans curate a vibrant marketplace of offerings from independent providers (like an "app store" but enterprise cloud style). These providers promise to enhance the enterprise cloud experience. This eco-system demonstrates the natural, constantly accelerating evolution of technology, as each new development engenders multiple offspring.

Lawyers may or may not be aware that their e-discovery platforms are increasingly likely to reside in Azure. Relativity, the standard tool for large scale legal review, launched a subscription software-as-a-service (SaaS) product, RelativityOne, in 2017. RelativityOne is in Azure, neighboring many clients whose discovery data will now reside in proximity to components of their IT environment (*e.g.*, Office 365). Service providers offering Relativity have embraced the move, even those who invested in proprietary data centers — their infrastructure cannot match the scale of Microsoft.

The staggering scale of the infrastructure big providers are building, aggressive competition among these providers and improvements in systemic efficiency benefit customers in the form of rapidly decreasing prices, tied to precisely measured units of time (seconds), volume and other such unit pricing structures. AWS, created in 2006 and sprinting miles ahead of potential competitors from inception, by far dominates the market. However, AWS cannot rest on its laurels in the enterprise sector with Azure wielding the inherent advantage that the longstanding Windows empire provides. A note of caution, an increasing number of businesses routinely run critical applications in AWS and Azure, which if knocked out of operation for any reason could cripple unprepared operations. Many businesses also store their most sensitive customer data with these providers.

A Taste of Modern Enterprise Cloud IT

To get a small taste for how different modern IT environments based on enterprise cloud services are from traditional environments, consider the basic storage services offered by AWS and Azure. For these and other services, clients interact with them through application programming interfaces (APIs). The APIs allow systems of different types, which use different languages, to communicate and interact (*e.g.*, a mobile shopping app and a payment system regarding a customer/user order). APIs are not unique to the cloud; they are also used internally within IT environments to accomplish similar cross-system purposes. Given the central role of the API in the cloud and the fact that it is a point of information transfer between customer and provider, API security is particularly important.

The basic storage service in AWS is called S3 ("simple storage service"), while Azure's is plainly "Blob." S3 can be used for many different types of data and for a wide variety of different uses. It can handle virtually unlimited data and numbers of concurrent users.

However, unlike most systems lawyers are used to dealing with, S3 does not use files and folders to store data. Instead, "objects" uniquely identified by "keys" are stored in "buckets." While objects and buckets are analogous to files and folders, they are not the same and they work differently. This type of object storage is used in the cloud because it provides advantages in terms of performance, including scalability and ease of replication. However, the details are important in terms of security, as well as other critical functions for lawyers such as search and retrieval.

Objects can be up to 5TB in data and AWS states that an unlimited number of objects — an *infinite* amount of data — can be stored in a bucket. An object is made up of data and metadata. AWS states that the data

portion of the object is "opaque" to AWS. This presumably means that it is unreadable by AWS employees. The metadata includes system-generated, default metadata such as last modified data, but the user can also provide "custom" metadata when the object is stored in S3.

Azure Blob also stores "object" data. However, Microsoft defines a blob as a "*file* of any type and size" (emphasis added). In the language of Azure, blobs are stored in "containers" (not buckets) and come in three flavors optimized for different uses: "block" blobs for files like documents and media (each block blob can contain a maximum of 50,000 blocks, each not more than 100MB, which means the maximum total size is slightly above 4.75TB); "append" blobs for logging (maximum of 50,000 blocks of not more than 4MB, total size just over 195GB) and "page" blobs for efficiency where frequent read/write operations are performed, for example, Azure Virtual Machines uses page blobs as OS disks (up to 8TB maximum).

These are a few basic tidbits of information about a basic resource, storage. Already, there is a new vocabulary of concepts lawyers will have to grasp. Getting to the point of understanding the full range of storage services and all of the other *hundreds* of other types and variations of services, even for one of these providers (especially AWS) is a major undertaking for anyone. In this always shifting environment, tomorrow could bring another full slate of new information technology — so read faster!

