TAG CYBER

ANALYST REPORT: INTRODUCING CLOUD DATA FRAGMENTATION (CDF)

EDWARD AMOROSO
TAG CYBER

TAGCYBER

ANALYST REPORT: INTRODUCING CLOUD DATA FRAGMENTATION (CDF)

EDWARD AMOROSO

The resiliency of cloud-hosted data can be optimized through a fragmentation process, resulting in a distributed representation that is effective against cyber threats such as insider attacks from public cloud hosting teams. Prominent commercial vendors supporting this capability are included in the report.

INTRODUCTION

Enterprise data security is best achieved in the context of the organizational mission. For example, banks that handle sensitive customer records must implement controls that prevent this data from being read by unauthorized parties. In contrast, industrial factories that depend on stored machine telemetry and configuration files require controls that prevent such data from being tampered with or altered.

One data security requirement that aligns with many different organizational objectives involves the need to ensure the resiliency of cloud-stored data. Companies that export sensitive or critical files and records to public cloud or SaaS services now require that such data be protected from disclosure, integrity, or blocking threats to the third-party infrastructure. These threats could result from external attacks or cloud service insiders with administrative access.

In this report, we introduce a new category of commercial data security method known as *cloud data fragmentation (CDF)*. The protection strategy that underlies modern CDF platforms and their practical implementation involves breaking up data into discrete components that can be separated, processed, and stored in diverse locations. The approach is well-suited to scattering fragments or shards of data into multiple public clouds.

The CDF concept is introduced and explained below from the perspective of an enterprise security practitioner team with responsibility to protect important data being used across hybrid infrastructure, including use of the major public cloud services. Several commercial vendors are also listed and shown to provide good options for IT and security teams looking to purchase a platform that can be integrated into their local cyber risk framework.

WHAT IS DATA RESILIENCY?

Stated simply, data resiliency references how well your data holds up to cyber threats. For many years, the term would include the caveat that resiliency is all about dealing with integrity threats, but with new techniques such as CDF, resiliency can be extended to the disclosure threat as well. This is a profound advance because it addresses the oft-claimed challenge that once viewed, data confidentiality cannot be restored.

The idea behind data resiliency in a CDF context is that distributed objects are more difficult to attack than combined ones. A common non-technical analogy illustrating the concept is that it is much easier to knock over an elephant with a large truck than it is to knock over a swarm of mosquitos. The truck might hit a small subset of the bugs, but no matter what it does, the majority will evade being targeted. This visual image can be extended to data resiliency.

Specifically, data stored in public clouds can be distributed across different storage domains to improve its resiliency. The algorithms for breaking up the stored data must obviously be designed to support not only the distribution of the data, but also reassembly when the data is needed. In addition, it must provide support for the various tasks that might be performed on the individual pieces including encryption.

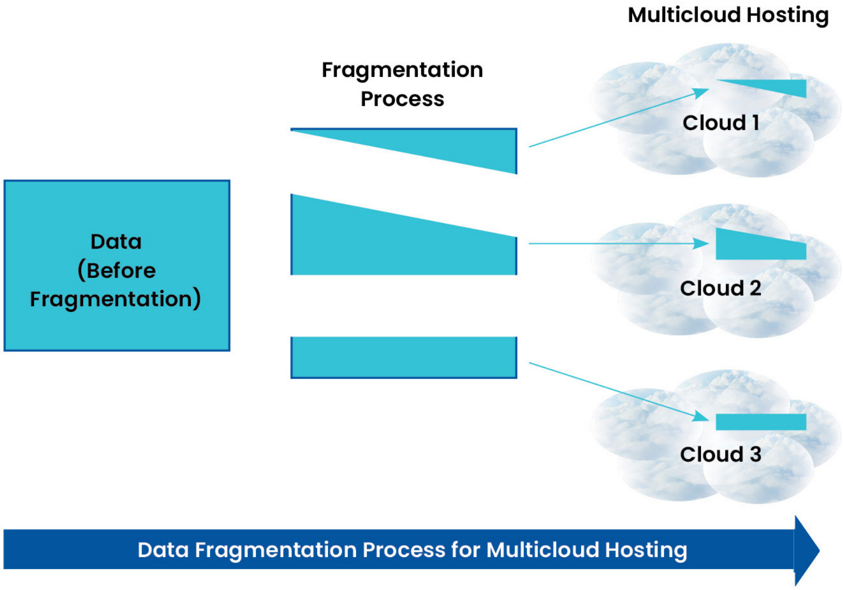
Multicloud Hosting
Fragmentation Process
Data (Before Fragmentation)
Cloud 1
Cloud 2
Cloud 3
Data Fragmentation Process for Multicloud Hosting

Figure 1. Distributing Data to Multiple Clouds to Improve Resiliency

The general method of distributing data has been in existence for many years, and infrastructure providers have considerable experience fragmenting data for the purpose of data storage control and optimization. It is only recently, however, that commercial vendors have emerged that allow for such distribution to be performed at the application level via CDF into multiple clouds to address cyber threats.

