Secure File Fragment Allocation in Distributed Storage Systems with Heterogeneous Vulnerabilities

AUBURN UNIVERSITY

SAMUEL GINN COLLEGE OF ENGINEERING

Yun Tian and Xiao Qin

Computer Science and Software Engineering Department

Auburn University

Auburn, AL 36849

AUBURN UNIVERSITY

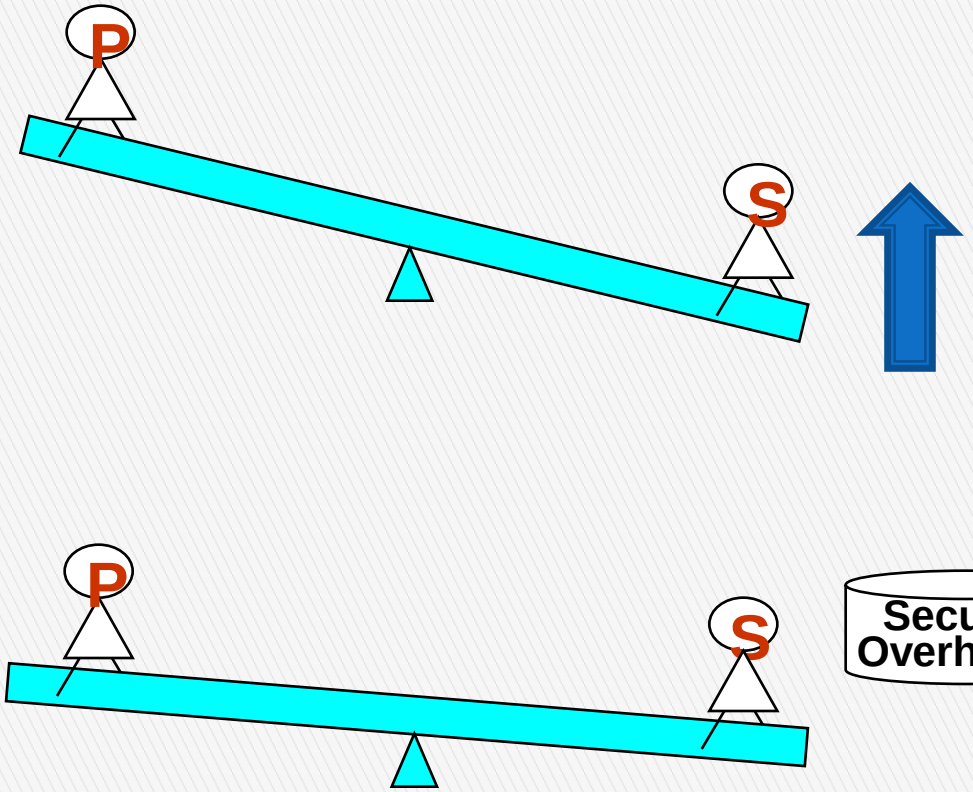
SAMUEL GINN COLLEGE OF ENGINEERING

Distributed Storage Systems

- Large-scale data processing
- Scalability, availability, performance, security

Security vs. Performance

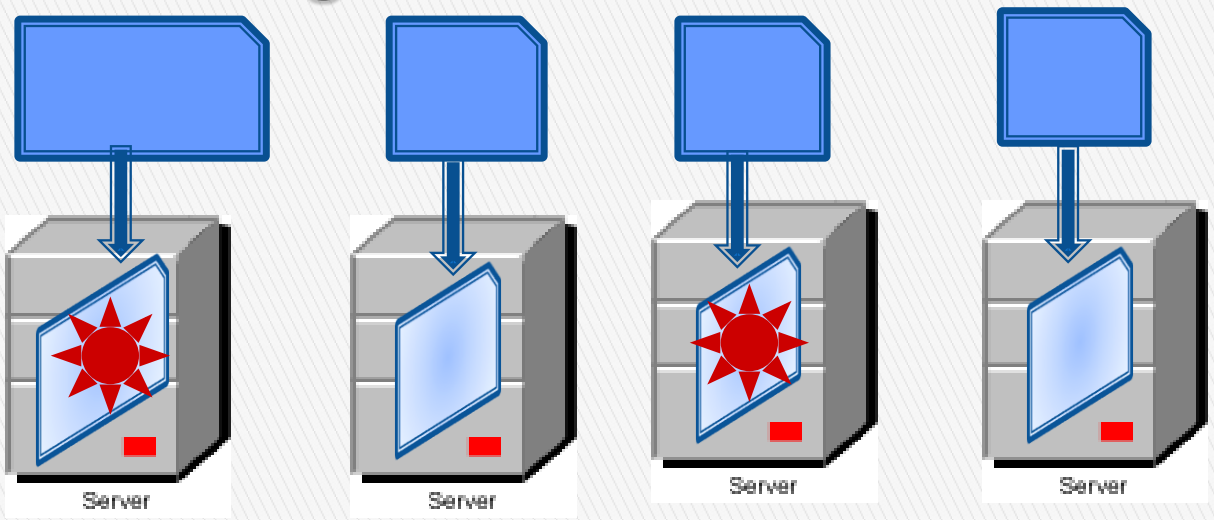
Security is achieved at the cost of performance degradation



AUBURN UNIVERSITY
SAMUEL GINN COLLEGE OF ENGINEERING

Secret Sharing

- *(m, n)* Secret sharing - e.g., (2, 3)
- File Fragmentation



Motivation: Storage nodes in a distributed system have heterogeneous vulnerabilities.

