How to Store a Secret

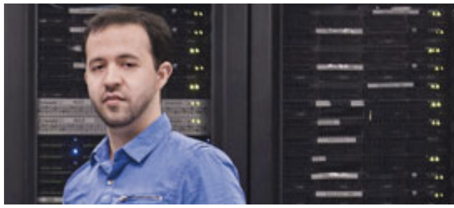
Salim El Rouayheb

Illinois Institute of Technology

A Brief History of Codes for Storage According to Emina

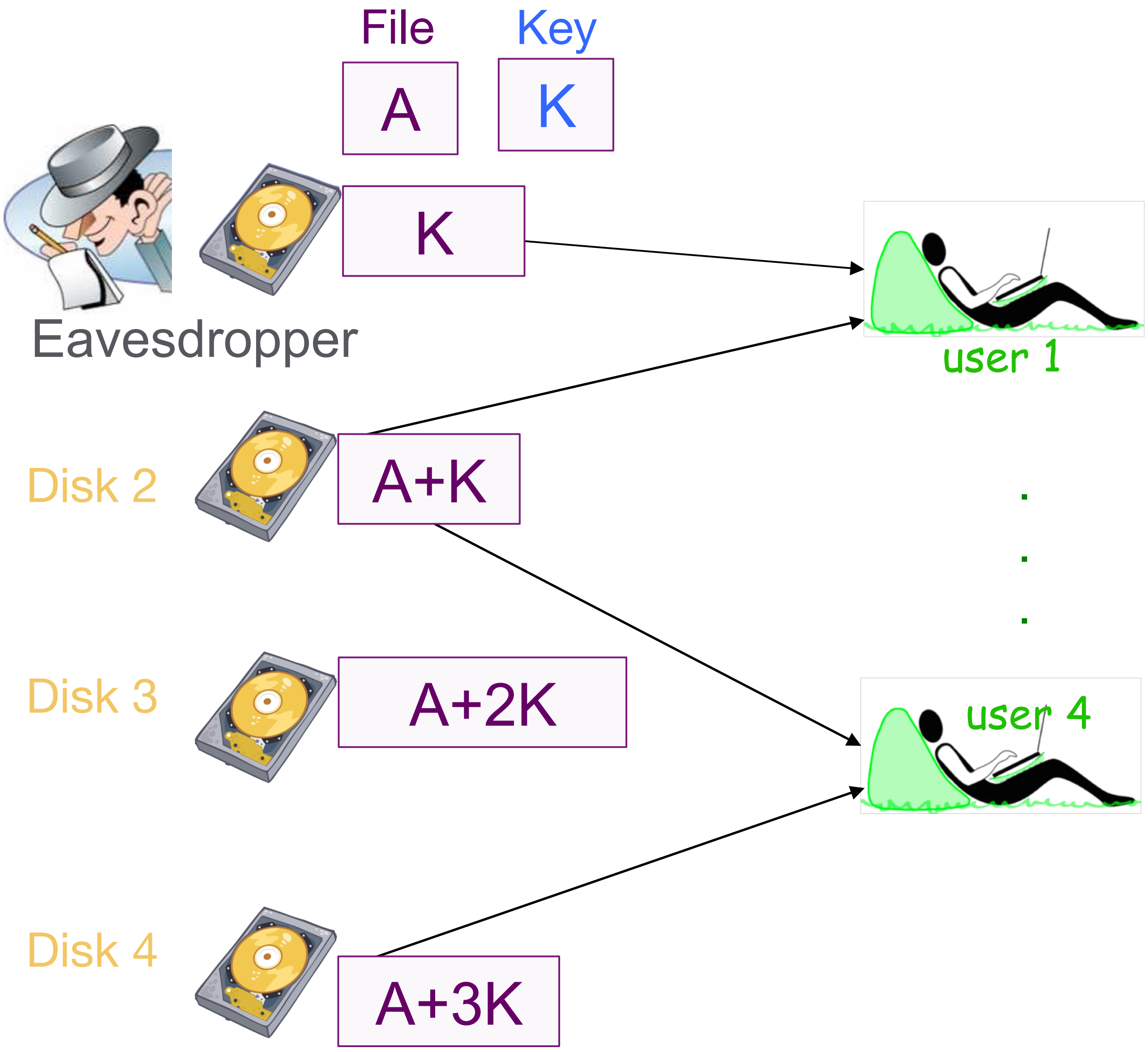
1982

Reed Solomon paper (1960)



What if some nodes cannot be trusted?

Adversary (passive for now) controls one node



File

Key

A

K

K

Eavesdropper

Disk 2

A+K

Disk 3

A+2K

Disk 4

A+3K

user 1

user 4

Secret Sharing [Shamir ’79]

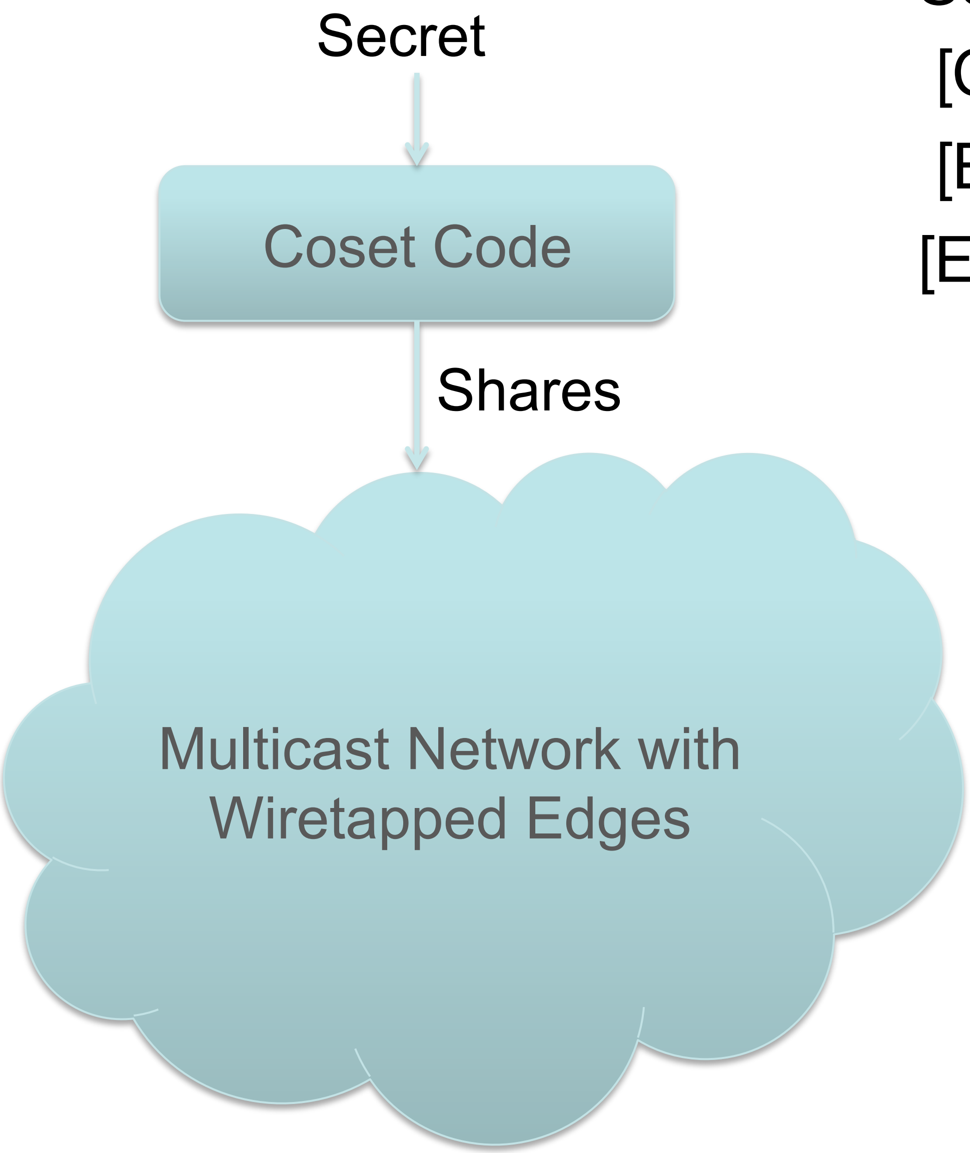
Wiretap channel II

Coset Codes

[Ozarow & Wyner ’84]

