INformation Storage Industry Consortium

Roadmap of Data Storage Devices and Systems Research

Giora J. Tarnopolsky, TarnoTek

INSIC

Information Storage Industry Consortium

HEC-IWG File Systems and I/O R&D Workshop

DFW Airport, Texas

16 August 2005

INSIC Members . . . and Universities

HITACHI GLOBAL STORAGE TECHNOLOGIES
HUTCHINSON TECHNOLOGY
SEAGATE TECHNOLOGY
ADVANCED RESEARCH
DUPONT-TEIJIN FILMS*
HEWLETT- PACKARD
AGERE SYSTEMS
MAGNECOMP
CERTANCE
IMATION
SONY
IBM
IDC*
MAXELL
FUJIFILM
SAMSUNG
QUANTUM
STORAGETEK
MEMS OPTICAL
WESTERN DIGITAL
TORAY INDUSTRIES
VEECO INSTRUMENTS
TEIJIN-DUPONT FILMS*
ADVANCED MICROSENSORS*

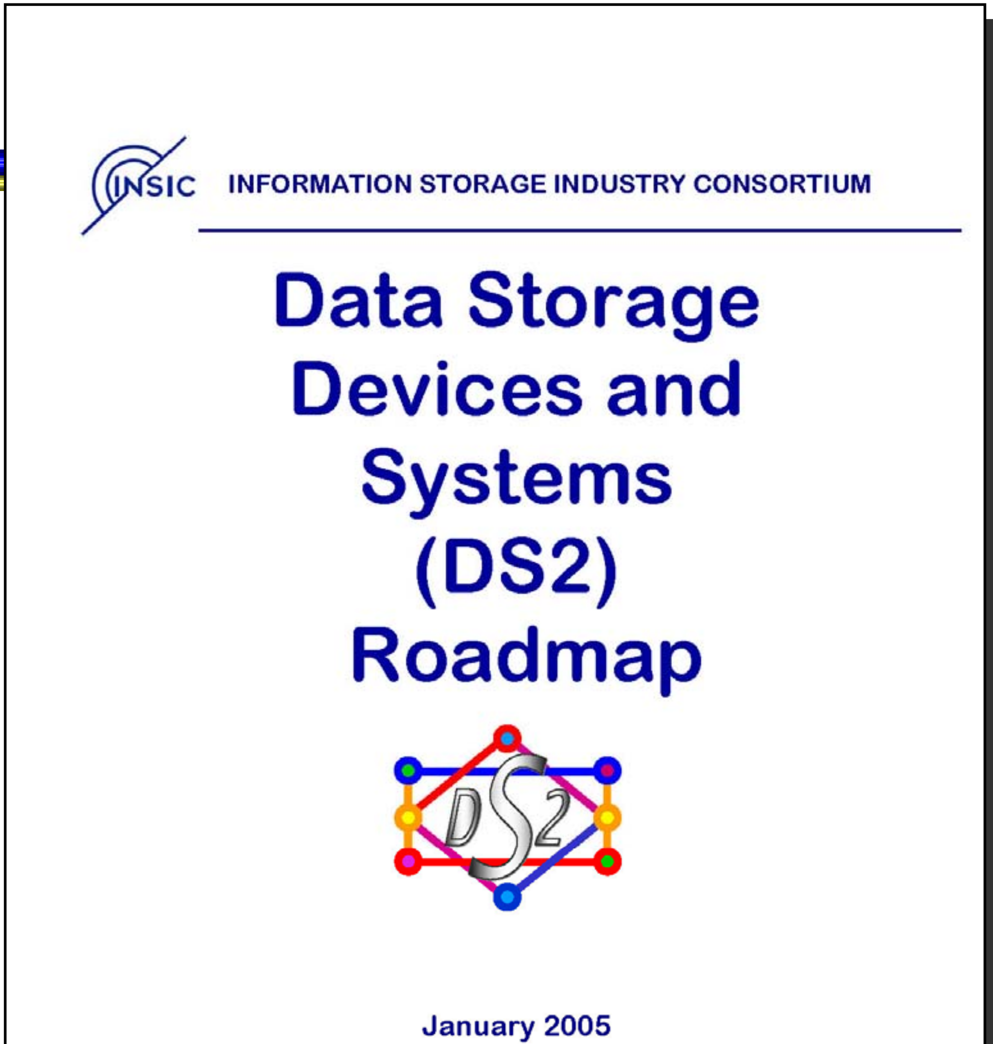
* Limited Member

During 1999-2005, INSIC Research Programs have supported research at a total of 26 Universities:

Massachusetts Institute of Technology
Data Storage Institute, Singapore
University of California, Berkeley
Georgia Institute of Technology
University of Washington
Northwestern University
University of the Pacific
University of Colorado
University of Alabama
University of Missouri
University of Arizona
University of Illinois
Harvard University
Stanford University
University of Alberta
Vanderbilt University
University of Virginia
University of Houston
University of Nebraska
University of Minnesota
University of Manchester
Colorado State University
Carnegie Mellon University
University of Central Lancashire
National University of Singapore
University of California, San Diego