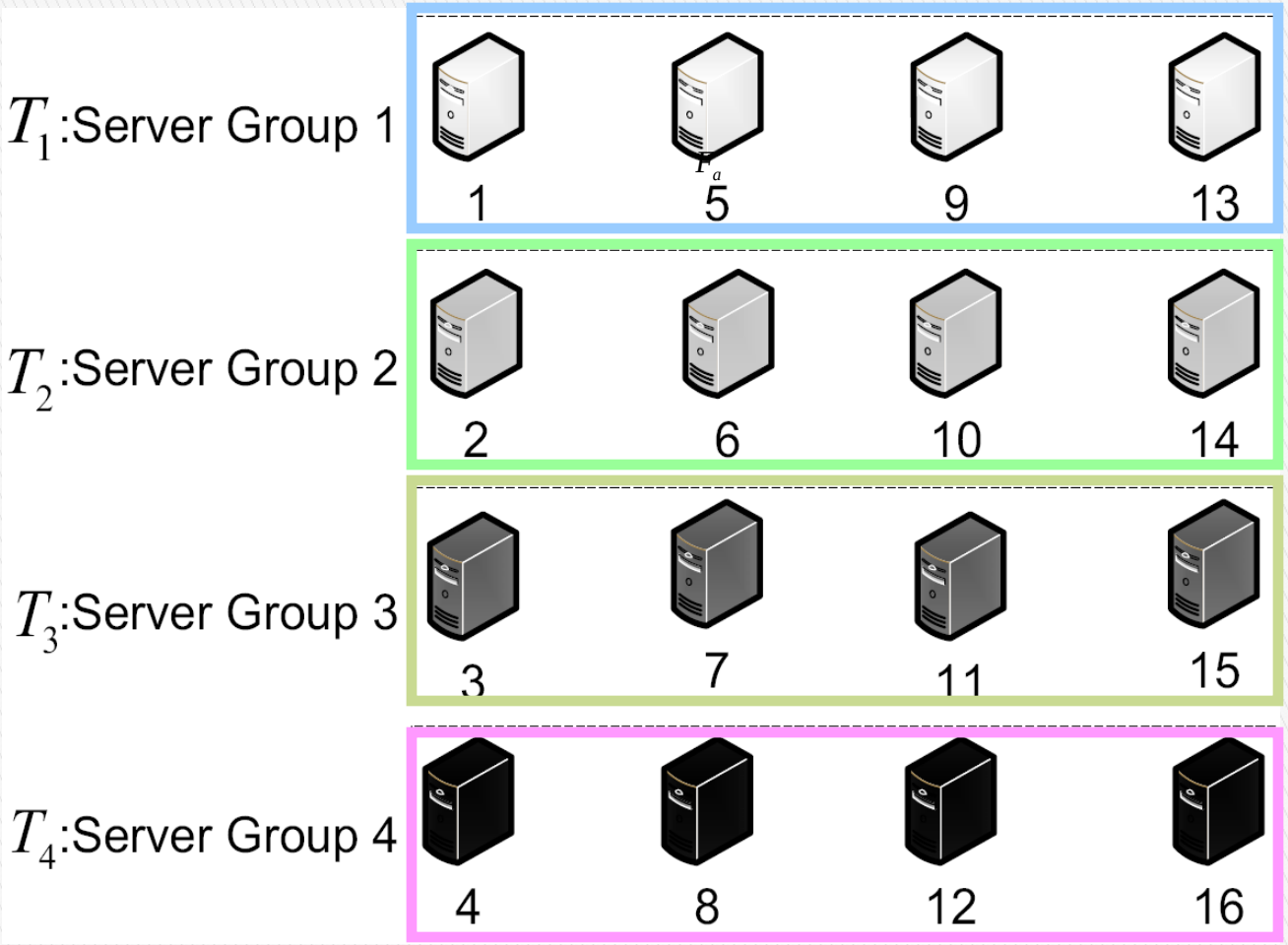
Heterogeneities in Large-Scale Distributed systems

- A wide variety of:
  - Hardware (e.g., SSDs, HDDs, Tapes)
  - Software (e.g, HDFS, Lustre, PVFS)

- Heterogeneities affect performance

Can we leverage heterogeneity features to improve security for distributed systems?

Considering Heterogeneity Issues



File F has 3 fragments: a, b, c

Fragment a

Fragment b

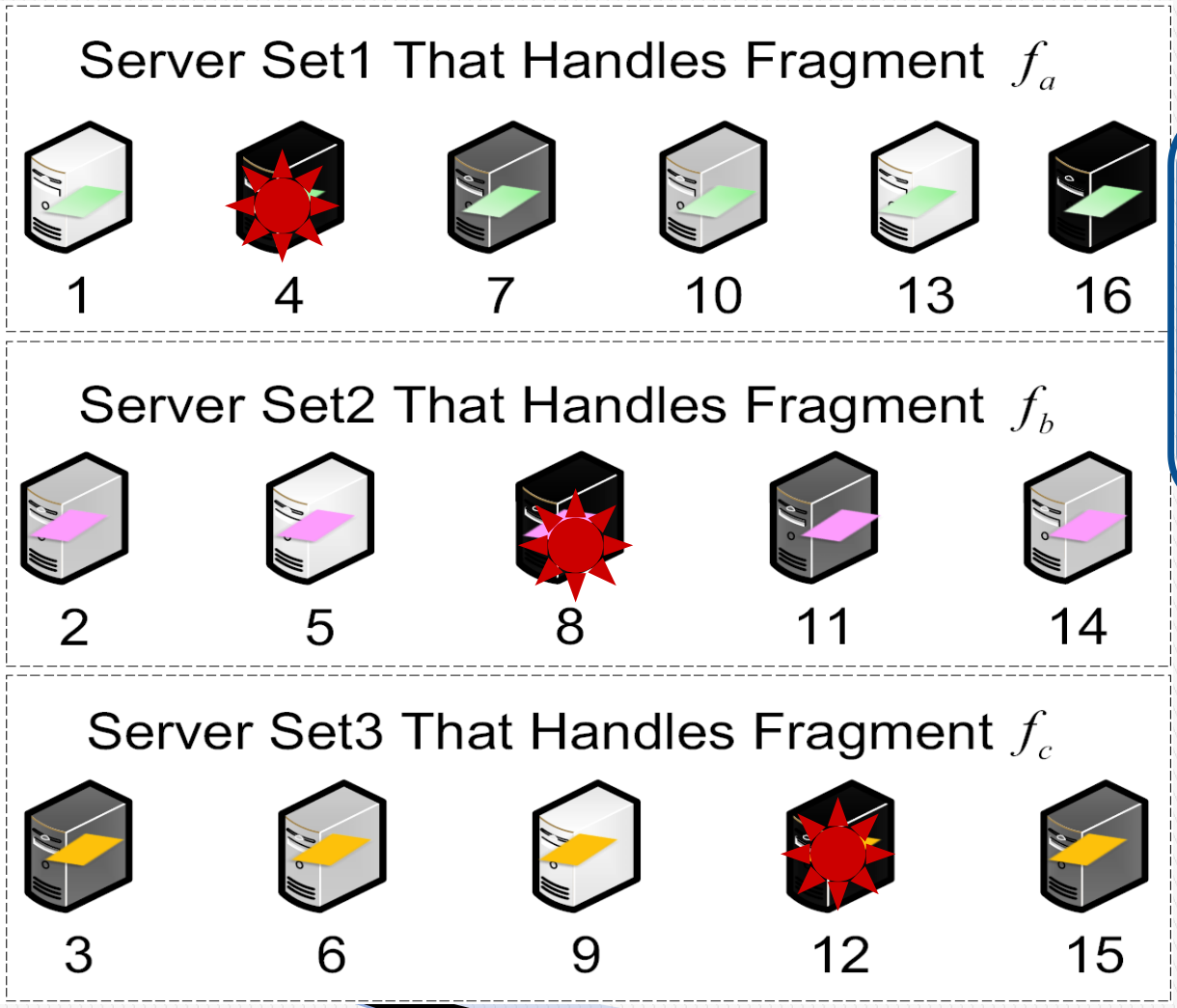
Fragment c

$$F(i) = j \bmod n$$

AUBURN UNIVERSITY

Samuel Ginn College of Engineering

Motivation: File Fragment Allocations



Problem?