WHAT IS THE INSIDER RISK FOR CLOUD-STORED DATA?

When sensitive data for some significant enterprise or agency is being hosted into a public cloud service, this engagement would be guided by a data hosting agreement. For example, if some large bank decides to use Amazon Web Services (AWS) for hosting of its sensitive data, the respective security teams from the bank and the cloud provider would work through the details of an agreed-upon set of controls for the cloud environment.

It is important to note that, in stark contrast, any mid-sized or smaller company using a public cloud service will have to accept whatever commercial terms and conditions are published by the provider. These might be acceptable or not, but the customer will have little or no options other than comparing the services of another provider. Microsoft and Google, for example, will not negotiate terms of its data hosting with anyone other than a massive buyer.

As a result, the risk emerges that a customer's cloud hosted data might be mishandled either accidentally or deliberately by an administrator working on behalf of the provider. Intentional compromise of hosted data might come from a disgruntled or coerced person with privilege to access hosted data. This is a non-trivial risk because insiders are tough to identify and cloud service providers offer a natural means for such bad actors to access valuable data.

To that end, cybersecurity controls, whether procedural, policy-based, or functional, must be put in place to avoid such insider risk. Typical methods are listed below, along with their respective pros and cons:

- *Multi-Person Controls* – By requiring multiple individuals to participate in the approval that some data-related action be taken, the risk of an insider is reduced to those cases where sufficient collusion is present.
- *User Behavioral Analytics* – Monitoring the behavior of insiders provides a means for highlighting activity that does not match their normal profile and that prompts deeper review and analysis for potential data misuse.
- *Cloud Data Fragmentation* – Breaking up data into fragments and scattering the pieces removes the possibility of an individual cloud service administrator with privileged access having the means to gain unauthorized access to hosted data.

Obviously, all methods for reducing insider risk are recommended, but it should be evident that the use of multi-person controls and user behavioral analytics has not been sufficient to reduce data risk in most environments. For this reason, the new technique introduced here known as cloud data fragmentation (CDF) represents a high-priority functional control that can complement these more mature means for addressing cyber risk.

HOW DOES CDF WORK?

The technique of cloud data fragmentation follows the general technical strategy outlined above for distributing data. At a high level, some data element, usually represented as a file, artifact, binary, or other resource, is subjected to a process in which it is fragmented into pieces. This process is also sometimes referred to as a data sharding activity, where the sharded pieces are derived from the whole using some algorithm.

While it would seem obvious, the fragmentation process must in fact represent a function whereby the original data can be reconstructed or interpreted from its pieces. Furthermore, the process must not involve data loss as is found in some compression algorithms that optimize storage for large files such as multimedia. Instead, fragmentation must preserve the essential properties of the original file while also creating shards for separate placement.

The steps involved in most commercial implementation of CDF include options for the IT, cloud, or security team to select based on local requirements. As such, a stepwise conditional methodology emerges for CDF that includes the following commonly found decision steps:

Step 1: Data Fragmentation
In this first step, an algorithm is used to break up the data into fragments. Different vendors will decide on the size of the individual shards (e.g., four bytes), and the use of compression in this step is a common approach. As standards emerge for CDF, one would expect more commonality in the algorithmic paths taken in this step.

Step 2: Data Packaging
The second step involves packaging the individual fragmented shards using an algorithm that might introduce salted or poisoned data. While such adjustment of the data helps with confidentiality, it is usually not a full encryption step. This step can, however, include native encryption as part of the packaging.

Step 3: Data Distribution
The third step allows for distribution of the fragments into whatever targeted infrastructure is desired by the user. This might include cloud storage services, hybrid cloud, or even legacy storage infrastructure. Configuration will be handled by the user, in conjunction with the public cloud services being used.

Step 4: Data Access
This step involves accessing the fragmented data via collection, unpackaging, and combination of the stored shards. It is a design decision whether the pieces are ever arranged back into an aggregate physical object or are just maintained separately for access. Any over-the-top encryption that might have been imposed on the fragments must be addressed here as well.

Certainly, the methodology associated with CDF will vary between enterprise teams (e.g., some adding additional encryption) and also across the major cloud service providers or legacy IT hosting team. In addition, the CDF vendor will play a prominent role obviously, and in the next section, we outline some prominent cybersecurity vendors, mostly start-ups, that offer acceptable capability in this area.

WHO ARE THE PROMINENT VENDORS SUPPORTING CDF?