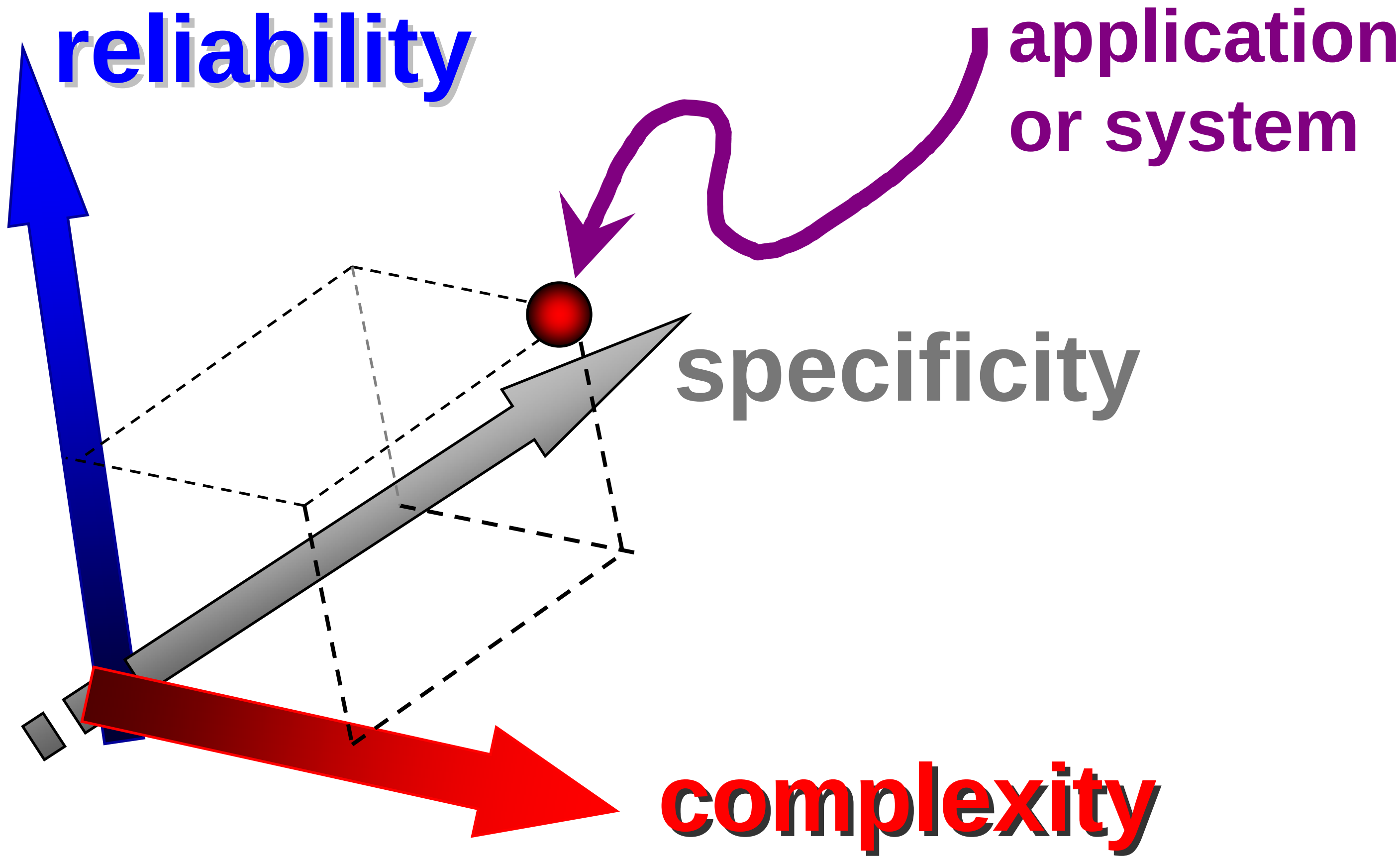
DS2 Roadmap

Roadmap of Data Storage Devices and Systems Research



http://www.insic.org/2005_insic_ds2_roadmap.pdf

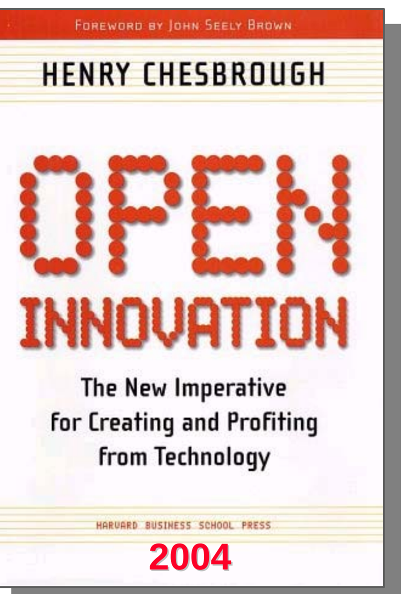
Storage Systems and Storage-attached World



- Balance complexity, reliability, and specificity

Pre-Competitiveness Becomes Mainstream

- Pre-competitive research addresses matters of undisputed relevance for which cooperative efforts are both possible and desirable.
  - For instance, subjects far removed from marketplace implementation, or a business model based on non-exclusive ownership of technology



- INSIC, incorporated in April 1991, “wrote the book” on pre-competitive research programs *before* the book was written!
- Pre-competitive research enhances research ROI, specific to storage management
  - Leverage: project funding is shared among the industrial sponsors of the program and other entities
  - Encourages academic researchers to coalesce their talents, mitigates redundancies while fostering creativity

Research Thrusts

Thrust	Issues addressed	Leaders
Active Storage Devices	General purpose data processing by the storage device	Erik Riedel Seagate Research
Application-aware Storage	Device or system behavior depends on data or users' characteristics	Michael Mesnier Intel & CMU
Autonomic Storage	Storage system manages itself	Remzi Arpaci-Dusseau U. Wisconsin
Long-term Storage	Preservation of digital assets	T. Ruwart/ G.Tarnopolsky U. Minnesota / INSIC
Pervasive Storage	Devices everywhere, data consistency, preservation, security	C. Harmer/ P. Massiglia VERITAS
Privacy and Security	Data access rights, data integrity, IP, security	James Hughes StorageTek

Application-aware Storage

- Application-aware storage devices are those which possess knowledge about the environments in which they operate, and enhance their performance as a result of that knowledge.

Examples: - aggregation information
- relationships among data, users, apps

Bytes don't change

Active Storage Devices

- Active storage devices are those which run application-specific processes to perform application-specific functions upon the data. These devices apply their own capabilities to improve application performance.

Bytes may change

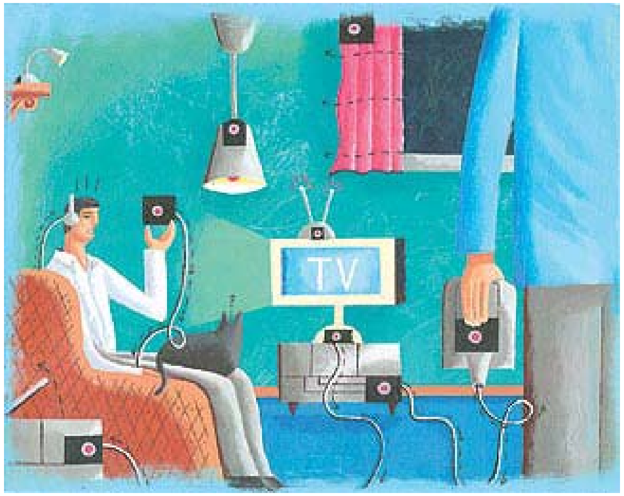
Example: data mining

Long-term Storage

- *Long-term preservation* assures the availability of tangible data records, digitally stored, over periods of time that vastly exceed the lifetime of the physical and logical system used to store and retrieve the record initially.
- A *tangible data record* is information that is sensorially evident to all users, visually or in natural languages, although certain information, such as hyperlinked documents, may require machinery for its display.

