

“Trustworthiness is a holistic property, encompassing **security** (conventionally including confidentiality, integrity, and availability), **correctness, reliability, privacy, safety, and survivability.”**

Fred. B. Schneider

HINDSIGHT AND FORESIGHT ON CODING THEORY WITH A SYSTEMS PERSPECTIVE

Vero Estrada-Galiñanes

Associate Professor
University of Stavanger



Contact: veronica.estrada@uis.no Twitter: @GalinanesVero

ON-GOING AND FUTURE WORK

ON-GOING AND FUTURE WORK

BBCHAIN PROJECT   

STORAGE      


    



HEALTH, QUALITY OF LIFE & WELL-BEING    



OUTREACH   

TEACHING 

CRAZY IDEAS FOR CREDENCE... 

BBCHAIN PROJECT 

STORAGE  

HEALTH, QUALITY OF LIFE & WELL-BEING  

OUTREACH

TEACHING

CRAZY IDEAS FOR CREDENCE...

PREVIOUS RESEARCH, INDUSTRIAL AND GOVERNMENT PROJECTS

INTRUSION DETECTION
SYSTEMS &
MACHINE LEARNING



VESSEL MONITORING SYSTEM



ENERGY INDUSTRY



SYSTEMS PERSPECTIVE

in the real world!



Efficient system or machine

Achieving maximum productivity with minimum wasted effort or expense

Example: a datacenter, a system for more efficient processing of information



Effective system or machine

Successful in producing a desired or intended result

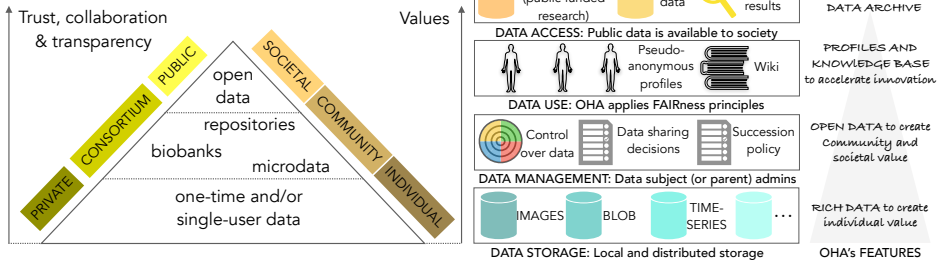
Example: privacy by design, censorship-resistant, and user-centric systems



COLLECTING, EXPLORING, AND SHARING PERSONAL DATA*



OPEN HEALTH ARCHIVE

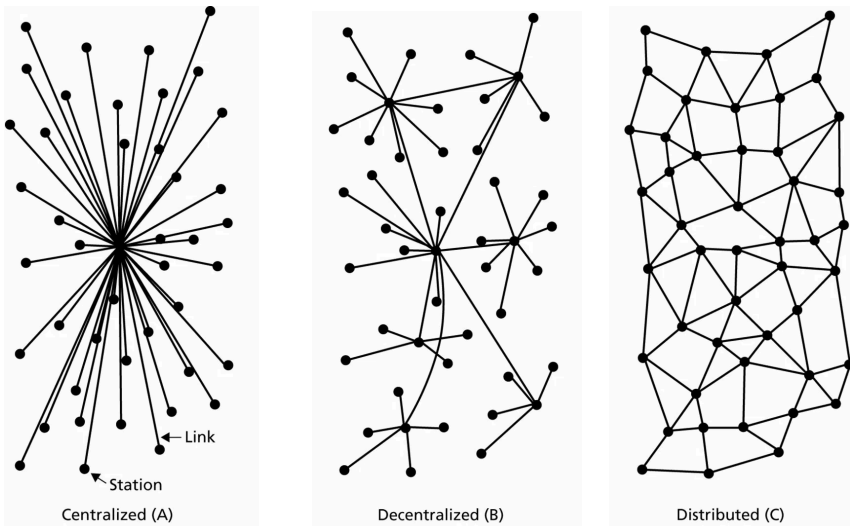


Find out more here * V. Estrada-Galiñanes, K. V. Wac. "Collecting, exploring and sharing personal data: why, how and where" (journal manuscript under peer-review, 2019)