Reconstruct

Fragment a

Fragment b

Fragment c

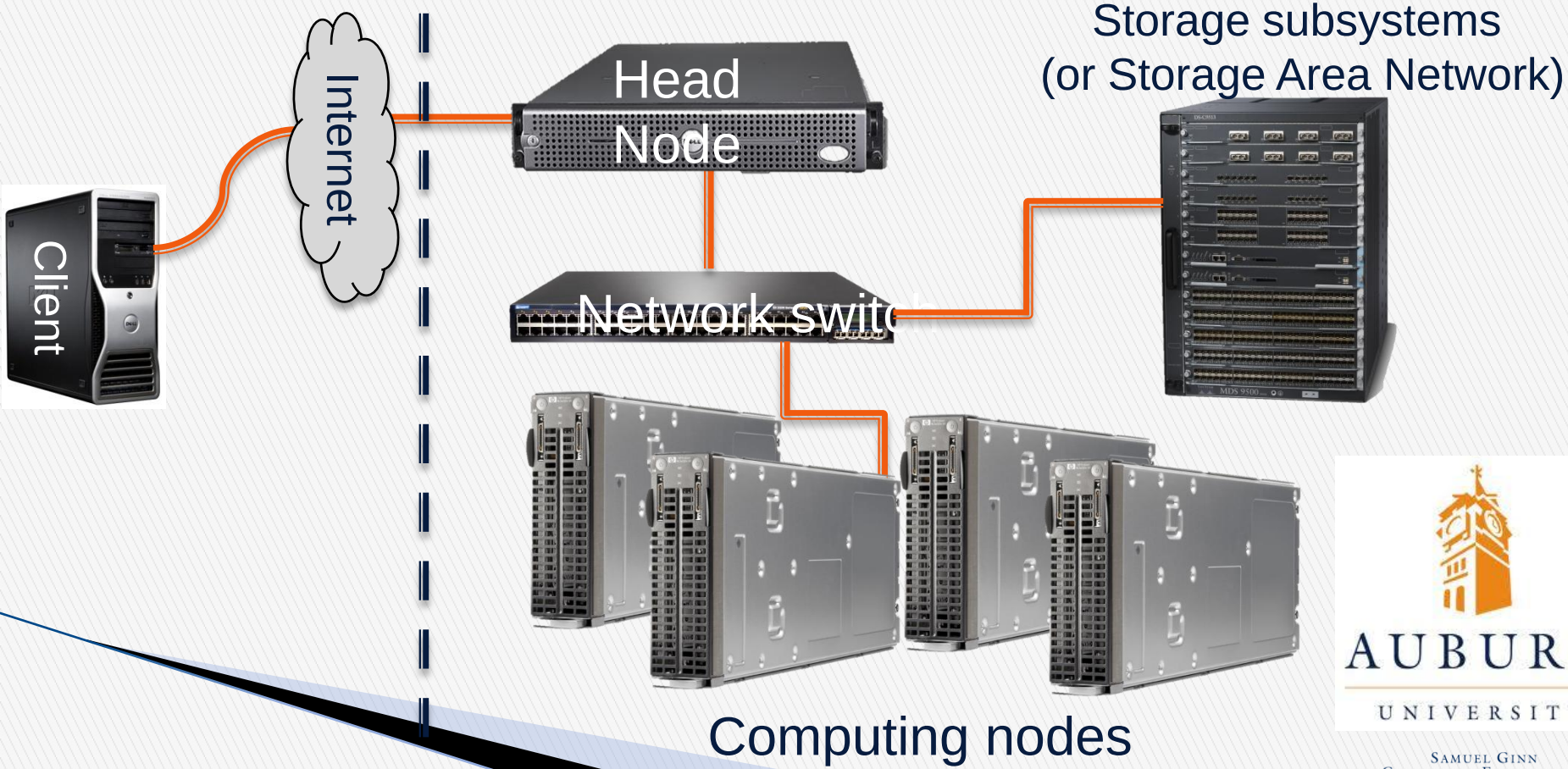
File F

AUBURN UNIVERSITY

Samuel Ginn College of Engineering

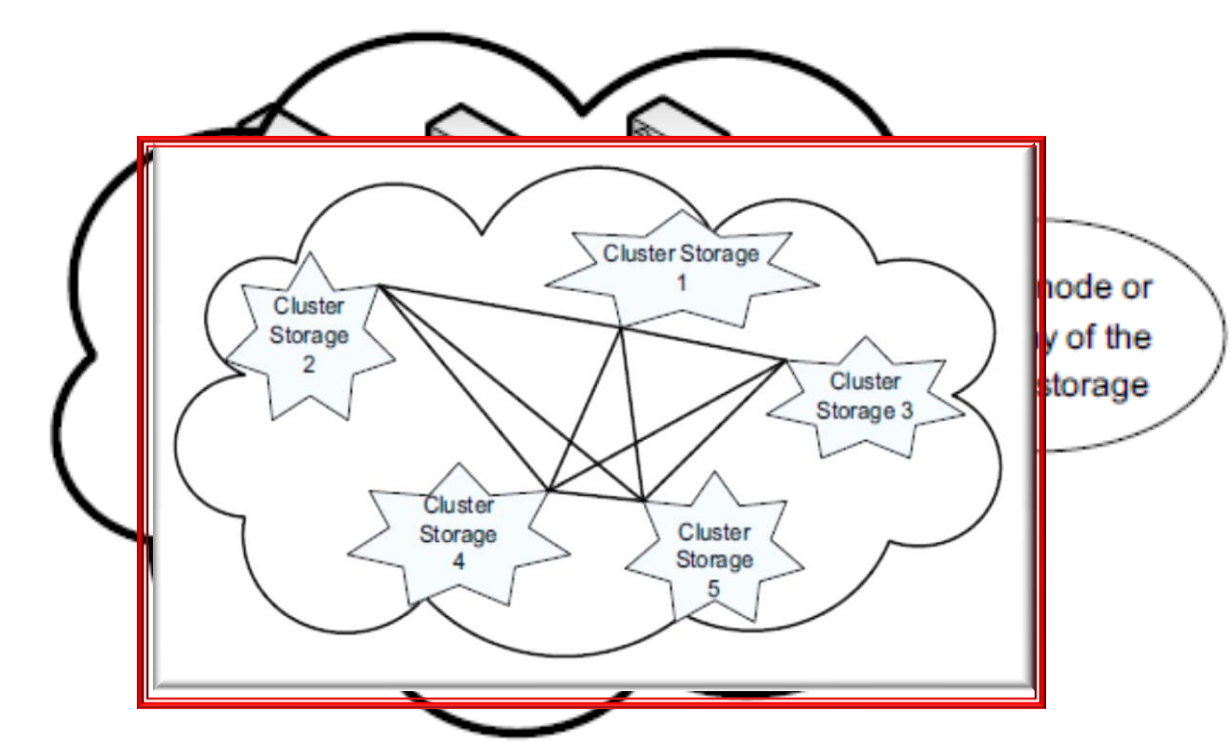
Cluster Storage Systems

- The Architecture of a Cluster



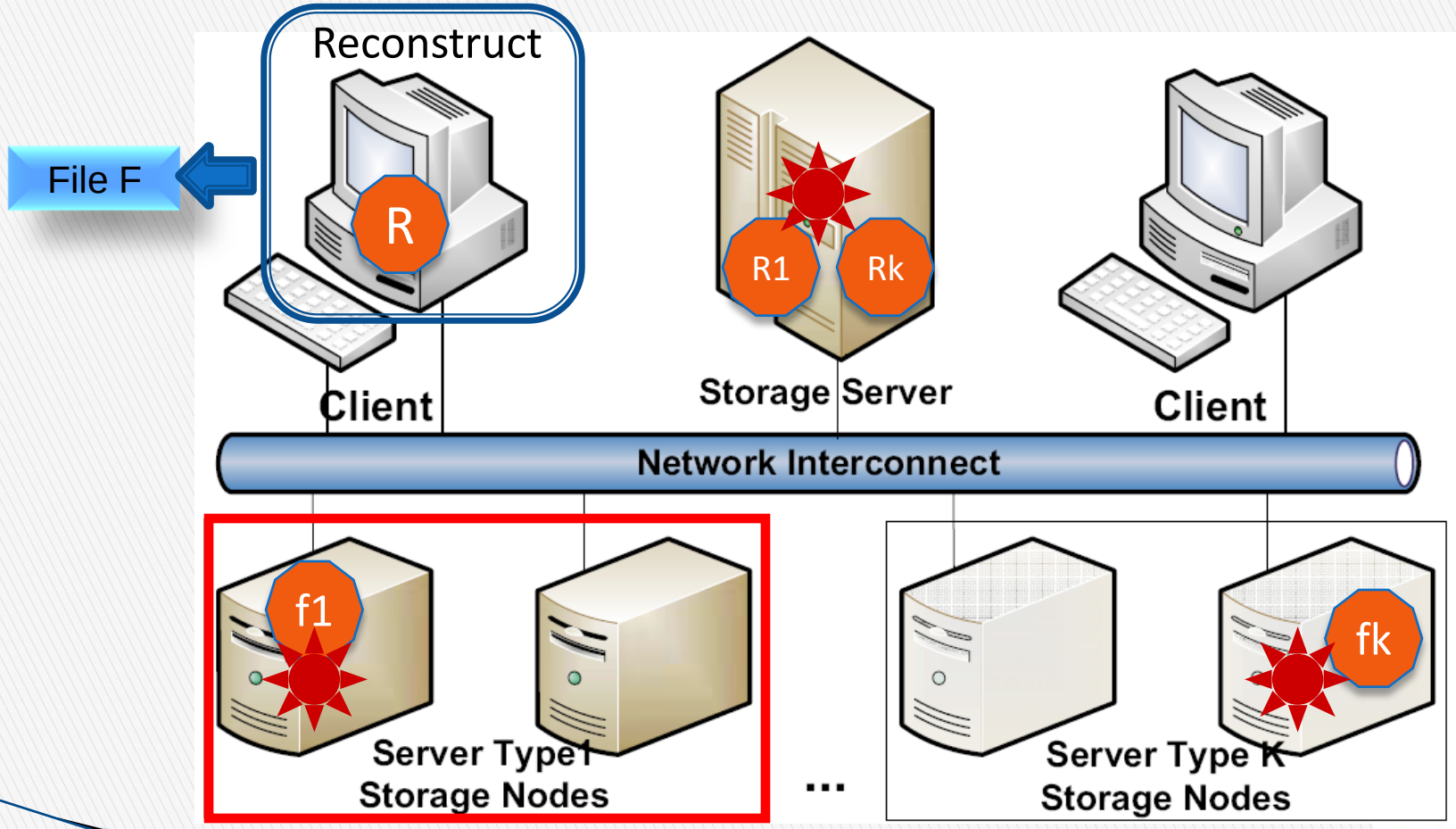
8

System Model



AUBURN UNIVERSITY
SAMUEL GINN COLLEGE OF ENGINEERING

A Prototype System



AUBURN UNIVERSITY
SAMUEL GINN COLLEGE OF ENGINEERING

The S-FAS Scheme

A file's fragment-allocation decisions are guided by four policies

- Multiple server groups
