Cloud Security Using Information Dispersal

Julie Bellanca

Jason Resch

Cleversafe

September 2010

Agenda

- Challenges in Securing Data in the Cloud
- Approaches for securing data
  - Encryption
  - Secret sharing
  - Advanced secret sharing
- Secure Information Dispersal – an approach for advanced secret sharing
- Leveraging Information Dispersal for Cloud

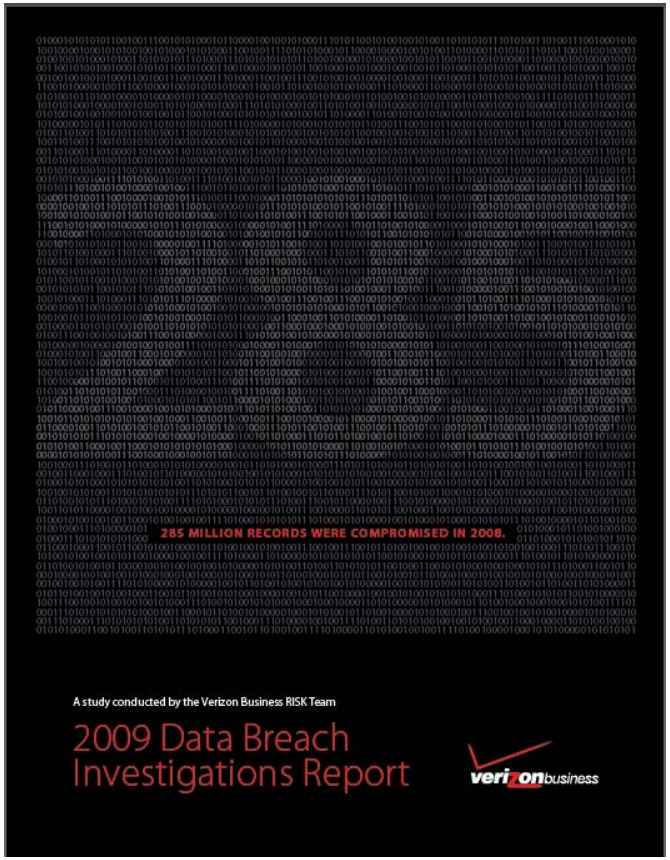
Challenges for Securing Cloud Data

- Cloud computing:
  - Offered as a service
  - Network enabled
  - Shared resources
  - Dynamically scalable
- New approach presents new challenges
- Industry says "Encrypt it"
  - If end user is responsible for encryption…
    - End user takes on burden of key management
  - If cloud provider is responsible for encryption…
    - How are they protecting the encryption keys?
- Industry standard of RAID & replication for data protection increases risk of data exposure

Verizon's Security Breach Analysis

"The best defense against data breaches is, in theory, quite simple – don't retain data."

- Verizon Investigative Response Team



One of Verizon's recommendations:
Clearly, knowing what information is present within the organization, its purpose within the business model, where it flows, and where it resides is foundational to its protection. Where not necessitated by valid business needs, a strong effort should be made to ***minimize the retention and replication of data****.*

900+ Breaches and 900 Million Records compromised 2003-2009

Assessing Security – CIA model

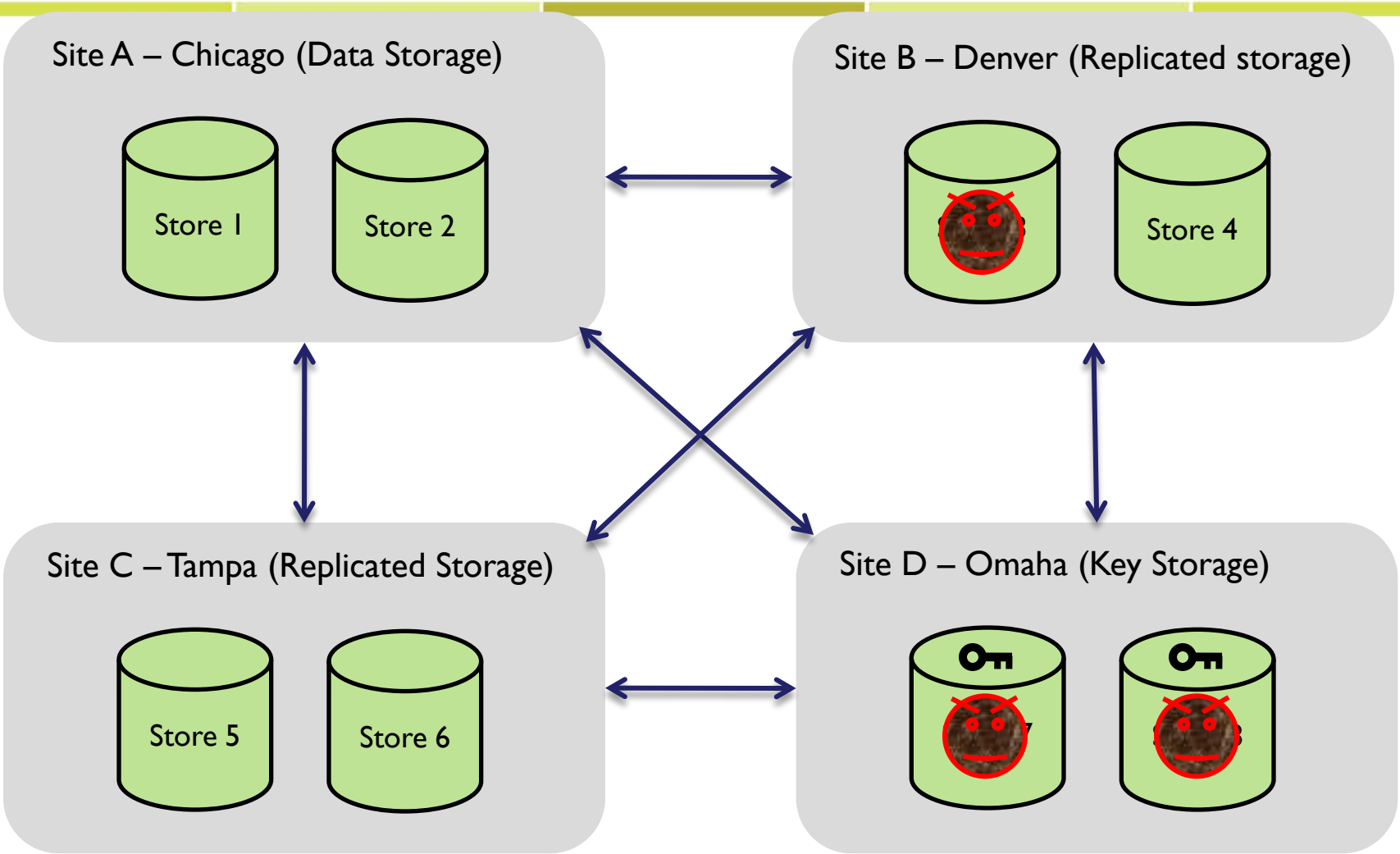
Objectives	Requirements	Example Threats
Confidentiality	Data is never accessed by unauthorized parties	• Key or credential mismanagement. • Accidental loss of media or devices. • Malicious access. • Remote compromise or theft. • Interception of packets.
Integrity	Data cannot be modified without authorization	• Bit errors in drives, memory, connections, or flash. • Physical read and write errors. • Accidental data corruption. • Malicious data tampering.
Availability	Data is always available to authorized parties	• Drive, location, server, and connection failures. • Maintenance operations. • Denial of service attacks.