"Digital information lasts forever – or five years, whichever comes first." Jeff Rothenberg

Pervasive Storage



Nokia N91 4GB phone

500 M units @ 10 GB = 5 Exabytes

❶ Pervasive *access* to information, supports either "disconnected" or connected operation.

❷ Pervasive storage refers to the widespread availability of storage resources of practically unlimited capacity, over unbound geographic areas, concurrent with the consistent management of the stored assets and their immediate accessibility

Privacy and Security

- Privacy refers to the denial of access to stored records by unauthorized clients concurrently with the assurance of access by authorized ones
- Security refers to the assurance of the integrity of stored records concurrently with efficient access by multitudinous clients

Autonomic Storage: Why?

- Achieve predictable performance in complex systems
- Protect customer data against loss or corruption
- Reduce storage system management TCO
  - TCO: A result of complexity: Too many human-intensive tasks
- Respond more efficiently than human managers
  - Reaction times short as compared to physiological response
  - Spectrum of causes/responses overwhelms human analysis
- Find and fix problems (self-healing)
  - address system complexity
  - diagnose sources of error precisely, resolve problem

DS2 Thrusts & Business Interest

Thrust	Business Opportunity			
• Active Storage Devices	Massively parallel database search and data mining	Massive indexing and searching	ILM & automatic destruction of data	Sensor networks
• Application-aware Storage	QoS, efficient I/O	Reliability	Security	System management
• Autonomic Storage	TCO	Predictability	Data integrity	Self-healing
• Long-term Storage	TCO	Data integrity	Language development	Consumer markets
• Pervasive Storage	Consumer markets	Record preservation	Consistency	Distributed storage utilities
• Privacy and Security	Assurance of service	Dispersed storage systems	Consumer markets	Record management

INSIC's DS2: Storage Research Partnership

- Elicit agreement, funding, and coherence for supporting one or more research programs in data storage devices and systems
- The DS2 initiative has identified many avenues for pre-competitive research, where joint efforts by industry, academics, and government would benefit the global storage systems endeavor.
- "Open Innovation" as a win-win-win (industrial-academic-governmental) common endeavor
- This initiative deserves vigorous support.

INSIC

Thank you

Application-aware Storage

- Application-aware storage devices are those which possess knowledge about the environments in which they operate, and enhance their performance as a result of that knowledge.

Examples: - aggregation information
- relationships among data, users, apps

Bytes don't change

Application-aware Fundamentals

- Application-aware storage devices may gain knowledge through a combination of user hints, statistical accounting, and inductive logic.
- Similar to active disks, apps-aware moves the application *characteristics* closer to the data.
- The further down the stack the knowledge travels, the higher the performance reward.
- Storage needs to be aware of the business processes - that generate the data being stored or retrieved.

Research Opportunities in

- Active disks
- Archive, backup, and volume management
- Autonomic storage
- I/O scheduling and storage transports
- Measurement, modeling and characterization
- Naming, indexing and searching
- Data organization, prediction and grouping
- Consistency protocols and versioning

Active Storage Devices

- Active storage devices are those which run application-specific processes to perform application-specific functions upon the data. These devices apply their own capabilities to improve application performance.

Bytes may change

Example: data mining

Motivation

- "Active Storage Devices provide advanced functions that operate on non-volatile data either through a *fixed function* or through the use of *general-purpose programming* within the device."
  - *active functions* – portions of application or application-specific code
  - *fixed functions* – enhanced but fixed interface
  - *general-purpose functions* – generic programming capability
- Benefits of active functions
  - ability to compute close to data
  - pre-processing before data is sent along
  - take advantage of local device information
  - scalability of computation at many independent nodes, rather than centralized hosts

Active Devices Research Issues

- **A model of distributed computation**
  - a theory of how to flexibly distribute the functionality in a system around a computing environment.
- **Resource management for active functions**
  - handling multiple executing active functions at the same time
- ***Internal device API (Application programming interface)***
  - how active functions interact with the local hardware environment
- ***Correctness/reliability/stability***
  - in disk or disk array, most corner cases are tested and interface is limited; in active storage, now many more dimensions to the problem.
- **Specialized hardware for fixed functions**
  - hardware-optimized functions in some settings.

Long-term Storage

- *Long-term preservation* assures the availability of tangible data records, digitally stored, over periods of time that vastly exceed the lifetime of the physical and logical system used to store and retrieve the record initially.
- A *tangible data record* is information that is sensorially evident to all users, visually or in natural languages, although certain information, such as hyperlinked documents, may require machinery for its display.

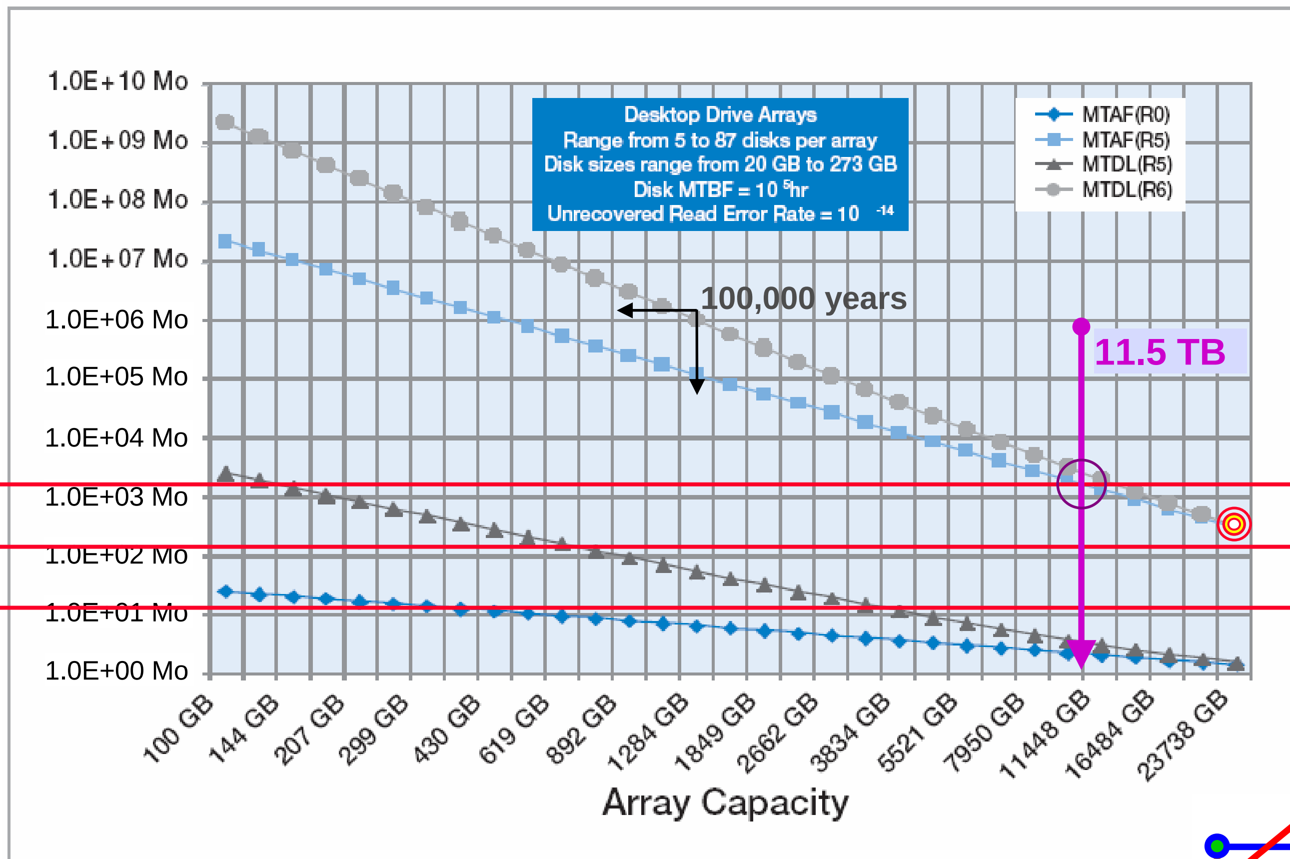
"Digital information lasts forever – or five years, whichever comes first." Jeff Rothenberg