(n,k)=(4,2)

Wiretap Network



Secure network coding

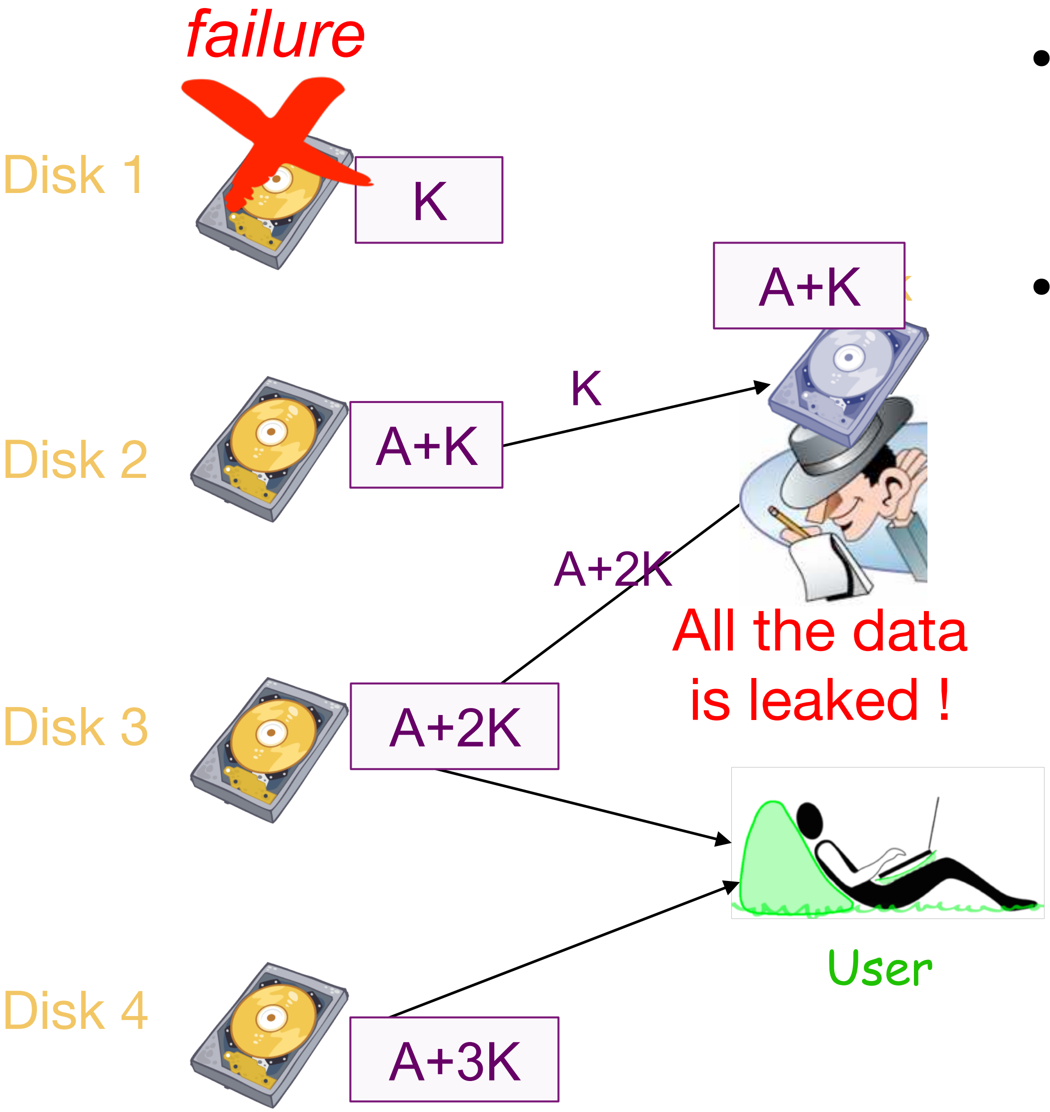
[Cai & Yeung ’02]

[ElRouayheb, Soljanin ’07]

[ElRouayheb, Sprintson, Soljanin ’10]

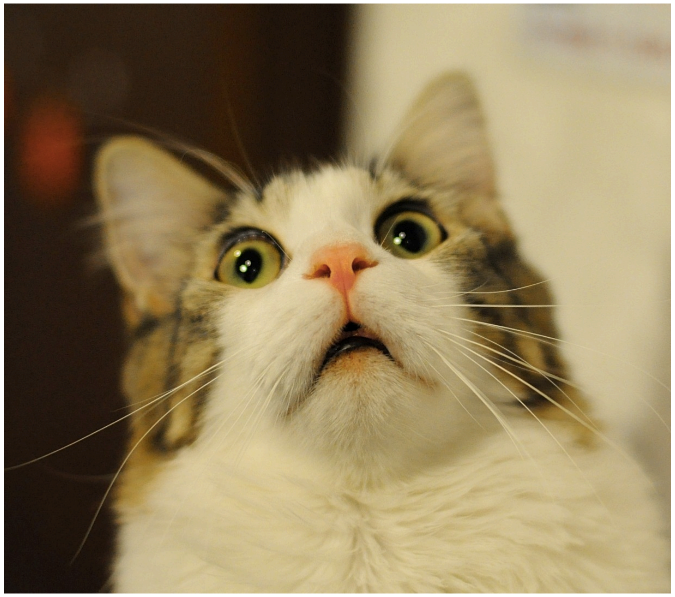
Main Message There:
Separation is optimal
Coset code + Network Code

Coset Codes/Secret Sharing are Not Enough

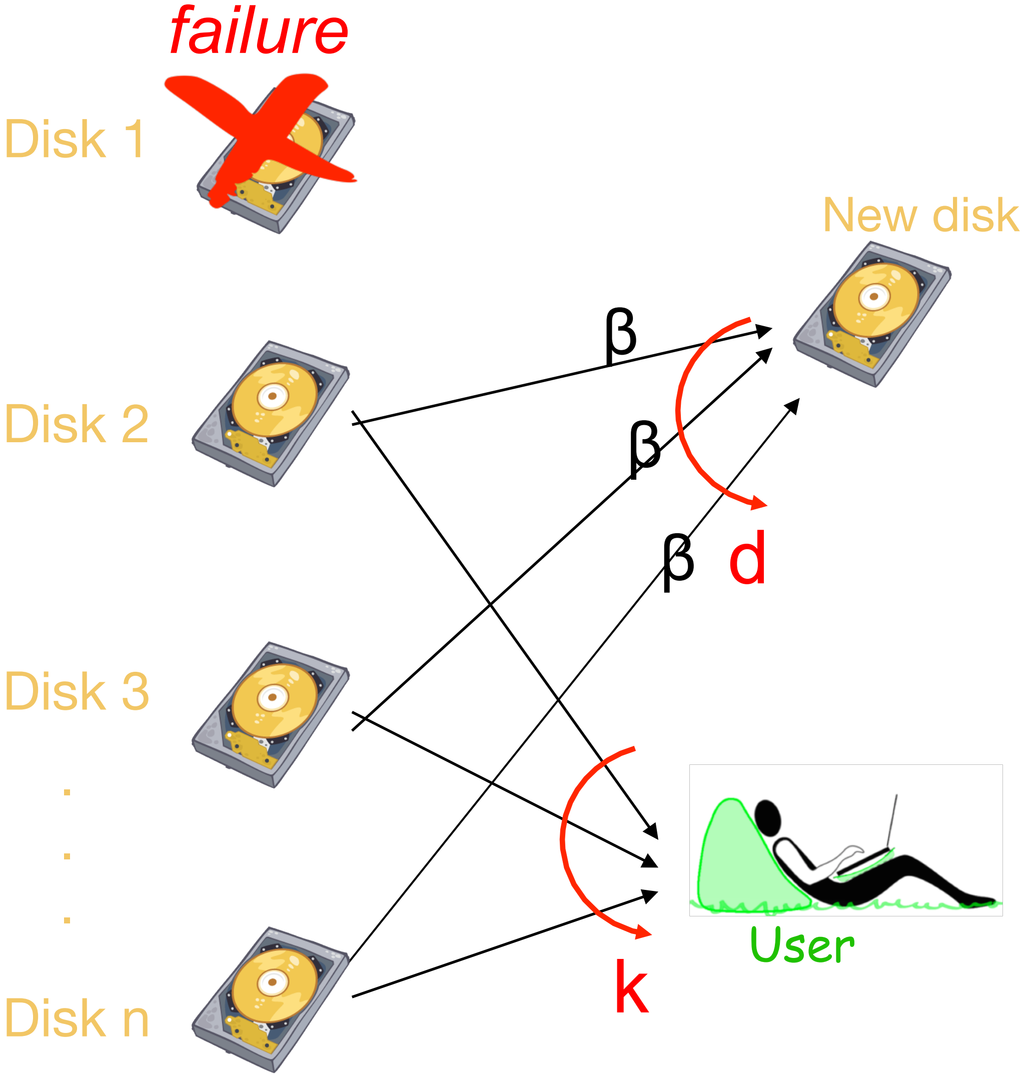


All the data is leaked !

