Getting the Most out of Erasure Codes

Jason Resch
Cleversafe

Agenda

- Erasure codes
- Parameters for erasure codes
- Modeling system properties
- Choosing optimal parameters
- Getting the best of all worlds
- Conclusions
- Questions

2013 Storage Developer Conference. Copyright © 2013 Cleversafe, Inc. All Rights Reserved.

What are erasure codes?

- Erasure codes are a type of forward error correction that can recover from “erasures”
- How do they do this?
  - They encode **K** inputs into **N** outputs
  - May recover input from any **K** outputs
    - K-of-N system, where $1 \leq K \leq N$
- Replication is a special case K-of-N system:
  - Where K = 1, and N = number of replicas

2013 Storage Developer Conference. Copyright © 2013 Cleversafe, Inc. All Rights Reserved.

How erasure codes work

- Erasure codes work by over-sampling the data
- Often based on the math of linear algebra
  - E.g., When solving for **5** unknowns, it takes the solutions from at least **5** equations
- To make an erasure code out of this:
  - Set **K** = number of variables
  - Set **N** = number of equations
    - Evaluate the **N** equations and store the **N** solutions
    - Can recover the **K** variables from any **K** solutions

2013 Storage Developer Conference. Copyright © 2013 Cleversafe, Inc. All Rights Reserved.

Benefits of erasure codes

- Erasure codes have many useful properties:
  - Storage efficient
    - Raw storage requirements are (N/K) × input size
    - Requirements for 1 TB in a 10-of-15 encoding?
  - Fault tolerant
    - The above 10-of-15 system can survive five simultaneous failures, equivalent to six copies
  - Secure
    - No copy exists at any single location
    - Threshold number of breaches are required

2013 Storage Developer Conference. Copyright © 2013 Cleversafe, Inc. All Rights Reserved.

Parameters in an EC system

Term	Definition
Width (N)	The configured number of outputs generated by the erasure code when processing some input
Threshold (K)	The number of outputs required by the erasure code to reassemble the original data, $1 \le K \le N$
Write Threshold (T_w)	The number of outputs that must be stored to consider a write successful, $1 \le K \le T_w \le N$
Site Count (S)	The number of unique locations to which outputs of the erasure code will be stored
System Capacity (C)	The total usable storage capacity of the erasure code based system
Drives per node (D)	The number of storage drives in each
Drive AFR	The annual failure rate of drives in the system, the inverse of the Mean Time to Failure (MTTF)
Drive MTTR	The average time to repair a failed drive and rebuild all the data that used to be on it
Drive Capacity	The total storage capacity of the drives used in the erasure code based system

2013 Storage Developer Conference. Copyright © 2013 Cleversafe, Inc. All Rights Reserved.

Metrics of an EC system

Term	Definition
Fault Tolerance	The number of outputs that can be lost without impacting the availability of the data
Site Fault Tolerance	The number of sites that can fail without impacting the availability of the data
Expansion Factor	The ratio of the size of all the outputs to the size of all the inputs, a measure of storage efficiency
CPU cost	The amount of processing required by the erasure code to encode a fixed amount of input
Rebuilding cost	The amount of network bandwidth required to rebuild a fixed amount of lost data
Read Availability	The probability that all data in the system can be read at any given time
Write Availability	The probability that a write in the system can succeed at any given time
Data Reliability	The probability that the system suffers no data loss over a given period of time
Worst Case Reliability	The reliability for data when only a write threshold number of outputs are stored

2013 Storage Developer Conference. Copyright © 2013 Cleversafe, Inc. All Rights Reserved.

Varying width and threshold

- Fault Tolerance is equal to (N – K)
  - Determines system reliability and availability
- Unlike replication, erasure codes can increase reliability without increasing storage costs
  - 2-of-3, 4-of-6, 10-of-15, 20-of-30 etc.
    - All have the same storage overhead of 1.5 ×
    - But have vastly different fault tolerances…
- Reliability and efficiency may even both improve:
  - Going from 10-of-15 to a 30-of-40

2013 Storage Developer Conference. Copyright © 2013 Cleversafe, Inc. All Rights Reserved.

Effects of these changes

- Increased width can offer improved reliability, availability, and storage efficiency.