Inherent conflict between Availability and Confidentiality

Approaches for data security

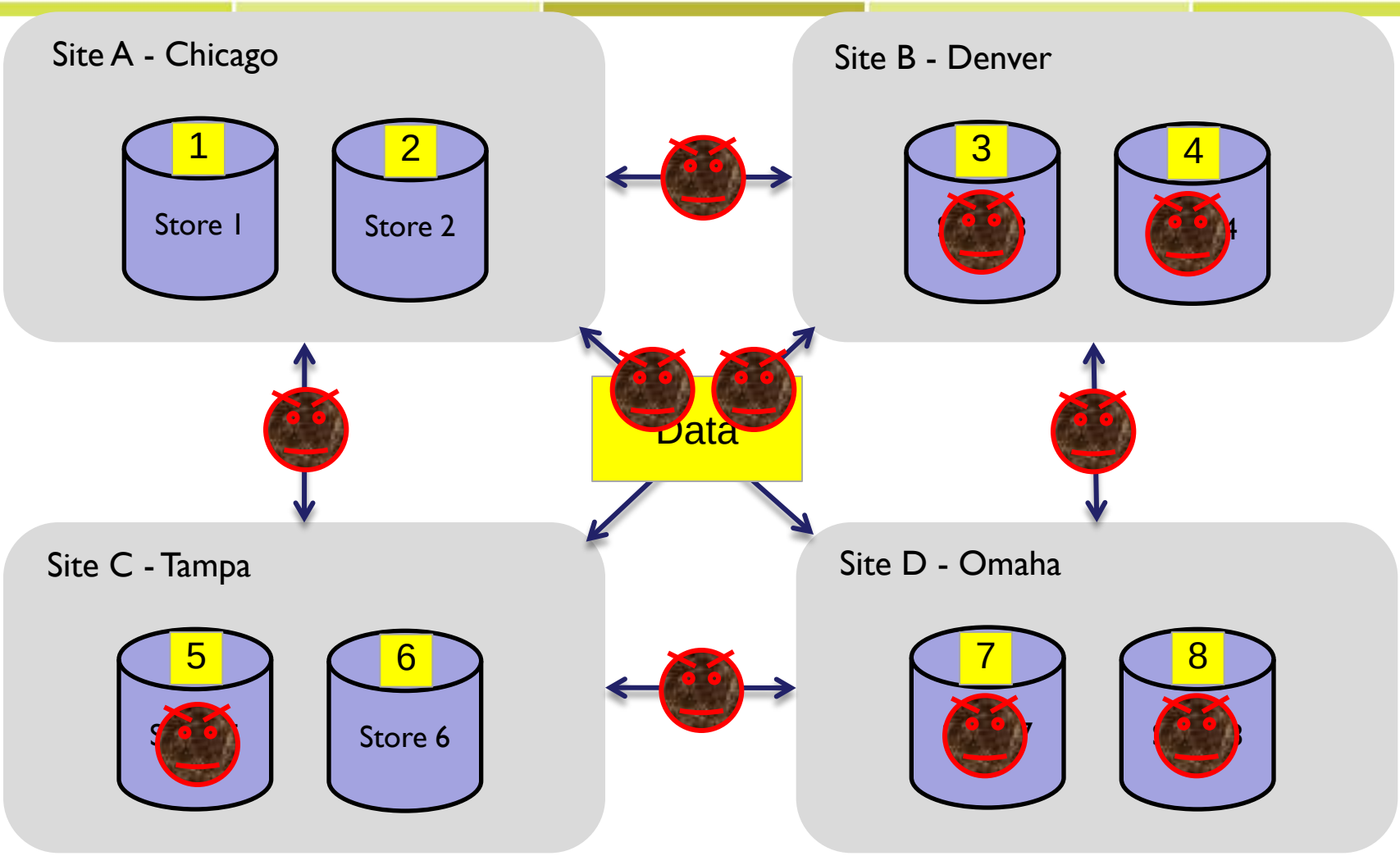
- Keyed encryption
  - Key and the encrypted data recreates data (2-of-2)
- Secret Sharing Scheme
  - Method for distributing a secret amongst a group of participants, each of which is allocated a *share* of the secret
  - The secret can be reconstructed only when a sufficient number of shares (threshold) are combined together; individual shares are of no use on their own
  - Each share is the size of original data
    - 10 of 16 with 1 PB of data = 16 PB raw storage required
- Advanced Secret Sharing Schemes
  - Benefits of secret sharing with lower cost overhead
  - Each share (slice) is 1/threshold the size of the data
    - Example: 10 of 16 with 1 PB of data = 1.6 PB raw storage required

