CS696 Research Project (2) — The whole process

Tao Xie

Department of Computer Science
San Diego State University

Content

- Establishing your goal and timing
- Identifying a research problem
- Summarizing existing related work
- Focusing on one particular technique
- Designing your algorithm
- Conducting experiments
- Writing your technical paper

Establish Your Goal

Before taking any action, ask yourself the following questions:

- In what areas I have strong background ?

  During my previous studies and research in CS, I have been good at ….

- What areas interest me ?

  System Resource Management (scheduling); Real-Time/Embedded Systems; Database Management; High Performance Computing; Parallel and Distributed Computing, Information Security, Programming Languages, etc.

- How much time I can put on this research project?

  how many hours/week and totally how many weeks?

- Will I work alone or collaborate with other people?

  For collaborative research project, I need to make a research schedule, which works for me and my partners. If working alone, I have to have a time schedule as well.

Deadline-Driven!!!

- Setup the deadline for your research project

1. Target on a conference where you will submit your research work.
2. For class research project, your instructor imposes a deadline for you. Start your work As Soon As Possible!
3. For your thesis, it's better to set the deadline one month before your defense

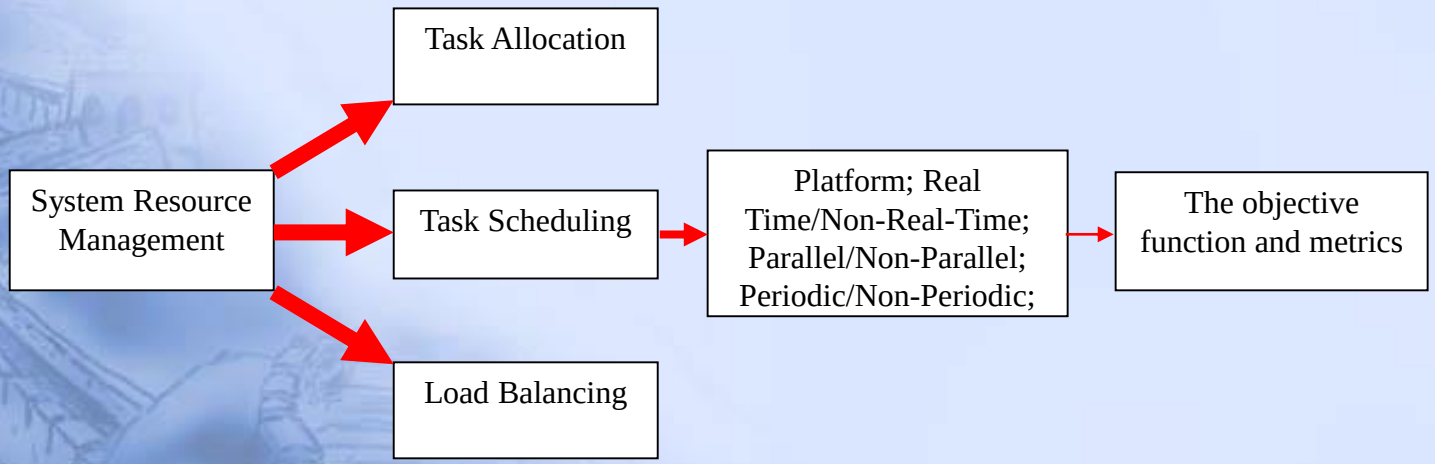
- Develop a list of milestones which demonstrate progress

1. You have a clear goal for each stage
2. Periodic results motivate you to continue until the end

- Without deadline-driven, it is very hard for you to accomplish something!!!

How to discover an interesting and achievable research issue?

- What are the research issues that I or my group have interests in?
- What I or my group can achieve before the deadline?
- Narrow down your research topic



Reading Phase 1: Extensive Reading

- Search recent years related papers using key words from top resources
- How many papers are enough? (10 ~ 20 for a class project)
- Reading these related papers can
  1. Educate you about the state-of-the-art techniques
  2. Motivate your own idea
- CS conference ranking list

  http://www-static.cc.gatech.edu/~guofei/CS_ConfRank.htm
- CS conferences & journals impact list

  http://citeseer.ist.psu.edu/impact.html

How To Conduct The Extensive Reading?

- Abstract

  What is it talking about?

- Introduction

  What the background and motivation of this work?

- Conclusion

  What had been achieved and what are future work?

Summarizing Existing Related Work

- After extensive reading, write a summary of the existing related work (one or two pages).
- This summary will be Related Work in your paper.
- Avoid a simple re-description of their abstracts

How To Summarize?