- But there are tradeoffs to wider systems:
  - As width increases, more nodes are required
  - As threshold goes up so does rebuilding cost
  - As fault-tolerance increases, erasure codes become computationally more expensive

2013 Storage Developer Conference. Copyright © 2013 Cleversafe, Inc. All Rights Reserved.

Write thresholds

- The write threshold is the number of outputs that must be written to consider the write successful
  - If equal to width, any node or drive outage will cause a loss of availability
  - If equal to threshold, then any node or drive failure can cause irrecoverable data loss
- An appropriately chosen write threshold will provide good reliability and availability
  - E.g. equal to 25 in a 20-of-30 system
    - Tolerates 5 failures without impact to system

2013 Storage Developer Conference. Copyright © 2013 Cleversafe, Inc. All Rights Reserved.

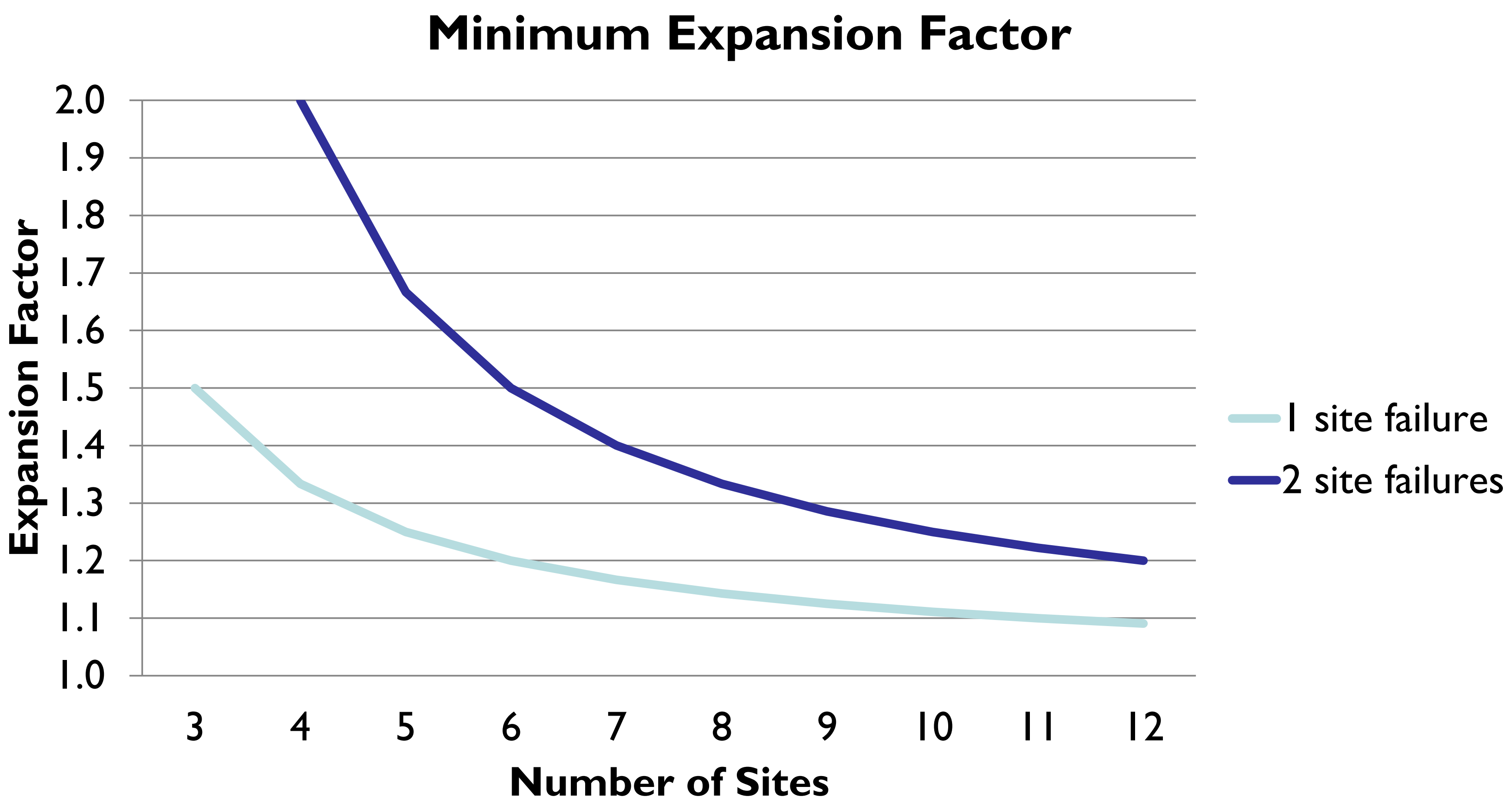
Effects of different write thresholds

- Assuming a 10-of-16 configuration with 99.9% node uptime, 5% disk AFR, ~60 hour MTTR:

Write Threshold	Write Availability	Worst Case Reliability (annual)
10	≥ 15 nines	0 nines (MTTDL = 2 years)
11	14 nines	2 nines (MTTDL = 547 years)
12	11 nines	5 nines (MTTDL = 274,458 years)
13	8 nines	8 nines
14	6 nines (0.999999445)	11 nines
15	3 nines (0.999881115)	14 nines
16	1 nine (0.984119442)	17 nines

2013 Storage Developer Conference. Copyright © 2013 Cleversafe, Inc. All Rights Reserved.

Site count and storage efficiency



Formula for minimum expansion factor: S / (S - F)
Where F is the number of tolerable site failures

2013 Storage Developer Conference. Copyright © 2013 Cleversafe, Inc. All Rights Reserved.

Summary of tradeoffs

Variable	Positive	Negative
↑ System capacity	↑ System throughput	↓ Data reliability ↓ Read availability