RAID6 time to data loss exceeds "lifetime"

- Statistics useful for short-term reliability assessment --- meaningless for long-term preservation

Figure 1 Time to Failure vs. Disk Capacity

Un-annotated was chart taken from:
"Intelligent RAID 6 Theory," Intel Corporation, 2005.
www.intel.com/design/storage/intelligent_raid.htm



Preservation of Digital Assets

- Preservation of an extant bit stream
  - Hardware, firmware, software means of assuring data integrity, including disaster recovery, within a single technology generation
- Preservation of a bit stream representative of the tangible data record over generations of hardware and software migrations. Invariant or adaptive.
- Preservation of the ability to re-create the sensorial representation, the tangible data record itself
  - Semantic continuity
  - Record aggregation, curatorial metadata
  - Emulation: future computer emulates O/S, application
    - Universal Virtual Computer approach

Date: 05 Aug 2205. CrtlAltDel Emulator v. 99999 SP 999

Illegal operation. Application will be shut down. Click "Continue" to close the application. Click "Cancel" to attempt recovery.

Continue

Cancel

Date: 05 Aug 2205. CrtlAltDel Emulator v. 99999 SP 999

Illegal operation. Application will be shut down. Click "Continue" to close the application. Click "Cancel" to attempt recovery.

Continue

Cancel

Date: 05 Aug 2205. CrtlAltDel Emulator v. 99999 SP 999

Illegal operation. Application will be shut down. Click "Continue" to close the application. Click "Cancel" to attempt recovery.

Continue

Cancel

Date: 05 Aug 2205. CrtlAltDel Emulator v. 99999 SP 999

Illegal operation. Application will be shut down. Click "Continue" to close the application. Click "Cancel" to attempt recovery.

Continue

Cancel

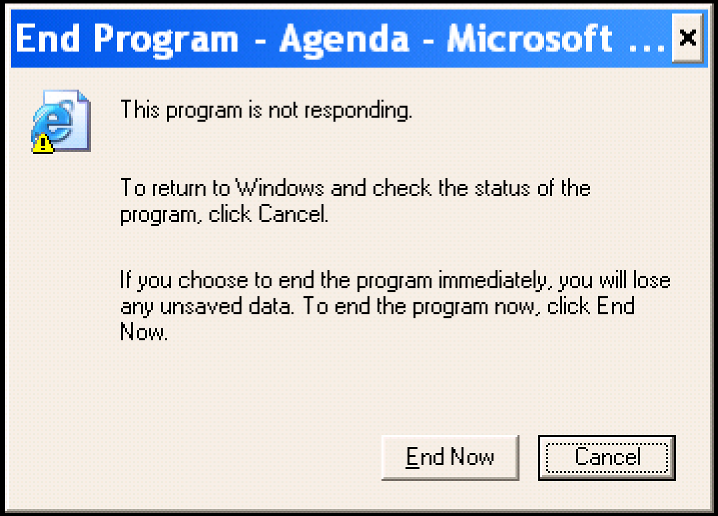
Date: 05 Aug 2205. CrtlAltDel Emulator v. 99999 SP 999

Illegal operation. Application will be shut down. Click "Continue" to close the application. Click "Cancel" to attempt recovery.

Continue

Cancel

Reliable Emulation Must Replicate Even the Dreaded Blue Screen.



It may be possible in principle, but unlikely

National Archives Outsources Long-term Storage

- **Press Release August 3, 2004**
- **National Archives Names Two Companies to Design an Electronic Archives**
- "Washington, D.C. . . Today, Archivist of the United States John W. Carlin announced the two companies that will lead the way in designing a technological solution to the challenge of preserving electronic information across space and time. These design contracts are valued at $20.1 million. At the end of the one-year design competition, the National Archives will select one of these two contractors to build the Electronic Records Archives, a revolutionary system that will capture electronic information, regardless of its format, save it permanently, and make it accessible on whatever hardware or software is currently in use. ***Over the life of the contract, it is potentially worth hundreds of millions of dollars***† with countless positive implications for individuals, private businesses, and government organizations alike.
- The two companies are:
  Lockheed Martin, Transportation and Security Solutions Division - $9.5 million
  Harris Corporation, Government Communications Systems Division - $10.6 million"
- **Reference: http://www.archives.gov/media_desk/press_releases/nr04-74.html**