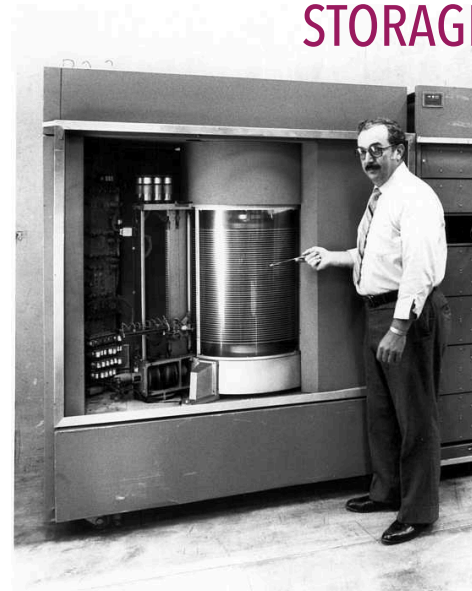


DECENTRALISED SYSTEMS

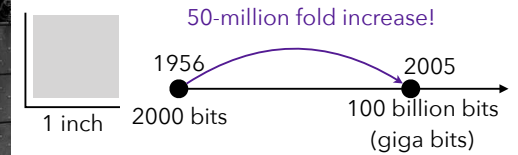
PAUL BARAN: CENTRALIZED, DECENTRALIZED AND DISTRIBUTED NETWORKS (1964)



STORAGE SYSTEMS



IBM 350 Disk Storage System
3.75MB - 1956



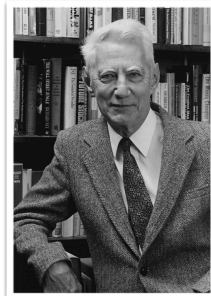
CRYPTO AND INFORMATION THEORY



Alan Turing - Modern crypto
1912-1954

Share many concepts and methods

- Measure of information
- Coding and decoding methods



Claude Shannon - Information theory
1916-2001

Protect a message against eavesdropper who may manipulate the messages

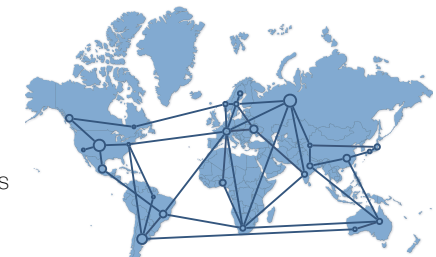
data integrity may refer to different things

Protect a message for transmission errors



High Availability, Scalable Storage, Dynamic Peer Networks: Pick Two Blake, Rodriguez (2003)

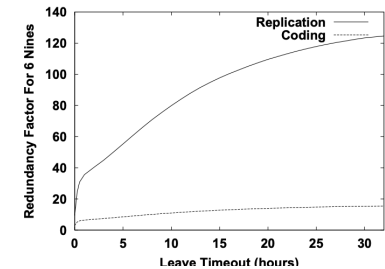
- Hosts are distributed around the globe
- They are unreliable, low-available...
- High availability (99.9999% or 6 "nines") is very expensive



Using replication (~120 copies)

Using RS codes (~15 copies)

- Maintaining redundancy may require too much bandwidth



PEER-TO-PEER SYSTEMS

WUALA: ONLINE STORAGE WITH THE POWER OF P2P



1GB free



Users can get more by sharing local disk space

low node availability (p) - min. requirement 4 h per day

FILE AVAILABILITY?

REPLICAS: Use k copies

$$1 - (1 - p)^k$$

$$1 - (1 - 0.25)^5 = 0.763$$

$$1 - (1 - 0.25)^{24} = 0.999 \quad \text{"3 nines availability"}$$

24x storage overhead

ERASURE CODES: encode m fragments into n fragments

$$\sum_{i=m}^n \binom{n}{i} p^i (1 - p)^{n-i}$$

Reed Solomon codes (optimal codes)

m=100, n = 517

3 nines availability

5x storage overhead



PEER-TO-PEER SYSTEMS

WUALA: ONLINE STORAGE WITH THE POWER OF P2P

2008-2015

"Wuala discontinued the P2P storage and moved completely to cloud storage in part motivated by software complexity and instability."*

more complicated to describe. It would be treacherously easy for the casual reader to dismiss the entire concept as impractically complicated--especially if he is unfamiliar with the ease with which logical transformations can be performed in a time-shared digital apparatus. The temptation to throw up one's hands and decide that it is all "too complicated," or to say, "It will require a mountain of equipment which we all know is unreliable," should be deferred until the fine print has been read.