- Store fragments of a file across as many different server groups as possible
- Integrate the (m n) secret sharing scheme
- Allocate file fragments within a sub-system to improve I/O performance

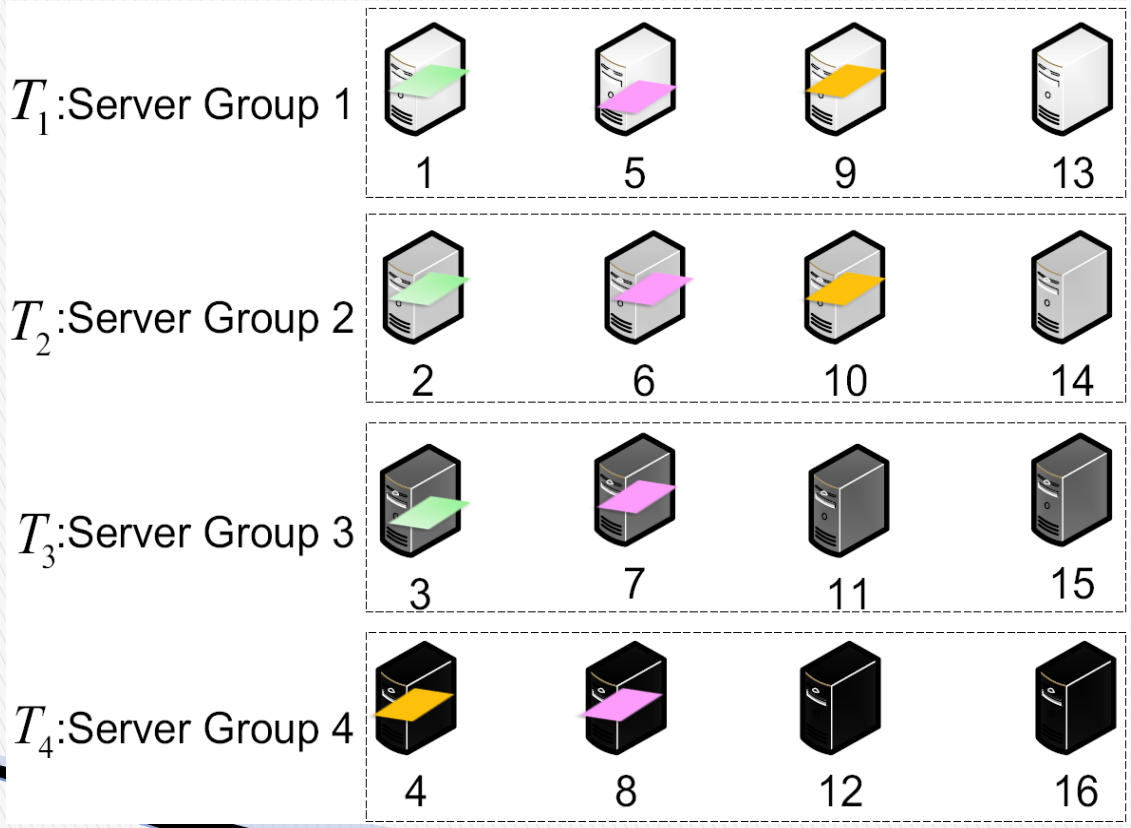
AUBURN UNIVERSITY
SAMUEL GINN COLLEGE OF ENGINEERING

Multiple server-type groups

- Diversity make sense?
  - A team with diversity make creativity;
  - A system with diversity may improve security;
  - We divide storage nodes of a system into different “server type” based on their different security level or strategy caused by hardware or software;

