CARBONITE® + WEBROOT®
opentext™ Business Solutions

Carbonite® Endpoint

Global deduplication across encrypted data

Introduction

Current data deduplication techniques fall into two broad categories depending on where the deduplication process takes place.

- Client-side (source-side) deduplication occurs at the source (where the data is created and stored).
- Target-side (post-process) deduplication takes places on the server (after the data has already been transported to its archival storage location).

While both forms of deduplication generally provide the same level of storage savings, client-side deduplication provides additional efficiencies through reductions in network bandwidth consumption. (See figure 1 on p.2)

Carbonite Endpoint's client-side data deduplication process goes one step further and provides additional benefits with an enhanced security model that keeps data safe and encrypted during the deduplication process.

Figure 1: Target-side deduplication vs. Client-side deduplication

Problem statement

Data deduplication is implemented through the use of a data block analysis algorithm that seeks to eliminate duplicate data blocks (also referred to as “chunks”) across an entire data store. Before data is backed-up, the process identifies chunks that already appear in the target vault. Once duplicates have been eliminated, the resulting data stream is reduced in size, resulting in reduced transfer times and reduced storage requirements.

The backup data footprint can be further optimized by implementing the data deduplication scope across multiple data vaults (or “stores”) depending upon specific business compliance and security requirements. However, encrypted data can present a problem.

Due to its very nature, data that is encrypted using different encryption keys results in data that cannot be deduplicated as the resultant encrypted data is unique. Consequently, to benefit from deduplication, data must be processed in an unencrypted format, which poses some security concerns.

Previous options

Limited by their inability to cope with encrypted data, traditional data deduplication techniques have forced IT departments to pit security needs against storage and networking budgets. An enterprise that wished to take advantage of the benefits offered by data deduplication would have had to forgo the level of security and privacy offered by data encryption.

One solution to the problem is to deploy channel encryption, which protects all communication between the client and server by encrypting the data path. At a minimum, this is required if no source data encryption is performed, or if any “secrets” – such as data encryption keys – are passed between the client and the server.

Another solution is to deploy transient data encryption, where the original data on the client is encrypted using symmetric or asymmetric methods. Encrypted data will then be decrypted on the server before it is deduplicated and subsequently archived.

Data could be deduplicated on the client side before being encrypted and uploaded, but in order to deduplicate data across multiple clients, the server would need to be able to decrypt the data for processing.

Transient data encryption alleviates the need for an encrypted channel architecture, but the technique includes a potential pitfall. To implement a transient data encryption scheme, the server needs to have access to and control of the encryption key and the process for storing encrypted data on disk.

This means that if the server’s security is breached, the security and privacy of data could also be compromised.

Carbonite Endpoint Solution

As one of the core pillars of the Carbonite Endpoint solution, addressing the end-to-end security and privacy of data is a primary requirement. By utilizing our automated key management and encryption technology in conjunction with our unique data deduplication technology, Carbonite Endpoint solves the problem of deduplicating encrypted data.

Scoping

The Carbonite Endpoint platform allows for granular access control to stored blocks for a particular data source (vault). Access can be scoped at the “vault” level down to the “device” level, where data from a source configured with highly restrictive scoping rules will only be deduplicated with itself.

Scope is set in policy and it can be applied at any level. You can create a new policy with specific scope and assign the policy to just one device, or assign it to all devices in a user group so it's scoped to the entire group.

In situations where a Managed Service Provider (MSP) is using shared infrastructure for multiple enterprises within

the same server infrastructure, they can set different default policies with unique scoping rules for each company to allow for explicit data separation between organizational boundaries in a shared environment.

Encryption

After scoping rules have been applied to a data block, a unique Block Encryption Key is deterministically generated.

This key is then used to encrypt the block using AES 256-bit encryption. The result is an encrypted data block.

The block encryption process ends after each data block has been encrypted. And as a final step, the Block Encryption Key is then itself encrypted and any clear text representation of the key is removed from the system.

Deduplication

As all processing thus far has been deterministic, data from any data source within the same scope will coalesce to identical encrypted data blocks. The data is then deduplicated so that only a single instance of any particular block is archived on the data store.

Following data duplication, each file can be represented by a simple index that associates a list of unique data blocks required with their order of arrangement and the Block Encryption Key required to completely reassemble an instance of the original data.

Benefits

Client-side deduplication of encrypted data
Every data source will maintain its own unique index of its data, but will share all of the encrypted data blocks (subject to scoping rules).

Security and privacy
All data, including all of the metadata about user files, is encrypted using the same process. So even an administrator with physical access to the backend server hardware cannot read archived data. In fact, an administrator would not even be able to determine what data resides in the archive.

