EMC²
where information lives®

# HAIL
(High-Assurance and Integrity Layer)

Timothy K. Harder
September 23, 2009

Getting back control

- The Law of Large Numbers is not on your side
  - Even unlikely events will happen given enough opportunity
  - 281 Exabytes of data is lots of opportunity
    - Drop 1 nine ; drop 281TB of data

- Assessment gap on dispersal techniques, erasure codes and ecc

- "striped write" + "correct code" + aggregator = "hail'

- M of N can be applied for 2 layer obfuscation and crypto

Sampling
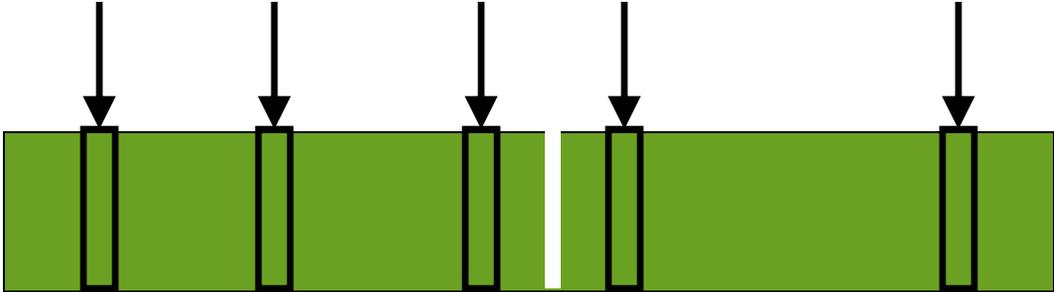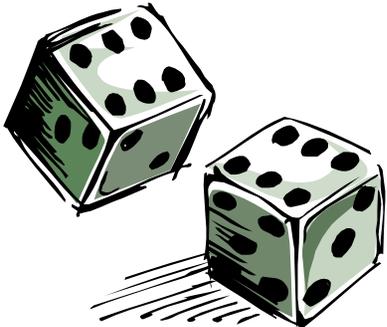
File

Proof of Retrievability (POR)

- Adds additional encoding information to the file

- Makes the file resilient to small errors

- Samples the file to check for errors

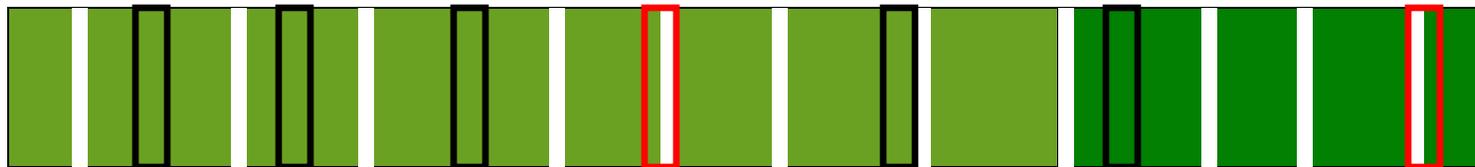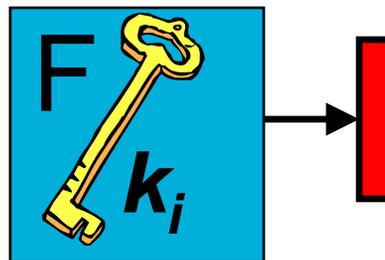- Encoding amplifies the effectiveness of sampling