Spread fragment storage of a file

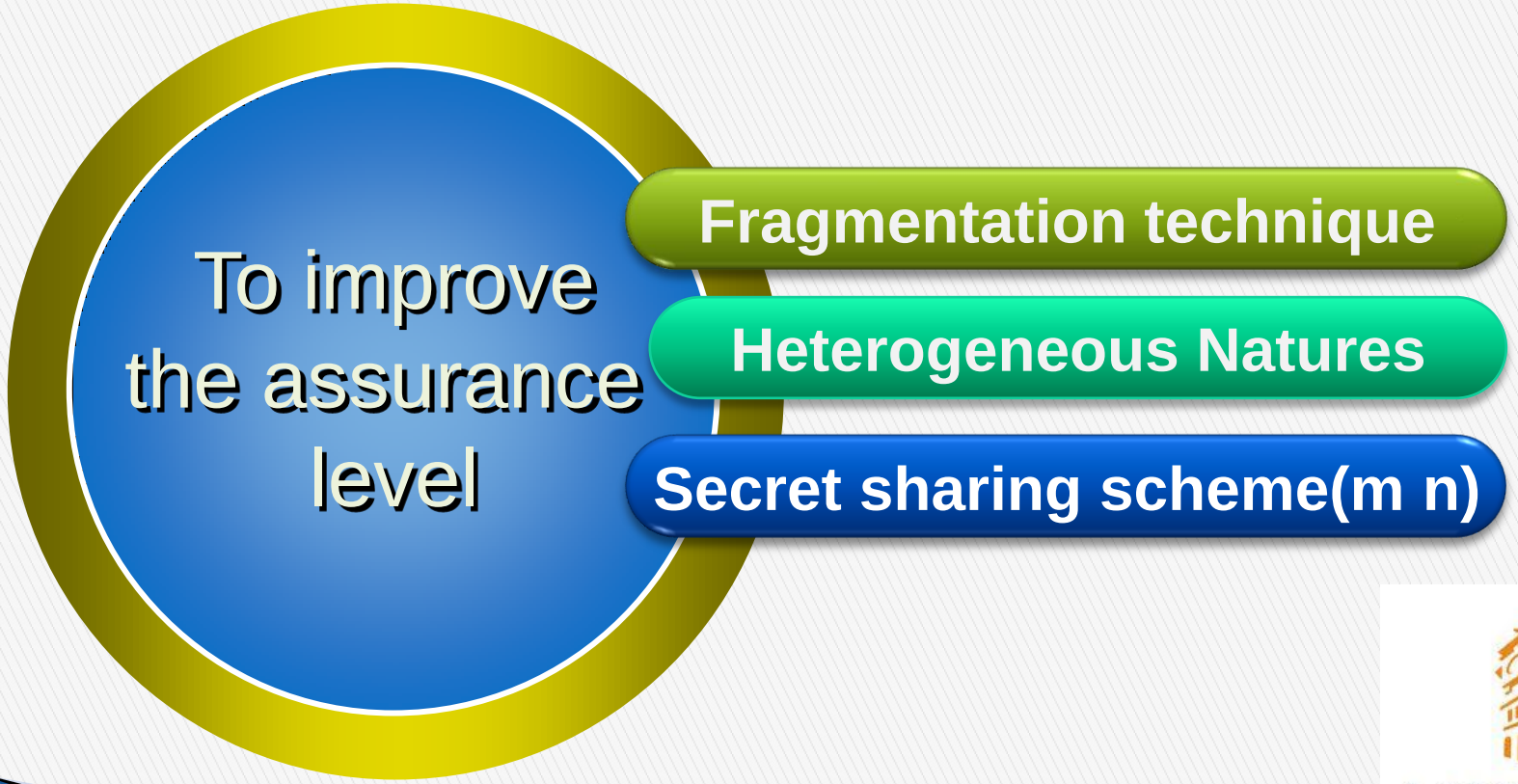
- Store fragments of a file across as many different server-type groups as possible



Allocate fragments of a file Within a Sub-system

- Allocate file fragments to improve I/O performance
  - Allocating fragments of a file into different storage clusters can degrade performance.
  - Our S-FAS scheme attempts to allocate fragments to storage nodes within a cluster.

Integrate the (m n) secret sharing scheme



AUBURN UNIVERSITY
SAMUEL GINN COLLEGE OF ENGINEERING

Notions and Meaning

Notations	Meaning
N	Number of storage nodes
K	Total number of server types
Sj	Size of a certain server type
(m,n)	Parameters of secret sharing scheme
α	Allocation mapping of file F

AUBURN UNIVERSITY

Samuel Ginn College of Engineering

Assurance Models

Static

$$\mathrm{SA}(\alpha) = 1 - \mathrm{P}(\mathrm{V}) = 1 - \sum_{\mathrm{j}=1}^{\mathrm{K}} \left(\frac{\mathrm{S_j}}{\mathrm{N}} \mathrm{P}(\mathrm{Z}) \sum_{\mathrm{i}=\mathrm{m}}^{\mathrm{n}} \frac{\mathrm{C}_{\mathrm{S_j}}^{\mathrm{i}} \mathrm{C}_{\mathrm{N}-\mathrm{S_j}}^{\mathrm{n}-\mathrm{i}}}{\mathrm{C}_{\mathrm{N}}^{\mathrm{n}}} \right)$$

Dynamic

$$DA(\alpha) = 1 - \left(P(V) + \left(\sum_{g=(m-i)}^{q} P_q(g) \right) \sum_{j=1}^{K} \left(\frac{S_j}{N} \times \sum_{i=0}^{m-1} \frac{C_{S_j}^{i} \, C_{N-S_j}^{n-i}}{C_N^n} \right) \right)$$