PAUL BARAN:
CENTRALIZED,
DECENTRALIZED AND
DISTRIBUTED
NETWORKS (1964)

* Pedro García López, Alberto Montresor and Anwitaman Datta. "Please, do not decentralize the Internet with (permissionless) blockchains!" In Proc. of the 39th International Conference on Distributed Computing Systems, volume abs/1904.13093 of ICDCS'19, 2019



[tahoe-dev] erasure coding makes files more fragile, not less* - Zooko Wilcox-O'Hearn

I've heard many stories of people losing their files from a Tahoe-LAFS grid even though they had erasure coding parameters that provide massive fault tolerance such as 3-of-10 or 4-of-8. In fact, I think approximately 90% of all files that have ever been stored on a Tahoe-LAFS grid have died. (That's excluding all of the files of all of the customers of allmydata.com, which went out of business.)

My conclusion: if you care about the longevity of your files, forget about erasure coding and concentrate on monitoring. (Go ahead and use 3-of-10 because everyone does, and it adds a reasonably low level of storage overhead.)

CEO & Founder of ZCash, Zooko has more than 20 years of experience in open, decentralized systems, cryptography and information security, and startups. He is recognized for his work on DigiCash, Mojo Nation, ZRTP, "Zooko's Triangle", Tahoe-LAFS, BLAKE2, and SPHINCS.

$$\sum_{i=m}^n \binom{n}{i} p^i (1-p)^{n-i}$$

Reed Solomon codes (optimal codes)
 m=3, n = 10, p = 0.25
 0.474 availability

p=0.75
 0.999 availability

* <https://tahoe-lafs.org/pipermail/tahoe-dev/2012-March/007185.html>

[tahoe-dev] erasure coding makes files more fragile, not less* - Zooko Wilcox-O'Hearn

In closing we point out a challenge that we faced in testing our system for which we have no systematic solution. By their very nature, fault-tolerant systems try to mask problems. Thus they can mask bugs or configuration problems while insidiously lowering their own fault-tolerance. For example, we have observed the following scenario. We once started a system with five replicas, but misspelled the name of one of the replicas in the initial group. The system appeared to run correctly as the four correctly configured replicas were able to make progress. Further, the fifth replica continuously ran in catch-up mode and therefore appeared to run correctly as well. However in this configuration the system only tolerates one faulty replica instead of the expected two. We now have processes in place to detect this particular type of problem. We have no way of knowing if there are other bugs/misconfigurations that are masked by fault-tolerance.

ZOOKO'S TAKE AWAY: the more powerful your fault-tolerance technology is, the more powerful you need your monitoring technology to be

* <https://tahoe-lafs.org/pipermail/tahoe-dev/2012-March/007186.html>

Paxos Made Live - An Engineering Perspective (2006 Invited Talk)

Tyler Chartre, Robert Griesemer, and Joshua Riedinger
 Google Inc.

ABSTRACT
 Distributed replication is a difficult problem. In this paper, we describe the design and implementation of Paxos, a distributed replication algorithm that is simple, efficient, and fault-tolerant. Paxos is a distributed replication algorithm that is simple, efficient, and fault-tolerant. Paxos is a distributed replication algorithm that is simple, efficient, and fault-tolerant. Paxos is a distributed replication algorithm that is simple, efficient, and fault-tolerant.

MEANWHILE, IN SWITZERLAND

There should be a way to mix data in the system to increase reliability...

Previous worked used "entanglement" to protect data against censorship but the approach didn't work.

