L50 - Lab 4, System Characterisation

Prof Andrew W. Moore

Michaelmas, 2021/22

This lab, as well as the following one, will explore the characterisation of an artifact. Students receive different artifacts, and each student defines and executes a characterisation plan for the artifact.
The goal of these labs is to demonstrate your knowledge in system and network performance measurements, building upon the lectures and your previous lab experiments.

1 The artifact

This year the artifacts are pseudo-cloud based (docker containers on our local research cloud). Each student is assigned one "cluster" of VMs, and needs to characterise them. Students' clusters are not identical.
Your goal is to characterise the clusters based on the practices learned in the course. There is no single definitive set of tests that need to be conducted, but your tests should cover most of the course's topics. We recommend to start with a basic set of experiments, and then to extend on a single topic. further guidance is provided below.

2 Characterisation plan

Before starting your experiments, you should prepare a characterisation plan. The characterisation plan should include all the aspects that you intend to cover as part of your tests, and which experiments you are going to conduct as part of the characterisation.

This plan is not for submission, and will not be graded, however instructors will be available to review the plans with you to make sure you are on the right track. One extra lab availability slot will be scheduled for Next Monday (15th of November, 2022)
For ease of discussion, it is best if your plan is formatted as a list or a table.

3 Measurement tools

You can use any measurement tool discussed in class, as well as any other open source tool available. This may include also tools not discussed in class.

The evaluation of your work will focus on your ability to demonstrate an understanding of obtained results, rather than a multitude of those, and in particular the understanding on tools' limitations, and the analysis of unexpected results. You should pick your measurement tools accordingly.

4 Reproducibility

In your characterisation, you should employ proper reproducibility methodologies, and in particular the use of scripts, repositories etc. You can use Jupyter notebooks as a means to script your experiments, and you can reuse code from previous experiments. You can also reuse code and test environments openly released by the measurements community. You need to provide instructions for reproducing your experiments, and your peers will use those instructions to repeat your tests.

5 Use of the cloud environment

Our container network is hosted in the department; by now you should have received all the login information required. Please contact the course's team if you haven't.

5.1 Login

Each student has access to a cluster of five machines (h1,h2,...,h5) the email will give you the IP addresses for each – these machines have the same access limitations as the nf-test machines you have used earlier in this module (access via CL-internal, or vpn2 or a computer lab based machine). A unique ssh key is assigned to each cluster of five machines; the IP addresses, ssh key, and it's pass phrase are contained in the email. The username for login is root.

Each host may be rebooted and is a non-persistent virtual machine. Thus, you need a place for results and each of the hosts has a unique directory `\persistent` Furthermore, you need to pursue your own backup procedures, to copy a remote directory onto your local machine:
`sftp root@128.232.97.XXX` and `get -r <directory>`.
There are also other ways to copy a remote directory, you are welcome to use those as well.

5.2 Practical tools available

You should be able to install on you machine any tool that you wish to use. We restrict this to the use of only free or open source tools.
All software based tools used through the labs should be available, either pre-installed or from a suitable repository, for your tests, such as (and not limited to) ping, traceroute, iperf, tcpdump, tcpreplay etc. You can also use variants of the tools used in class - e.g. udp or tcp based "ping" (a.k.a. tcpping).
Before installing any tools, you will need to run "sudo apt-get update".

Some of the tools studied in class will require using sudo. For example, ping flood and ping using an interval of less than a millisecond will return an error unless sudo is used.

6 Lab report

A lab report summarising your work is required.

Submission Deadline: 03/12/2021, 12:00
Submission: submit through Moodle. Four files are required:

1. The lab report, as a single pdf file.
2. The reproduction environment, as a single compressed file (tar, zip).
3. A dump of the measurements' results, as a single compressed file (tar, zip)
4. The (paper) artifact review form.

The reproduction environment and the results files should include a README file, explaining the organisation of the folder, file name conventions and the meaning of different files.
There is a size limitation for file submission. Please contact the course's team if your measurements result is bigger than that. The reproduction environment and the report must be (significantly) smaller than the file size limitation.

6.1 Structure

The report must be **No more than 5000 words**. Longer reports will not be accepted nor graded. Figures, graphs and citations, referenced within the text, are not counted toward the word count. Please use a font size of at least 10pt.

If you encounter exceptional results, which can lead to a report longer than the word limit, you can include those in an appendix. The appendix will not be graded, but the course's team is happy to discuss and follow up on such results.
We request that you indicate in your report if you are happy that we share any interesting measurement results with the cloud provider.
While not mandatory, we suggest the following format for the report:

- A description of the artifact(s) and all relevant metadata.
- Platform information and environment.
- Latency experiments.
- Throughput experiments.
- Advanced and focused experiments.

- Instructions for reproducing the experiments.

Each experiment should detail, shortly, the following:

- The goal of the experiment.
- Setup and tools used.
- The methodology used.
- Results.
- Analysis and discussion of the results.

There is no requirement on the formatting of the pages or the sections.
Latency and throughput experiments are intended to provide basic information about your artifact. You can also experiment with topology experiment, OS profiling and other types of basic experiments, but those are not mandatory due to limitations of the environment.
As always, you should look for odd or surprising results, and try to explain them. Note that sometimes exceptional results indicate a problem in your setup or scripts.

6.2 Advanced and Focused Experiments

This section include some ideas for advances and focused experiments. These ideas are not comprehensive nor mandatory. They are only intended to provide scope and guidance.

- The repeatability of an experiment within a cloud environment.
- Variability in network performance over time.
- The effect of indirect cross-traffic on network performance.
- The effect of reallocation on network performance.
- Packet loss under normal and congested operating modes, and inferring underlying network properties.
- Comparing different measurement methodologies of one network property (e.g., latency, throughput) or proposing new methodologies.

  Note that any anomalous results must be discussed and (where possible) explained.

7 Reproducibility Report

Each student is required to provide a reproducibility report for a research paper. The template for the report as well as the papers are provided on Moodle. The paper number corresponds with the artifact number.
There is no need to reproduce an experiment from the paper, just to evaluate the artifacts up to the functionality point. Properly reproducing (or replicating) one experiment from the paper will earn you an extra mark.