A Closer Look at Data Destruction in the Cloud

by Kenneth G. Hartman

Market research indicates that 93% of organizations have concerns about the security of the public cloud [1]. This healthy distrust likely stems from a lack of information. Cloud service providers know their customers; they understand these concerns and have developed a plethora of documentation and sales collateral to earn our trust. One very welcome documentation improvement by the leading cloud providers is the amount of transparency pertaining to data destruction. This article surveys this documentation to form a more complete picture of what exactly happens when we tell our cloud service provider to delete our data.

To frame the problem, let us take an inductive look at the question of *why do we care about data destruction?* Well, it is one of many security controls that the Cloud Security Alliance (cloudsecurityalliance.org) teaches us to review before engaging with a cloud service provider (CSP) [2]. When I teach about cloud security [3], I remind my audience that a given security control only mitigates *specific* risks and that in cybersecurity, there are no magic security bullets. Hence, it is essential to understand what a particular security control mitigates and what it does not.

Okay, but why is data destruction an important security control? Sensitive data must be destroyed when it is no longer needed to prevent unauthorized access to it. Until data is destroyed, it must be properly secured. How could an unauthorized person access sensitive information in the cloud that was not properly destroyed? They could:

- Use forensic tools to extract data from the cloud service provider's hard drives
- Encounter remnants of data from another tenant
- Use insider privileges at the CSP to access the data
- Recover sensitive data from a backup

Forensic examination of a hard drive

For many people, the first tactic that comes to mind to gain unauthorized access to sensitive information would be to obtain a hard disk somehow and use forensic tools to extract data from the drive. Now, let's break this down into its two parts — getting access to a hard disk and then extracting meaningful data from it.

Physical security of hard disks

The large, mature cloud service providers excel at physical security. The only personnel that have access to a CSP data center are the few people that have job duties inside the data center, and only a subset of those employees is responsible for the lifecycle of the hard disks. Hard disk drives (HDD) have a limited lifespan, and cloud service providers consume them by the thousands. The CSP uses software to track each HDD by serial number and accounts for its exact location at any point in time. When the drive has reached the end of its useful life, the cloud service provider will shred it or use a similar means of complete physical destruction. Independent audit firms closely scrutinize this process.

Data extraction

If an attacker is somehow able to obtain access to a physical hard drive, they may attempt to use various forensic techniques to extract sensitive data from the device. However, unlike the disk drive in your

laptop, each hard drive used by cloud service providers contains fragments of data (called shards [4]) from potentially hundreds of different tenants. Even if these fragments are not encrypted, it would be nearly impossible for an attacker to associate a fragment with a specific tenant. Note that I stated "nearly impossible" because the fragment could contain an identifying data element. Likewise, lacking the mapping information, it would be impossible for an attacker to identify all the drives for a specific target. I cover the benefits of encrypting customer data with tenant-specific encryption keys later in this article.

Data remnants from another tenant

Many of us have had the experience of renting an apartment only to find that the previous occupant left us with cleaning supplies, trash, and possibly even a lost diamond earring. We certainly do not want that to happen when we become a tenant in the cloud. Amazon Web Services, Microsoft Azure, and Google Cloud Platform have designed their cloud systems to prevent this from happening. The AWS Security Whitepaper [5] states:

> "When an object is deleted from Amazon S3, removal of the mapping from the public name to the object starts immediately, and is generally processed across the distributed system within several seconds. Once the mapping is removed, there is no remote access to the deleted object. The underlying storage area is then reclaimed for use by the system."
>
> "Amazon EBS volumes are presented to you as raw unformatted block devices that have been wiped prior to being made available for use. Wiping occurs immediately before reuse so that you can be assured that the wipe process completed."

```
sansforensics@siftworkstation -> ~
$ lsblk
NAME    MAJ:MIN RM  SIZE RO TYPE MOUNTPOINT
xvda    202:0    0  100G  0 disk
├─xvda1 202:1    0 80.3G  0 part /
├─xvda2 202:2    0    1K  0 part
└─xvda5 202:5    0 19.7G  0 part [SWAP]
xvdf    202:80   0    1G  0 disk
sansforensics@siftworkstation -> ~
$ sudo xxd /dev/xvdf | head
00000000: 0000 0000 0000 0000 0000 0000 0000 0000  ................
00000010: 0000 0000 0000 0000 0000 0000 0000 0000  ................
00000020: 0000 0000 0000 0000 0000 0000 0000 0000  ................
00000030: 0000 0000 0000 0000 0000 0000 0000 0000  ................
00000040: 0000 0000 0000 0000 0000 0000 0000 0000  ................
00000050: 0000 0000 0000 0000 0000 0000 0000 0000  ................
00000060: 0000 0000 0000 0000 0000 0000 0000 0000  ................
00000070: 0000 0000 0000 0000 0000 0000 0000 0000  ................
00000080: 0000 0000 0000 0000 0000 0000 0000 0000  ................
00000090: 0000 0000 0000 0000 0000 0000 0000 0000  ................
sansforensics@siftworkstation -> ~
$
```