Keyed Encryption & Replication CIA



Confidentiality Lost | Integrity Lost | Availability Lost

Secret Sharing CIA



Confidentiality Lost | Integrity Lost | Availability Lost

8 shares, 5 required

Secret Sharing vs Keyed Encryption

Secret Sharing Scheme	Keyed Encryption System
Each segment of data is uniquely protected.	One key may protect a large amount of data.
If someone leaves the organization, or if a credential is exposed: **permissions can be changed instantly** to restore security.	If someone with access or a key leaves the organization, many TB or perhaps **PB would have to be re-encrypted** with a new key.
There are no keys to be lost. Multiple simultaneous losses may occur yet **data remains recoverable**.	**If key is lost, so is the data.** Keys must be stored as reliably as the data. This requires storing copies in multiple locations.
Organization is in control of data. It is not possible for someone to leave with the key.	Individuals may be able to **walk off with keys** or passwords necessary to recover the data.
Secret sharing schemes don't sacrifice confidentially for reliability. They achieve high levels for each. **No critical locations exist.**	Every copy of the key or data is a critical location which must not be compromised. **Each copy is another attack vector** and opportunity to be hacked or exposed.

Encryption Versus Secret Sharing

System	# to Compromise Confidentiality	# to Compromise Integrity	# to Compromise Availability
Encryption & Replication	Always 2	Always 2	All Copies Or Key
	2	*2*	*3 or lost Key (1)*
Secret Sharing	Threshold	Threshold	1 + Width – Threshold
	10	*10*	*7*

- **Keyed Encryption + Replication -** 3 data copies, and 1 encryption key
  - Limited Confidentiality and Integrity
    - Attacker only needs to compromise two locations (key and data)
  - Increasing availability hurts confidentiality
    - May make copies of keys or data, but this increases attack vectors
- **Secret Sharing -** 16 shares, 10 needed (10 of 16)
  - Offers arbitrarily high levels of confidentiality, integrity and availability

Storage efficiency

System	Storage Overhead	Drives Needed / PB	Drive Cost / PB
Encryption & Replication	275%	1,920	$576,000
Traditional Secret Sharing	1500%	8,192	$2,457,600
Advanced Secret Sharing	60%	820	$246,000

Model parameters:

- Encryption & Replication - 3 data copies, RAID 5 (4+1)
- Traditional Secret Sharing - 16 shares, 10 needed (10 of 16)
- Advanced Secret Sharing - 16 shares, 10 needed (10 of 16)
- 2 TB drives
- $300 / drive

Secure Information Dispersal Approach

- One approach to accomplish advanced secret sharing is to combine two established algorithms:

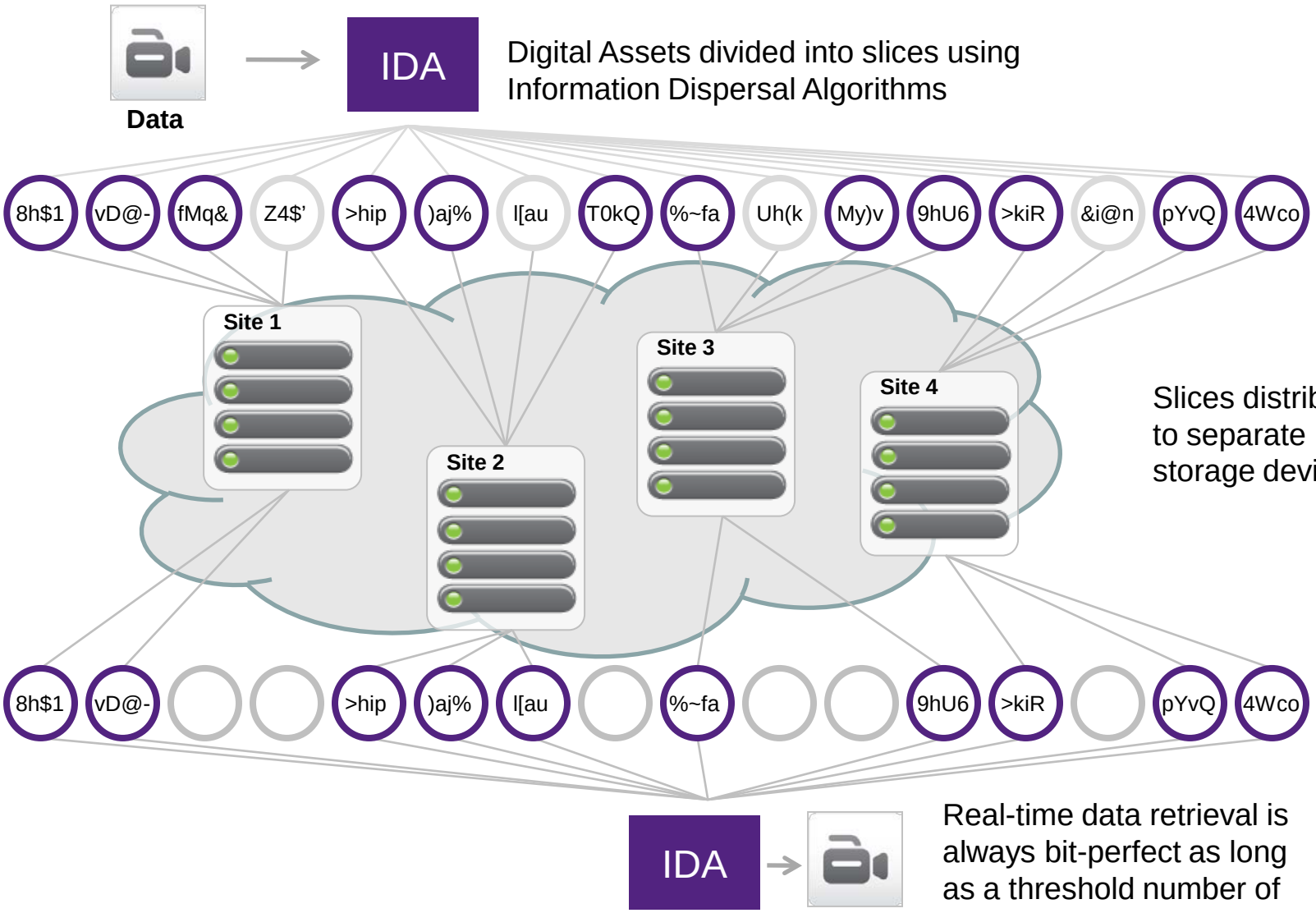
Availability & Reliability

- **Information Dispersal Algorithms (IDAs)**
- Forward error correction (AKA Reed Solomon) that forms data into n segments where k are needed to recreate data (k of n)
- Store segments on separate storage nodes to increase availability & reliability