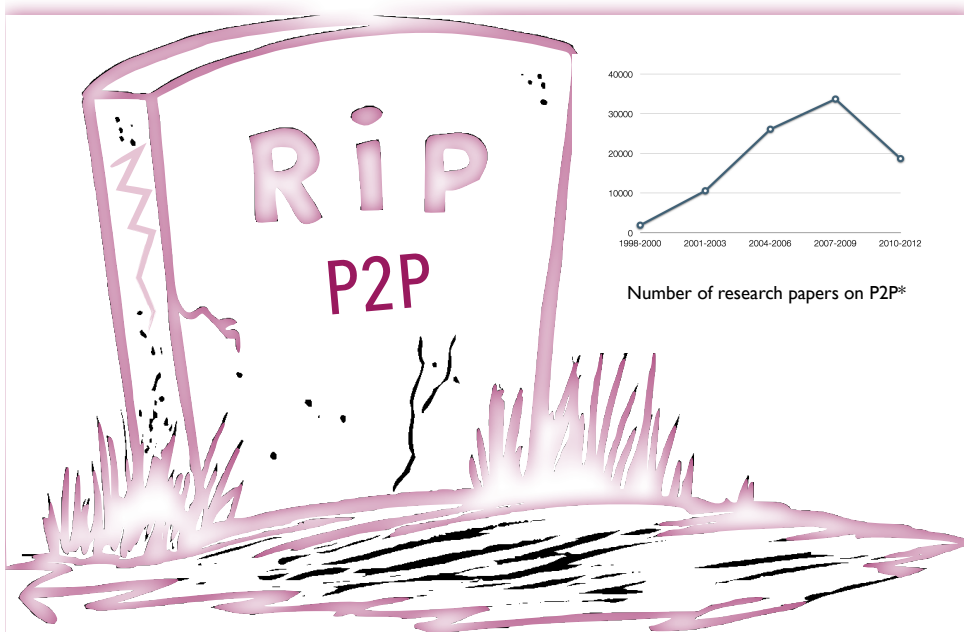


Jedi master

"We show that entanglement as provided by Dagster and Tangler is not by itself sufficiently strong to deter a dishonest storage provider from tampering with data, because not enough documents get deleted on average when destroying a block of a typical document." **Towards a Theory of Data Entanglement, Aspnes et al.**



Jedi apprentice



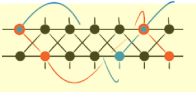
* Li, B., Feng, Y., & Li, B. (2012). Rise and fall of the peer-to-peer empire. *Tsinghua science and technology*, 17(1), 1-16.

THE STUDY OF ENTANGLEMENT CODES

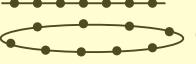
How to provide more protection with a practical and fair algorithm?

How to run fair evaluations/comparisons?

Helical entanglements codes (p-HEC), SSS'13
Estrada, Felber



Open/Closed simple entanglements, IPCC'16
Estrada, Páris, Felber




Research playground to compare codes, SYSTOR'17
Estrada

Stripe	Block	Location	Available	Repaired
100	0	34	True	False
100	8	38	True	False

Left Id	Right Id	Block Type	Location	Available	Repaired
26	29	d	41	True	False
26	31	h	7	True	False

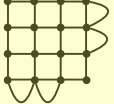
Practical erasure codes for storage systems: The study of entanglement codes, an approach that propagates redundancy to increase reliability and performance, **PhD Thesis'17**
Estrada, Thesis Director Pascal Felber



Failure-repair dynamics, a new approach to compare erasure/network codes
Estrada

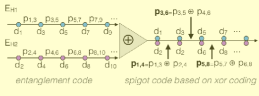
2012

Evaluation and comparisons of p-HEC, **CLUSTER'15**
Estrada, Felber



Two-dimensional entangled square arrays, **IPCC'17**
Páris, Estrada, Amer, Fincon

Spigot codes, See/cite PhD Thesis while full paper version is not ready



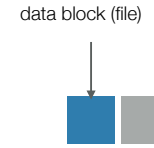
Alpha entanglement codes, **AE(α,s,p), DSN'18**
Estrada, Miller, Felber, Páris

Eliminating triple-failures in two-dimensional arrays
Páris, Estrada



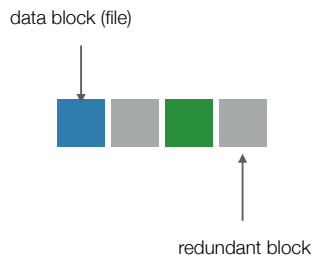
ALPHA ENTANGLEMENTS: SINGLE CHAIN ($\alpha=1$)

XORing blocks propagates redundant data



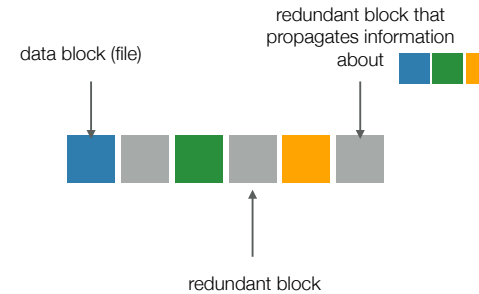
ALPHA ENTANGLEMENTS: SINGLE CHAIN ($\alpha=1$)

XORing blocks propagates redundant data



ALPHA ENTANGLEMENTS: SINGLE CHAIN ($\alpha=1$)