- Because storage systems are dynamic
- Can we still protect the stored secret?
- **Two surprising results**



General Problem Formulation



- (n,k) system
- d: repair degree
- α: storage per node
- β: repair bandwidth
- b: nbr of compromised nodes
- Adversary: passive/active

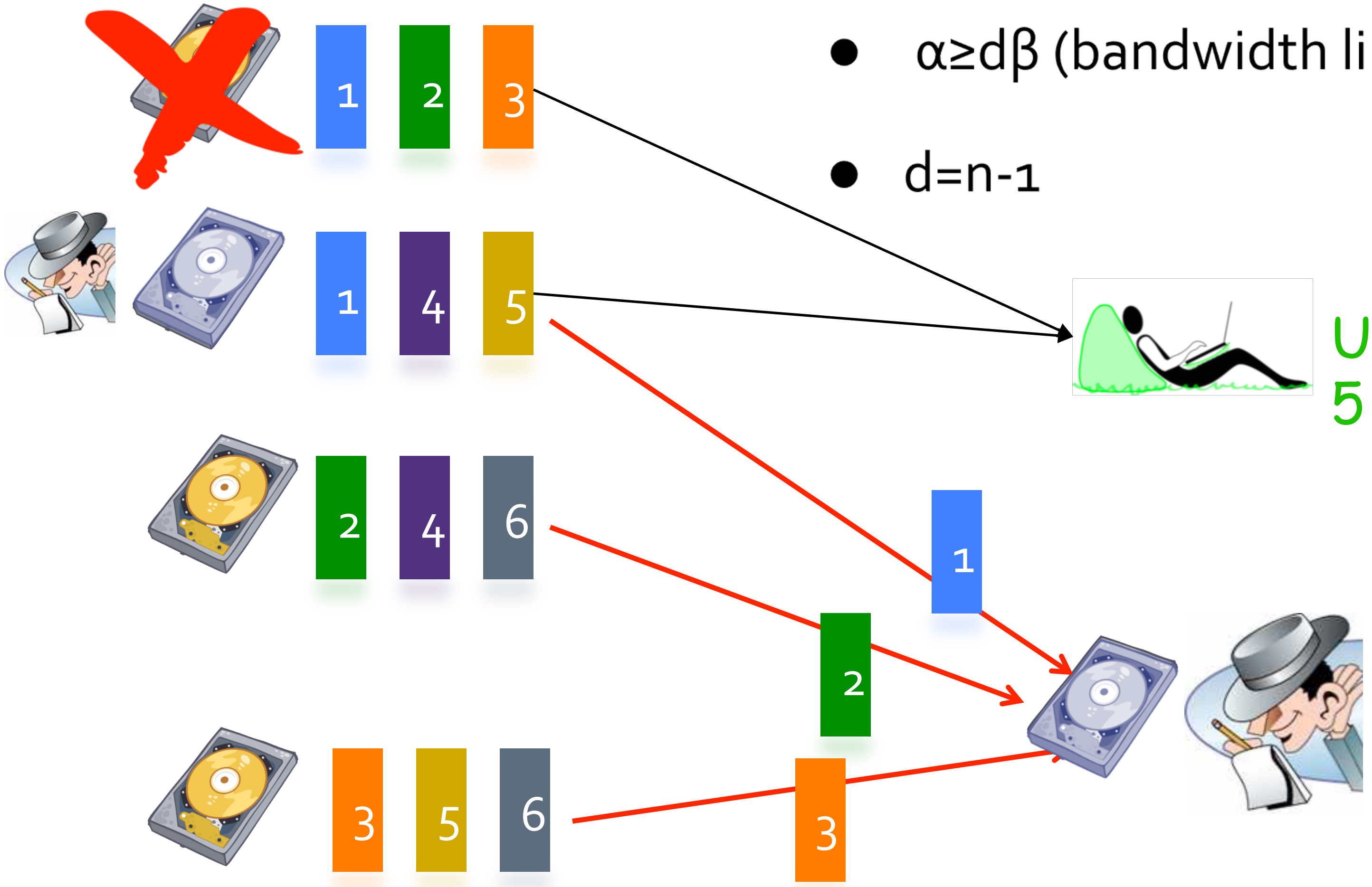
Pawar, ElRouayheb, Ramchandran, ’10

What is the largest secret I can store in this system without loosing it or revealing it?

A Divide and Share Scheme

b=1 compromised node

- Transformed dynamic system into a static system
- Transformation possible if
  - $\alpha \geq d\beta$ (bandwidth limited regime)
  - d=n-1