Security

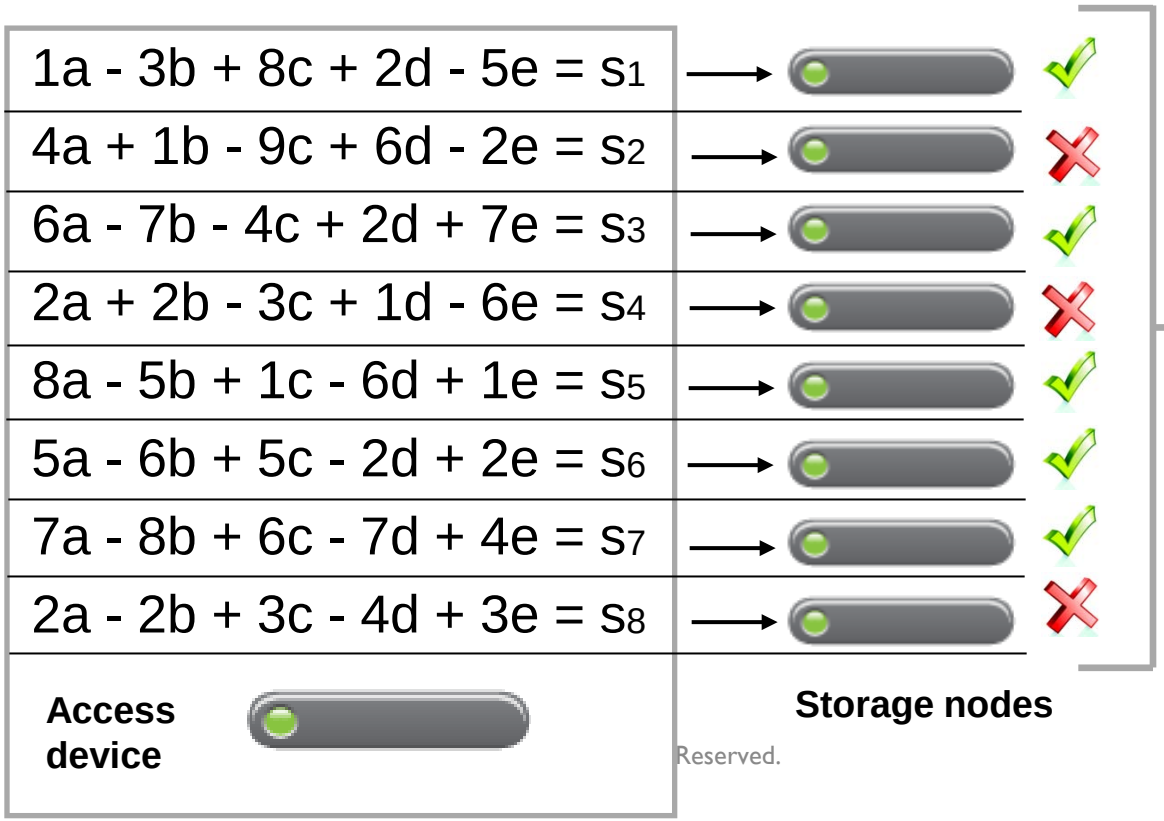
- **All Or Nothing Transform**
- Encryption mode which allows data to be understood only if all of it is known

Information Dispersal Introduction



Information Dispersal Math

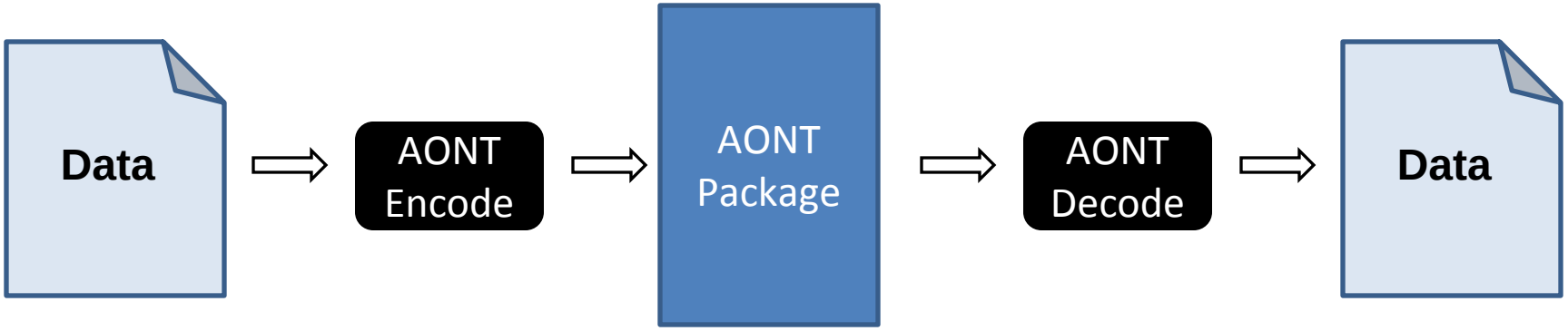
- Reed-Solomon is Linear Algebra - solving a system of equations
- Reed-Solomon as an FEC code
- Perfectly efficient in storage space
- Supports any desired fault tolerance
- Example encoding, k=5, n=8



