ACM Computing Surveys **28**(4es), December 1996, http://www.acm.org/pubs/citations/journals/surveys/1996-28-4es/a209-gibson/.

This article derives from a position statement prepared for the Workshop on Strategic Directions in Computing Research.

Self-Managing Network-Attached Storage

Garth Gibson

Carnegie Mellon University Computer Science Department
5000 Forbes Avenue, Pittsburgh, PA 15213-3890, USA
garth.gibson@cs.cmu.edu

John Wilkes

Hewlett Packard Laboratories
1501 Page Mill Road, Palo Alto, CA 94304-1126, USA
wilkes@hpl.hp.com

Abstract: We outline the critical issues impacting storage access technology, our view of appropriate responses and our vision of the outcome of re-architecting storage access systems.

Categories and Subject Descriptors: B.4.0 **[Input/Output and Data Communications]**: General; C.0 **[Computer Systems Organization]**: General - *System architectures, Systems specification methodology*; D.4.2 **[Operating Systems]**: Storage Management - *Allocation/deallocation strategies, Secondary Storage, Storage hierarchies*; D.4.3 **[Operating Systems]**: File Systems Management - *File organization, Maintenance*; D.4.4 **[Operating Systems]**: Communications Management - *Input/Output*; D.4.5 **[Operating Systems]**: Reliability - *Fault-tolerance*; D.4.6 **[Operating Systems]**: Security and Protection - *Access controls, Authentication*; D.4.8 **[Operating Systems]**: Performance - *Modeling and prediction, Monitors*; H.2.7 **[Database Management]**: Database Administration; K.6.2 **[Management of Computing and Information Systems]**: Installation Management - *Computing equipment management, Pricing and resource allocation*; K.6.4 **[Management of Computing and Information Systems]**: System Management.

General Terms: Algorithms, Design, Management, Measurement, Performance, Reliability.

Additional Key Words and Phrases: I/O, communication, disk drives, secondary memory, network attached storage, storage management.

Critical I/O issues

An organization's most valuable asset in the information age is its ability to obtain, maintain and manipulate an information base. It is our position that storage access technology is a critical component of an information base; that storage access technology is experiencing dramatic change; and that a re-architecting of storage access systems is beginning. Needless to say, we see storage as an important and rich domain for the efforts of computer science researchers. In this position statement we outline the critical issues impacting storage access technology, our view of appropriate responses and our vision of the outcome of re-architecting storage access systems.

Underlying technology

- Magnetic disk, optical disk and tape continue to improve density (hence capacity) at or above DRAM rates, insuring the continued viability of these technologies and the continued drop in storage cost.
- Data rates, traditionally nearly static, are now increasing significantly, causing interface technologies and delivery software to increasingly limit end-to-end performance.
- The ever problematic access gap, the 5-6 orders of magnitude difference between the access time of DRAM and disk, continues to grow, causing the overall performance of a transparent memory hierarchy to be increasingly sensitive to I/O access times.

Use profiles and customer requirements

- Storage systems represent a huge business that is growing more rapidly than the PC industry (4/22/96 Wall Street Journal article).
- The sizes of data objects are growing faster than memory sizes, demanding that storage access play a more active role in computation.
- At the forefront of storage requirements is support for richly typed objects such as hypertext, dynamically linked and remotely executable code, image, audio, video, telemetry, simulation state and data/graphics animation.
- The ratio of peak to average bandwidth demand is growing, leading to expensive bandwidth-delivery infrastructures that are often underutilized.
- Wide geographic availability is increasingly important, causing wired and mobile networking to be critical to customer requirements.
- Addressable storage, spurred by the World Wide Web, is growing at a phenomenal rate, making obsolete traditional namespace management, indexing, and search technologies and inducing profuse heterogeniety in storage mechanisms, delivery systems, and access control systems.
- Fault-tolerant availability is increasingly important because of the growing interdependence of computing in all aspects of society.
- The availability of specific performance levels, demanded by continuous media applications for example, is increasingly required of storage.

System architecture

- Memory hierarchies are getting deeper, placing more reliance on caching and further stressing the latency of non-local accesses.
- Operating system configurability and stability are more important than performance; the resulting inertia retards the introduction of new technology and impedes the delivery of higher peak performance.
- The administration and management of storage, including manual balancing of capacity and load as well as backup, costs substantially more each year than the capital cost of the storage.
- Storage-embedded functionality is growing, benefiting from the ubiquitous microprocessor cost/performance improvements, and is used to smooth technology introduction, overcome operating system shortcomings, and exploit subsystem specializations.
- The addressability explosion and the cost-effectiveness of systems composed of various off-the-shelf components means that the overall system is highly heterogeneous (subsystems, interconnects/networks, file systems, operating systems), ensuring that security and integrity must be end-to-end rather than provided by the cobbled-together infrastructure.