4

Aggregation

- Sample only 0.1% of the file to achieve 99.9999% accuracy

- 0.1% of 10 TB = 10 GB

- Combine the blocks to send less data
  - Use a keyed aggregation function
  - Return a small output value (32 bytes)
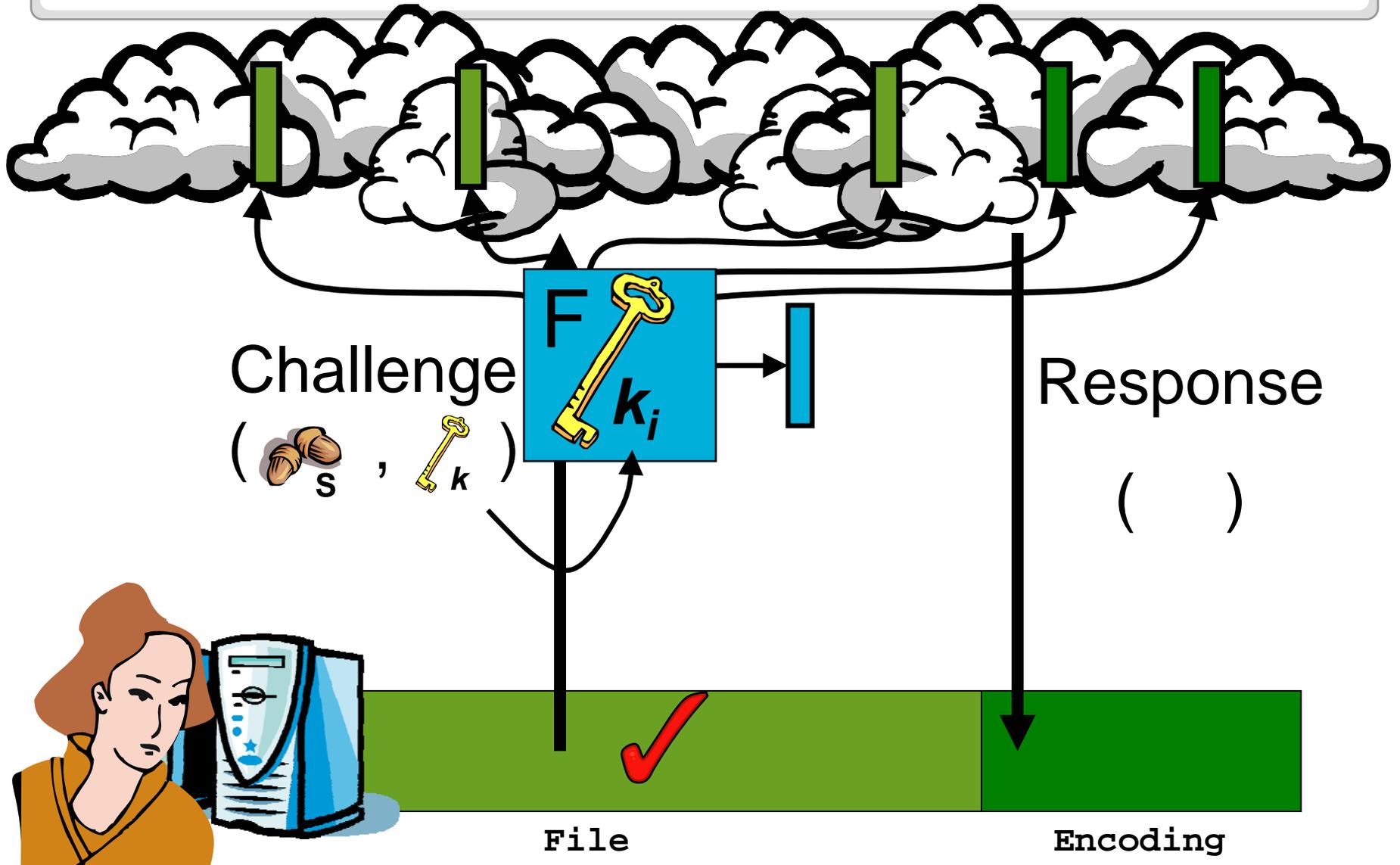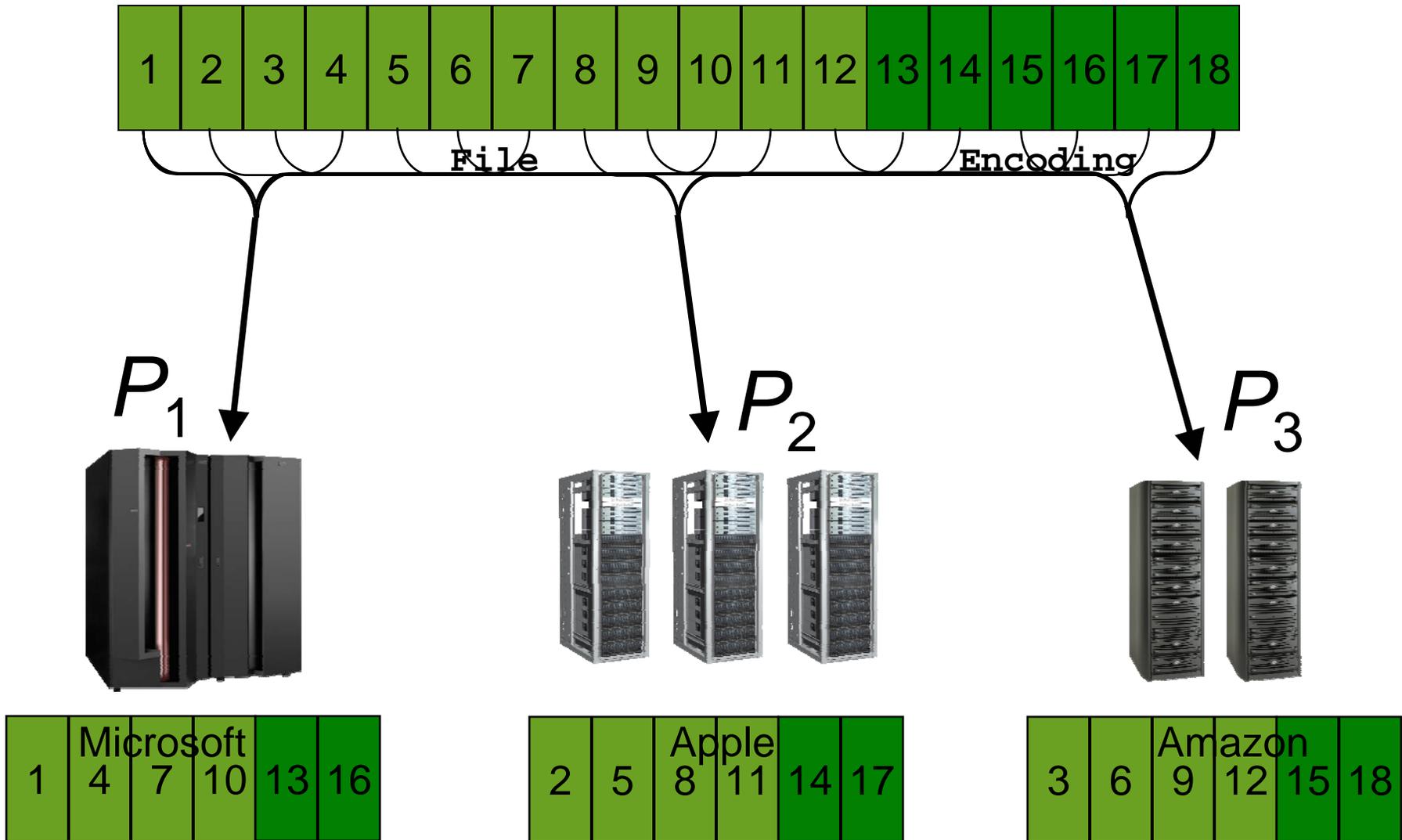
$$F_{k_i}$$
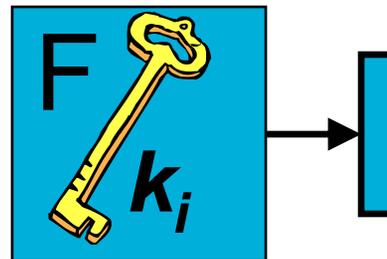
Aggregation