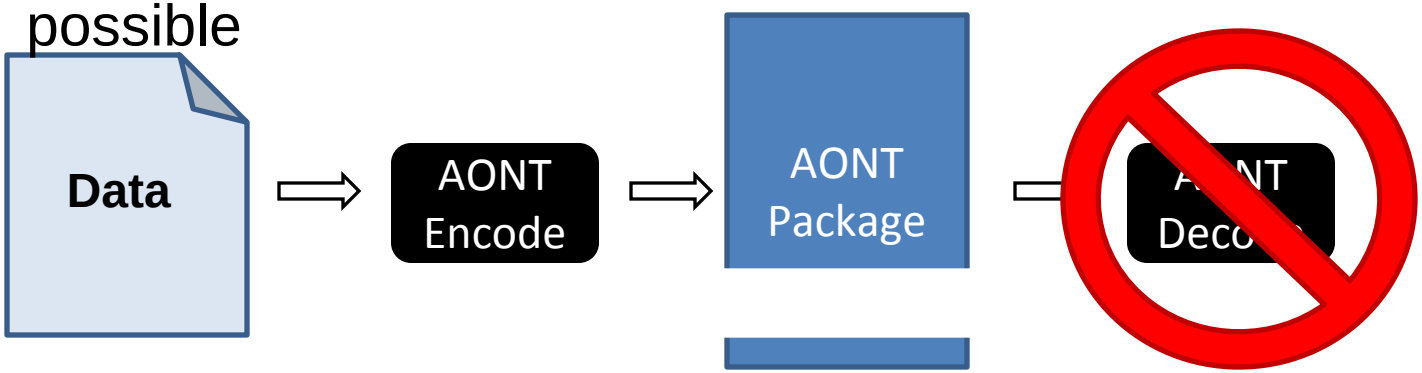
- Solving for ***k*** variables requires knowing ***k*** solutions (slices)
- Therefore we can lose ***(n-k)*** solutions (slices)
- Data overhead on the disk and on the wire is ***(n/k)*** (n slices stored, for input variables)

AONT Introduction

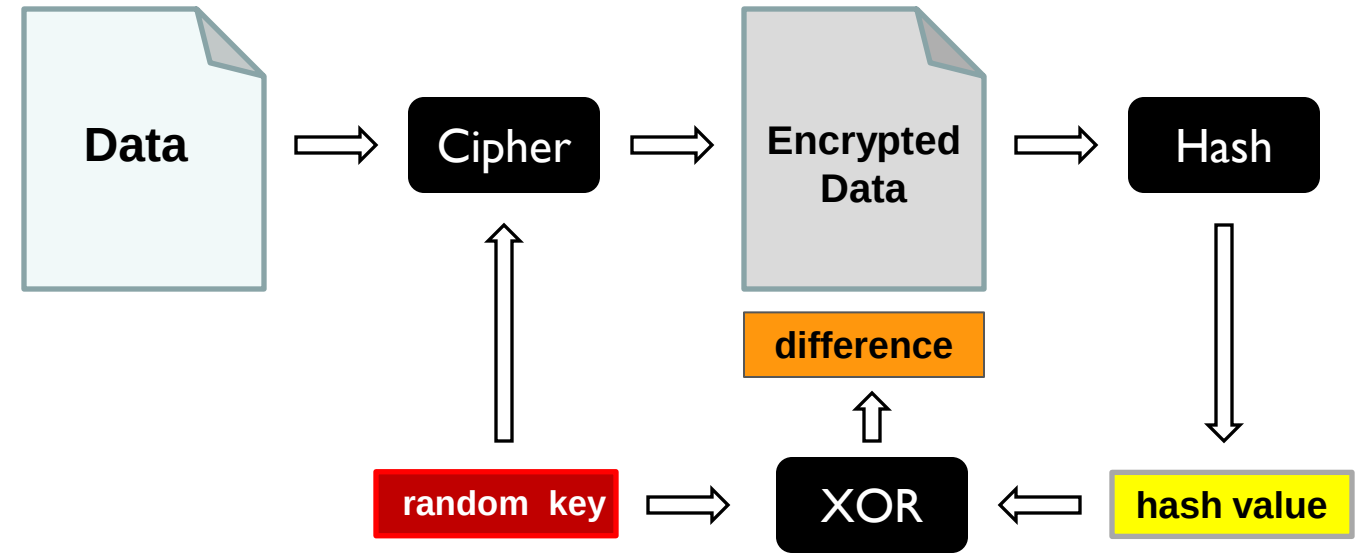
When one has the entire AONT package, decoding is trivial