Responding to these observations, we see profitable research in

- Out-of-core computation support in libraries and compilers.
- Better exploitation of available parallelism.
- Aggressive and accurate prefetching.
- Revamping hardware and software interfaces for lower overhead, better streaming, broader interoperability.
- Transparent and application-assisted adaptivity to dynamic workload, network and subsystem characteristics.
- Tighter interdependence with networking technology.
- Global resource management of multiple-level caches and storage parallelism.
- Application and datatype customization at storage, in the network and at the client.
- Reduced dependence on the client's operating system, enabling storage technology to evolve independently and limiting the security risk inherent in trusting a client operating system.
- Strategies- and protocols-assisted- or self-management of storage.
- More efficient client software interfaces with features such as rich scatter-gather, zero-copy access, disclosure of future accesses, relaxed ordering of access among dynamically defined objects, and function shipping for storage-intensive operations.
- Subsystem support for specific data types such as continuous media, and more generic support for yet-to-be-defined data types.
- Subsystem support for name space, indexing and search.

Re-architecting for storage-centric computing

Integrating many of these emerging storage technology advances, our vision of a re-architected storage access system stresses a closer relationship between storage subsystem capability and client/application requirements. Specifically, accessing storage is analogous to a communication connection and should be endowed with analogies to quality-of-service, fastpathing, intergrated layer processing, application specialization, and embedded management support. Our vision of storage-centric computing redraws the boundaries among application, file system, network, and storage subsystem, enabling applications to have an expressive interface to storage subsystems, decoupling policy and primitives to allow separate optimization and directly connecting storage to the networking infrastructure.

Advantages of this approach include:

- high bandwidth end-to-end data transfer cost shared with other network applications
- offload the gory details of the storage system from its clients
- exploitation of sophisticated storage-specific optimizations
- simplified, more effectively scalable high-performance services (file systems, databases)
- migration of storage-specific function out of operating system
- rapid deployment of new storage devices (good for manufacturers and customers)
- higher-level compatibility promoting richer competition (more value at lower cost) and less extensive integration and upgrade effort (good for system houses and customers)
- transparent and application-assisted dynamic adaptation to global state or data type
- custom tailoring of heterogeneous, cost-effective, off-the-shelf system

Technical enabling issues

In addition to the research responses outlined above, our approach to re-architecting storage systems highlights a variety of related technical issues, including:

Networking infrastructure:

- cost-effective scalability
- cheap device connections

- low latency (for control traffic)
- (guaranteed) high bandwidth (for data transfers)
- third-party transfers: initiated by A, sent to B, delivered to C

Security protocols:

- minimal-cost protocols, maximum flexibility for storage manager requirements
- which classes of service: authenticated control traffic only, encrypted data on network, encrypted data on network and disk?

Storage subsystems:

- migrate into the storage system responsibility to provide service properties
- which high-level "attributes" best express needs and how are they mapped to service?
- which performance guarantees and mechanisms meet these needs?
- automated capacity planning, load monitoring and trend analysis
- dynamic data rearrangement for capacity, load balancing and specific application needs
- ``plug and go" model: new device is added, rest of system exploits it without further prompting

Storage interfaces and protocols:

- fastpathing of storage device to client (both management and data interfaces) with guaranteed and negotiated service
- network cooperation and negotiation interfaces
- storage-device to storage-device (e.g., peer-to-peer resource/load balancing)
- file system, database, language run-time, common library and other application-specialization interfaces (e.g., video editors)

Fault resilience:

- no single point of control/failure: fully distributed fault-tolerant control
- scalable across nodes, sites, continent

Non-technical enabling issues

Changes like these require more than just technical innovation; the commercial infrastructure must also be brought along. This means finding imaginative ways for the technology to be delivered through multiple channels simultaneously: device manufacturers, storage-network designers, interface-card builders, OS/file system/dbms designers, system intergators, and application providers. There's a need for coordinated, incremental delivery of components of this new technology, in a way that allows marketplace acceptance to drive the development of later phases. This may be as large a challenge as the technical aspects.

For further information

- National Storage Industry Consortium working group on Network-Attached Storage Devices (NASD) web page.

Last modified: Mon Feb 24 15:08:33 EST 1997

John Wilkes *<wilkes@hpl.hp.com>*