Storage—The Unnoticed Revolution

Most emphasis in Computer Science is on Processing:
- Fast processors getting cheap; SpecMarks, Etc
- Multiprocessor architecture
- Programming

Except for the Data Communications people:
- Gigabit networks getting cheap
- Organizing Video streams

There are equally significant challenges in STORAGE:
- RAM getting cheap
- Magnetic Disk getting cheap

Notes for a talk by Jerome H. Saltzer, given at an open session of the Computer Science and Telecommunications Board of the National Research Council on May 24, 1993.

Here is the curve for Magnetic Disk.

There is a similar curve for Random Access Memory (RAM)

The upper edge represents a doubling time of one year, which is what street prices have been doing since 1983.

The lower edge of the fuzzy area is based on the progress rate of Magnetic Areal Density, which has been 40%/year for more than 10 years, and promises to continue for perhaps 20 more years. We are getting closer to the MAD ceiling, so will probably switch over to that curve in a year or two.

Most people are familiar with the numbers in the first two boxes of this chart, representing 1983 and 1993. Both RAM and Magnetic Disk storage capacities in typical large storage systems have increased by factors of about 100; in personal computers the factor is 25 to 50 because the physical size of the system has decreased significantly.

But look what happens when we extrapolate ten more years. I want to call your attention particularly to the information service column, where capacity per server will be in the 1-10 Terabyte range.

What is happening here is that the cost of on-line storage is declining faster than the adaption rate of applications; when things change by an order of magnitude it takes a while to realize all the things that become feasible.

One of the things that is about to cross the feasibility threshold is static image storage, of books and papers—the thing sometimes called the “Electronic Library.”

Let’s look at the driving forces involved.

The cost of storing a scanning image of a page on magnetic disk is currently about ten times that of the corresponding piece of paper. But in 1998 the two costs will be about equal, and in 2003 it will be ten times cheaper to use magnetic disk.

An even more striking crossover occurs on volume. The space occupied by a scanned image of a page on magnetic disk is currently about EQUAL to the space occupied by the corresponding piece of paper. (A 1 Gbyte disk will hold approximately one shelf of books in scanned image form.) By the year 2003 the magnetic disk will be 100 times smaller.

Since the two scarcest things in libraries are dollars and space, these two crossovers can’t help but attract a lot of attention.

(SKIP THIS SLIDE IF WE ARE RUNNING OUT OF TIME)

Here is a curious, but little-recognized set of facts:

The cost of silicon storage has hovered around 25 times the cost of magnetic disk storage for some time now, and it promises to stay there through the next decade.

The space required to storage the image of a page is about 25 times the space required to store the tagged ASCII representation of the words on that page.

When you put these two facts in front of you at the same time, it suddenly is apparent that any image you can afford to place in magnetic disk storage, you can afford to index in RAM.

This observation has profound implications for people working on information retrieval and discovery.

So where is the action, and what might CSTB do in this area?

- Finding what you want in a sea of information (Discovery)
- Getting someone else to tell you about information you care about(Alerting)

- Mobility + Information access leads to killer applications:
    - working at home at your mountain cabin
    - every automobile is on the network
    - every walkman is on the network

- For CSTB: Help in discerning boundaries among the players.

Role for CSTB:

Stay on top of the intellectual property issue—it could stall the technological opportunity.

copyright: intellectual property = revenue
liability
revenue flows
jockeying for position

Old regulations are becoming technically irrelevant
Pragmatic barriers are tumbling.

The boundaries among these entities are very fuzzy already, and the cheaper storage gets, the fuzzier they will be. Two more factors of ten in fuzziness could be deadly.

Note that off the edge of this slide, stirring in the wings, are a couple of other ideas that are ready to build on cheap storage: (1) Instant CD's and (2) Video on Demand.

It may help to take a very deep historical perspective. I am indebted to a contributor from Western Australia to a medieval history discussion list for the observation that there may be some useful lessons from the middle ages. On-line documents, for which there is a master copy that everyone consults, resemble medieval manuscripts in several ways:

- The concept of "copyright" is non-existent.
- There is no clear concept of "publication," which may be defined as the production of a widely distributed edition of many identical, fixed copies. Instead, the texts change with time and are massaged for the purpose at hand.

With that thought I had better sit down.

Assumptions used in preparation of the slides:

1. A book has a paper/printing component of about one penny/page.

2. A scanned page image, of 300 DPI compressed with G4 Fax standards, takes up about 100 KBytes.

3. A page of ASCII, with tags showing what is where, occupies about 4 Kbytes.

4. A 2 Gbyte disk costs about $2000 today. That is $1/MByte, or 10 cents/page-image.

5. RAM sells for about $25/MByte today. That is 2.5 cents/Kbyte, or 10 cents/ASCII-page.

6. A shelf of (40) books, in “compact storage,” takes up about 3 cubic feet and contains about 20,000 pages.

7. A 2 Gbyte disk, with power supplies and space for access, takes up about 3 cubic feet. At 0.1 Mbyte per page, there is room for about 20,000 pages.