User always sees all the 5 packets

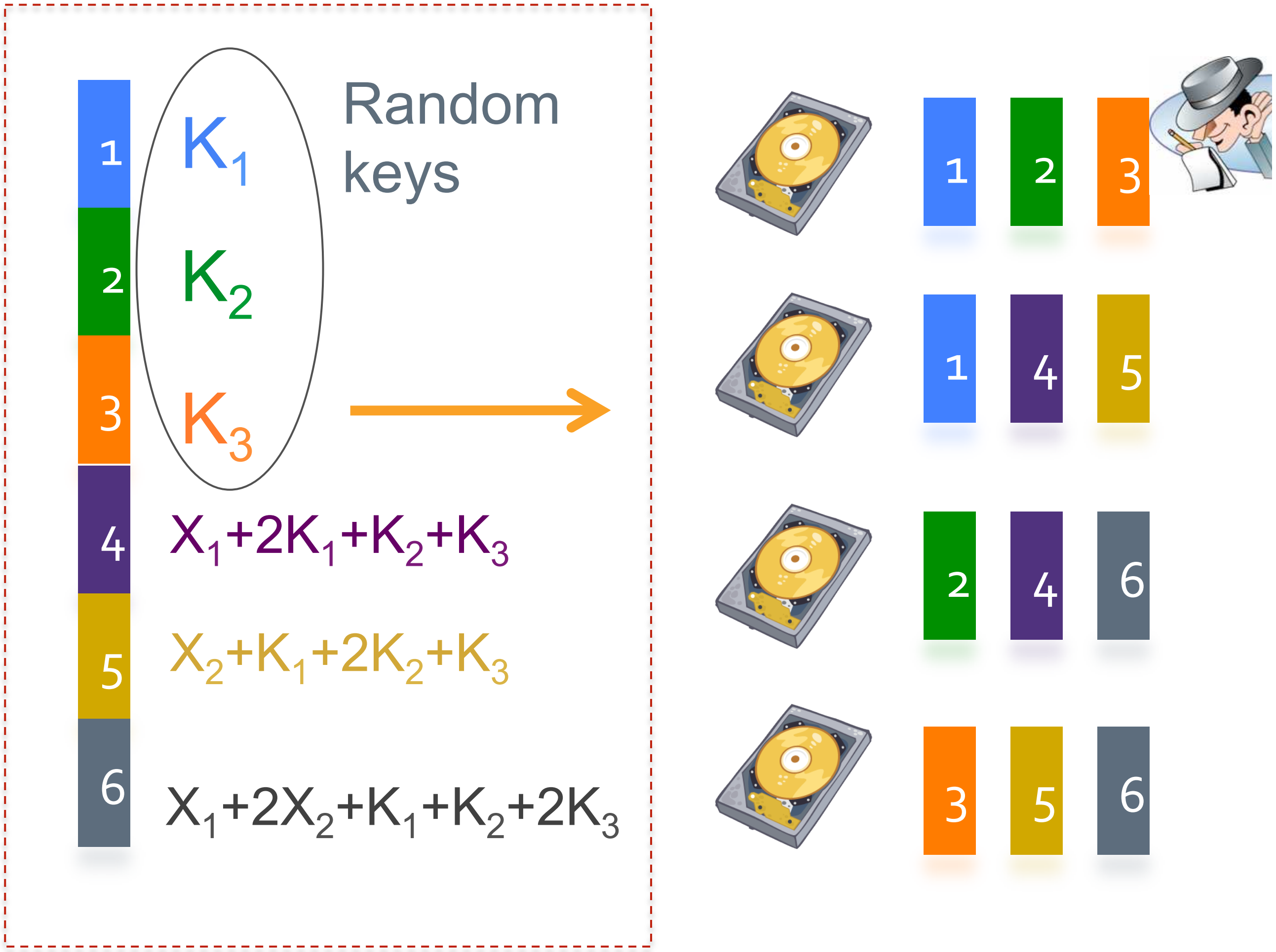
Eavesdropper always observe 3 packets

(n,k,d)=(4,2,3)

Rashmi, Shah, Kumar & Ramchandran '09

Secure Code

Secret:
X1 X2 X3



Random keys

K_1

K_2

K_3

$X_1+2K_1+K_2+K_3$

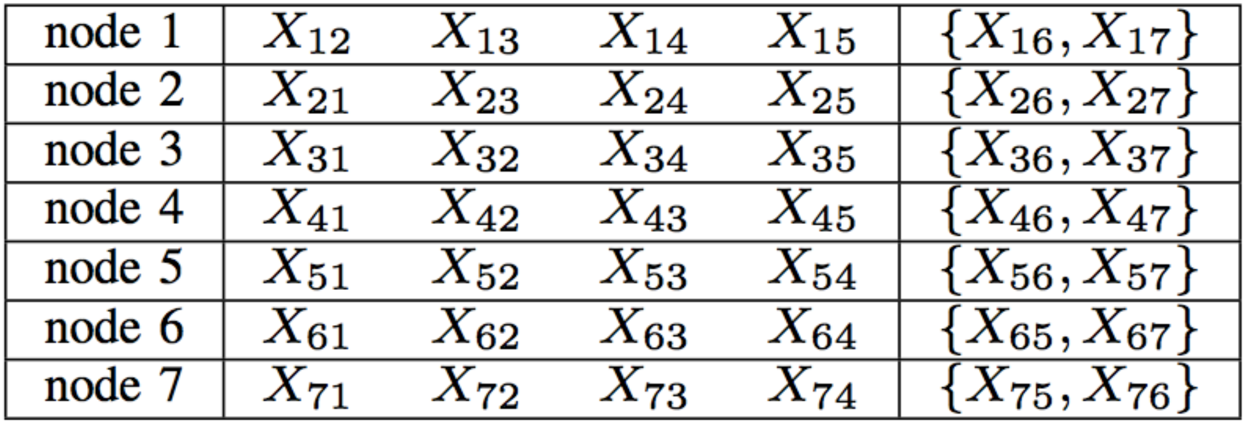
$X_2+K_1+2K_2+K_3$

$X_1+2X_2+K_1+K_2+2K_3$

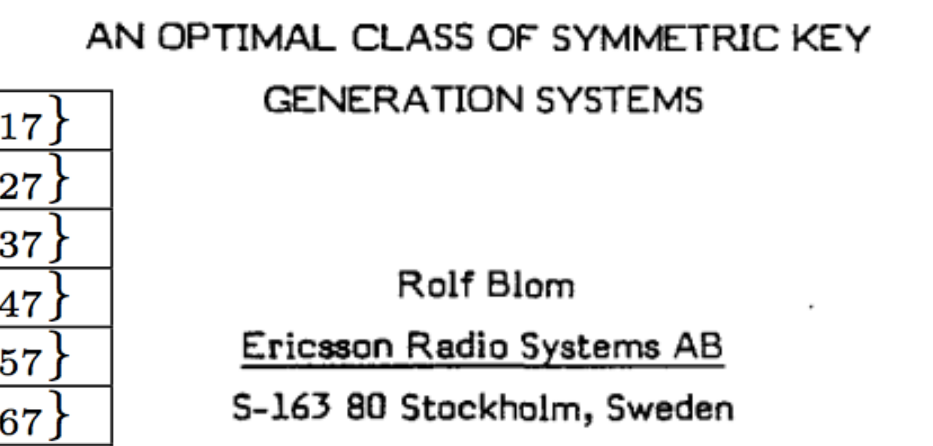
Coset Code

Secure Code in Bandwidth-Limited Regime and d<n-1

(n,k,d)=(7,3,4)



node 1	X_{12}	X_{13}	X_{14}	X_{15}	$\{X_{16}, X_{17}\}$
node 2	X_{21}	X_{23}	X_{24}	X_{25}	$\{X_{26}, X_{27}\}$
node 3	X_{31}	X_{32}	X_{34}	X_{35}	$\{X_{36}, X_{37}\}$
node 4	X_{41}	X_{42}	X_{43}	X_{45}	$\{X_{46}, X_{47}\}$
node 5	X_{51}	X_{52}	X_{53}	X_{54}	$\{X_{56}, X_{57}\}$
node 6	X_{61}	X_{62}	X_{63}	X_{64}	$\{X_{65}, X_{67}\}$
node 7	X_{71}	X_{72}	X_{73}	X_{74}	$\{X_{75}, X_{76}\}$



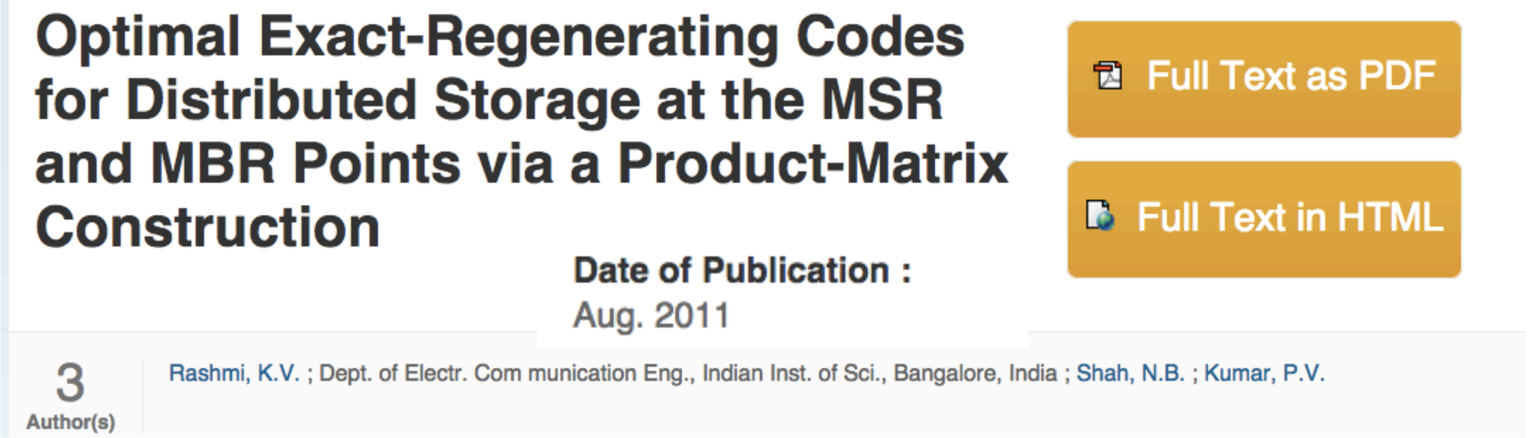
AN OPTIMAL CLASS OF SYMMETRIC KEY GENERATION SYSTEMS