Heterogeneity Impact

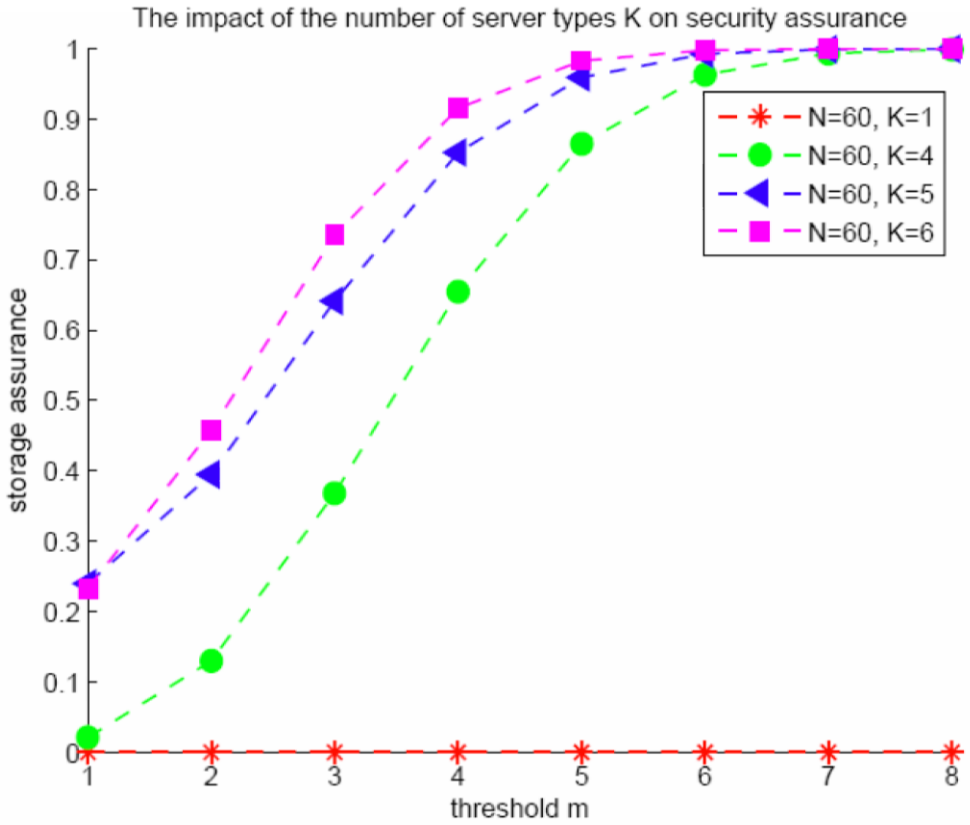


Figure 1.5: Heterogeneous system and homogeneous system using secret sharing scheme.In all the four test cases, N is set to 60. K is set to 1, 4, 5, and 6, respectively. When K is 1, there is only one server group in the system.

Impact of System Size

We gradually increase system size from 45 to 70 by increments of 5, keep *k at 3,and also* vary *m from 4 to 8.*

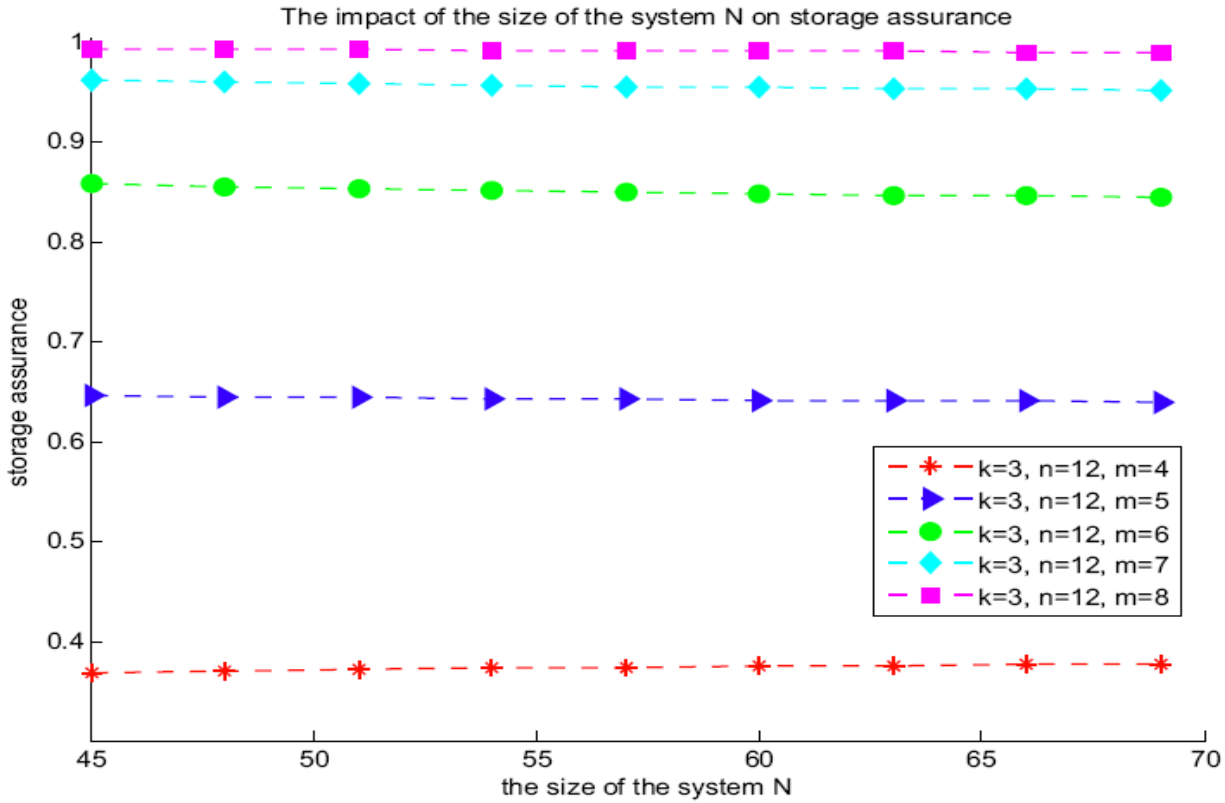


Figure 1.6: The impact of the system size N on storage assurance.

