Storage Systems

NPTEL Course

Jan 2012

(Lecture 14)

K. Gopinath

Indian Institute of Science

Types of Storage

- “Mobile” Storage
  - Memory Stick, Camera, Smart Phone, Laptop
- Personal Storage
  - PC, “Home” RAID systems, “Home” NFS server
- Dept/Organizational Storage
  - NFS/CIFS server
- Cloud Storage
- Highly Available Storage
- Parallel Storage
- Web-scale storage
- Secure Storage
- Attribute-based Storage (“QoS”)
- Long term storage
  - DNA storage!?

Parallel Storage

- Access to storage in parallel in HPC appls
- Multiple threads (cores) working on a large distr file across many nodes
- Scalability issues critical
  - Access/Updating of metadata
  - Bottlenecks in file name processing

Web-scale storage

- Google
  - GFS/BigTable, gmail, ...
- Facebook
- Dropbox, ...
- Cloud storage
  - Amazon, Azure storage services

Facebook Modular Storage

- Web/Chat
  - High CPU, low memory 16GB, low capacity 250GB disks
- Database
  - Med CPU, high memory 144GB, High IOPs: 3.2TB flash
- Hadoop
  - Med CPU, med memory 48GB, high capacity: 12x3TB SATA
- Haystack (photos, video)
  - Low CPU, low memory 18GB, high capacity: 12x3TB SATA
- Feed (ads, search, multifeed (to feed the “Wall”))
  - High CPU, high memory 144GB, medium 2TB SATA
- “Flash sled”: 500GB-8TB, 600kIOPS
- “Storage sled”: 15 drives, 3kIOPs

Storage Security

- Appl
  - Can be selective wrt data but need to know what is sensitive
  - Each appl has its own mgmt of security
- FS/OS
  - Can be selective but costly in time in sw
  - Changes in FS/OS?
- Device driver (HBA), netw interface
  - Only for data in flight (SAN)
- Network
  - Heterogenous devices OK (disk/tape)
- Storage Controller
- Device

Secure Storage: Disk

- Entire drive (incl MBR) encrypted
  - MBR+OS unmodified (transparency); MBR cannot be corrupted
  - Auth occurs before OS/any malicious software loaded

- BIOS attempts MBR read
  - drive redirects to pre-boot area
- Drive loads pre-boot OS
- User enters auth credentials for drive to verify
- If auth successful, drive loads original MBR
- Normal operation commences

Attribute-based Storage

- Latency, BW, reliability guarantees
  - May be power, longevity, ... also!
- End2end guarantees typ
  - Difficult!
- In research stage still!

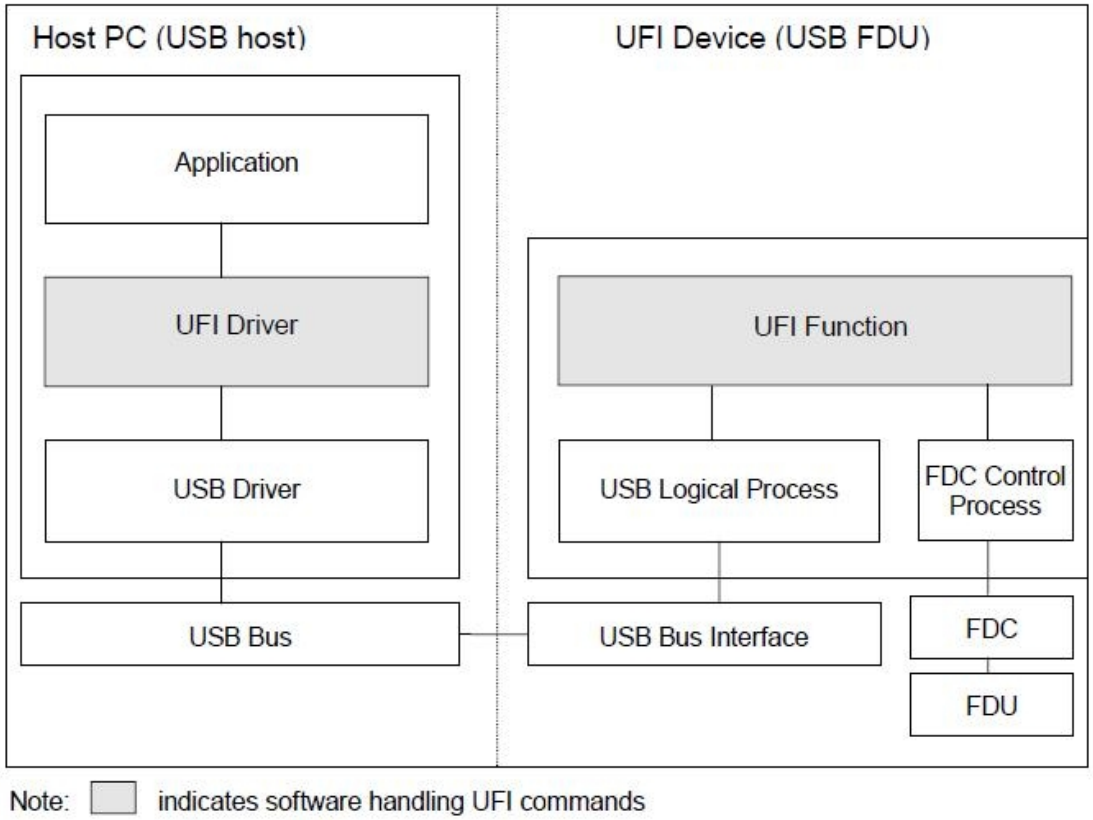
Saving Current Documents for the Next Millennium