Caption: Screenshot of a new EBS volume attached to a forensic workstation. The Linux 'lsblk' command shows the new, unformatted 1 gigabyte EBS volume is attached as `xvdf' while the 'xxd' command shows that the virtual device has been zeroed. For more info see https://forensicate.cloud/ws1/.

To continue with the apartment building analogy, it is as if the apartment is completely obliterated, (walls, floors, ceiling, and all) and the elevator (which provides the access control) will no longer stop at that floor. In the case of Amazon Elastic Block Storage (EBS), the data is not securely wiped until a new EBS volume is provisioned for a tenant and sized according to the cloud user specifications.

Some readers may be initially concerned that Amazon waits to wipe the data until it is re-provision for a new user, that is most efficient and preserves the life of the solid-state hard drives. Also, do not make the inaccurate assumption that an EBS volume is hosted on a single physical hard drive. The AWS documentation states, "Amazon EBS volume data is replicated across multiple servers in an Availability Zone to prevent the loss of data from the failure of any single component" [6].

Microsoft Azure makes a similar statement:

> The sectors on the disk associated with the deleted data become immediately available for reuse and are overwritten when the associated storage block is reused for storing other data. The time to overwrite varies depending on disk utilization and activity but is rarely more than two days. This is consistent with the operation of a log-structured file system. Azure Storage interfaces do not permit direct disk reads, mitigating the risk of another customer (or even the same customer) from accessing the deleted data before it is overwritten [7].

I appreciate the additional information provided by Microsoft in the above blog excerpt because this is the type of disclosure we need from our cloud service providers. I applaud the relatively recent acknowledgments that data is not wiped until it is provisioned by for a new customer and appreciate the additional context provided by Azure that reminds us that these highly optimized resources are overwritten naturally within days due to high utilization. It should be noted that GCP [8] also uses log-structured file systems [9]. I would like to see all cloud service providers provide additional technical details of these systems along with the relevant security implications. Given the fact that the cloud service provider maintains strict physical security over its hard drives, my professional belief is that this data handling is acceptable for any classifications of data suitable for storage in the public cloud.

Insider Privileges

Cloud customers expect their data to be protected throughout its lifecycle until the data is destroyed and can no longer be accessed. The previous sections have dealt with the protections in place to protect customer data from external parties prior to its destruction, but what about protecting data from trusted insiders? AWS, Azure, and GCP have security documentation [5, 10, 11] that covers the applicable security controls, including background checks, separation of duties, supervision, privileged access monitoring, etc.

The primary concern with insider threats is that employees and contractors have detailed system knowledge and access to lower-level systems that are not exposed to public cloud customers. The CERT National Insider Threat Center has detailed guidance [12], and cloud customers should explore what

controls are in place to protect data that has been deleted and is still pending destruction. As thoughtful technical customers ask probing questions of their cloud service providers, en masse, the best cloud providers listen and respond with increasingly transparent documentation.

Encryption is a security control that can mitigate unauthorized insider access when appropriately applied. Unfortunately, encryption is often used as a Jedi mind trick. Some customers stop asking the tough questions once they hear that the service uses encryption. Encryption is a technique to control access. The person or system that controls the encryption key controls the access. For example, with transparent database encryption, the database management system controls the key and therefore controls the access. A database administrator (DBA) can query the data in the clear, but the administrator of the storage system that the database uses can only see the ciphertext. On the other hand, if the application controls the key, both the DBA and the storage system administrator can only see ciphertext.