- Sample only 0.1% of the file to achieve 99.9999% accuracy

- 0.1% of 10 TB =

- Combine the blocks to send less data
  - Use another error correcting code
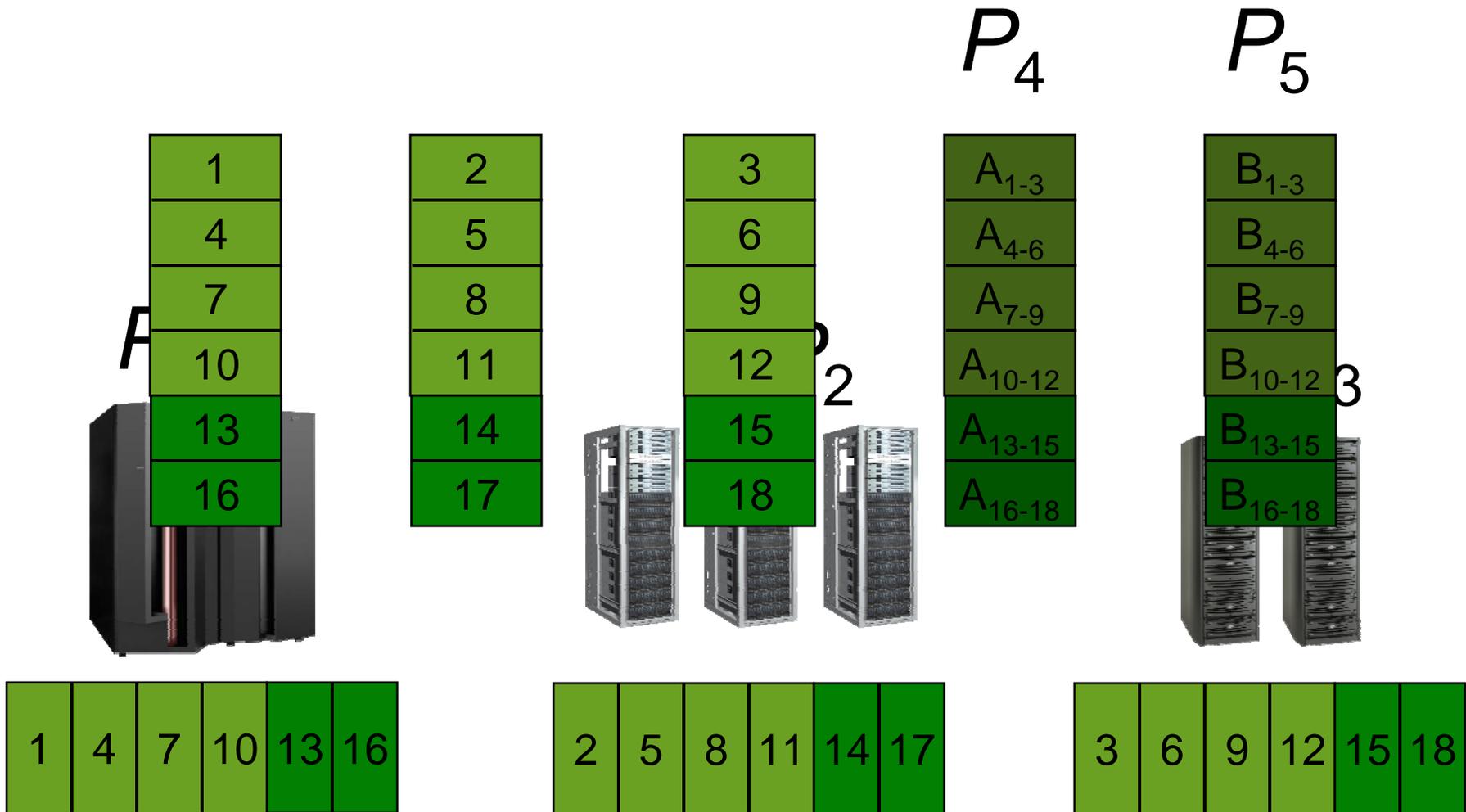 - Return only 1 output value (32 bytes)

Proof of Retrievability

Challenge

$(\quad _s \ , \quad _k)$

$F \quad k_i$

Response

$(\quad)$

File

Encoding

HAIL Encoding - File code

HAIL Encoding - Dispersal Code

P_4 P_5

P_1	P_2	P_3		
1	2	3	A_{1-3}	B_{1-3}
4	5	6	A_{4-6}	B_{4-6}
7	8	9	A_{7-9}	B_{7-9}
10	11	12	A_{10-12}	B_{10-12}
13	14	15	A_{13-15}	B_{13-15}
16	17	18	A_{16-18}	B_{16-18}

1	4	7	10	13	16

2	5	8	11	14	17

3	6	9	12	15	18

$$(\boxed{R_1}, \boxed{R_2}, \boxed{R_3}, \boxed{R_4}, \boxed{R_5}) = \text{Consistent Set}$$

- Use inter-provider redundancy
- Check one provider against all the others

References

- http://www.networkworld.com/news/2008/081108-linkup-failure.html?hpg1=bn

- http://www.emc.com/collateral/analyst-reports/diverse-exploding-idc-exec-summary.pdf

- http://www.informationweek.com/news/services/storage/showArticle.jhtml?articleID=211200905