As part of our research at TAG Cyber, we interview cybersecurity vendors including start-ups, and then create customized strength, weakness, opportunity, and threat (SWOT) analyses based on our investigation. This information is integrated into our Research as a Service (RaaS) portal which is used by enterprise teams, vendors, and investors around the world to inform their own source selection of the best vendors for their specific mission.

Using this research base, we offer below a brief summary of the vendors we found to be offering solutions consistent with our introduction of CDF in this report. Considerable additional detail on these vendors is available to TAG Cyber RaaS customers including tailored support for specific questions or challenges being addressed by an enterprise or other team. The summaries below should thus be used as a starting point for CDF planning.

Calamu™

Business Information: Paul Lewis serves as founder and CEO of Calamu. Prominent advisors include John Stewart, former SVP Chief Security & Trust Officer for Cisco. Headquartered in New Jersey, the company recently raised a $16.5M Series A round of funding.

Brief Solution Description: The Calamu Protect platform supports securing data at rest in a highly scalable and resilient geo-fragmented cloud environment call a data harbor. Calamu offers an agent called Calamu Drive for Windows and MacOS which provides transparent protection for user endpoints. Calamu Drive monitors file activity, performing encryption and fragmentation as needed, and then storing the protected fragments in the data harbor. Calamu Connectors provide API-based integration with premise and cloud-based workloads and databases, providing the same encryption, fragmentation, and storage features as the endpoint agents. Finally, the Calamu Console provides administrative access to the data harbor to enable management of storage and users and to monitor the health of agents and API's.

COHESITY

Business Information: Mohit Aron serves as Founder and CEO of Cohesity. Investors include Sequoia Capital, Wing Ventures, SoftBank, and others. Headquartered in San Jose and founded in 2013, the company completed a Series D round of $250M in funding in 2018.

Brief Solution Description: The Cohesity Helios solution protects enterprise data with machine learning-based backup and data deduplication through four vertical solutions. Cohesity Helios SaaS provides consolidated archive for premise and cloud-based enterprise data. The Data Management as a Service solution offers data backup, consolidation, disaster recovery, and governance across storage platforms. Cohesity DataProtect supports autonomous backups of critical workloads with SLA support within the workflow toolset, thus providing global search and restore, storage space efficiency, and ransomware protection. Cohesity SmartFiles offers native integration with NAS, SMB, and S3 file storage solutions, enabling near real-time file redundancy. The Cohesity SiteContinuity supports disaster.

Leonovus

Business Information: Michael John Gaffney serves as Chair and CEO of Leonovus. Headquartered in Ottawa and trading on the Toronto Stock Exchange and TSX Venture Exchange (TSXV: LTV), the company was founded in 2010.

Brief Solution Description: Leonovus Smart Filer automatically moves infrequently used data to less expensive cloud storage, thus leaving a symbolic link in its place. Smart Filer leverages any combination of cloud storage offerings and provides access to archived files as if locally stored. The Leonovus Vault provides a FIPS140-2 encrypted certified storage vault. The Vault shreds the encrypted vault container and distributes the data fragments across multiple clouds to ensure data sovereignty and confidentiality. The Leonovus XVault is a byproduct of the Vault, but one focused on secure file sharing. XVault allows enterprise teams to store files in the same highly secure manner as the Vault, but allows them to grant specific users access, thus providing a secure Dropbox mechanism.

myota™

Business Information: Steve Wray serves as CEO of Myota. Founded in 2017 and headquartered in Pennsylvania, the company raised $3.65M in funding in 2021 in a Series A2 round led by investor Ira Lubart.

Brief Solution Description: The Myota solution is a cloud-only file management and data governance platform that chunks, encrypts, and shards critical files to provide both security and redundancy. Any file that is uploaded to the service is chunked, encrypted, and sharded within the Myota infrastructure, with a link provided back to the end-user. Key features supported by Myota includes protection of unstructured data with encryption, shredding of data to render them unusable by third-party administrators or insiders, spreading of data to various repositories to avoid extortion and ransomware, and enablement of data recovery in the event of an attack targeting the resiliency of critical enterprise data. The solution integrates with anti- malware platforms, data leakage prevention solutions, and cloud storage.

SHARDSECURE

Business Information: Bob Lam serves as CEO and co-founder of ShardSecure. Prominent technologist and ShardSecure co-founder Lou Steinberg, former CTO of TD Ameritrade, serves as chairman of the company. The company closed an $11M Series A funding round in May, 2022, led by Grotech Ventures, Gula Tech Adventures, Tom Noonan, and KPMG in addition to EPIC Ventures and Industrifonden. (Full disclosure notice: Dr. Edward Amoroso participates as an unpaid advisor to Lou Steinberg's firm which funds and advises ShardSecure.)

