Energy-Efficiency of Long-term Storage

Irina Tolokonnikova

Seminar "Energy-Efficient Programming"
Arbeitsbereich Wissenschaftliches Rechnen
Fachbereich Informatik
Fakultät für Mathematik, Informatik und Naturwissenschaften
Universität Hamburg

2015-01-14
show the audience what and how much data an archive has to handle with. I want to use the data of the DKRZ given on its homepage and the xkcd-what if? about google https://what-if.xkcd.com/63/

What is an archive for digital data? How is data stored? and about how much data are we talking about?
storage of digital data for many years

requirements:
- preservation
- retrieval
- auditing

archival data ≠ backup data

needs to be cheap to obtain, cheap to operate, easy to expand

high costs for energy consumption

→ room for improvement

main requirements of an archive but the main problem of archives are the high costs for energy consumption, and in this talk, we want are going to see, how we can reduce this costs improvements are important in this field, because we want to match the requirements for big data that means, we want to work with huge amounts of data, compare it and search for association rules e.g.
How much data are we talking about?

DKRZ: > 100 PetaBytes total capacity [1]

Google: ∼ 15 ExaBytes (in 2013) = 15000 PetaBytes (only estimation)

what if-comic, where people can send in absurd but interesting questions, and Randall Munroe will try to answer it. That’s more than a lot of data, because in 2013 only 8 ExaBytes of Hard Drives were produced for sale in total worldwide.

NSA probably stores 1 YottaByte
How much data are we talking about?

- **DKRZ:** > 100 PetaBytes total capacity [1]
How much data are we talking about?

- DKRZ: > 100 PetaBytes total capacity [1]
- Google: ~ 15 ExaBytes (in 2013) = 15000 Petabytes (only estimation)

how much data is stored in such an archive? first the data from the DKRZ homepage, to have an anchor for further comparison. This estimation based on the published energy consumption. Google doesn’t publish how much data they store.

what if-comic, where people can send in absurd but interesting questions, and Randall Munroe will try to answer it. that’s more than a lot of data, because in 2013 only 8 ExaBytes of Hard Drives were produced for sale in total worldwide.

NSA probably stores 1 YottaByte
Google

but which device should we choose, if we want to store as much data as Google does?
show some devices, that are not used for archives and which are
Show a real cassette, make clear, why LP or Punch cards are not suitable for long term storage

Figure: 15 Exabytes of punch cards would be enough to cover New England, to a depth of about 4.5 kilometers
probably not LPs - not easy to search and not as much data, BUT it lasts very long without errors, because it is engraved

LPs were sent to space, so alien life forms get an impression of the earth, music and human life. It was well chosen, because it had to have a long lifetime.
as we already learned, punch cards would not be suitable for masses of data
Figure: a United States National Archives Records Service facility in 1959. Each carton could hold 2000 cards [wikipedia.org]

... because than out storage would look like this
Show, how "normal" cassettes look like and where we know them in everyday life. Maybe you still know cassettes, but do your younger siblings still know how to use them? Explain that they are still used in another region.
maybe some of you can remember, that cassettes were used not only for music but also for data in such computers like C64 but the cassettes used for data storage have changed.
Tape

notetoday cassettes have a single reel and can store as much data as Hard Discs

- used as a cartridge with a single reel
- holds several tens to thousands of GB (state wikipedia.org 13.01.15)

they look like this and are well protected by the case
Tape

notetoday cassettes have a single reel and can store as much data as Hard Discs
- used as a cartridge with a single reel
- holds several tens to thousands of GB (state wikipedia.org 13.01.15)
- Oracle StorageTek T10000 T2 hold 8.5 TB

they look like this and are well protected by the case
7 automated Oracle/StorageTek SL8500 tape libraries
8 robots per library
over 67000 slots for magnetic tape cassettes

Figure: Inside the Tape library of DKRZ [1]
lifetime and costs

- lifetime: 30 years
- costs: less than 1 cent per GB
- 238X less energy over 12 years than HDD

main point, why tape is used for long term storage: When not in use, tape doesn’t produce any energy and doesn’t need any electricity. It is important to upgrade the device, even if it has a long lifetime, because you can save much space, as shown in this graphic.
### Pros and Cons

<table>
<thead>
<tr>
<th>Pros</th>
<th>Cons</th>
</tr>
</thead>
<tbody>
<tr>
<td>cheap</td>
<td>needs special equipment</td>
</tr>
<tr>
<td>long lifetime</td>
<td>sequential access pattern</td>
</tr>
<tr>
<td>no power needed when not accessed</td>
<td></td>
</tr>
</tbody>
</table>

Tape has to be rewinded when searched and needs much more time to get to information, even if you know, where it is stored.
Even more time is needed, if you want to search for data or compare it.
Hard Drives

- easy and fast to access data storage
- searching, consistency checking and inter-media reliability operations
- costs: 0.07 $ per GB and falling
- lifetime: 10 years, but easy to break mechanics

Same for Hard drives, but in this case a real object is not needed, because every one of you probably has one or more at home the costs get higher, when we need to change the Discs more often and when we need redundancy to save data, in case one disc breaks normal HDD, like in most of your laptops probably
### Pros and Cons

<table>
<thead>
<tr>
<th>Pros</th>
<th>Cons</th>
</tr>
</thead>
<tbody>
<tr>
<td>easy access, simply system</td>
<td>needs much power, even when turned off</td>
</tr>
<tr>
<td>matches requirements of big data</td>
<td>easy to break</td>
</tr>
<tr>
<td>higher bandwidth (200X)</td>
<td>needs extra space for redundancy</td>
</tr>
</tbody>
</table>

When storing on HDD the main point to remember is that extra space for redundancy is needed, and that’s why more space is needed for the devices, than we would estimate from raw data amount.
Colarelli, Grunwald et al. (2002)