- Categorize the current related techniques from a high level view.
- Then concisely point out their respective cons and pros.
- Finally give the shining features of your research.

Related Work Summary Example (1)

- Use a few sentences to give a background message.

  Energy conservation has been an important research topic in storage systems and there have been extensive studies on developing energy saving techniques for hard disk drives. Compared with numerous single-disk energy saving schemes [11][15][20][28][64], energy conservation techniques for parallel disk arrays used by data centers are new and more complicated because they have to take load balancing, data and access parallelism, data availability, reliability, scalability, and performance into account.

- Give a summary of current related techniques

  Typical energy conservation techniques for parallel disk arrays can be categorized into four broad categories as shown in Table 1.

Related Work Summary Example (2)

Category	Representative Techniques
Power Management	DRPM [18], Multi-speed [8], Hibernator [71], CMTPM and CMDRPM [55]
RAID Configuration	RAID Tuning [17], PARAID [61]
Caching	MAID [9], PA-LRU and PB-LRU [72], PDC [37], CBSM [7]
Data Placement	PDDL [56], PF+[54]

Related Work Summary Example (3)

- Point out the differences between yours and others

  Our techniques are significantly different from the above schemes. First, we are targeting mobile disk arrays where a medium size (from several dozens to several hundreds) disk array can provide not only a large capacity but also a high disk I/O bandwidth for data-intensive applications. Second, a practical energy-centric reliability model to be built for reliability is at least equally important as energy saving in mobile computing environments. Next, we propose several novel energy conservation techniques such as Energy-aware Data Placement strategies and Buffered Disk framework, which aim at saving energy while satisfying other criteria like reliability, availability, and performance.

Reading Phase 2: Comprehensive Reading (Standing On The Shoulders Of Giants)

- Identify one or two papers from the extensive reading.
- Reread it and gain a thorough understanding.
- Ask the following questions:

1. Is this a good work representing the new trend?
2. What are the future work pointed by the authors?
3. Can I extend it to make a new research ?

a) I just go one step further based on the future direction.

b) I want to extend it by integrating a new objective, which were not considered by the authors.

c) I can apply their techniques in a new environment with some modification.

d) I can come up with a brand new idea motivated by their work. (Terrific but rarely happens in class research project)

Reading Phase 2: Comprehensive Reading (Implementing Their Algorithms)

- Make sure that you completely understand the paper.

  (You can even re-produce its results)
- Find the drawbacks of their methods.

  (In conclusion part)
- Construct your algorithm to either overcome the drawbacks or add new functions.

  (On paper)

Designing Your Strategy

There are normally two ways.

- Develop a new one, which can be a peer of the existing idea.

- Complement the existing one and make it better in your new performance metric.

Write Down Pseudo Code

Algorithm: Sort Partition

Input: m = number of disks

n = number of files

h_i = heat of file i

s_i = expected service time of file i

Output: assignment of files to disks $\{I_1, I_2, \ldots, I_m\}$

Step 1: Compute the average disk utilization ρ: $\rho = \frac{1}{m} \cdot \sum_{i=1}^{n} h_i$

Step 2: Sort all files into list I in descending order of their service times s_i

Step 3: Allocate to each disk d_j the next contiguous segment of I until its load, $load_j$, reaches the maximum allowed level ρ:

```
i = 1
for j = 1 to m do
        load_j = 0; I_j = ∅
        while (load_j ≤ ρ or i ≤ n) do
                I_j = I_j ∪ {i}          // assign file f_i to disk d_j
                load_j = load_j + h_i
                i = i + 1
        od
od
if (i ≤ n) then
        I_m = I_m ∪ {i, i + 1, ..., n}          // assign remainder to disk d_m
fi
```

Conducting Experiments (1)

- Implement your algorithm
  - Implementing a prototype
  - Using simulator: (1) well-known existing; (2) your own
  - Using any languages appropriate to construct your simulator
- Collect data
  - Synthetically generate data by using some data generators
  - Traces: http://iotta.snia.org/traces?cookies_enabled=testing
- Obtain experimental results
  - Using the data to drive your simulator

Conducting Experiments (2)

- Make an experiment plan
- Determine the objective metrics (e.g., mean response time, overall energy consumption.)
- Determine the parameters to be evaluated
- (request arrival rate, workload request skew degree, number of disks in a disk array)
- All parameter settings should be realistic and can be justified

Make Your Experimental Results Convincing

- Use traces to validate your synthetic results
- Validate your simulator
- Implement your strategy in real applications