- Across time (written now 2012 and read then 3012) & space (written here on Earth & read there on Mars!)
  - Say document written in Postscript/ Wordstar/ Word!
  - Stored on a SCSI 36GB 15K rpm disk drive on, say, a ext3 filesystem on Linux...
- What is the ***hard*** technical problem?
  - Drives/device driver/filesystem/kernel/application may become obsolete
  - Along with document, the associated model of device/software may need to be saved (recursion problem!)
- Any current technologies useful?
  - ASN.1? Virtual machines? Virtualization?!
- *Till today no effective solution!*

Evolution

- Design and architecture of subsystems (say, USB) may need to be modified over a period of time.
- Consider a (new) development where we have thin clients with a non-local server storing data and configuration.
- Users would still want the convenience of USB devices; if this is not possible, thin clients will not succeed.
- Though USB devices were expected to be electrically connected to the PC, it is possible to have remote USB functionality with appropriate redesign of the layers, using Internet as the connection.
- We need to “virtualise” some aspects of the previous legacy design.
  - also happened with other devices
    - terminals resulting in pseudo-terminals
    - SCSI resulting in iSCSI designs that allow remote access to disks through Internet.

Emulation of Floppy Drive (from UFI std)



UFI: USB Floppy Interface
FDU: Floppy Disk Unit
FDC: Floppy Disk Controller

Indus Script

- Yet un-deciphered: meaning across time not yet accomplished
- Compare with hieroglyphics (Egyptian Rosetta stone): three scripts side by side

- What is the problem? Not enough contextual info:
  - can see the script (human "readable")
  - no mapping between symbols and phonemes
  - need interpretation of sequences of symbols
  - a problem in archaeology, history, society,...

Vedas

Transmitted across atleast 3500-5000 years without differing versions

Including exact pronunciation!

“UNESCO proclaimed the tradition of Vedic chant a Masterpiece of the Oral and Intangible Heritage of Humanity on November 7, 2003”

What "technology" used? Redundancy!

Various "pathas" of Samhita text: can recover from a corrupted text due to added redundancy: RAID-like! (Redundant Array of Indep Disks)

Pada-patha: each word in its separate form

Krama-patha: connects a word in pairs

ABCD becomes AB BC CD DE... (“2-mirroring”): 2 copies

Jata-patha: ABBAAB (“3-mirroring”): 3 copies of A, B, ...

a Ghana-patha (ABBA ABCCBA ABC BCCB BCDDCB BCD...)(“10x”)

Metrical (similar to checksums!) & Musical

"Information dispersal"

Human Reproduction! (Oral transmission)

Use efficient “virtualizers”!

पदपाठ

1,2,3

तम् । भागधेयेन । वि । मुञ्चति । प्रतिष्ठित्यै । यया । रज्ज्वा । उत्तमां । गाम् । आजेत् । ताम् । भ्रातृव्याय । प्र । हिणुयात् । निर्ऋतिम् । एव । अस्मै । प्र । हिणोति ॥ तै सं २-२-६-५ ॥

घनपाठ

1,2,2,1,1,2,3,3,2,1,1,2,3.

तं भागधेयेन भागधेयेन तं तं भागधेयेन वि वि भागधेयेन तं तं भागधेयेन वि ॥

भागधेयेन वि वि भागधेयेन भागधेयेन वि मुञ्चति मुञ्चति वि भागधेयेन भागधेयेन वि मुञ्चति । भागधेयेनेति भागऽधेयेन ॥

वि मुञ्चति मुञ्चति वि वि मुञ्चति प्रतिष्ठित्यै प्रतिष्ठित्यै मुञ्चति वि वि मुञ्चति प्रतिष्ठित्यै ॥

मुञ्चति प्रतिष्ठित्यै प्रतिष्ठित्यै मुञ्चति मुञ्चति प्रतिष्ठित्यै यया यया प्रतिष्ठित्यै मुञ्चति मुञ्चति प्रतिष्ठित्यै यया ॥

प्रतिष्ठित्यै यया यया प्रतिष्ठित्यै प्रतिष्ठित्यै यया रज्ज्वा रज्ज्वा यया प्रतिष्ठित्यै प्रतिष्ठित्यै यया रज्ज्वा । प्रतिष्ठित्या इति प्रतिऽस्तित्यै ॥

यया रज्ज्वा रज्ज्वा यया यया रज्ज्वोत्तमामुत्तमꣳ रज्ज्वा यया यया रज्ज्वोत्तमाम् ॥

रज्ज्वोत्तमामुत्तमꣳ रज्ज्वा रज्ज्वोत्तमां गां गामुत्तमꣳ रज्ज्वा रज्ज्वोत्तमां गाम् ॥