↑Width	Enables ↑ Threshold	↓ Expansion granularity ↑ Seeks per write
↑ Threshold	↑ Throughput ↑ Storage efficiency	↑ Rebuilding cost ↑ Seeks per read
↑ Write Threshold	↑ Worst case data reliability	↓ Write availability
↑ (Width – Threshold)	↑ Data Reliability ↑ Availability	↑ CPU cost
↑ (Width / Threshold)	↑ Site Failure Tolerance	↓ Storage efficiency
↑ Drives per node	↓ Cost per unit of storage	↓ Expansion granularity
↑ Drive Capacity	↓ Cost per unit of storage	↓ Seeks capacity / throughput

2013 Storage Developer Conference. Copyright © 2013 Cleversafe, Inc. All Rights Reserved.

Modeling system attributes

- System availability using binomial distribution:

$$Availability_{K-of-N} = \sum_{i=K}^{N} \binom{N}{i} \cdot p^i \cdot (1-p)^{(N-i)}$$

- Data reliability using model by John E. Angus:

$$MTTDL_{K-of-N} = \frac{MTTF_{disk}}{K \cdot \binom{N}{K}} \times \left(\frac{MTTF_{disk}}{MTTR_{disk}} \right)^{N-K}$$

- Average rebuild traffic:

$$Rate_{K-of-N} = K \times \left(\frac{SystemCapacity}{MTTF_{disk}} \right)$$

2013 Storage Developer Conference. Copyright © 2013 Cleversafe, Inc. All Rights Reserved.

Availability model

- Using the binomial distribution, we can estimate the probability that at least a threshold number of nodes are available:

$$Availability_{K-of-N} = \sum_{i=K}^{N} \binom{N}{i} \cdot p^{i} \cdot (1-p)^{(N-i)}$$

System Configuration	Estimated Availability	Annual Downtime
1-of-3 (triple replication)	9 nines	31.56 milliseconds
3-of-4 (raid 5)	5 nines	31.51 minutes
6-of-8 (raid 6)	7 nines	1.760 seconds
10-of-15 (erasure code)	14 nines	1.577 nanoseconds

2013 Storage Developer Conference. Copyright © 2013 Cleversafe, Inc. All Rights Reserved.

Reliability model

- Using the below formula, we can estimate the Mean-Time-to-Data-Loss and annual reliability:

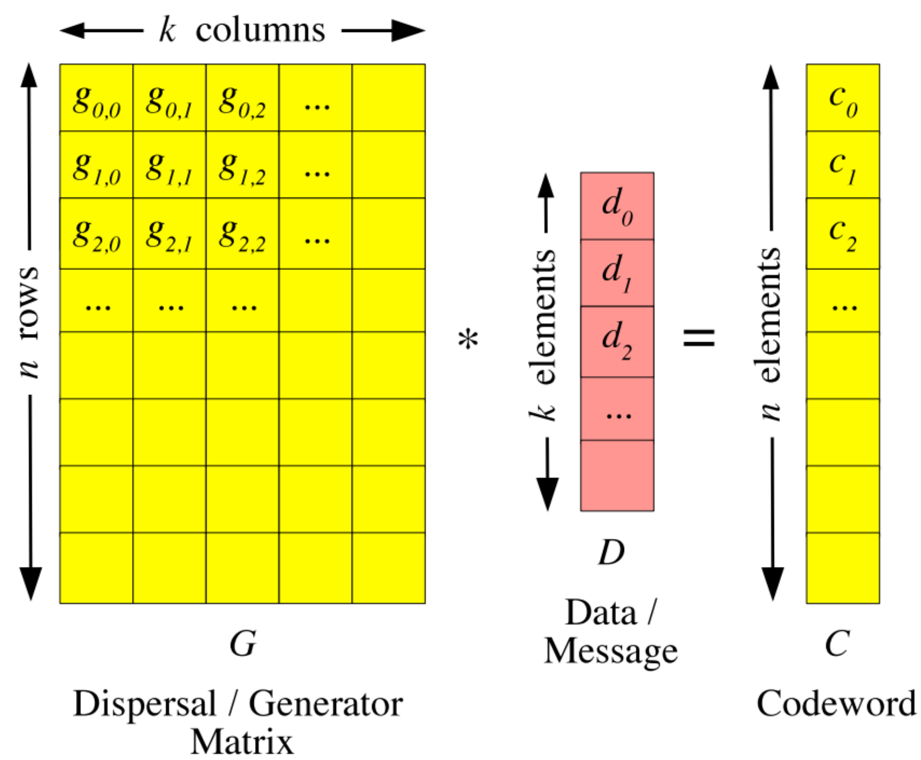
$$MTTDL_{K-of-N} = \frac{MTTF_{disk}}{K \cdot \binom{N}{K}} \times \left(\frac{MTTF_{disk}}{MTTR_{disk}} \right)^{N-K}$$

System Configuration	MTTDL	Annual Reliability
1-of-3 (triple replication)	60,380,721.31 years	7 nines
3-of-4 (raid 5)	5,015.84 years	3 nines
6-of-8 (raid 6)	1,078,227.17 years	6 nines
10-of-15 (erasure code)	164,417,694,101,199.00 years	14 nines