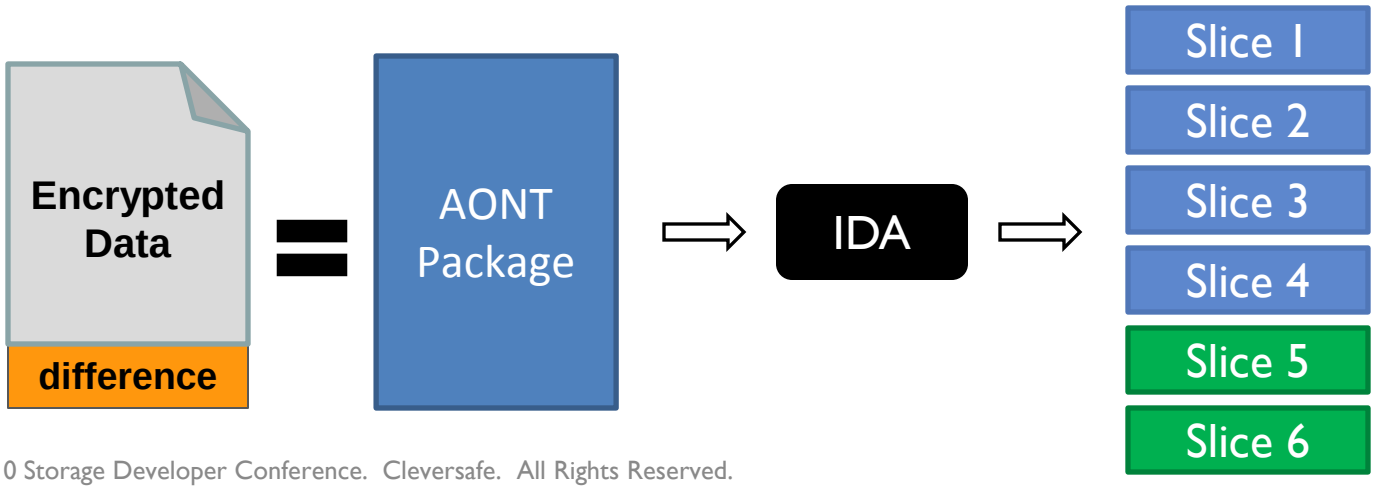


If any part of the package is unknown or missing, decoding is not possible

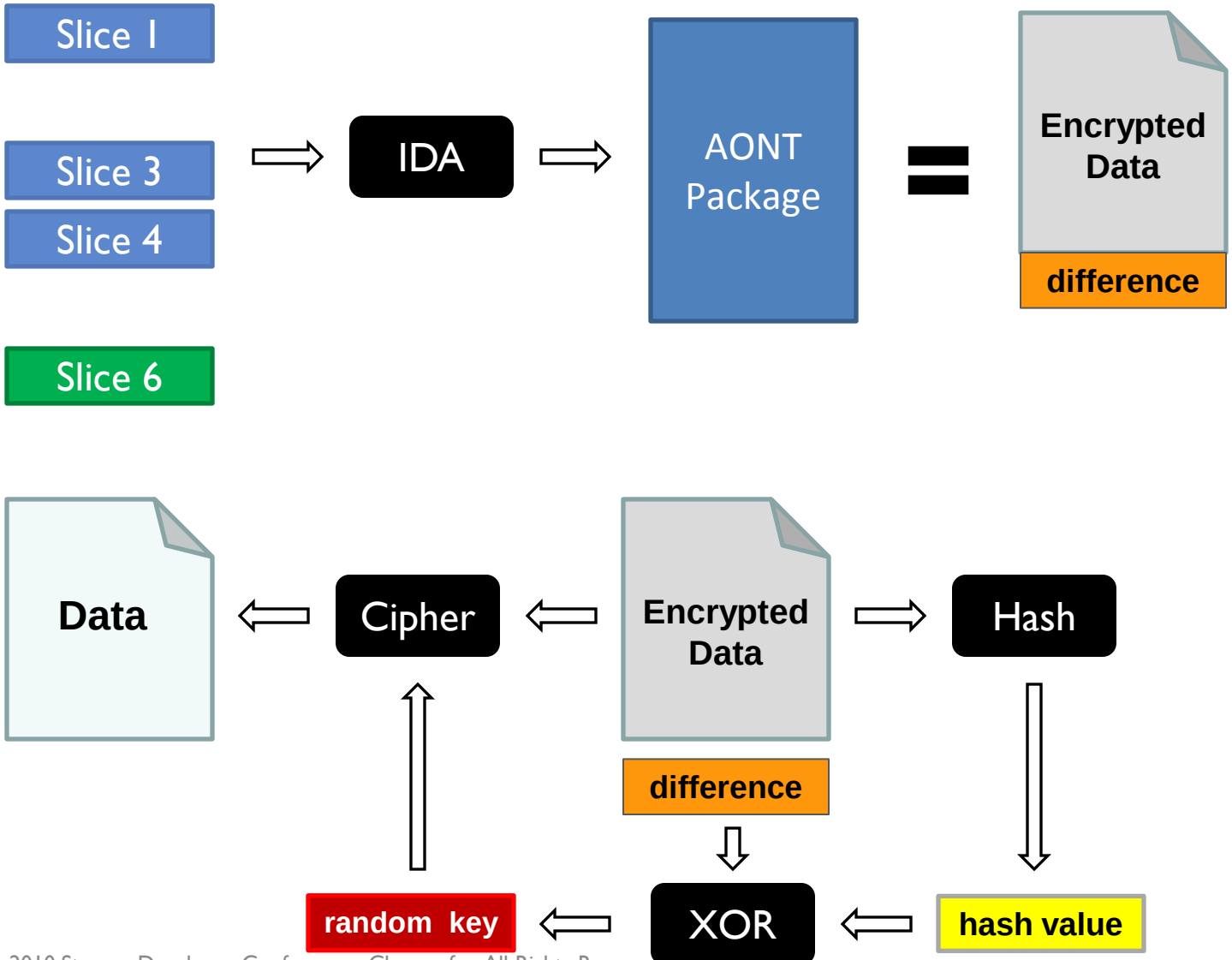


Combining AONT & IDAs

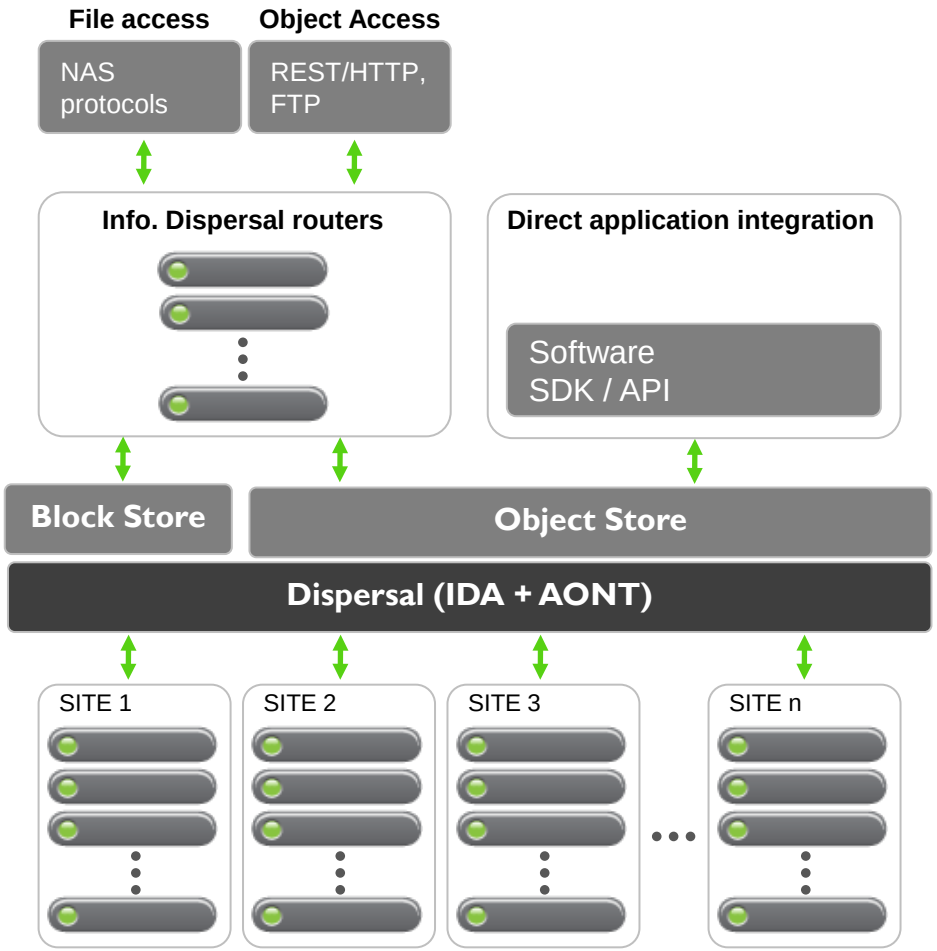




AONT & IDA Reconstruction



Secure Information Dispersal Architecture



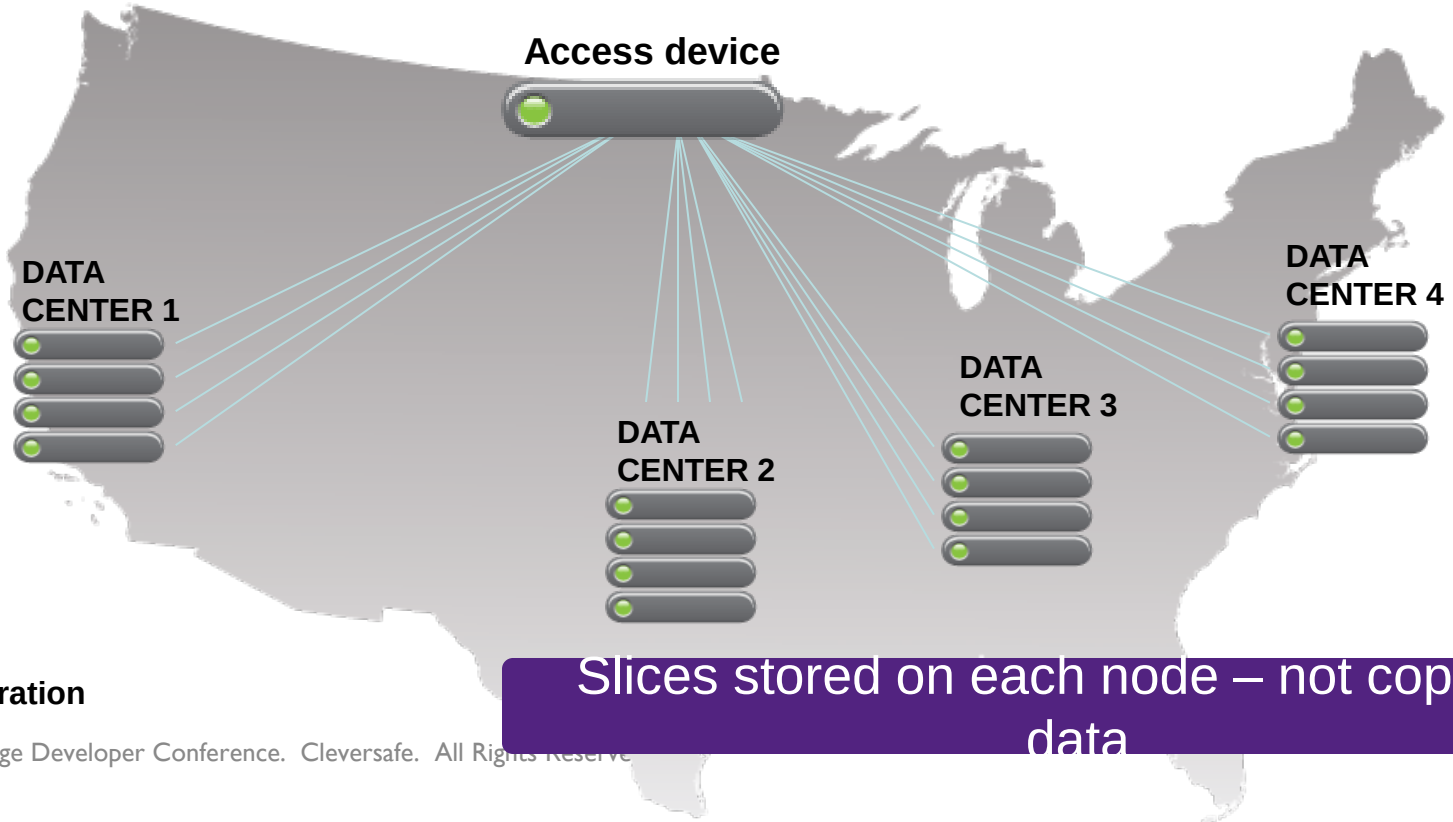
Access Control mechanism may be used in conjunction with the access protocol

Dispersal fits in between access and storage protocols

Access to storage nodes is over an authenticated and encrypted channel (SSL / TLS)

Secure Information Dispersal for Cloud

- Access device acts as gateway to cloud
  - It's only data where and when end users want it to
- Disaster recovery are achieved without the overhead of replication
- Ideal for unstructured content
- Caching can be utilized to achieve higher performance at any site

