

Exercise 3

Goal: Gain a deeper understanding of networking research through in-depth analysis of a recent paper and reproduction and extension of its key results. In project 3, you will select a paper to focus on, analyze it, and develop a plan for recreating and extending (some of) its results. In the final project, you will execute your plan, conducting your proposed experiments and producing a final report describing your results and explaining your findings.

You are strongly encouraged to work on this assignment in groups of exactly 2 students, but you may also work alone if you prefer. Since this project builds into the final project, you will not be able to change partners between project 3 and the final project.

You must select the paper you will focus on and let the instructor know by Tuesday, November 2, 2021, 11pm.

The final submission date for project 3 (paper analysis and experimental plan) is Wednesday, November 10, 2021, 11pm.

Selecting a Paper

You should select a paper from a high-quality conference that focuses on networking or networked systems. In order to conduct your experiments and reproduce/extend the paper's results without needing to reimplement the entire work, you should select a paper where the authors have made their code publicly available.

For conferences, I would suggest:

SIGCOMM

2021: <https://conferences.sigcomm.org/sigcomm/2021/program.html>

2020: <https://conferences.sigcomm.org/sigcomm/2020/program.html>

2019: <https://conferences.sigcomm.org/sigcomm/2019/program.html>

IMC

2020: <https://conferences.sigcomm.org/imc/2020/paper-access/>

2019: <https://conferences.sigcomm.org/imc/2019/program/>

NSDI

2021: <https://www.usenix.org/conference/nsdi21/technical-sessions>

2020: <https://www.usenix.org/conference/nsdi20/technical-sessions>

2019: <https://www.usenix.org/conference/nsdi19/technical-sessions>

Earlier years are also fine. Note that the SIGCOMM 2020 and 2021 programs include "Artifact available" badges to indicate whether the authors have made their code and/or data available -- papers with these badges can be good candidates for your projects.

If you're not sure where to start, here is a list of recent papers that are related to topics we've discussed in class, and have publicly available code and/or data (or base their results on existing, publicly available software, e.g. papers conducting experiments using BBR, which is implemented in the Linux kernel):

1. "When to use and when not to use BBR: An empirical analysis and evaluation study" (IMC 2019)
<https://dl.acm.org/doi/10.1145/3355369.3355579>
2. "Modeling BBR's Interactions with Loss-Based Congestion Control" (IMC 2019)
<https://dl.acm.org/doi/10.1145/3355369.3355604>
3. "Profiling BGP Serial Hijackers: Capturing Persistent Misbehavior in the Global Routing Table" (IMC 2019)

<https://dl.acm.org/doi/10.1145/3355369.3355581>

4. "RPKI is Coming of Age: A Longitudinal Study of RPKI Deployment and Invalid Route Origins" (IMC 2019)

<https://dl.acm.org/doi/10.1145/3355369.3355596>

5. "Practical Intrusion-Tolerant Networks" (ICDCS 2016)

<https://ieeexplore.ieee.org/document/7536504>

6. "Timely, Reliable, and Cost-Effective Internet Transport Service Using Dissemination Graphs" (ICDCS 2017)

<https://ieeexplore.ieee.org/document/7979950>

These are fairly "safe" options, since either I was involved with them (5,6) and can help you out with running the code if you run into problems, or I've taken a look at them and they either have good infrastructure for reproducing their results (3,4) or they mostly use standard Linux tools (1,2).

You are welcome to choose your own paper though (it does not need to be one of the 6 above)! Even if you do choose one of those 6, I'd encourage you to spend some time looking at the conference programs above to get a sense of current research topics. If you have a paper that you're interested in but are not sure if it can work for the project, just ask.

Note: If you are interested in choosing one of the BBR papers, you may find this GENI exercise helpful in planning your experiments:

<https://www.cs.unc.edu/Research/geni/geniEdu/09-queues.html>

Paper review

You should start by reading your chosen paper and writing a review (similar to the ones you have submitted throughout the semester for our paper discussions).

Your review should include:

1. Summary
2. Strengths
3. Weaknesses
4. Justification

You do not need to include discussion points / questions (but please contact us if you have questions about your paper that you'd like to discuss!).

Analyzing claims and experiments

Next, you will analyze the main claims of the paper, and the experimental evidence or analysis that is used to support those claims.