2013 Storage Developer Conference. Copyright © 2013 Cleversafe, Inc. All Rights Reserved.

CPU cost

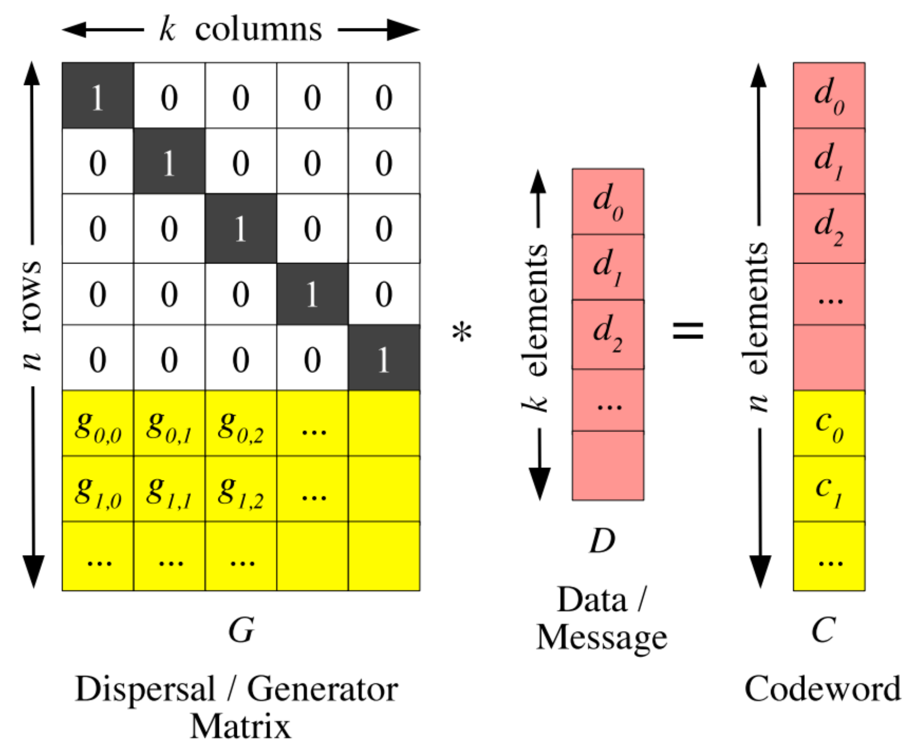
- Proportional to width
- Requires a matrix-vector multiplication
  - K multiplications and additions for each of the N outputs
  - Encodes all K inputs
  - Performance per MB is proportional to (1/N)



2013 Storage Developer Conference. Copyright © 2013 Cleversafe, Inc. All Rights Reserved.

Systematic encoding

- Encodes N-K outputs
- First K outputs = the K inputs, no processing
  - Performance per MB is proportional to 1/(N-K)
  - This means 10-of-15 same cost as 40-of-45
  - 84-of-100 costs same as 10-of-16 does without systematic erasure codes!



2013 Storage Developer Conference. Copyright © 2013 Cleversafe, Inc. All Rights Reserved.

Storage cost

- Each output value is 1/K the size of all the inputs
- By information theory, this is smallest possible size, and hence the best possible efficiency
  - If they were any smaller, then the K outputs would be smaller than the original input
- Therefore, the N outputs, each 1/K the input size, total up to N/K times the size of all inputs
  - To make the erasure code efficient, K needs to be about the same size as N