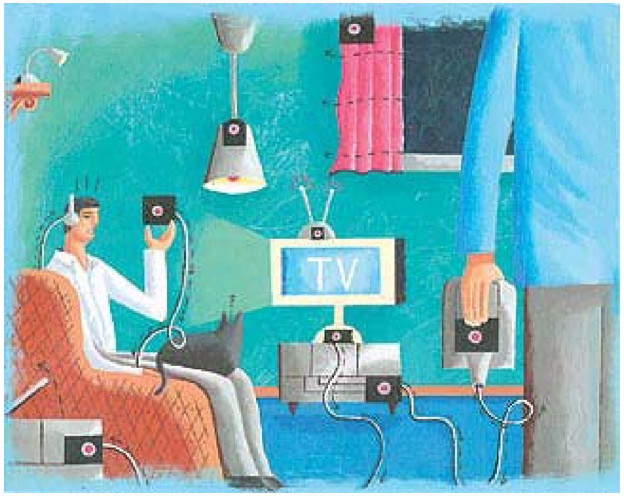
†) emphasis added

Digital Preservation Research

- Preservation of integrity extant bit streams
  - Security - assurance of the integrity of stored records
  - Privacy - assurance of only legitimate access
- Preservation of a bit stream representative of the tangible data record over generations of hardware and software migrations. (*)
  - capture, storage, retrieval, management, presentation and re-presentation of metadata encapsulation-based solutions.
  - metadata identification, creation or captured at the time of the creation of the records
  - metadata storage in inviolable conjunction with the record contents
  - assurance of access to record by authorized users over time
  - by whom, where, and at what costs the infrastructure will be constructed and maintained.
- Mechanisms for the re-presentation of the tangible data object: Universal Virtual Computer? (Semantic continuity)
- Emulation. Is it possible to "emulate" Windows XP Professional? How to emulate systems of high complexity?
- Cost models of digital preservation. Institutional costs, consumer costs. Include the capability of rendering intellectual content accurately, regardless of technology changes over time

(*) David Bearman, D-Lib Magazine, Vol. 5, No. 4, April 1999

Pervasive Storage



❶ Pervasive *access* to information, supports either "disconnected" or connected operation.

❷ Pervasive storage refers to the widespread availability of storage resources of practically unlimited capacity, over unbound geographic areas, concurrent with the consistent management of the stored assets and their immediate accessibility

Nokia N91 4GB phone

500 M units @ 10 GB = 5 Exabytes

DS2

Pervasive Storage Research Issues

- Storage cells vs. pure storage farms
- Name space management
  - – universal, unique identifiers regardless of home location for $O(10^{15})$ objects
- Privacy and security
- Architecture of the required metadata
- Data consistency
  - – multiple users share data object
- Intermittent connectivity operation
  - Economics - mass deployment of storage

Autonomic Storage: Why?

- Achieve predictable performance in complex systems
- Protect customer data against loss or corruption
- Reduce storage system management TCO
  - TCO: A result of complexity: Too many human-intensive tasks
- Respond more efficiently than human managers
  - Reaction times short as compared to physiological response
  - Spectrum of causes/responses overwhelms human analysis
- Find and fix problems (self-healing)
  - address system complexity
  - diagnose sources of error precisely, resolve problem

Issues in Autonomic Storage

- It's not just middleware
  - Components will likely have to change too
- It's not just storage in isolation
  - Applications and their needs must be considered
- The industrial psychology of autonomics is important
  - Have to influence how managers manage systems ("stick-shift" syndrome)
- Security matters

Some Research Directions

- Transparency
  - How to "explain" autonomic decisions to system manager?
- Evaluation and Metrics
  - How to compare how "autonomic" systems are?
- Study of Processes and Practices
  - What are the processes that we are automating?
- Management Policies
  - What are the policies and support machinery needed?
- Evolution, Growth, Scale
  - How to adapt over time as systems change?
- Specialized Storage Systems
  - How to build less general systems that are more autonomic?

DS2 Thrusts & Business Interest

Thrust	Business Opportunity			
• Active Storage Devices	Massively parallel database search and data mining	Massive indexing and searching	ILM & automatic destruction of data	Sensor networks
• Application-aware Storage	QoS, efficient I/O	Reliability	Security	System management
• Autonomic Storage	TCO	Predictability	Data integrity	Self-healing
• Long-term Storage	TCO	Data integrity	Language development	Consumer markets
• Pervasive Storage	Consumer markets	Record preservation	Consistency	Distributed storage utilities
• Privacy and Security	Assurance of service	Dispersed storage systems	Consumer markets	Record management

INSIC's DS2: Storage Research Partnership

- **Elicit agreement, funding, and coherence for supporting one or more research programs in data storage devices and systems**
- **The DS2 initiative has identified many avenues for pre-competitive research, where joint efforts by industry, academics, and government would benefit the global storage systems endeavor.**
- **"Open Innovation" as a win-win-win (industrial-academic-governmental) common endeavor**
- **This initiative deserves vigorous support.**

INSIC

DS2

Thank you

"Devices and Systems"

- In the context of this program, "devices" go together with "systems"
- "Devices" here are understood in the context of a managed storage system, not discrete independent devices
- "Systems" may not refer to devices at all, but to issues such as consistent file systems, content-addressable storage, or semantic continuity, that are not directly linked to a device.

DS2 Work Now – Summer 2005

- Formulated a roadmap for the evolution of data storage devices and systems
  - **Applications' opportunities**
  - **Technology trends**
  - **What could happen**
  - **What should change**

Elicit agreement, support

- Have determined that there are ***pre-competitive*** research subjects
  - **These are areas of research of undisputed relevance for which cooperative efforts are both possible and desirable. For instance, subjects far removed from marketplace implementation, or a business model based on non-exclusive ownership of technology.**

DS2 Workshop

- **Attendees: 62**
  - **Industry: 38 Universities: 19 Government: 5**
- **Organizations: 39**
  - **Industry: 23 Universities: 12 Government: 4**
- **Organizations Represented:**

IBM	StorageTek	Hitachi GST	*California Institute for Telecomm & IT**
IDC	Qualcomm	U. Colorado	*Information Storage Industry Center **
RPI	Pillar Data	U. Minnesota	*Digital Technology Center, U. Minn **
SGI	Sun Micro	Penn State U.	*Center Mag. Rec. Research, UCSD **
EMC	Quantum	U. Wisconsin	Hitachi Data Systems
SNIA	Seagate	BAE Systems	Library of Congress
NASA	Xyratex	Santa Clara U.	Data Mobility Group
UCLA	Google	Agere Systems	Los Alamos Nat. Lab
UCSD	Ciprico	UC Santa Cruz	Carnegie Mellon U.
INSIC	Veritas	Hewlett-Packard	Fermi National Lab
INTEL	Oracle	Johns Hopkins U.	SD Supercomputer

