Our paper reading up to this point has been for the purpose of gaining a general understanding of a research paper topic. But gaining a deep enough understanding of a paper to make use of it in your research often requires going beyond the reading needed to write the abstract. For instance, if you are required to present a paper to others, either orally or in writing, you will need to dig deeply, to challenge the paper’s arguments until you understand it fully. The purpose of the final project is to write a paper that synthesizes three closely related papers. In the process, and almost as a byproduct, you will gain a rich understanding of a topic of your choice.

Your task will be to select three high-quality conference or journal publications that are related works to one another. The papers can be on any distributed systems topic of your choice. You will write a paper that discusses all three publications. I strongly suggest that you select a topic about which you do not know much and are interested in learning more. Use the assignment to your advantage to strengthen your background on something of interest to you.

Choosing the Papers

The topic can be of your choosing but must have a link either to what is covered in the Tannenbaum book, or what we cover in class. Your topic must be approved.

One of the three papers must be from either one of the conferences listed below, from a distributed systems journal such as IEEE TPDS, or from the bibliography in the back of Tannenbaum’s book. Selecting a topic and finding the first paper is the difficult task. After the first paper is selected, papers two and three can be found by reading the ‘related work’ section of the first paper. The related work section appears in every well-written paper; it is where the authors identify closely related projects and discuss how their work differs.

My recommendation for where to start is to scan the bibliography of Tannenbaum. If nothing strikes you, the second place to start is to scan the past 3 years of proceedings for the relevant conferences amongst those listed. Pick papers with titles that look to be in your subject of interest. If the paper is not available on the conference site, type the exact title into Google and see if it comes up in Citeseer. If not, go to the home page of the first or second author and look for it there. USENIX papers are not available to the general public from the USENIX site. Similarly, many IEEE or ACM conferences will not have their proceedings generally available to the public. If you have identified a paper but cannot locate a digital copy after searching Citeseer, conference proceedings, and the author’s home pages please let me know and I’ll download a copy for you.
Please be aware that this offer applies to USENIX, IEEE, or ACM papers only and the expectation is that you’ve exhausted all other sources before asking me.

- International Conference on Distributed Computing Systems (ICDCS)
- High Performance Distributed Computing (HPDC)
- IEEE Symposium on Security and Privacy
- ACM/IFIP/USENIX International Middleware Conference
- International Parallel and Distributed Processing (IPDPS)
- USENIX MobiSys – mobile systems
- USENIX USITS – Internet Technologies and Systems
- SC – www.supercomp.org
- SRDS – IEEE Symposium on Reliable Distributed Systems

Challenge What you Read

Paper reading often requires making multiple passes. In the first pass, expect to gain a high level understanding of the paper. Skip parts of paragraphs that are too dense or contain terms that are not clear to you. Your goal is to pull away a general understanding but not necessarily to understand all the details. I’ve found through years and years of paper reading experience, that if a paper is unintelligible, it is either because the paper is poorly written, or because it contains terms or concepts that I do not know. Because it takes experience to be able to identify a poorly written paper (but poorly written in subtle ways); I’ve minimized the chances of your selecting of a bad paper by restricting the selection of papers to respected conferences and journals. The second pass of the paper should then be made with the goal of understanding the details, cognizant all the while of the phrases or terms that are stumbling blocks to your comprehension. If the term is important to overall understanding, turn to other sources to define or explain the concept or phrase that is confusing. It is not unusual to have to make three passes over a paper before you gain a full understanding of the paper’s contents.

As you are reading, view the paper with scientific skepticism. That is:

Examine the assumptions: do their results rely on any assumptions about trends or environments? Are the assumptions reasonable?

Examine the methods: did they measure what they claim? Can they explain what they observed? Did they are adequate controls? Were tests carried out in a standard way?

Examine the statistics: were appropriate statistical tests applied properly? Did they do proper error analysis? Are the results statistically significant?
Examine the conclusions: do the conclusions follow logically from the observations? What other explanations are there for the observed effects? What other conclusions or correlations are there in the data that they did not point out?

Writing Your Results

Your task in writing is to come up with a framework for discussing three papers that likely have very different presentations and perspectives. A trivial and extremely uninteresting way to do this is to write three abstracts of the form we’ve been doing, and glue them together with transition sentences. If you encountered such a paper in your reading, you would likely consider it to be a waste of your time to read, so we’ll reject that approach. One worthwhile approach is to develop a taxonomy. A taxonomy is a identification of the main traits of a system, algorithm, paradigm, or model. The taxonomy, once developed, can be the organizing basis for a discussion of the three systems. For instance, the taxonomy of a distributed file system might consist of the following orthogonal issues: stateful or stateless? Caching? Security? Consistency guarantee used? API? Architecture? The paper could be organized as follows: introduce and defend the taxonomy, brief introduction to three works, then a discussion organized by the items in the taxonomy. Another way to organize it would be as follows: introduce and defend the taxonomy, then a discussion organized by works. Within each work, the items of the taxonomy are discussed in turn. A synthesis paper could also take the form of an overview of each work, followed by a discussion organized by assumptions, methods, statistics, and conclusions.

Paper Regulations:

The paper will be 8 pages in length, and formatted as a standard research paper with numbered sections and subsections and reasonable margins. Use the papers you have read as a guideline. Your paper must include citations in the body of the text to any papers you consulted that helped you to write your paper. The font will be 12-point, the spacing will be single spacing, single column. If you know LaTex2e, there’s a style file on the course web page that you can use that will do formatting for you. The paper must include a bibliography of references. References must be complete. That is, they must include the title, authors, date, and venue in which the paper was published. The three selected papers must appear in the bibliography and cited in the text. Sloppy bibliographies, spelling, and formatting will be held against you.

Deadlines:

Wednesday April 7: single sheet of paper is due that lists your topic, and the first of the three papers (where ‘first’ is as described above.) The paper must be a full citation.

Monday April 26: final paper due at beginning of class.

Monday April 26, Wednesday April 28: student presentation in class. Each student will make a 5-minute presentation of their work.