Brief Solution Description: ShardSecure offers on-premises and cloud-based solutions that shred, mix, and distribute enterprise cloud data in multi-cloud and hybrid-cloud environments. The solution's self-healing data offers protection of data from cloud-based ransomware attacks and other data breaches. Once a file has been microsharded, the solution will rebuild that file if any unauthorized attempts of modification or deletion are detected, neutralizing cloud storage ransomware attacks and bolstering business continuity. Additional use cases include secure cold storage migration to the cloud, Microsoft 365 privacy and security, and secure cloud adoption.

STORJ

Business Information: Ben Golub serves as CEO of Storj. Founded in 2015 by Shawn Wilkinson, who also serves as Chief Strategy Officer, the Atlanta-based company has over 70 employees working in many different countries. Storj is an Ethereum token.

Brief Solution Description: Storj is an open-source software development platform that enables developers to encrypt, shard, and distribute objects across public or private cloud containers. The solution is primarily intended for video streaming, software distribution, decentralized backups, and integration into cloud-native apps. The Storj software development kit (SDK) consists of three primary components, the Uplink CLI, the S3-compatible Gateway, and the Storj client Libraries. Additionally, several pre-built integrations have been developed for Fastly, FileZilla, MongoDB, Duplicati, and Filebase. Many different offers exist for developers to obtain low-cost access to cloud storage.

ACTION PLAN FOR ENTERPRISE

Enterprise teams considering CDF for sensitive data are advised to create a local initiative to begin planning. While plans will vary between different groups, we recommend that the following steps be included in whatever process is being used to integrate CDF into existing and planned sensitive data storage architectures and associated cloud hosting processes.

Step 1: Data Storage Inventory
Developing an accurate view of the data being stored in legacy, hybrid, and cloud infrastructure is an obvious first step in any plan to integrate CDF. This should include third-party storage and should separately identify SaaS-based data, which is usually not in scope for present CDF solutions.

Step 2: Data Security Policy
Creation of policies for data marking and categorization is important because not all data will likely be subjected to the fragmentation process. Without clear prioritization policies, CDF will be unevenly applied in an ad hoc manner. Instead, it should be a required step for data that rises above some criticality threshold.

Step 3: Vendor Selection and Test
The selection of vendors should be informed by the list included in this report, but enterprise teams might have other options, including even developing the method internally. As always, TAG Cyber analysts are available to assist buyers in this important source selection process, usually leading to a proof-of-concept (POC) test.

ABOUT TAG CYBER

TAG Cyber is a trusted cyber security research analyst firm, providing unbiased industry insights and recommendations to security solution providers and Fortune 100 enterprises. Founded in 2016 by Dr. Edward Amoroso, former SVP/CSO of AT&T, the company bucks the trend of pay-for-play research by offering in-depth research, market analysis, consulting, and personalized content based on hundreds of engagements with clients and non-clients alike—all from a former practitioner perspective.

IMPORTANT INFORMATION ABOUT THIS PAPER

Contributor: Edward Amoroso

Publisher: TAG Cyber LLC. ("TAG Cyber"), TAG Cyber, LLC, 45 Broadway, Suite 1250, New York, NY 10006.

Inquiries: Please contact Lester Goodman, (lgoodman@tag-cyber.com), if you'd like to discuss this report. We will respond promptly.

Citations: This paper can be cited by accredited press and analysts but must be cited in context, displaying the author's name, author's title, and "TAG Cyber". Non-press and non-analysts must receive prior written permission from TAG Cyber for any citations.

Disclosures: This paper was commissioned by ShardSecure©. TAG Cyber provides research, analysis, and advisory services to many cybersecurity firms mentioned in this paper. No employees at the firm hold any equity positions with any companies cited in this document.

Disclaimer: The information presented in this document is for informational purposes only and may contain technical inaccuracies, omissions, and typographical errors. TAG Cyber disclaims all warranties as to the accuracy, completeness, or adequacy of such information and shall have no liability for errors, omissions, or inadequacies in such information. This document consists of the opinions of TAG Cyber's analysts and should not be construed as statements of fact. The opinions expressed herein are subject to change without notice.

TAG Cyber may provide forecasts and forward-looking statements as directional indicators and not as precise predictions of future events. While our forecasts and forward-looking statements represent our current judgment and opinion on what the future holds, they are subject to risks and uncertainties that could cause actual results to differ materially. You are cautioned not to place undue reliance on these forecasts and forward-looking statements, which reflect our opinions only as of the date of publication for this document. Please keep in mind that we are not obligating ourselves to revise or publicly release the results of any revision to these forecasts and forward-looking statements considering new information or future events.