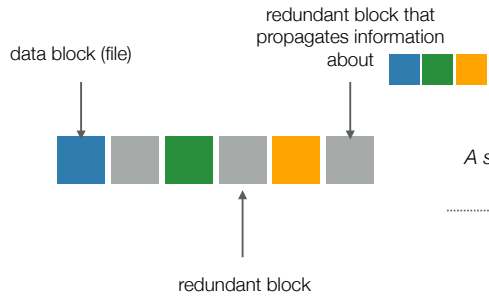
XORing blocks propagates redundant data



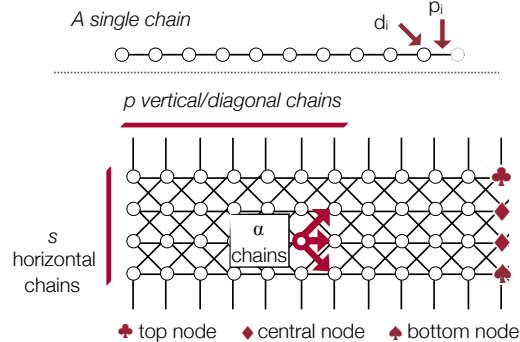


ALPHA ENTANGLEMENTS: MULTIPLE CHAINS ($\alpha > 1$)

XORing blocks propagates redundant data

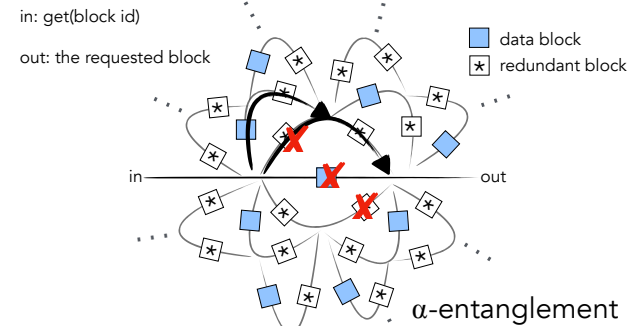


- Each data block belongs to multiple entanglement chains
- All the chains are intertwined.
- The encoder only needs to keep the last elements of all of the chains.



RETHINKING REDUNDANCY: ALPHA ENTANGLEMENTS

Redundancy Propagation Quasi-Sphere



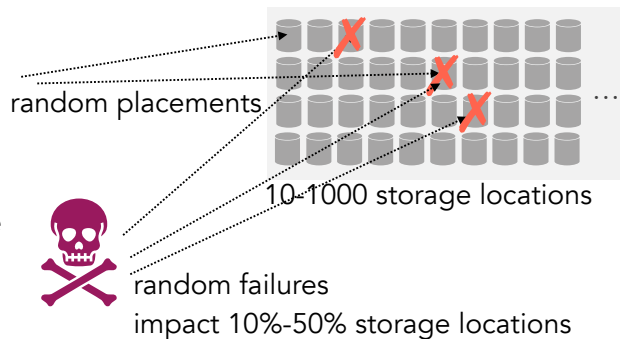
- Paths that are closer to the centre have less elements in serial combinations
- Repair effort for a single block scales with the size of a failure