Virtualization, Aggregation & Legal Risk

Certain legal risks arise directly from the fundamental characteristics of these services that enable their superpowers. The systemic efficiencies that enable delivery of cloud services start with the powerful magic of virtualization. Virtualization is neither new nor unique to the cloud, but the way the major cloud service providers (CSPs) have deployed is a far cry from its humble origins. Virtualization allows different systems, such as operating systems, to share the same hardware resources. A simple example of the traditional use of virtualization would be to run the Windows Operating System on your Apple computer. In the language of the cloud, running virtual machines from different customers on the same physical hardware is "multi-tenancy."

More recently, the technology known as containerization has exploded onto the enterprise IT cloud scene, offering even greater efficiency in wringing more from the same underlying resources. With containers, applications are packaged with all of the accouterments that allow them to run on different hosts, except without their own OSs. Instead, multiple containers share the resources of the underlying OS. This means digital "real estate" in terms of memory demand is reduced substantially in comparison with virtual machines. There are security implications to this arrangement, however, insofar as acquiring access to the OS potentially provides a way to access all of the containers sharing its resources. Virtual machines are protected by another layer of isolation, independent actors with their very own operating systems (and the higher volume of data to store that comes with them).

You can, of course, pay extra to keep other tenants of the landlord off your computerized lawn, but only by a step back towards the traditional data center where expensive resources sit idly waiting for a surge in activity. The major CSPs are masters at the most effective, automated shifting of "workloads" dynamically within their massive, global infrastructure to assure maximum efficiency. But this becomes optimized at the scale of the tech titan providers only where many customers, separated virtually, are co-habiting physically. You don't have to be a pessimist to see that multi-tenancy raises at least the concern, if not the practical possibility, of attacks that utilize access to shared resources to hop over virtual fences ("VM hopping" or "VM escape").

Virtualization and related containerization support the broader movement towards "software-defined everything" in enterprise IT Think about this as the "mind-body" relationship except for enterprise IT, where the mind is the software and the body is the hardware. According to a recent *Wired* magazine article:

> The movement towards a software-defined infrastructure is about decoupling the bare metal that executes the point data transactions from the software layer that orchestrates them. The hope is that by separating the smarts from the brawn, the underlying hardware can become cheaper and interchangeable (avoiding vendor lock-in) while the overarching software becomes more capable and faster-evolving …. Rather than individual elements (compute, storage, and networking), infrastructure will be treated as a set of resources required for specific workloads. In this world, the application, the end user, and hopefully the business are king.

Mat Mathews (Plexxi), "Are You Ready for Software-Defined Everything?" *Wired*.

The buzzword for the "cheaper and interchangeable" underlying hardware is hyper-converged. The use of virtual machines and containers, as well as the automated management of such resources, called orchestration, is what allows the "decoupling" the author describes.

More recently, "serverless" or "Function-as-a-Service (FaaS)" has burst onto the scene threatening to upend the new, incredibly young cloud order. Introduced by AWS with a service called Lambda and with characteristic Azure marketing flair as Functions, it enables a cloud customer to run code without managing server systems or applications. Of course, there is a server somewhere, but not in the way we are accustomed to as a dedicated and persistent machine standing by waiting to process requests. AWS, Azure and their ilk can spin servers up or down in milliseconds, which at least to the human eye doesn't look persistent (thus, "serverless").

On a technical level, getting this magic to work is of course more complicated. On a conceptual level for human language, components of the tasks the server performs are split up and distributed between the client and third parties like AWS or Azure, triggered and springing to life based on programming. The code is run and the value is returned. There is no need to worry about whether the server instance can handle the number of users; scaling is handled automatically behind the curtain and the customers are charged only for the time the code is running, not for the server to run it. There are infinite potential use cases, but examples typically involve a compute operation triggered by a condition or event, with unpredictable spikes in traffic. Get ready for serverless e-discovery.

Part two of this article will continue the examination of the security of platforms such as AWS and Azure, and what the notion of "sharing responsibility for security" should really mean.

*****

Adam Cohen, Esq. CISSP CEH CCSP, helps business clients practice Digital Discipline, harmonizing information technology with security and legal compliance. He is also an author cited in federal court opinions on electronic evidence, computer forensics and data privacy and has been teaching law school courses in e-discovery for more than a decade.