2013 Storage Developer Conference. Copyright © 2013 Cleversafe, Inc. All Rights Reserved.

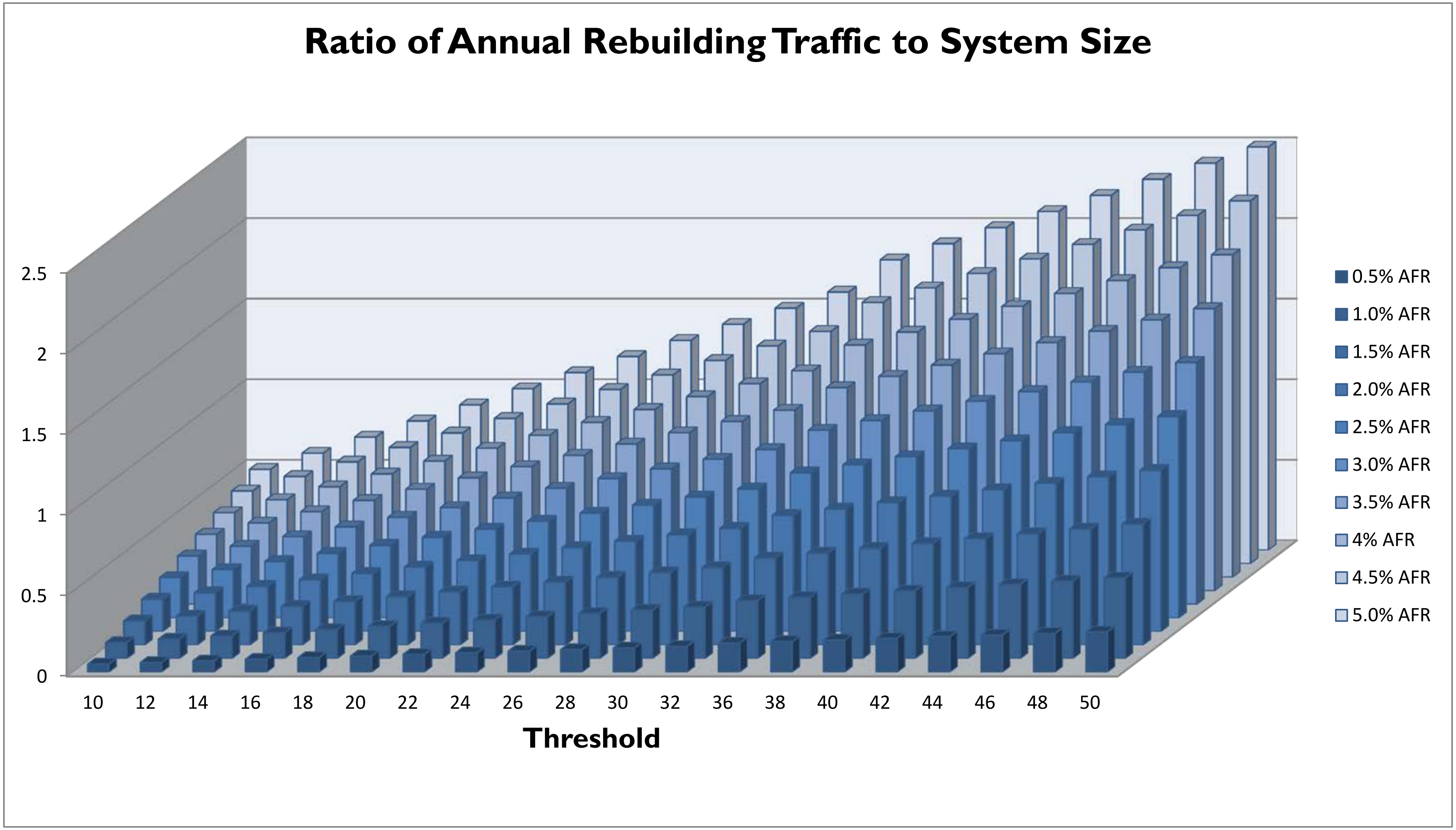
Rebuilding cost

- Rebuilding is the recovery of lost outputs
- To recover a lost output requires information from a threshold number surviving output values
- In a 10-of-15 system, if a 4 TB drive containing output values fails, then 40 TB worth of other output values must be read to recover them
  - Total rebuilding cost in the system is:

$$Rate_{K-of-N} = K \times \left(\frac{SystemCapacity}{MTTF_{disk}} \right)$$

2013 Storage Developer Conference. Copyright © 2013 Cleversafe, Inc. All Rights Reserved.

Expected rebuilding cost



2013 Storage Developer Conference. Copyright © 2013 Cleversafe, Inc. All Rights Reserved.

Design goals of a storage system

- Designers contend with many, sometimes mutually competing or opposing goals:
  - High Reliability and Availability
  - Low cost and high efficiency
  - Secure and easy to use
  - High performance and commodity hardware
  - Economy of scale and low entry cost
- Parameters in an EC system are highly interrelated and affect all of the above attributes

2013 Storage Developer Conference. Copyright © 2013 Cleversafe, Inc. All Rights Reserved.

Example goals for an EC system

- Target goals:
  - Worst case annual reliability of 7 nines
  - Write availability of at least 5 nines
  - Tolerate one site outage (out of 5 sites)
  - Expansion factor of less than 1.75
  - Total usable capacity of 10 PB
  - Annual rebuild traffic of less than 5 PB
- Is this possible?

2013 Storage Developer Conference. Copyright © 2013 Cleversafe, Inc. All Rights Reserved.

It is possible, almost…

Cleversafe dsNet Calculator

The dsNet® Availability & Reliability can be modeled with this tool.

Enter the system parameters in the left column, and the resulting Availability & Reliability performance will be shown in the right column. Place the cursor over a data field. Press HELP for more information.