Rolf Blom

Ericsson Radio Systems AB

S-163 80 Stockholm, Sweden

Iwan's Observation



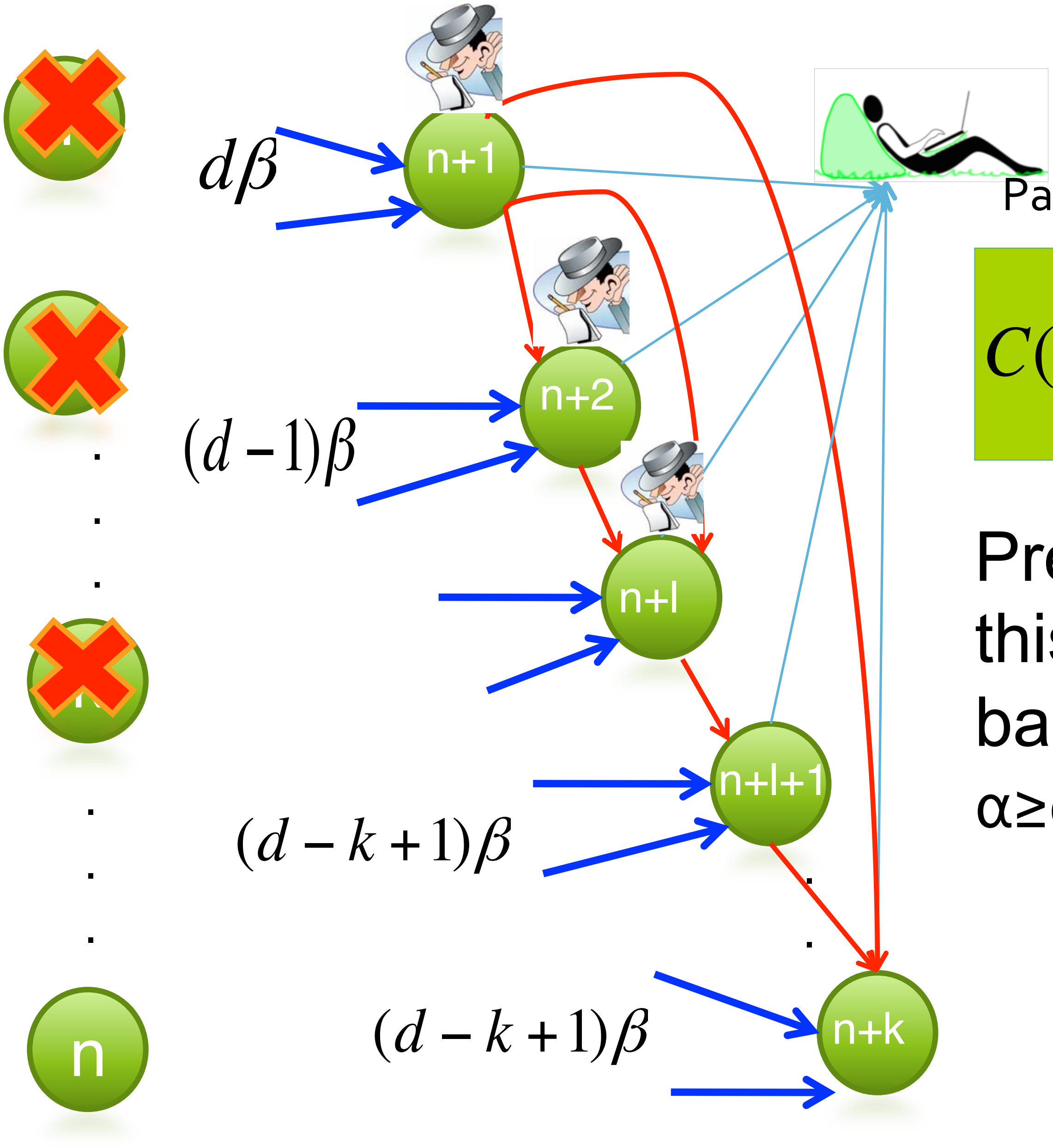
Optimal Exact-Regenerating Codes for Distributed Storage at the MSR and MBR Points via a Product-Matrix Construction

Date of Publication : Aug. 2011

3 Author(s)

Rashmi, K.V. ; Dept. of Electr. Com munication Eng., Indian Inst. of Sci., Bangalore, India ; Shah, N.B. ; Kumar, P.V.

Upper Bound on Secrecy Capacity



Pawar, ElRouayheb, Ramchandran, '10

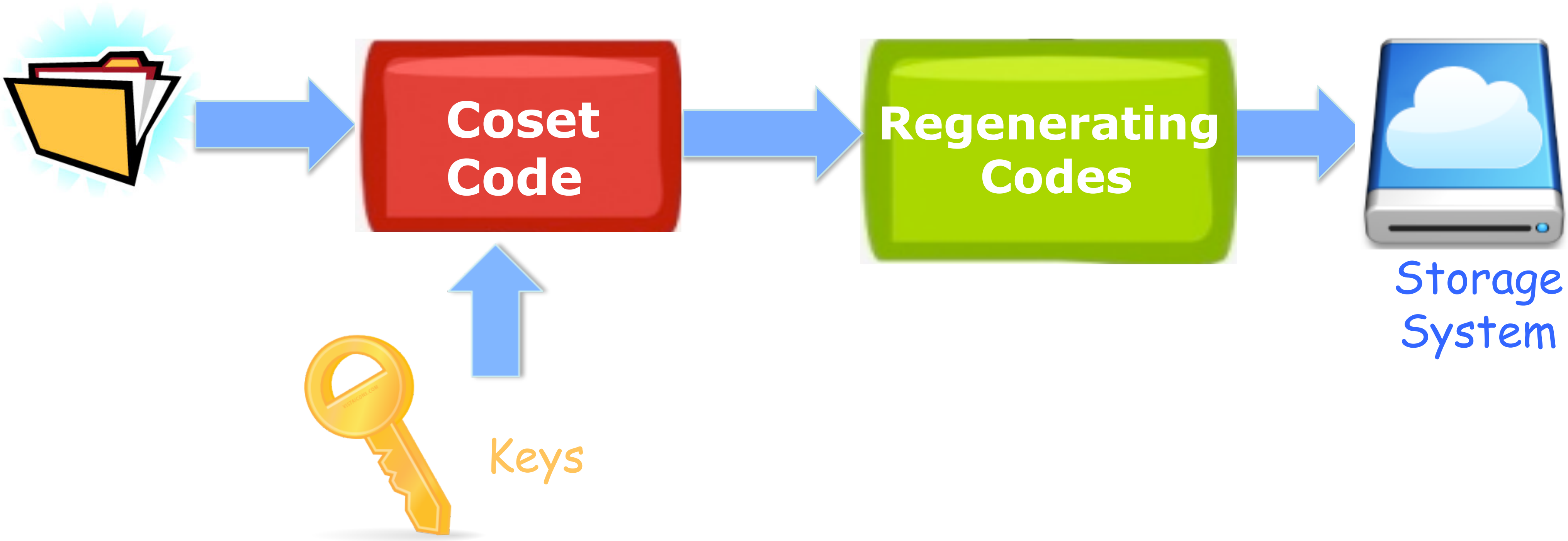
$$C(\alpha,\beta) \leq \sum_{i=l+1}^{k} \min\{(d-i+1)\beta,\alpha\}$$