- massive array of idle disks = MAIDs
- aim: storage densities matching those of tape, with reduced energy consumption
- but operating same data volume in disks costs 10X more than in tape
- idea: use a cache manager to keep only part of disks in an array powered up
- varying spin-down delays

a newer approach on handling HDD energy efficiently. An important article, cited in following research articles dealing with the question, how we can store on HDD more efficiently, maybe someday as eedicient as on tape when using only the disks already spinning, we can save energy.

previos observation: mostly searched information is those just saved or added. So by keeping this information in cache we probably can manage most of the requests ...to save extra energy
Results

- Good trade off in performance and energy efficiency
- Read performance still effected by the spin-down delay
- But 82% of read requests were satisfied by the cache
- Least energy consumed with 4 sec spin-down delay

What are the main points about MAIDs, using MAIDs the HDD might still not be as efficient as tape, but it's a good approach to work on.
SSD costs: 0.66 $ per GB, yet too expensive.

Lifetime depends on usage, ~10 years.

Yet unclear, how unused data behaves on SSD coming soon?

Looking forward: will we use SSD in some years?
Pergamum tomes by Storer et al. (2008)

- interfaces and protocols change slowly
- using inter- and intra-device redundancy
- work energy efficient, by not spinning up idle disks
  → intelligent, self managing storage device

Figure: Pergamum tome, redrawn

newer research and what will maybe be the future of long term storage talk about, how HDDs can be made more efficient but still fit Big Data

Named after the library of Alexandria

main ideas of this paper are based on those about MAIDs
Use the fact, that interfaces and protocols take much more time to change. Just think of http... not only for restoring data, which was saved on broken devices, but also for reading
Results of Pergamum

- size of the hard drive
- nonvolatile RAM handles many types of requests (e.g., hashes) without spinning up the disk
- using signatures for redundancy checking in entire inter-disk group
- using trees of hash values to reduce signature data
- once added to the network, the tome automatically joins a redundancy group or builds new one
  → makes storage management easier
- using intra-device redundancy, recovering from small errors without other devices
- aim to be price-competitive with tape

main points about the Pergamum tome, that can help improve it in future small and cheap energy efficient itself a good algorithm helps reducing the energy only a minimum trained administrator needed to change the broken devices once a month.

the new devices will organise themselves in existing groups or start new ones this goal is not reached (yet?)
Problems and improvements

- still not included in data archives(?)
- redundancy overhead, but much energy saved
- "disposable" tomes
- encoding time 10X longer than on laptop processor BUT 10X less power consumed
- future work:
  - better algorithms
  - parallel processes (distributed searching)

I don’t know much about this point, and I couldn’t find anything about this are disposable devices really more energy efficient and saving money? Do we really want so throw that much HDDs away? future tasks named in paper
A Spin-Up Saved is Energy Earned, Greenan et al.(2008)

- idea: use redundancies on active devices instead of waking up inactive ones

→ Power aware coding

- three conditions needed:

![Diagram: Three conditions for a power-aware system]

Figure: Three conditions for a power-aware system

Based on "A spin-up saved is Energy earned" paper by Greenan et al. (2008). It’s slightly newer and it also presents some algorithms. But I find it a little confusing and I think, it would be too much, if I explained it in detail, so here is only a summary of the main ideas, with pictures I made to illustrate them.
Power Aware Techniques

- rules known from Pergamum tome
- Power Schedule
  - each code instance should have own write policy
  - write parallel across disk groups
- Power-Aware Read Algorithm
  - minimize the number of disk activations
  - first find out, if lost data is recoverable
  - like solving a matrix where inactive devices are treated as erased
- Disk Activation Algorithm
  - perform search to find best activation
  - how and when is a spin-down performed?
mind the trade-off trilemma!

**Figure:** the trade-off when trying power aware coding

- open questions:
  - which environments will benefit from power aware coding?
  - how to find optimal policies?
  - robust metrics have to be developed for evaluation the power-reliability-performance trade-off

it’s not called trilemma in the paper, that’s my interpretation of it and the main point I want the audience to keep in mind: You always have trade-offs, be aware where you want to have them. Important point for future research
### Conclusion

<table>
<thead>
<tr>
<th></th>
<th>Disk</th>
<th>Tape</th>
</tr>
</thead>
<tbody>
<tr>
<td>Max shelf life (bit rot)</td>
<td>10 years</td>
<td>30 years</td>
</tr>
<tr>
<td>Best practices for data migration to new technology</td>
<td>3-5 years</td>
<td>8-12 years</td>
</tr>
<tr>
<td>Uncorrected Bit Error Rate, Probability (avg 1 error in x TB)</td>
<td>$10^{-14}$ (~10's of TB)</td>
<td>$10^{-19}$ (~1 million TB)</td>
</tr>
<tr>
<td>Power and cooling</td>
<td>238X</td>
<td>X</td>
</tr>
</tbody>
</table>

**Figure:** Disk compared to Tape [3]
In the summary I want to present the pros and cons of tape and HDD and on which data or in which fields which seems to be the better choice. Also I want to remind, that the life time of the device chosen should not be forgotten this would be the basic slid for discussion this quote shows, that you always have to keep in mind, how and how often you want to reuse the stored information if you want to search for data or compare random files or look for association rules e.g., you will have to calculate with higher energy consumption
How would you store...

- (your own) private medical data?
- research data of a medical study?
- data of all patients of a hospital?

Instead of a slide with "Any questions? Thanks for listening" I want to end with a question, the audience should answer for themselves, which device they would choose.
How would you store...

Instead of a slide whit "Any questions? Thanks for listening" I want to end with a question, the audience should answer for themselves, which device they would choose.
References