



"The goal of He's research is to build a robust I/O and data storage systems research program at VCU. He's passion for this work stems from its potential impact in today's information age where data is a critical component."

# Data Storage: **Demand Performance**

**D**ata storage is the blood of modern computer systems. However, the gap between I/O and computation power is increasing and many large-application computer systems crash because their storage systems are vulnerable. To address this issue, Dr. Xubin He, Associate Professor in the Department of Electrical and Computer Engineering, and his team are working toward solutions for high performance and reliable data storage systems.

Performance and reliability are critical issues for storage systems. During the last decade, high-end computing (HEC) and cloud computing systems have become a necessary tool for scientists worldwide to understand and address complex problems, such as nuclear fusion, biological processes and nanotechnology. However, modern storage systems face several challenges to satisfy this demand, particularly reliability and performance, which is an increasing concern as we evolve toward petascale computing systems with tens to hundreds of thousands of processors. The robustness of the I/O system is crucial to large-scale applications that generate and analyze terabytes of data. Storage systems are vulnerable to numerous hardware and software failures (eg. I/O and metadata server crashes) and contribute to as much as twenty-five percent of all system failures. Building a robust high performing, reliable storage system is becoming increasingly more critical as high-end computing systems scale up in size.

Currently, He is working with visiting professor Huailiang Tan, Ph.D. students Chentao Wu and Guanying Wu, and undergraduates Vladislav Sorkin and David Lyons to conduct three NSF funded projects on high performance and reliable I/O and data storage systems.

The team is investigating different solutions to improve storage system performance and reliability. They are cultivating specific techniques such as developing an I/O and storage benchmarking and analysis framework to efficiently profile the storage systems in both native and virtualized environments so as to identify the performance bottlenecks and optimize the systems.

Another technique the team is working on is developing different data cache schemes adaptive to the I/O hierarchy, including cache for networked storage systems (iCache and STICS), cache for distributed systems (Hint-K) and cache for solid state devices (BPAC). These cache schemes dramatically improve the data I/O performance.

The team also proposes and generates different schemes to increase the storage system reliability. As a proactive approach, active-active metadata service provides a robust solution to ensure the data is available even when partial metadata servers crash. They are also developing different coding schemes (Code-M, H-Code, and HDP Code) to tolerate multiple disk failures.

His successful research has earned him international reputation in the area of data storage systems. Through his research, He has collaborated with national institutions such as Oak Ridge National Lab, University of Florida, and RPI as well as

schools in China including Huazhong University of Science and Technology, Dalian University of Technology, and the Institute of Computing, Chinese Academy of Sciences. Since 2007, He has been an invited speaker to schools throughout China and Thailand. He was the keynote speaker at the Third International Symposium on Parallel Architectures, Algorithms, and Programming in December of 2010. He has also co-chaired several international conferences including the research track of the 26th IEEE Symposium on Massive Storage Systems and Technologies (MSST) which is a premier international conference in data storage systems.

The goal of He's research is to build a robust I/O and data storage systems research program at VCU. He's passion for this work stems from its potential impact in today's information age where data is a critical component. He also relishes the opportunity to work with the talented students in the data storage field, a crucial growth area in the twenty-first century.

