Virtualizing Disk Performance with Fahrrad

Anna Povzner, Scott Brandt, Richard Golding†, Theodore Wong†, Carlos Maltzahn
Computer Science Department, University of California, Santa Cruz
†IBM Almaden Research Center

Storage systems are frequently shared among multiple clients—customers, host computers, or virtual machines—to reduce hardware and management costs. In such circumstances, it is desirable to provide some or all clients with guarantees on I/O performance to ensure desired throughput, sharing, latency, etc., independent of how other clients are accessing the device. Ideally, we would like to provide each client with the equivalent of a "virtual disk", i.e. the illusion of having a disk all to itself.

In a shared storage system where workloads compete with each other for access to common storage resource, the primary challenge of providing guaranteed performance is guaranteeing performance isolation. Isolating clients' workloads from the behavior of others is challenging because of mechanical nature of disks: performance maybe affected by seeks introduced by competing workloads.

Existing approaches for storage performance virtualization provide soft guarantees. Façade [2] provides statistical performance guarantees by controlling the load from multiple clients. Argon [4] attempts to provide clients with at least a configured fraction of a standalone throughput. Our goal is to provide throughput equivalent to a standalone throughput, which we define as follows: Given a configured time interval p and integer n, the amount of data from a given I/O stream transfered by a virtual disk with x% share during time $t = n * p$ must be equal to the amount of data transfered by the same stream when using the disk alone during time $x * t$.

The Fahrrad real-time disk I/O scheduler [3] guarantees disk time utilization. Managing disk I/O in terms of utilization yields better control and more efficient use of disk resources than throughput-based schedulers [1]. The ability to reserve and guarantee a portion of disk time forms the basis for our virtual disk abstraction. Arbitrary shares of the disk time utilization may be granted to each client, and arbitrary reservation granularities allow clients to state arbitrary bounds on the latency of their virtual disks. Fahrrad minimizes interference between streams by servicing as many contiguous requests as possible from each stream without violating utilization guarantees.

This work extends Fahrrad to guarantee isolation between streams. Isolation is ensured by correctly accounting for all I/O latencies. Since some seeks between streams are unavoidable, we account for extra seeks caused by inter-stream seeking by reserving "overhead" utilization. Given a reserved disk share x%, we calculate the overhead utilization y% and reserve $u\% = x\% + y\%$ share of the disk. In Fahrrad, inter-stream seeking is caused by streams that must service I/Os to meet their utilization reservation. We account for the additional seeks in the reservation of streams responsible for inter-stream seeking and bill these streams for the additional seeks. Hard guarantees require worst-case seek time assumptions for overhead reservation, which could result in high overhead. However, we can use less then worst-case assumptions for soft requirements (*i.e.* using statistics of request time distributions), resulting in considerably lower overhead.

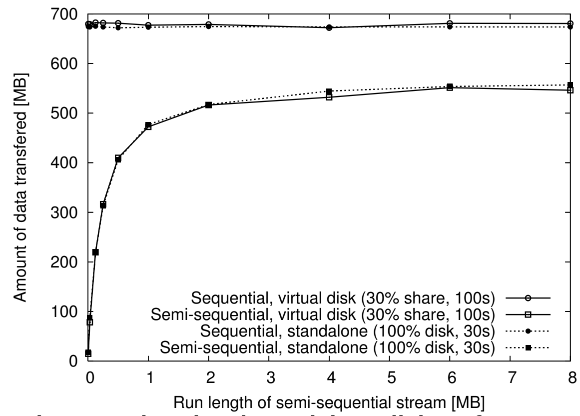

Figure 1. Virtual and standalone disk performance.

Figure 1 examines Fahrrad's virtual disk performance. In this experiment, two virtual disks share a storage system. Each virtual disk reserves 30% of the disk time with a granularity of 1 second. One virtual disk services a sequential workload and the other services a semi-sequential workload with a specified run-length. Both send I/Os as fast as possible. Each point in Figure 1 represents the amount of data transfered during 100s for a virtual disk and 30s for a standalone disk. As the sequentiality of the semi-sequential stream increases, it becomes more efficient, but it does not affect the behavior of the sequential stream. Figure 1 also shows that the performance of the workloads on Fahrrad's concurrently executing virtual disks is within 2% of that obtained running standalone, demonstrating the effectiveness of this approach.

Our approach for guaranteed performance isolation is to correctly account for inter-stream seeks caused by competing workloads. The time to perform these seeks is included in the "overhead" utilization so that I/O performance achieved from the reserved disk share depends only upon the workload behavior.

1. References

[1] T. Kaldewey, T. Wong, R. Golding, A. Povzner, S. Brandt, and C. Maltzahn. Virtualizing disk performance. In *RTAS'08*.

[2] C. R. Lumb, A. Merchant, and G. A. Alvarez. Façade: Virtual storage devices with performance guarantees. In *FAST'03*.

[3] A. Povzner, T. Kaldewey, S. Brandt, R. Golding, T. Wong, and C. Maltzahn. Efficient guaranteed disk request scheduling with Fahrrad. In *Eurosys'08*.

[4] M. Wachs, M. Abd-El-Malek, E. Thereska, and G. R. Ganger. Argon: performance insulation for shared storage servers. In *FAST'07*.