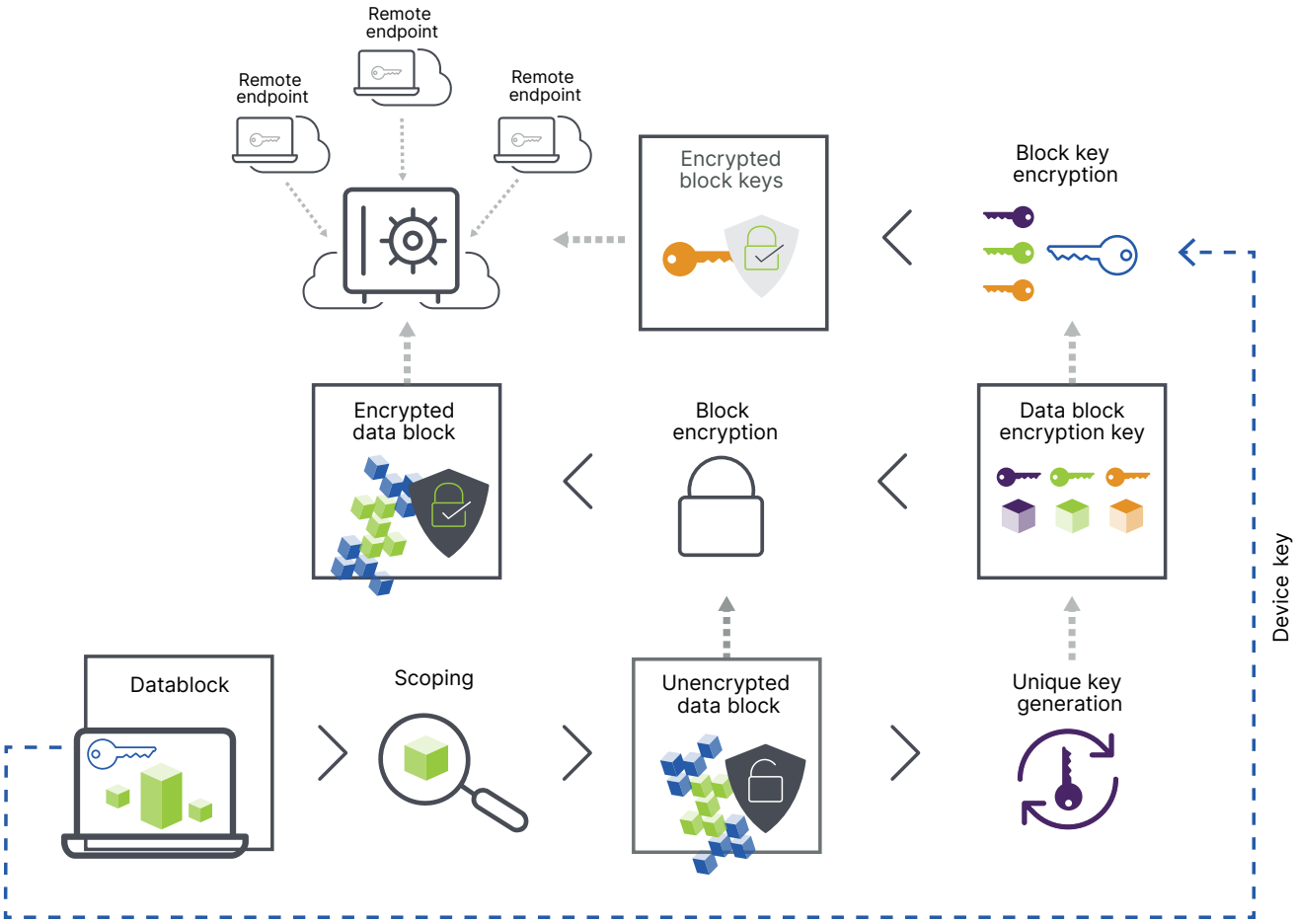

Figure 2: Carbonite Endpoint global deduplication of encrypted data

Scalability

With all the segmentation, encryption and deduplication that needs to take place, the scale of any storage solution is important. Based on the Carbonite Endpoint next generation data deduplication design, all of the heavy lifting is pushed out to the edge of the network.

Each client has an agent that runs as a low priority background task and utilizes spare CPU cycles. Automatic network bandwidth shaping and disk I/O throttling are combined with efficient CPU utilization to provide minimal impact on the client.

As data is opaque to the server environment, there can be no further analysis of the data once it has been dispatched by the client. This design provides for a more robust and scalable backend.

Carbonite Endpoint is a proven solution that has already been deployed in multiple public cloud environments as well as within on-premises data centers.

Multi-tenancy

For MSPs and larger enterprises, the ability to securely segment and share the same physical resources between disparate organizational entities (enterprises, subsidiaries, departments, etc.) is important to the economics of any data deduplication solution.

Carbonite Endpoint allows transient server resources to be allocated between different groups of devices. This allows the cost of a Carbonite Endpoint infrastructure to be amortized across all participating devices or for different SLAs to be made available to various classes of devices based on business need.

At the storage layer, physical segmentation and logical segmentation is available. For example, regulatory compliance may dictate that data from certain sources must be stored on physically separate storage. In most cases, however, logical separation (implemented with scoping rules) may suffice.

Through the Carbonite Endpoint storage routing engine, commodity storage nodes can be combined into a highly scalable, available and reliable storage network.

No compromises

With Carbonite Endpoint there's no need to choose between economic benefits derived from the storage-saving features of data deduplication and all the other benefits listed. You get them all.

Other forms of data deduplication technology require a choice between decryption on the server (which compromises security, privacy, multi-tenancy, etc.) or deduplication that is limited to data from individual data sources (instead of data across the enterprise).

Summary

Carbonite Endpoint is the only solution on the market that allows encrypted data to be deduplicated. Carbonite Endpoint provides the full economic benefit of data deduplication to be achieved across the enterprise without sacrificing data security or privacy. As growth in storage requirements continues to increase, new techniques in data deduplication are continuing to be sought after to help address this growth.

Contact us to learn more – Carbonite US

Phone: 877-542-8637

Email: carb-data_protection_sales@opentext.com

About Carbonite and Webroot

Carbonite and Webroot, OpenText companies, harness the cloud and artificial intelligence to provide comprehensive cyber resilience solutions for businesses, individuals, and managed service providers. Cyber resilience means being able to stay up and running, even in the face of cyberattacks and data loss. That's why we've combined forces to provide endpoint protection, network protection, security awareness training, and data backup and disaster recovery solutions, as well as threat intelligence services used by market leading technology providers worldwide. Leveraging the power of machine learning to protect millions of businesses and individuals, we secure the connected world. Carbonite and Webroot operate globally across North America, Europe, Australia, and Asia. Discover cyber resilience at carbonite.com and webroot.com.