You should write a short report that:

1. Lists the main claims or conclusions of the paper
 - a. For a system-building type paper (e.g. 5,6), think about what the authors claim their system is good at. What properties or performance do they claim it supports?
 - b. For a measurement-oriented paper (e.g. 1-4), what are the key findings or conclusions that the authors draw from their data collection / experiments and analysis? What do they claim to have learned?
2. For each of the main claims you identified, explains what evidence the authors use to support that claim. You should answer (at a minimum) the following:

- a. What testbed or data collection environment did they use?
- b. What experiments or analysis did they perform?
- c. What were the results?
- d. Do those results effectively support the claim?

Assessing existing code and data

To prepare to create your experimental plan for reproducing and extending key findings, you should familiarize yourself with the existing code/data/tools as relevant to the paper. This will help you understand what parts of the paper's experiments are likely to be feasible to reproduce, and how you might extend them. (Note: you should do an initial pass on this step when selecting your paper, to get a sense of what working with the particular codebase or tools associated with the paper will involve)

You should download any relevant code/data, and test that you can run (part of) it. You don't need to run full well-thought-out experiments at this stage, but should do enough to be confident that you'll be able to get something to work, e.g. to know you're not going to get stuck on just getting the code to compile (that can definitely happen when trying to use other people's research code!).

You should write a short description that includes:

1. A summary of what existing code/data/tools you'll be using in your project
 - a. Please include URLs for where the code/data is located if relevant
2. What you've done to test that you're likely to be able to work with these tools effectively

Experimental plan

You should write a detailed plan for the experiments you plan to perform to try to reproduce and extend the results of the paper.

There is not a set number of required experiments, but you must attempt to **reproduce** at least one result (i.e. a particular table or figure) from the paper as closely as possible, and you must attempt to **extend** the results of the paper in some way (i.e. try at least one experiment or analysis that the authors did not do). Additional experiments may focus on reproducing or extending results (or a combination of the two). Note that for some recent measurement-based papers, the authors do an excellent job making their results reproducible, providing their data and the scripts used to produce their figures. You may use these for the reproduction part of your work, but you *must* also extend the work in some way (i.e. it is not enough to *only* run the authors' own code on their own data.).

For the reproduction part, it may not be possible to replicate a particular result exactly (e.g. you may not have access to exactly the same testing environment as the authors, or it may be that some parts of the code or data are not publicly available). This is perfectly fine -- in this case you should explain which parts you expect to be possible to replicate, and which are not (and why). Note, for example, that for the "Timely, Reliable, and Cost-Effective Internet Transport Service Using Dissemination Graphs" paper (6), the data and simulator used are not currently publicly available, but the system implementation is (although if you really want access to the simulator let me know).

For the extension, there can be many possibilities. For example, you might try to run the same experiments as the authors, but vary a parameter that they didn't consider in the paper. Or for a measurement-based paper you might try collecting new data and running the same analysis as the authors to see if there has been change over time. Or, you might use the original data (if provided) but change the analysis slightly to try to answer a new question.

Your written plan should include:

1. The test environment you plan to use

2. The set of experiments you plan to run
3. The justification for each experiment - what do you plan to learn? What results from the paper would the experiment let you confirm/refute/extend?

Deliverables

You should select the paper you will focus on and let the instructor know by **Tuesday, November 2, 2021, 11pm**.

This should be done by email to babay@pitt.edu

Note: this does NOT mean you should wait to start working on the main part of the project until after November 2.

As part of deciding what paper to choose, you should take a look at the available code and data and think about what will be feasible for you to experiment with.

Your final submission is due **Wednesday, November 10, 2021, 11pm**. The submission should be done in Canvas as a single PDF document that includes all of the required components:

1. Paper review
2. Analyzing claims and experiments
3. Assessing existing code and data
4. Experimental plan

Each of these 4 parts should be a clearly labeled section in your report (with subsections within each section as needed to organize your ideas).

Note: For your final project, you will need to produce a report using the IEEE two-column format. You are encouraged to use this for your proposal (project 3) as well. You can find MS Word and Latex templates (including an Overleaf template) here:

<https://www.ieee.org/conferences/publishing/templates.html>