Previous codes achieve this upper bound for bandwidth-limited regime α≥dβ

General Secure Codes

SEVERE
SEVERE RISK OF TERRORIST ATTACKS
HIGH
HIGH RISK OF TERRORIST ATTACKS
ELEVATED
SIGNIFICANT RISK OF TERRORIST ATTACKS
GUARDED
GENERAL RISK OF TERRORIST ATTACKS
LOW
LOW RISK OF TERRORIST ATTACKS

file



Coset Code

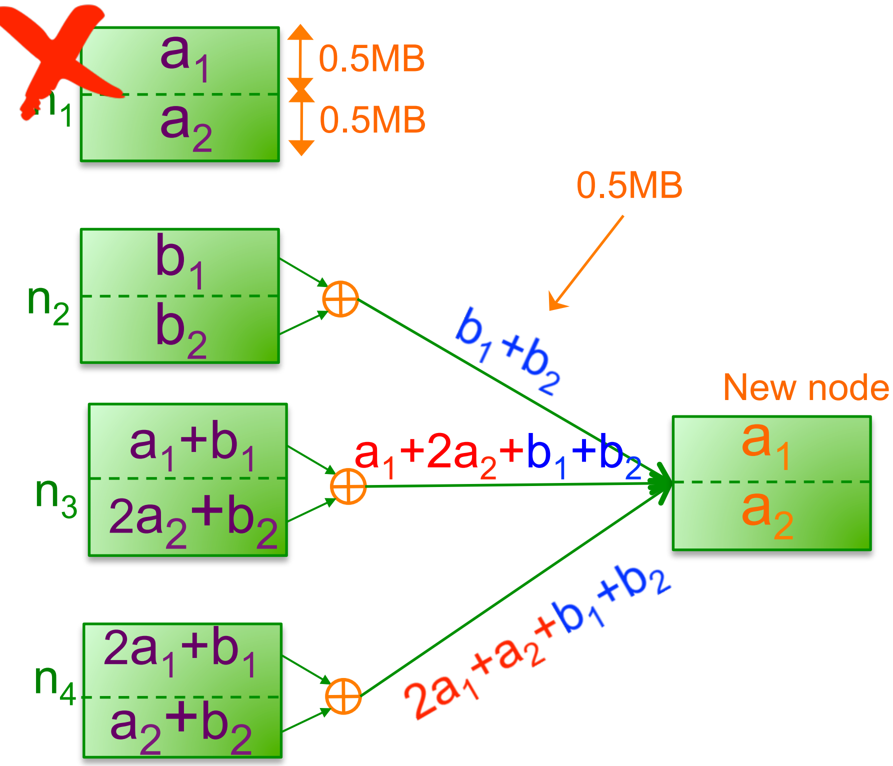
Regenerating Codes

Storage System

Keys

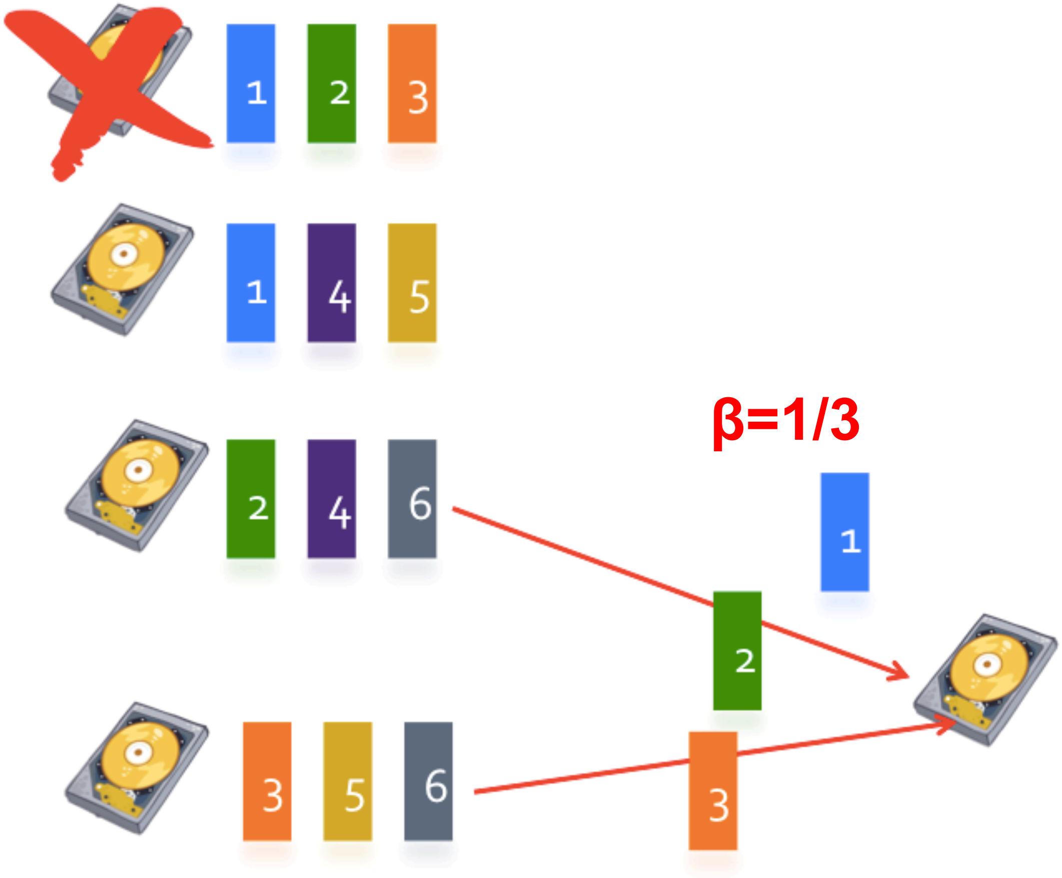
Separation is Optimal for Bandwith-Limited Regime

Surprising result #1: Separation is NOT Optimal



(n,k,d)= (4,2,3)

α=1 β=1/2



Secret Size=1/2MB

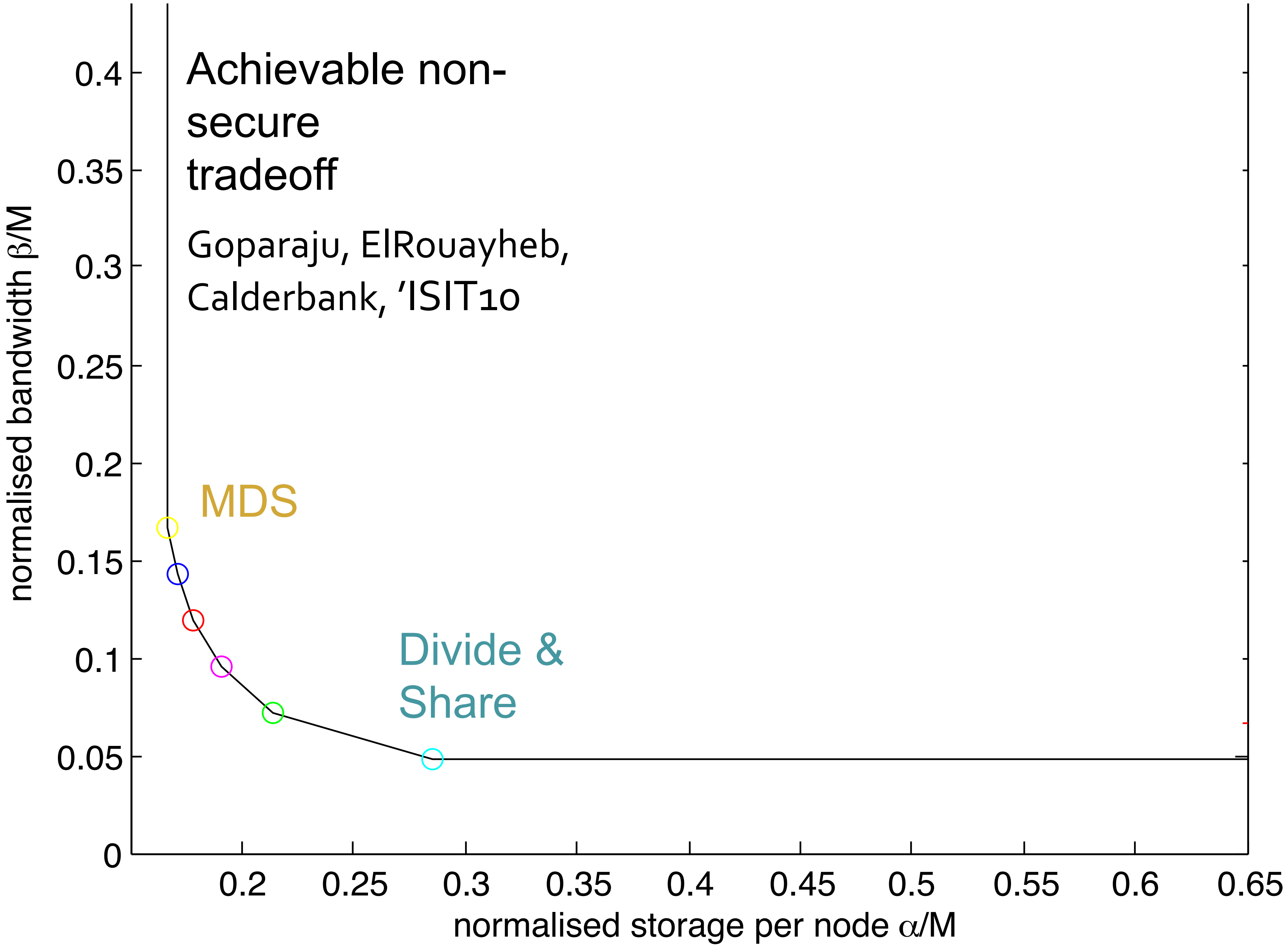
Secret Size=2/3MB

It may be better not to use all your budgeted bandwidth or storage!

