Work in Progress: Performance Evaluation of RAID6 Systems

Yan Li[†], Tim Courtney[‡], Roland N. Ibbett[†], Nigel Topham[†]

[†]Institute for Computing Systems Architecture,
School of Informatics, University of Edinburgh
Y.Li-24@sms.ed.ac.uk, {rni, npt}@inf.ed.ac.uk

[‡] Xyratex, 1000-80 Langstone Technology Park,
Langstone Road, Havant, Hampshire, PO9 1SA, UK
tim_courtney@xyratex.com

The aim of this on-going work is to study the performance of RAID6 protected storage systems under a Storage Performance Council-1 (SPC-1) benchmark based workload [3]. With the scale of modern storage systems becoming increasingly large, the probability of having double disk failure in a protection group is much higher than before. As a result, there is an increasing demand for products protected against double disk failure. Although a number of double disk failure protection algorithms have been proposed, RAID6 has not been used extensively in practice due to its poor small write performance. In addition, there have been few publications about its performance under real world workloads. Given the trend of using RAID6 protection, it is important to understand the performance of RAID6 under real-world workloads. This work will study the performance of P+Q-like RAID6 algorithm as an exemplar (other RAID6 algorithms deploying two redundant disks in a protection group include Row-Diagonal Parity (RDP) [2] and EVENODD [1]) under an SPC-1 benchmark based workload using simulation. A discrete-event driven simulation model called SIMRAID has been developed to model the storage sub-system. SIMRAID contains a benchmark workload generator, a detailed RAID controller coupled with cache model, a transport-level Fibre-Channel model and disk model. The accuracy of SIMRAID has been verified through simulation of a system for which there are published SPC-1 benchmark results. This simulation shows a maximum inaccuracy of 5%. We use the maximum Business Scale Unit (BSU) value for which the average response time does not exceed $30ms$ as the performance metric.

The study consists of three parts. First, we are going to study the performance of RAID6 under normal operation mode (fault-free mode) and explore its design space. In particular, we have looked at the performance sensitivity to the controller processing time, the optimum size of the stripe under such a workload, the impact of increasing cache size and number disks, and their interactive effects. The study also includes a comparison with the performance of RAID5 systems. Simulation results show that the maximum number of BSU with response time not exceeding $30ms$ for a 28-disk system coupled with 2G cache only differs 2 when the processing time vary from $27\mu s$ to $324\mu s$. There is only 3% difference on performance with a 400% difference on the processing time. Thus, we can say that the performance of RAID6 is insensitive to the controller processing time. The optimum size of stripe unit is $16KB$ for non-cached system and $32KB$ for cached system. The second part of the study is to study the performance of RAID systems under degraded mode. The third part of the study is to find out the best recovery schemes. After disk failures the storage system will experience three stages before it recovers to normal working status. The first stage is the degraded mode during which the system is only able to handle a reduced workload and there is a reduced level of protection. The second stage is the reconstruction stage during which the system rebuilds the failed disks. During this time the workload the system can handle is further reduced since some of the resources are used rebuilding disks. Assuming that the load continues at (or above) that the degraded mode can handle, the system is unable to handle the whole incoming command stream resulting in queuing of commands. Finally, the third stage is the recovering stage. After the failed disks are rebuilt, assuming the workload remains the same as during re-build the system can clear the queue of commands accumulated during re-build. After this stage, the system will be back to normal operation mode. It is of interest to find out how to allocate resources between rebuilding and serving of incoming requests during that second stage so that it needs the shortest time for the system to return to normal operation mode.

REFERENCES

[1] M. Blaum, J. Brady, J. Bruck, and J. Menon. Evenodd: An efficient scheme for tolerating double disk failures in RAID architectures. *Trans. on Computers*, 44(2):192–202, Feb 1995.

[2] P. Corbett, B. English, A. Goel, T. Grcanac, S. Kleiman, J. Leong, and S. Sankar. Row-diagonal parity for double disk failure correction. In *Proceedings of the USENIX FAST '04 Conference on File and Storage Technologies*, pages 1–14, San Francisco, CA, Mar 2004.

[3] Storage Performance Council. Spc benchmark-1 (SPC-1) official specification version 1.7, July 2003. available at http://www.storageperformance.org/Specifications/SPC-1_v170.pdf.