Input Parameters

IDA Width:	30
IDA Threshold:	18
Write Threshold:	23
Number of Sites:	5
Usable Capacity (TB):	10000
Average Site Downtime (Min):	5
Mean Time to Site Outage (Days):	90
Average Node Downtime (Min):	60
Mean Time to Node Outage (Days):	90
Drives Per Store:	48
Drive Capacity (TB):	4
Drive Annual Failure Rate (%):	4
Drive URE Rate (10^-x):	15
Rebuild Rate (MB/s):	10
Rebuild Detection Time (Hours):	24

Output Parameters

Expansion Factor	1.67
Required Slicestor® Devices	90

DATA RELIABILITY

Permanent Reliability	100 %
Nines (9's)	Infinity
Mean Time to Data Loss	Infinity Yrs
Instantaneous Reliability	99.9999923 %
Nines (9's)	7
Mean Time to Data Loss	12,959,107.1 Yrs

DATA AVAILABILITY	Read	Write
System	99.99999932 %	99.99976 %
Nines (9's)	8	5
Annual Downtime (sec)	0.2	77.2
Per-Site	99.9961 %	99.9959 %
Nines (9's)	4	4

2013 Storage Developer Conference. Copyright © 2013 Cleversafe, Inc. All Rights Reserved.

Getting the best of all worlds?

- We have designed a system with a great combination of properties:
  - Worst case reliability of 7 nines
  - Write availability of 5 nines
  - Tolerates a site outage + any 2 other nodes
  - Expansion factor of only 1.67
- But we exceeded the rebuild traffic of 5 PB/year
  - Lower thresholds would prevent us from reaching the availability or reliability goals…

2013 Storage Developer Conference. Copyright © 2013 Cleversafe, Inc. All Rights Reserved.

Solution: Deferred rebuilding

- A simple, but unintuitive result:
  - By waiting longer before rebuilding, we can make a system that is more reliable, more efficient, and with less rebuild traffic
- When we read T outputs to do a rebuild, we can rebuild any number of outputs for that data
  - E.g. a 60-of-80 system that rebuilds after 4 outputs are lost will have $1/4^{th}$ the rebuild cost
  - This enables much wider systems than before
    - Allowing even better efficiency and reliability

2013 Storage Developer Conference. Copyright © 2013 Cleversafe, Inc. All Rights Reserved.

Deferred rebuilding comparison

- Comparison of two 10 PB systems:
  - One is significantly wider, but defers rebuilding until four failures have occurred

System Attribute	18-of-30, T_w=23	39-of-60, T_w=46, defer=4
Worst Case Reliability	7 nines	10 nines ✓
Write Availability	5 nines	6 nines ✓
Expansion Factor	1.67	1.54 ✓
Rebuild Traffic	7.2 PB	3.9 PB (15.6 without deferring) ✓
CPU cost	12 ✓	21

2013 Storage Developer Conference. Copyright © 2013 Cleversafe, Inc. All Rights Reserved.

Other improvements to rebuilding

- Online Codes
  - Most outputs are produced from fewer than a threshold number of inputs (only need those)
- Partial Rebuilding
  - Combines information from multiple outputs into a smaller piece, resulting in less traffic
- LRC Codes
  - Stay for the next presentation on this subject!

2013 Storage Developer Conference. Copyright © 2013 Cleversafe, Inc. All Rights Reserved.

Conclusions

- Selecting the ideal parameters for an erasure coded system is a complex, multi-dimensional, non-linear optimization problem
- There are many tradeoffs, and often unexpected consequences when changing parameters
- Erasure codes provide a great deal of flexibility when it comes to finding solutions that meet all the goals of the storage system
  - It would be much harder to meet the same goals if one were restricted to 1-of-N systems!