Falling back to bandwidth-limited regime codes is always optimal for (n,n-1,n-1) systems

Tandon et al. ’10

Finding the Optimal Inner Code is not trivial



Achievable non-secure tradeoff

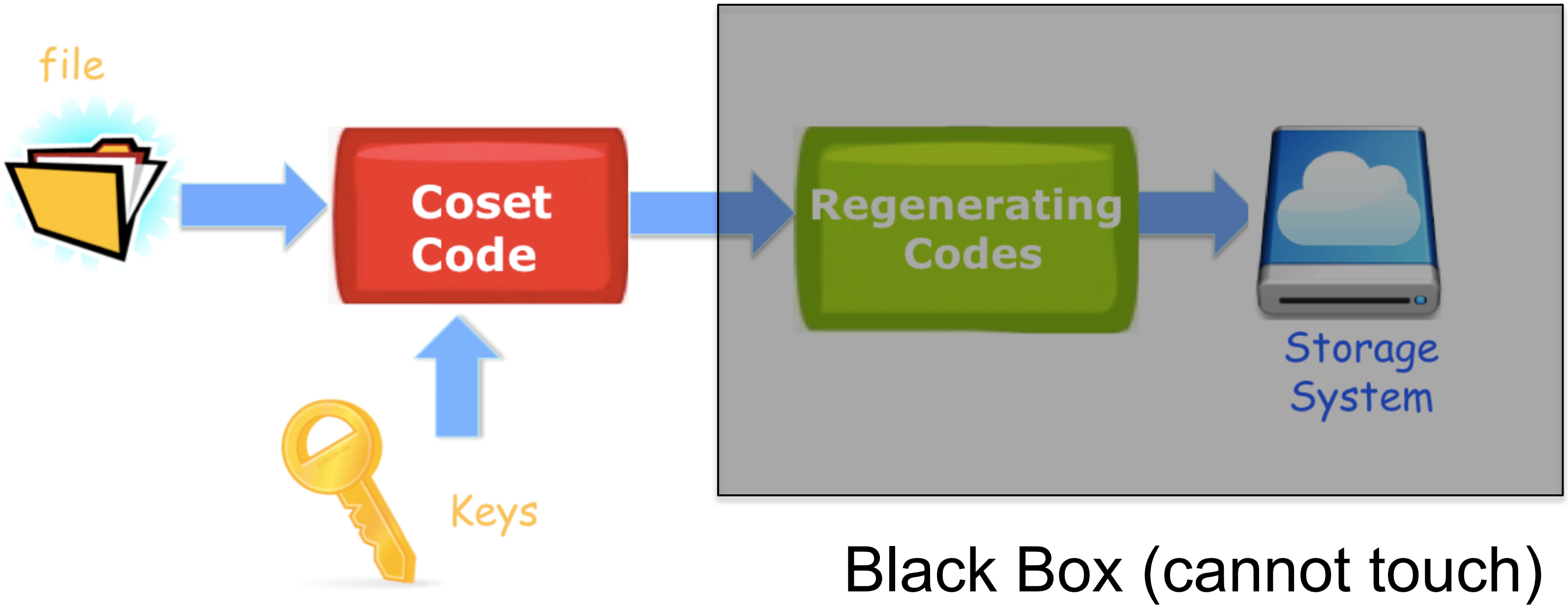
Goparaju, ElRouayheb, Calderbank, 'ISIT10

MDS

Divide & Share

(n,k,d)=(7,6,6)

What is the best we can do with a Separation Scheme



Black Box (cannot touch)

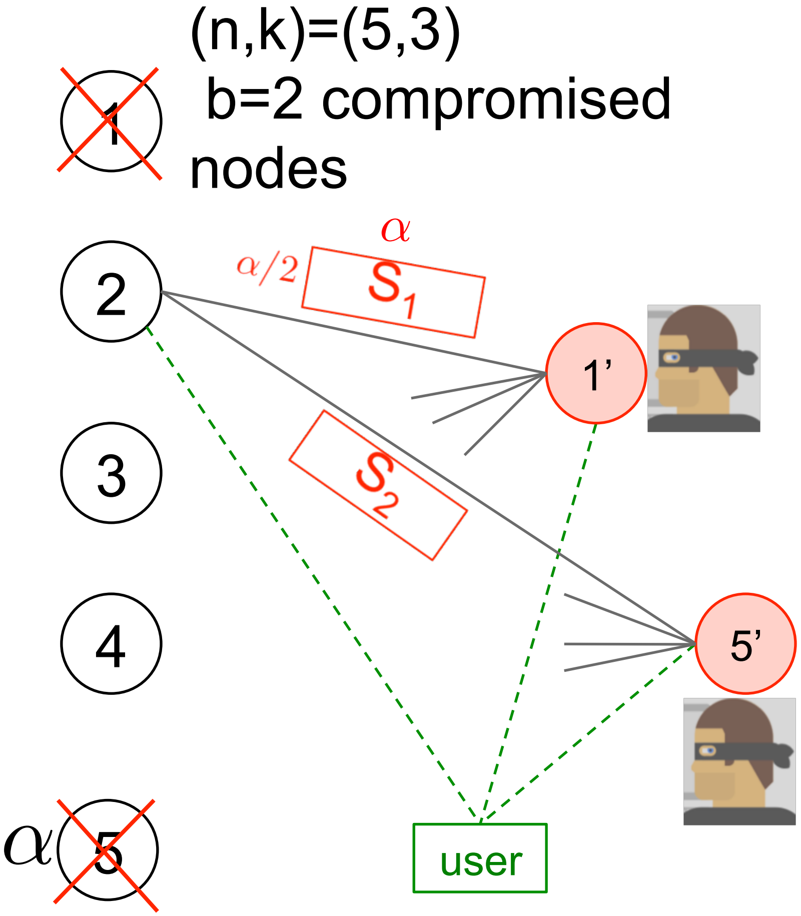
- Simpler design if we want different files with different security requirements
- Cloud user: does not have control over the code

Theorem: [Goparaju, R., Calderbank, Poor Netcod '13]

$$C_s^* = (k-b)\left(1 - \frac{1}{n-k}\right)^b \alpha$$

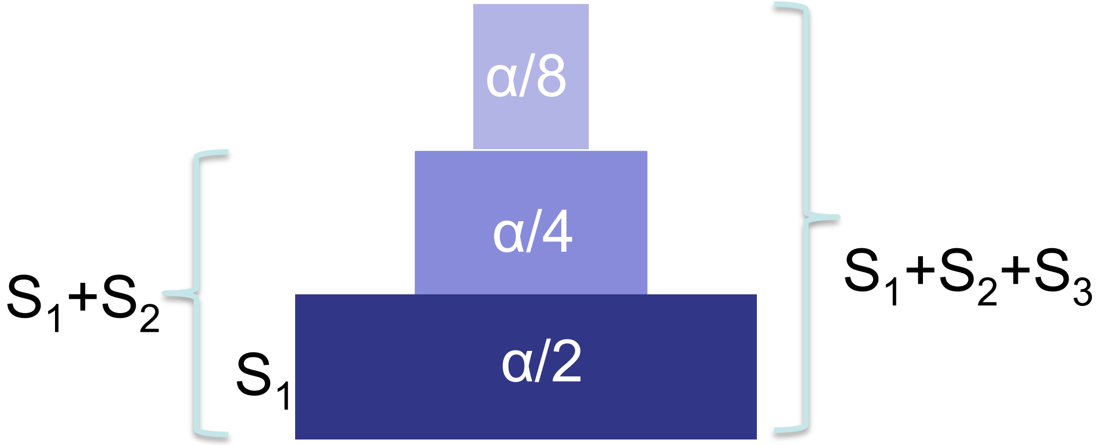
Surprising result #2

Proof based on Geometry of Repair Spaces



(n,k)=(5,3)
b=2 compromised nodes

Data observed by Eve = $b\alpha$
Data stored on nodes 1’ and 2’
+ $\dim(S_1+S_2)$
Data downloaded from node 2