(*) Co-sponsors

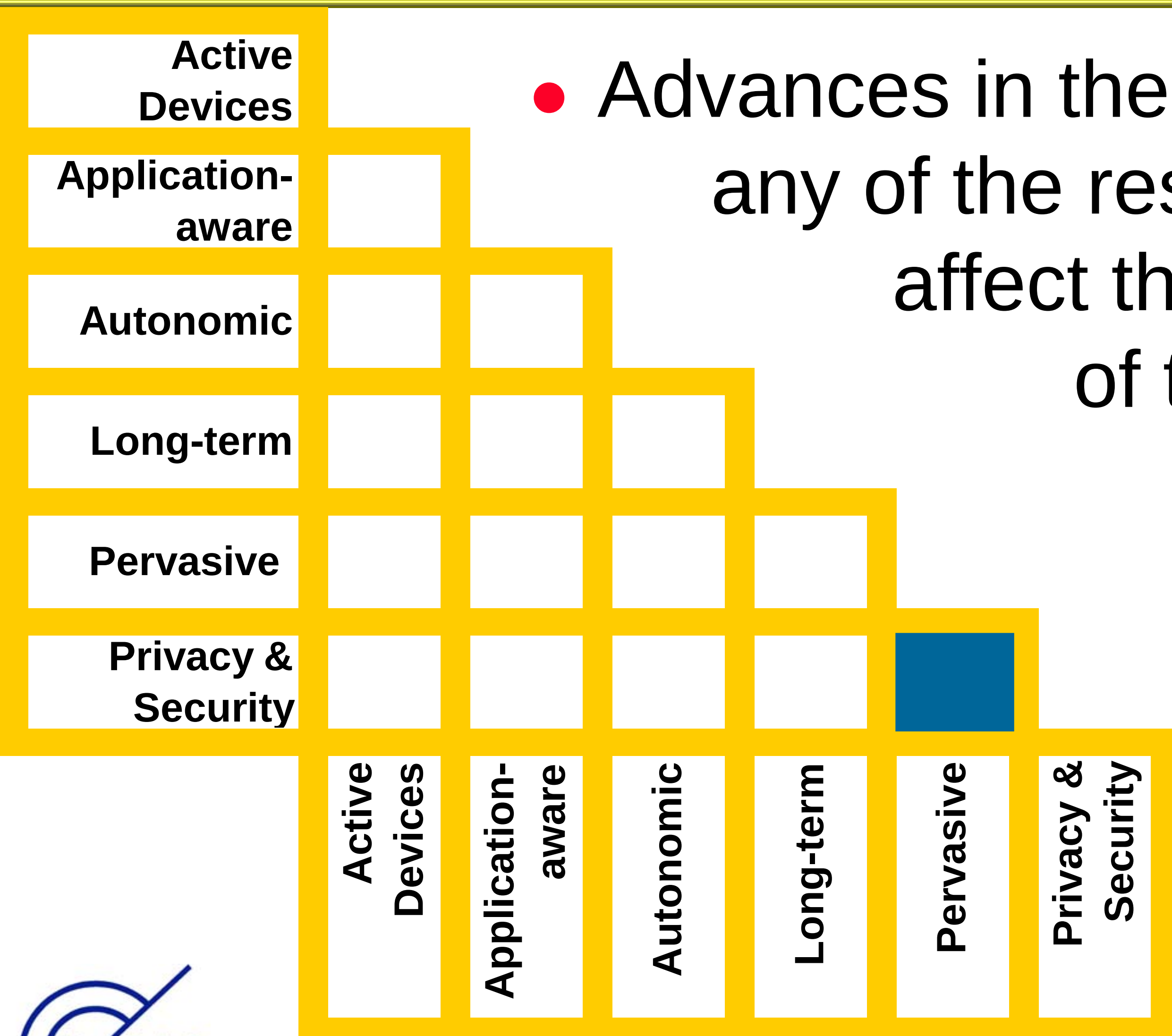
DS2 Research Program Development

Photo of DS2 Workshop Participants, UCSD, April 2004

DS2 Research Program Development

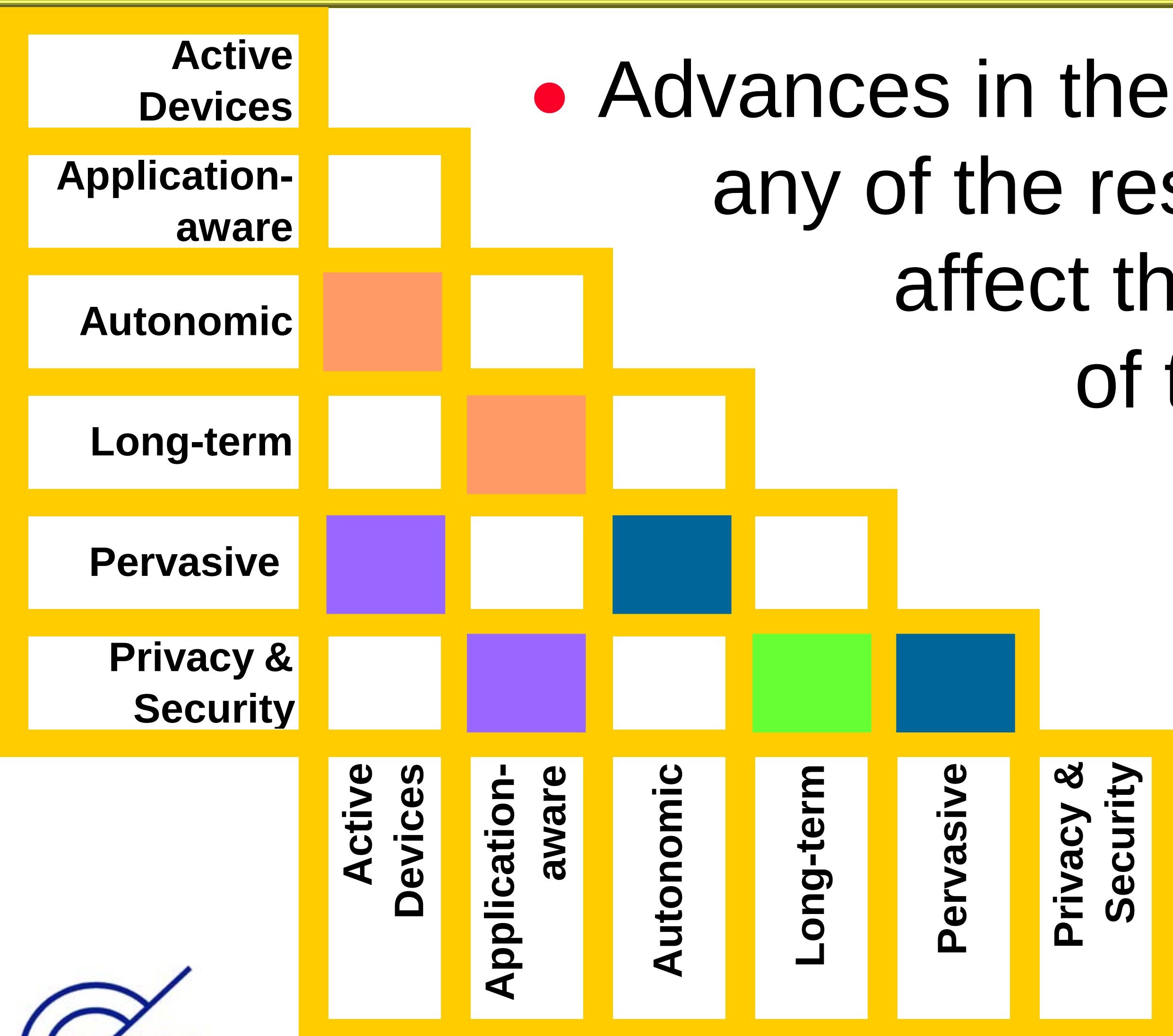
- Planning
  - Feasibility study, Summer ’03
  - Industry visits, Fall ’03
- Implementation
  - Workshop organization, Winter’03/04
  - Workshop @ UCSD, April 2004
  - DS2 Roadmap published, January 2005
- Research Program Execution
  - Validate Roadmap, Spring/Summer ’05
  - Obtain support from systems’ companies - NOW
  - Launch pre-competitive research program