AUBURN UNIVERSITY
SAMUEL GINN COLLEGE OF ENGINEERING

Impact of Size of Server Groups

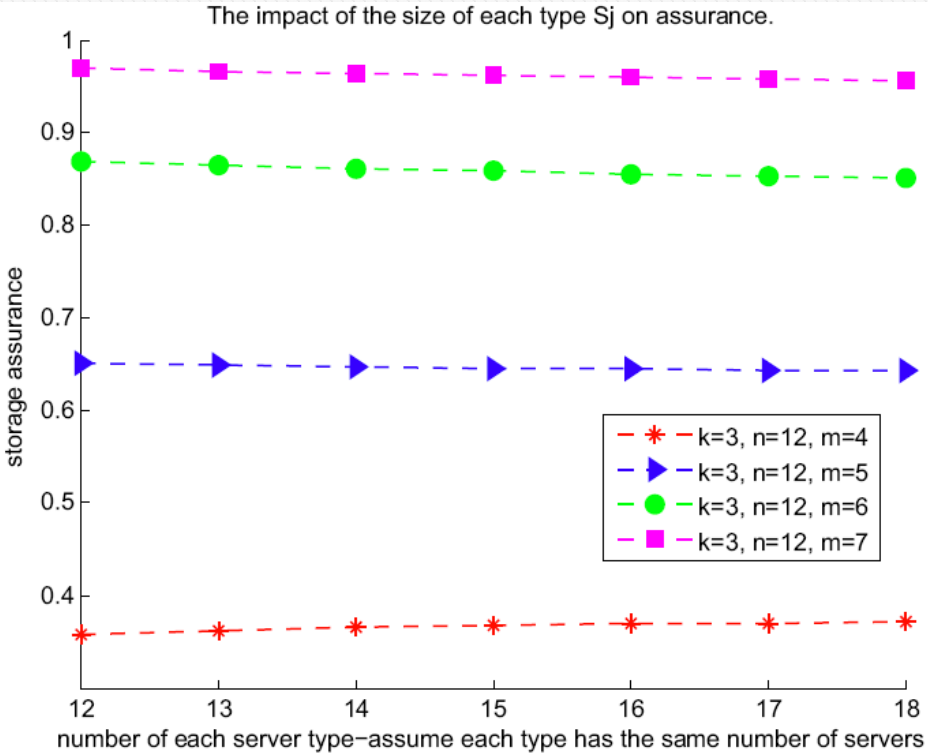


Figure 1.7: The impact of server-group size on data storage assurance. The server-group size means the number of storage nodes in a server-type group. Note that the storage nodes within a server group share the same level of vulnerability. The server-group size varies from 12 to 18 with an increment of 1.

Number n of Fragments of a File

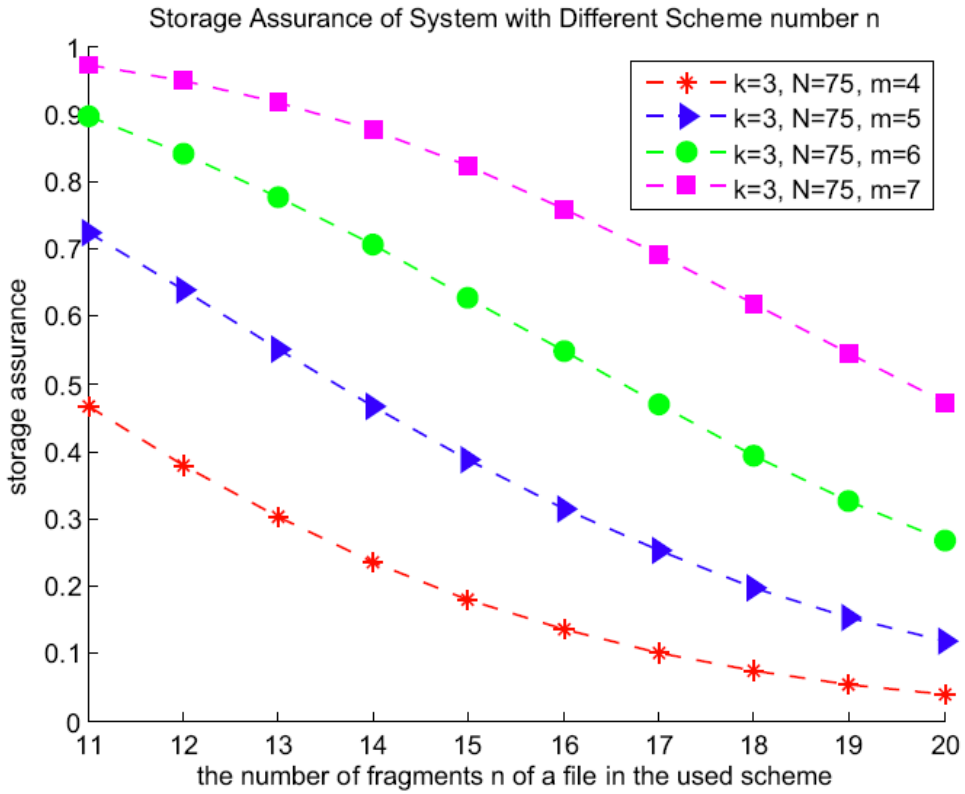


Figure 1.8: The impact of the number n of fragments of a file on storage assurance. We increase the number n of fragments from 11 to 20. The parameters k and N are set to 3 and 75, respectively.

AUBURN UNIVERSITY
Samuel Ginn College of Engineering

Impact of PL on Dynamic Assurance

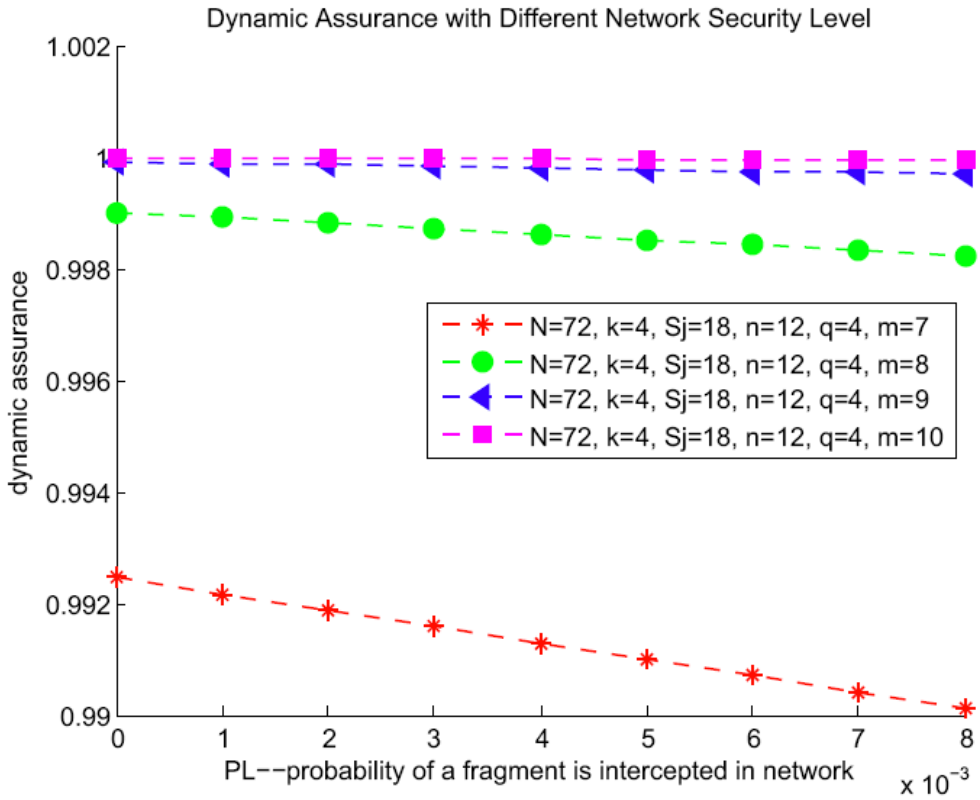


Figure 1.9: Impact of P_L - the probability that a fragment might be intercepted by an attacker during the fragment's transmission through an insecure link. P_L is set from 0 to $8 * 10^{-}3$ with an increment of $1 * 10^{-}3$. Threshold m is set from 7 to 10