उत्तमां गां गामुत्तमामुत्तमां गामाजेदाजेद्गामुत्तमामुत्तमां गामाजेत् । उत्तमामित्युत्ऽतमाम् ॥

गामाजेदाजेद्गां गामाजेत्तां तामाजेद्गां गामाजेत्ताम् ॥

आजेत्तां तामाजेदाजेत्तां भ्रातृव्याय भ्रातृव्याय तामाजेदाजेत्तां भ्रातृव्याय । आजेदित्याऽअजेत् ॥

तां भ्रातृव्याय भ्रातृव्याय तां तां भ्रातृव्याय प्र प्र भ्रातृव्याय तां तां भ्रातृव्याय प्र ॥

भ्रातृव्याय प्र प्र भ्रातृव्याय भ्रातृव्याय प्र हिणुयाद् हिणुयात्प्र भ्रातृव्याय भ्रातृव्याय प्र

What is needed?

from *Avoiding Technological Quicksand: Finding a Viable Technical Foundation for Digital Preservation by Jeff Rothenberg January 1998*

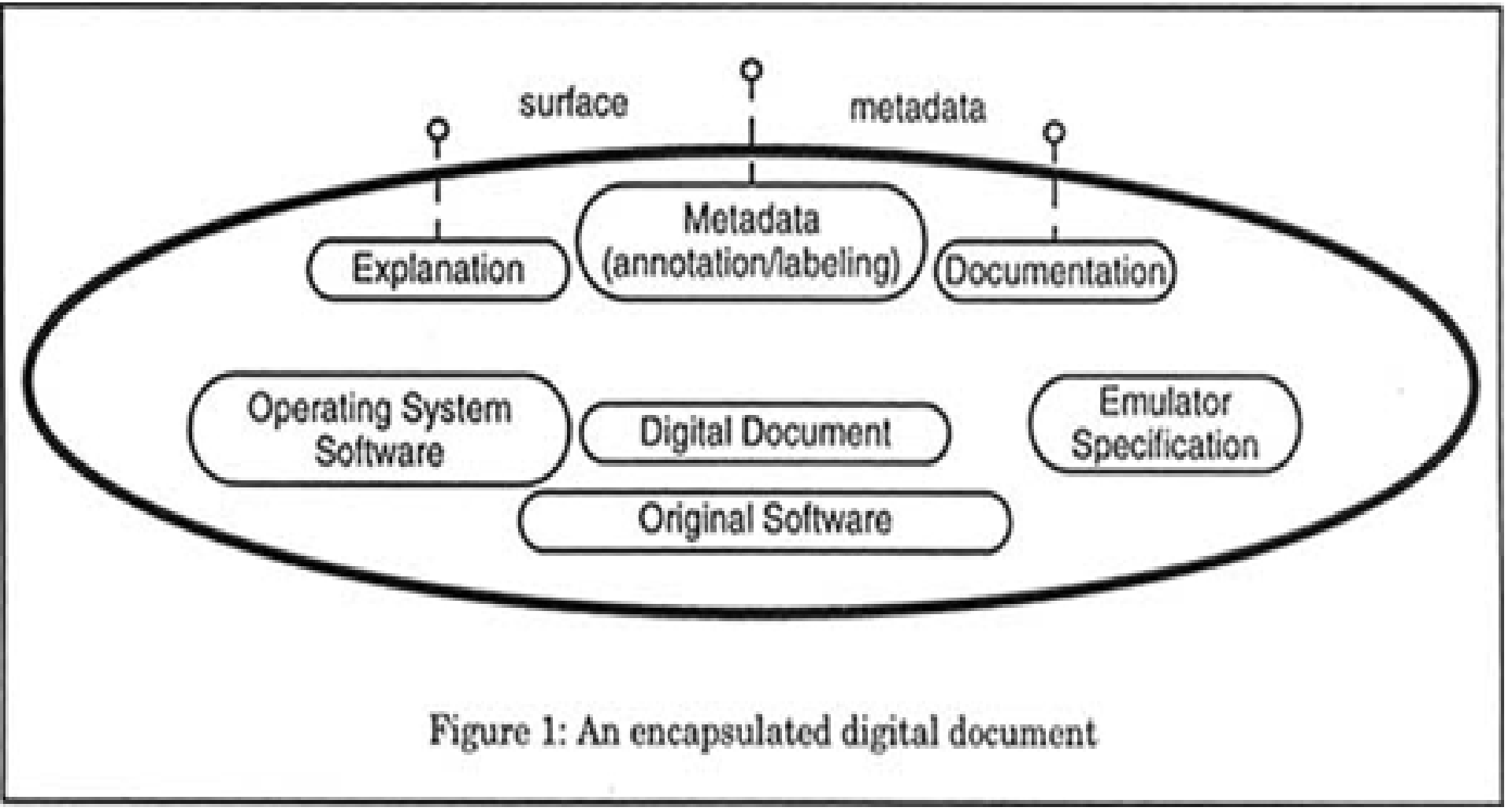


Figure 1: An encapsulated digital document