The Fabric of Storage Systems Research



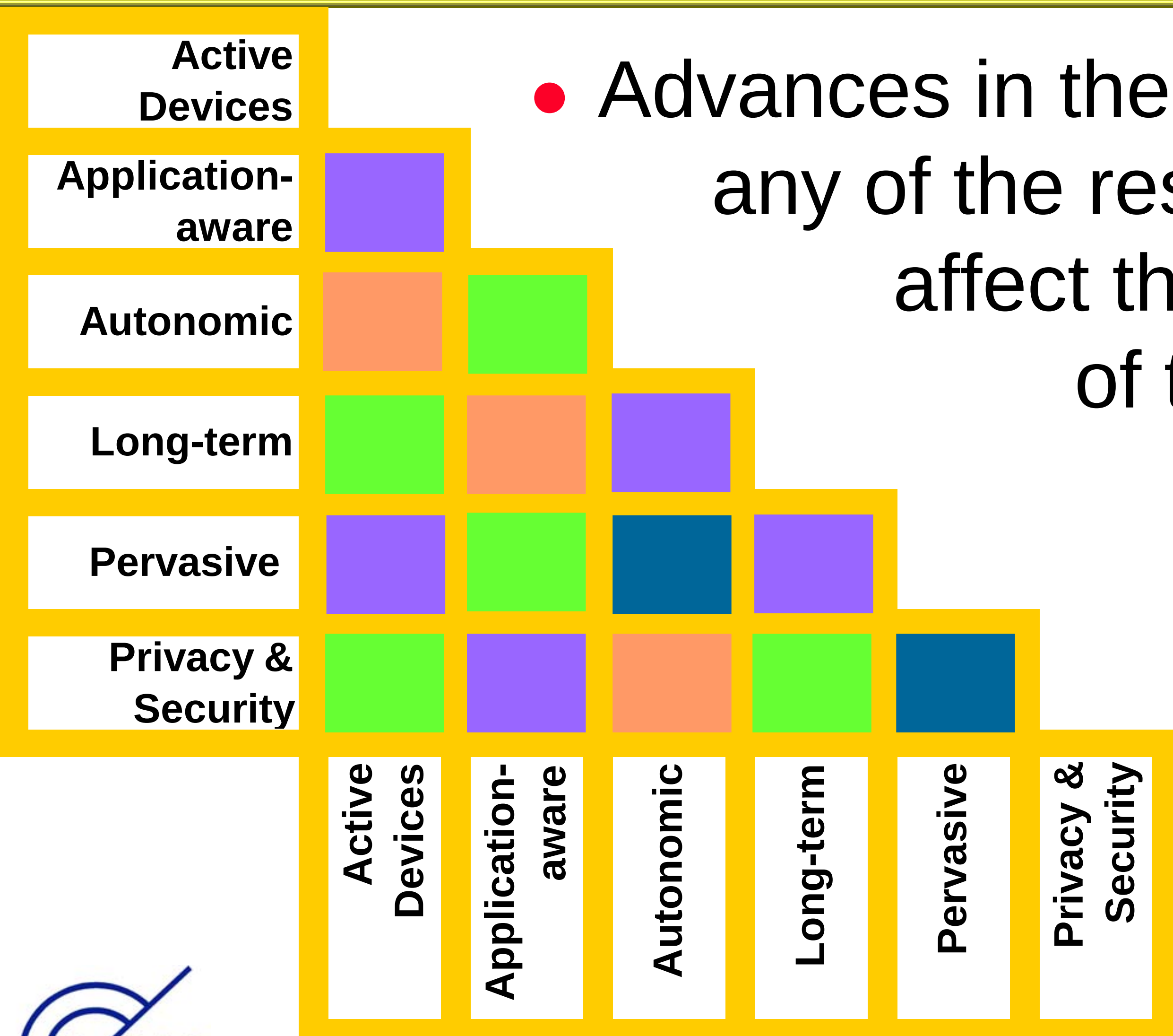
- Advances in the technologies of any of the research thrusts affect the performance of the system

The Fabric of Storage Systems Research



- Advances in the technologies of any of the research thrusts affect the performance of the system

The Fabric of Storage Systems Research



- Advances in the technologies of any of the research thrusts affect the performance of the system

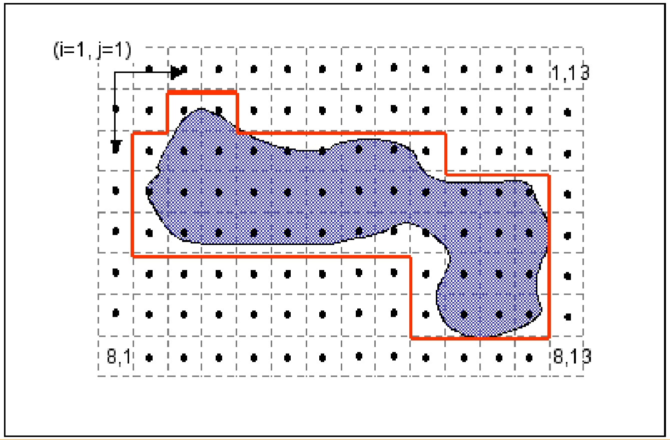
Application-aware Storage Opportunities

- Spatial and temporal access patterns
  - For better data layout and organization
- Relationships among data, users and apps
  - For improved indexing, searching, organizing
- Data replication factors
  - For higher availability and data reconstruction
- Access control lists and what I/O is “normal”
  - For device-resident anomaly detection
- Caching hierarchies
  - For exclusive and/or cooperative caching
- Application goals (e.g., latency, availability)
  - For autonomic storage