Cryptographic erasure is the name of the technique wherein the only copies of the encryption keys are destroyed, thereby rendering the encrypted data unrecoverable. NIST Special Publication 800-88 recognizes cryptographic erasure as a valid data destruction technique within certain parameters that are readily enforced in modern public cloud environments [13].

The Azure documentation states that encryption is enabled for all storage accounts and cannot be disabled [14]—the same for Google [15]. However, in AWS, it is a configuration option for services like S3 [16] and EBS [17].

Unfortunately, the security white papers from AWS and Azure fail to tout the benefits of the cryptographic erasure technique even though they are using it to destroy customer data. Also, it is often unclear when a CSP is using a tenant-specific encryption key to perform encryption at rest for their various services. When a tenant-specific encryption key is used in conjunction with cryptographic erasure, only the data belonging to a single tenant is destroyed. Cryptographic erasure is a very attractive alternative to overwriting data, especially for customers with hundreds of petabytes of data in cloud storage, like Netflix [18].

Recover Data from Backups

The last attack vector involves an adversary attempting to recover sensitive data from a backup. I always caution my clients to not assume that a CSP is backing up your data unless the contract clearly specifies it. Unless stated otherwise, cloud service providers are primarily using backups or snapshot techniques to meet service level agreements regarding data durability and availability.

If data is being backed up, the backup must be protected with at least the same level of security as the primary data store. Among the top three cloud service providers, Google's white paper titled "Data deletion on Google Cloud Platform [19]" provides the most transparency concerning how deleted data must expire and be rotated out throughout its 180-day regimen of daily / weekly / monthly backup cycles. To Google's credit, this document even covers the important role of cryptographic erasure in protecting the data until it expires from all the backups.

https://youtu.be/WB6xrgI9eGk

Video Caption: Data Deletion on the Google Cloud Platform

Conclusion

Without a doubt, the top three cloud service providers have expended great effort to make their system secure. All cloud service providers must balance the need to protect against leaking too much information that would aid an adversary while providing enough transparency to maintain the trust of their customers. As cloud customers dialogue with their cloud service providers and seek the appropriate information necessary to make intelligent risk decisions, the cloud service providers are improving their messaging that explains their security investments. Data destruction is one topic that had been inadequately discussed until recently, even by the top three CSPs. Regardless of which of the hundreds of cloud service providers you use, this review of the top three providers' documentation on data destruction should equip you to have an even more informed due diligence process.

---

[1] https://www.cybersecurity-insiders.com/portfolio/2019-cloud-security-report-isc2/
[2] https://cloudsecurityalliance.org/research/cloud-controls-matrix/
[3] https://www.sans.org/profiles/kenneth-g-hartman/
[4] https://docs.microsoft.com/en-us/azure/architecture/patterns/sharding
[5] http://d0.awsstatic.com/whitepapers/Security/AWS%20Security%20Whitepaper.pdf
[6] https://aws.amazon.com/ebs/features/#Amazon_EBS_availability_and_durability
[7] https://devblogs.microsoft.com/azuregov/data-security-qa-with-john-molesky-azure-security-engineering/
[8] https://cloud.google.com/security/deletion
[9] https://web.stanford.edu/~ouster/cgi-bin/papers/lfs.pdf
[10] https://download.microsoft.com/download/1/6/0/160216AA-8445-480B-B60F-5C8EC8067FCA/WindowsAzure-SecurityPrivacyCompliance.pdf
[11] https://services.google.com/fh/files/misc/security_whitepapers_march2018.pdf
[12] https://resources.sei.cmu.edu/library/asset-view.cfm?assetid=540644
[13] https://nvlpubs.nist.gov/nistpubs/SpecialPublications/NIST.SP.800-88r1.pdf
[14] https://docs.microsoft.com/en-ie/azure/storage/common/storage-service-encryption
[15] https://cloud.google.com/storage/docs/encryption
[16] https://docs.aws.amazon.com/AmazonS3/latest/dev/bucket-encryption.html
[17] https://docs.aws.amazon.com/AWSEC2/latest/UserGuide/EBSEncryption.html
[18] https://www.datanami.com/2018/04/30/how-netflix-optimized-flink-for-massive-scale-on-aws/
[19] https://cloud.google.com/security/deletion