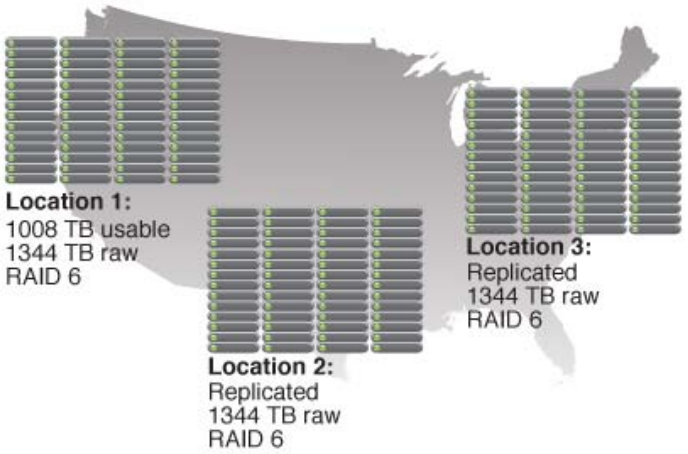


10 of 16 configuration

Slices stored on each node – not copies of data

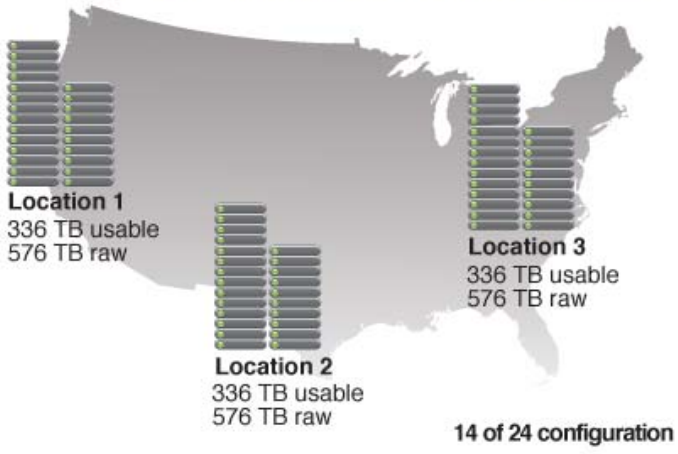
Security & Efficiency Advantages

Replication



Dispersal

Tolerates an entire location failure
Tolerates any 10 simulataneous hardware failures



14 of 24 configuration

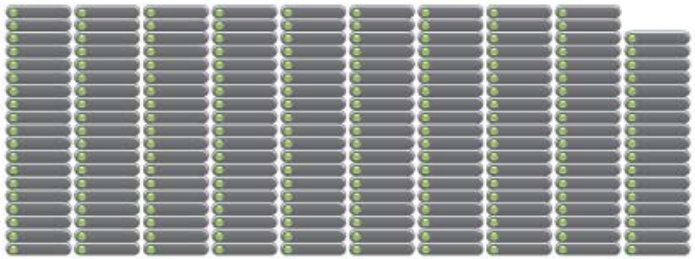
Replication: Total Raw Storage = 4 PB

1008 TB Usable	336 TB Parity
1008 TB Replicated	336 TB Parity
1008 TB Replicated	336 TB Parity

Dispersal: Total Raw Storage = 1.7 PB

1008 TB Usable	1728 TB Raw Capacity

Replication: Total Servers = 168



Dispersal: Total Servers = 72

Information Dispersal Conclusion

- Information Dispersal is a fundamental building block in addressing key challenges of cloud storage:
  - Don't have to sacrifice confidentiality to gain availability
    - No replicated copies of encryption keys or data
  - No encryption keys to manage
  - Enables people to gain the advantages of secret sharing without the typical storage overhead
  - Puts end users in control of their data since it only exists where and when they want it to

Provides a solution for secure storage over the public internet

Thank you

Julie - jbellanca@cleversafe.com

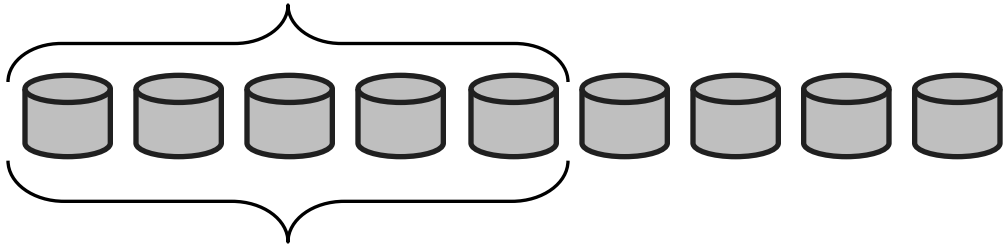
Jason - jresch@cleversafe.com

Backup

Secret Sharing Advantages

- Better Reliability – tolerates loss or unavailability of shares (n-k)
- Better Security – tolerates k compromises

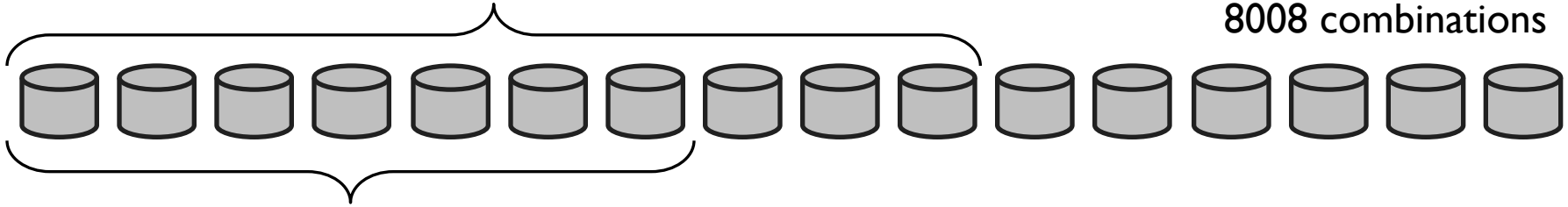
5 stores needed to break Confidentiality or Integrity



5 stores needed to break Availability

5-of-9 Configuration
126 combinations

10 stores needed to break Confidentiality or Integrity



7 stores needed to break Availability

10-of-16 Configuration
8008 combinations