RETHINKING REDUNDANCY: ALPHA ENTANGLEMENTS

1M data blocks

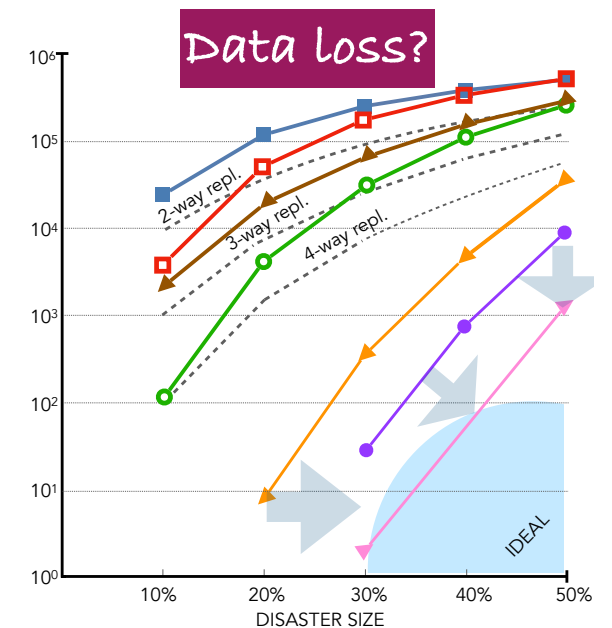
Assumption: minimum maintenance



ALPHA ENTANGLEMENTS

Evaluations

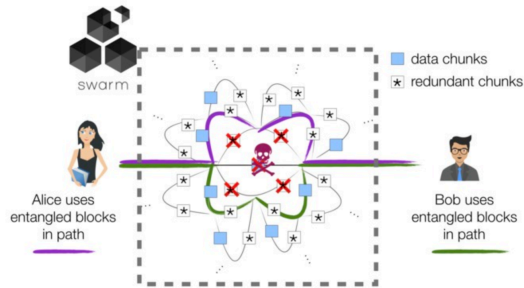
Code	Redundancy	Equivalent to
RS(8,2)	Low	mirroring 2x
RS(10,4)	Low	mirroring 2x
RS(5,5)	Equivalent to mirroring 2x	
AE(1,-,-)	Equivalent to mirroring 2x	
AE(2,2,5)	Equivalent to triplication 3x	
RS(4,12)	High	
AE(3,2,5)	High redundancy 4x	



DECENTRALISED STORAGE SYSTEMS



PROOF OF CONCEPT



SUBMITTED TO

Ethereummadrid Hackathon 2019
WINNER First Price

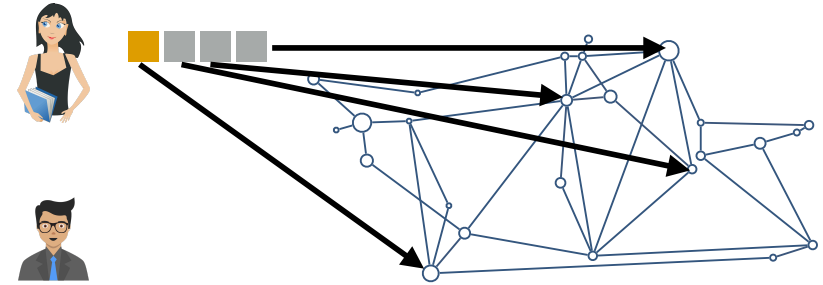
CREATED BY

- Vero Estrada-Gallinanes
- Edgar García de Pereda
- Racin Nygaard
- Gus

Entanglements in action: redundancy, repairs, load balance, integrity, ...

37

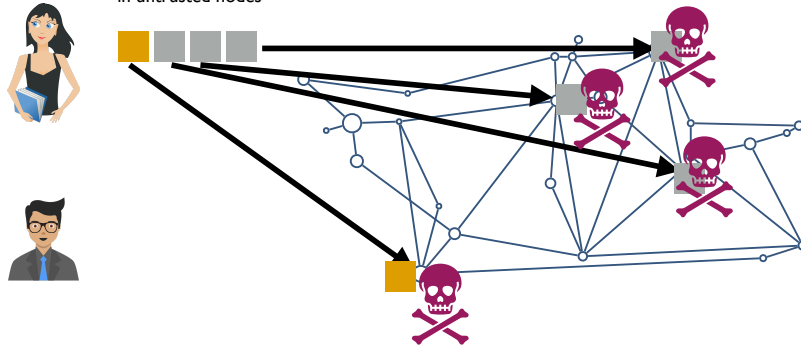
UNTRUSTED NODES



38

UNTRUSTED NODES

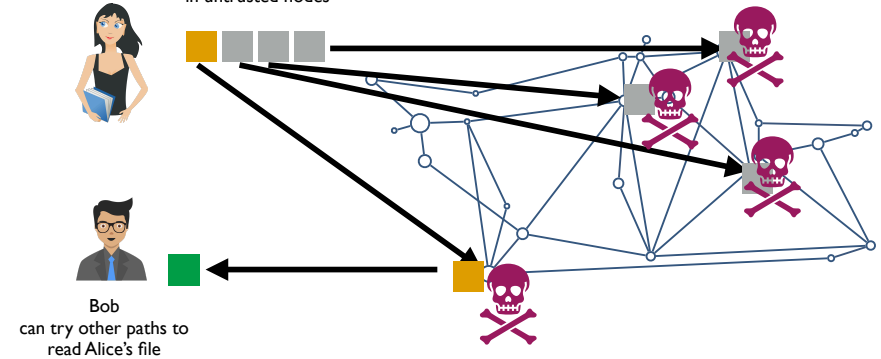
Alice (without knowing) stores the blocks in untrusted nodes



39

UNTRUSTED NODES

Alice (without knowing) stores the blocks in untrusted nodes



40



Discussing redundancy use cases in Swarm

Swarm Orange Summit
Ethereum Madrid Hackaton
2019

THANK YOU

@GalinanesVero