Impact of q on Dynamic Assurance

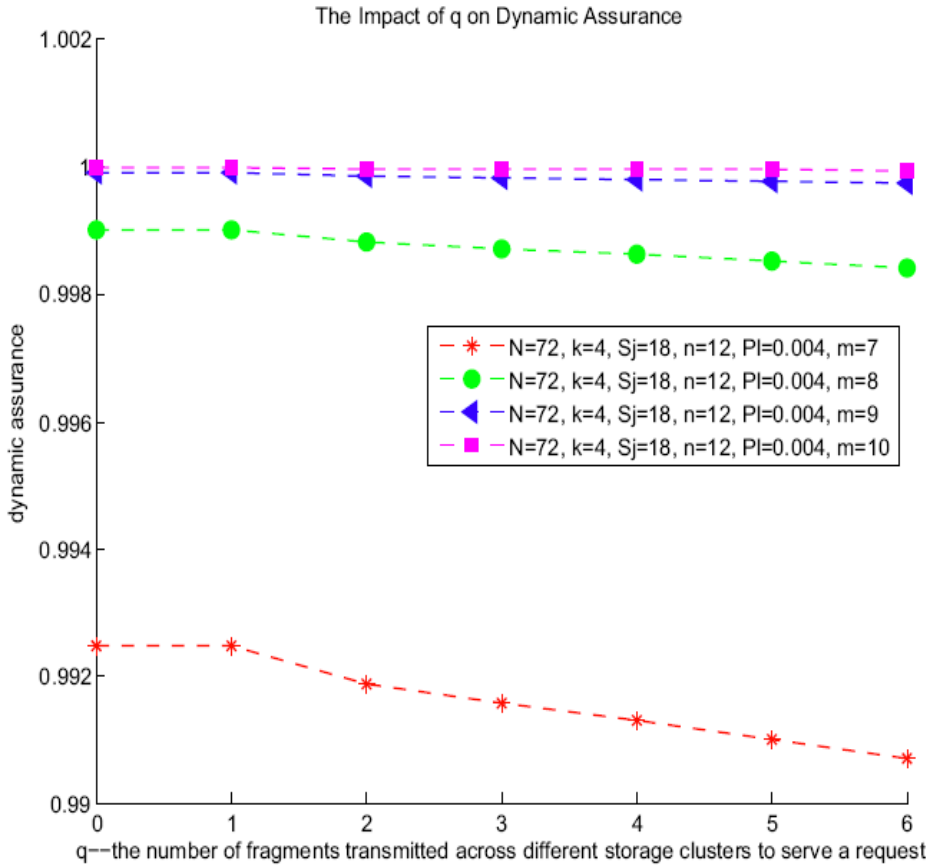


Figure 1.10: Impact of q - the number q of fragments transmitted to and from a storage cluster. q is chosen from 0 to 6 with an increment of 1. Threshold m is set from 7 to 10

AUBURN UNIVERSITY
SAMUEL GINN COLLEGE OF ENGINEERING

Conclusions

- Heterogeneous vulnerability: storage nodes are classified into different server groups based upon their vulnerabilities
- S-FAS : a secure fragmentation allocation scheme
- Storage assurance and dynamic assurance models
- A prototype in which S-FAS was implemented

New Directions

- Consider data replications to enhance reliability and performance
- Authorization/Authentication mechanisms in the prototype.
- Encryption/Decryption Mechanisms in the prototype.

Download the presentation slides
http://www.slideshare.net/xqin74

Google: **slideshare Xiao Qin**

AUBURN UNIVERSITY
SAMUEL GINN COLLEGE OF ENGINEERING

Other Research Presentations and Projects

http://www.eng.auburn.edu/~xqin

Xiao Qin, Associate Professor of Computer Science, Auburn University
Sep 3, 2010 ... UNL CSE graduate **Xiao Qin** received an NSF CAREER award in 2009 to investigate parallel disks that put substantial multicore computing power ...
www.eng.auburn.edu/~x**qin**/ - Cached - Similar

AUBURN UNIVERSITY
SAMUEL GINN COLLEGE OF ENGINEERING

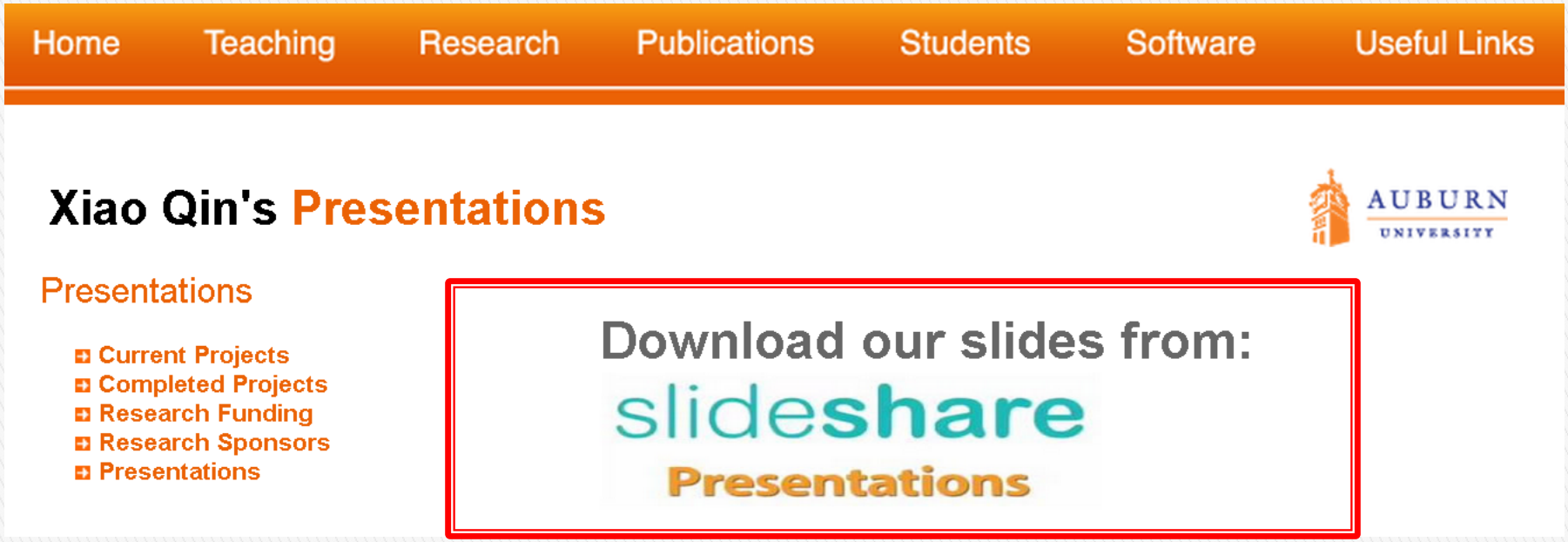
My webpage

http://www.eng.auburn.edu/~xqin

AUBURN
UNIVERSITY
SAMUEL GINN
COLLEGE OF ENGINEERING

Download Slides at slideshare

http://www.slideshare.net/xqin74



AUBURN UNIVERSITY
SAMUEL GINN COLLEGE OF ENGINEERING

Thank you!

AUBURN UNIVERSITY

Suggestions & Questions?

AUBURN
UNIVERSITY
SAMUEL GINN
COLLEGE OF ENGINEERING