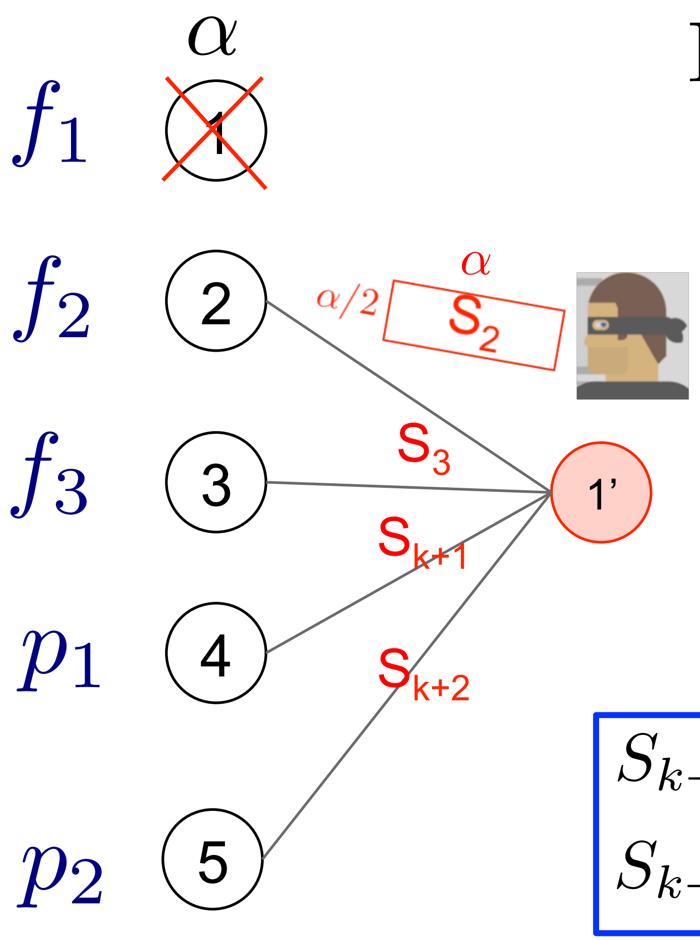
Secure (linear) capacity= kα – amount observed by Eve

$$C_s^* \leq (k-b)\frac{\alpha}{2^b}$$

Theorem: [Goparaju, R., Calderbank, Poor Netcod ’13]

$$\dim(S_{i_1}+S_{i_2}+\cdots+S_{i_b}) \geq \frac{\alpha}{2}+\frac{\alpha}{2^2}+\cdots+\frac{\alpha}{2^b}$$

A Taste of the Proof…



$$\text{File:}(f_1, \ldots, f_k) \;\; f_i = (f_{i1}, \ldots, f_{i\alpha})$$

$$p_1 = \sum_{i=1}^{k} A_i f_i, p_2 = \sum_{i=1}^{k} B_i f_i$$

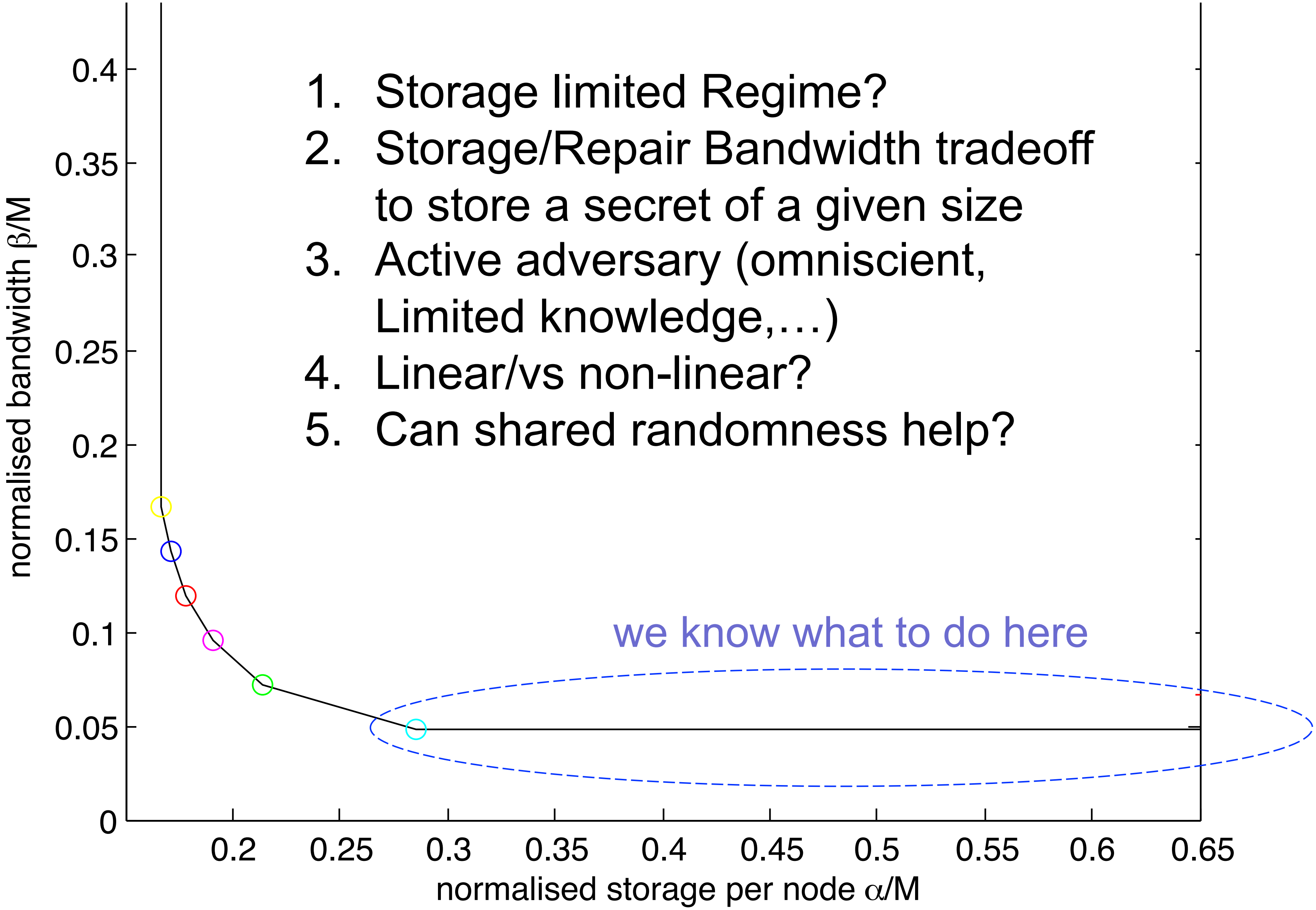
- Node 1’ downloads:

$$\begin{array}{ccccccc} & & S_2 f_2 & & S_3 f_3 & & S_k f_k \\ S_{k+1}A_1f_1 & + & S_{k+1}A_2f_2 & + & \cdots & + & S_{k+1}A_kf_k \\ S_{k+2}B_1f_1 & + & S_{k+2}B_2f_2 & + & \cdots & + & S_{k+2}B_kf_k \end{array}$$

$S_{k+1}A_1 + S_{k+2}B_1 = \mathbf{F}_q^n$ $S_2 = S_{k+1}A_2 = S_{k+2}B_2$ $S_k = S_{k+2}A_k = S_{k+1}B_k$

- Analogy to interference alignment
- Write these subspace conditions for all failures
- Use them to proof theorem by induction

Open Problems



1. Storage limited Regime?
2. Storage/Repair Bandwidth tradeoff to store a secret of a given size
3. Active adversary (omniscient, Limited knowledge,…)
4. Linear/vs non-linear?
5. Can shared randomness help?

QUESTIONS?