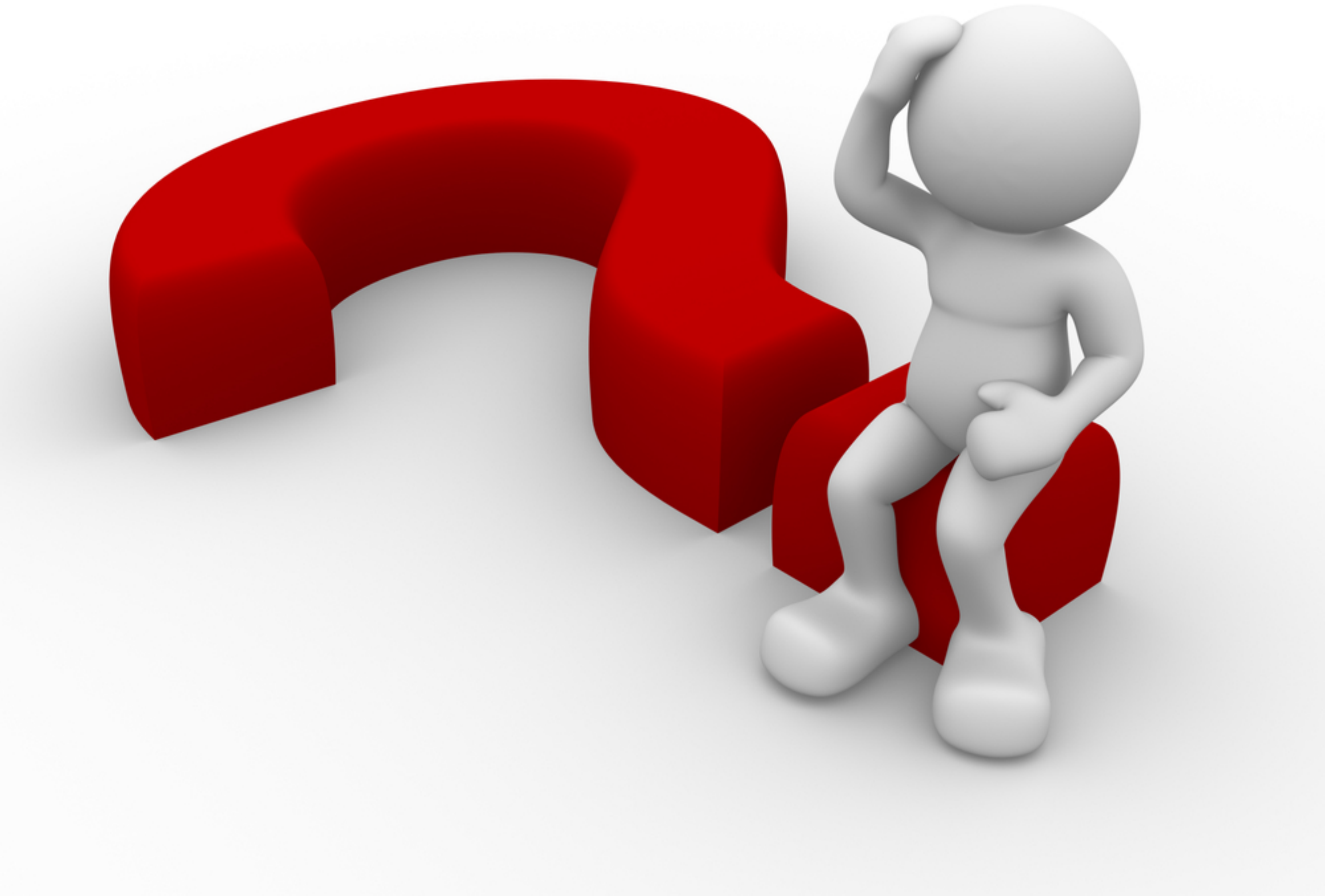
2013 Storage Developer Conference. Copyright © 2013 Cleversafe, Inc. All Rights Reserved.

The future of erasure codes?

- Erasure codes seem to offer the only practical solution to achieve reliability at scale
- Moreover, as CPUs continue to grow in power, the processing cost of erasure codes becomes increasingly marginal
  - See the “Screaming Fast Galois Field Arithmetic Using Intel SIMD Instructions” talk
- Finally, as growth in network speeds continues to outpace growth of disk speeds, erasure codes become an increasingly attractive proposition

2013 Storage Developer Conference. Copyright © 2013 Cleversafe, Inc. All Rights Reserved.

Questions



2013 Storage Developer Conference. Copyright © 2013 Cleversafe, Inc. All Rights Reserved.