Active Storage Devices Examples

High-performance computing

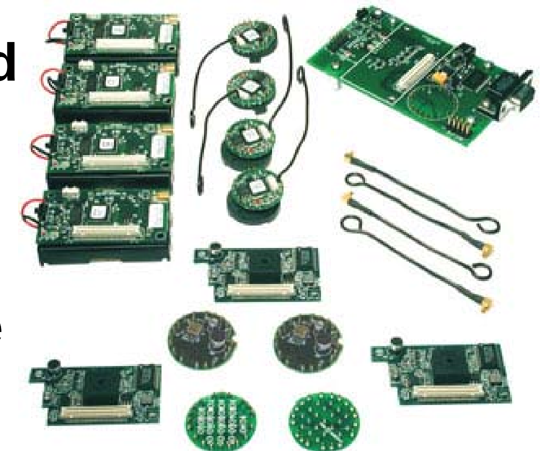
Mesh computa-tions performed by device

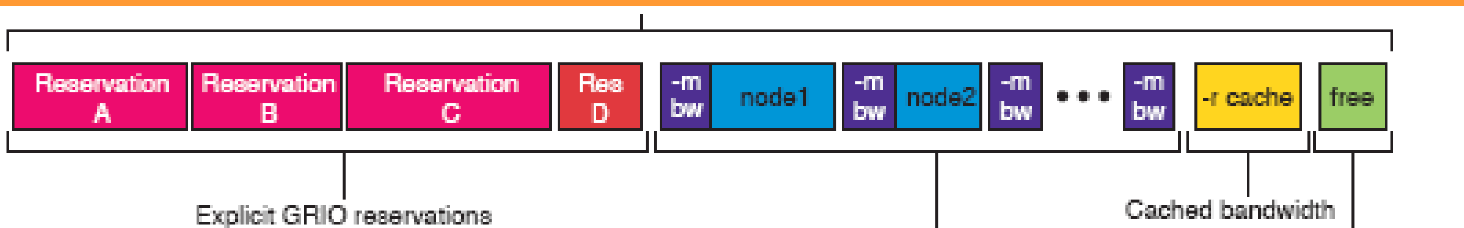


Sensor networks

Only processed data is reported.

Raw data may never leave the storage device



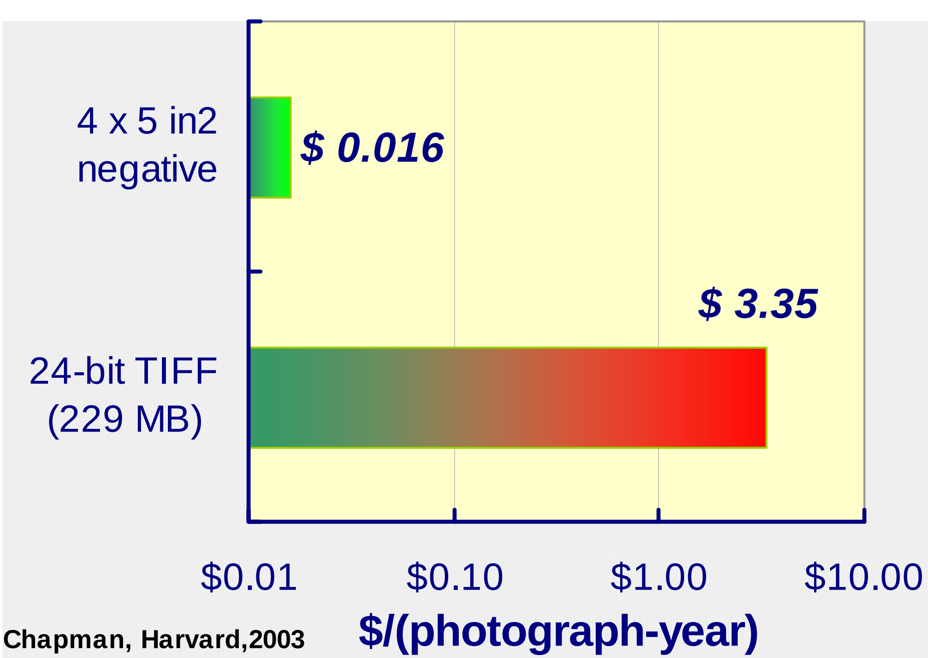


Quality-of-Service assurance, GRIO = guaranteed-rate I/O

scheduling of dedicated bandwidth on a SAN for guaranteed real-time performance

Preservation Cost Issues & ROI Models

- Comparison of costs between the Harvard Depository film vault and the Online Computer Library Center, Inc., Digital Archive(2003). **(Chapman, 2003)**



Chapman, Harvard,2003

- Factor 200 in favor of film plate
- Digital costs include extant bit preservation and exclude long-term preservation
- Raw capacity cost of disk 229 MB: $0.069 (2004)

Privacy and Security

- Privacy refers to the denial of access to stored records by unauthorized clients concurrently with the assurance of access by authorized ones
- Security refers to the assurance of the integrity of stored records concurrently with efficient access by multitudinous clients

Privacy & Security: Hard Problem

- **Balancing the efficient access to data against the overheads of safeguarding data integrity and safety**
  - **Number of Firewalls**
  - **Holes**
    - dial-in
    - carry-in
    - VPN-in
  - **Insiders**
    - If there are 1,000 (100,000) employees inside ...
  - **Breaches**
    - What happens when the inevitable happens

Privacy & Security Research Opportunities

- Data Integrity: protection and recovery
- Data Privacy
- Data Destruction
- Intrusion Detection
- Key Management
- Authorization
  - Authenticity
- Operational risk
- Economic issues

Pervasive Storage Architecture

- System configurations
  - Storage cells - caches for data
    - analogous to wireless telephony cells
  - Interconnected storage farms
    - Home locations for the data
    - Persistent data repositories - commitment to service
- Usage Modes
  - Consumers - acting on their own behalf
  - Corporate citizens - acting within an organization
  - Sensor networks
  - Data centers QoS enhancements by availability of a pervasive storage infrastructure

Autonomic Storage

- Autonomic storage is:
  - → Self-configuring
  - → Self-optimizing
  - → Self-healing
  - → Self-protecting
  - → "Self-*": important computing operations can run without the need for human intervention

Example: Detection, diagnosis, and avoidance of service interruption or system failure