L50 - Lab 4, System Characterisation

Dr Noa Zilberman, Dr Andrew Moore

Michaelmas, 2017/18

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. In the next (and last) lab, each student reproduces an experiment by another student.
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 artifact is the cloud, or rather a set of three virtual machines (VMs) running in the cloud. Each student is assigned a different set of VMs, and the VMs are different in their characteristics.

Your goal is to characterise the 3-nodes system 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 that you focus on a specific element, e.g., latency, but we request that you at least conduct rudimentary measurements of each type (e.g., bandwidth, topology, etc.). Once you have conducted this basic set of measurements you can move to a full-on characterisation of a selected aspect.

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 during the lab the instructors will the review the plans with you to make sure you are on the right track. 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

As in the following lab you will be requested to reproduce each other's experiments, you should employ proper reproducibilty methodologies, and in particular the use of scripts, repositories etc. You can use Jupyter notebooks as a mean 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 should prepare instructions for reproducing your experiments, and your peers will use those instructions to repeat your tests. We will provide during the lab a form to be completed with feedback on the reproduction.

5 Use of the cloud environment

Two cloud environments are being used: Google Cloud (Setup #1 - Setup #4) and Microsoft Azure (Setup #5 - Setup #7).

Your cloud machines were assigned a limited budget. Stop/Shut down your machines when not in use. Note that VMs may be relocated within the data centre.

You can only use the allocated setup resources. You are not allowed to add or delete resources from your setup or apply any services. Except for starting/Stopping machines you should operate completely within your VM.

5.1 Google Cloud

Login to Google Cloud using your google user name. If you have been assigned to Setup #1-Setup #4 then you should have been granted access to your setup.

1. To login, go to `https://console.cloud.google.com`
2. At the top of the page (next to "Google Cloud Platform") make sure that L50-Setup-< *setup number* > is selected.

3. Go to Compute Engine Dashboard (either through the menu on the left hand side, or through the dashboard - Compute Engine - Go to the Compute Engine Dashboard.

4. Go to VM Instances (by default you should see the VM instances).

5. Select all your VM instances and start them by pressing "Start" at the menu on the top.

6. Connect to your VM using SSH (under the "connect" column).

5.2 Microsoft Azure

Login to Microsoft Azure using Cambridge CRSid. If you have been assigned to Setup #5-Setup #7 then you should have been granted access to your setup.

1. To login, go to `https://portal.azure.com`

2. Press "Virtual machines" on the left side menu (9th from the top).

3. Make sure that the selected subscription is "L50-2017-Setup< *setup number* >"

4. Select all your VM instances and start them by pressing "Start" at the menu on the top.

5. Pressing on a specific VM will display information about this VM, including its public IP address.

6. Connect to a VM using ssh - either from your own terminal or using the cloud shell (Marked "> _" on the top toolbar). The username and password will be provided in class.

5.3 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. There are differences between the Google Cloud and Azure environments in terms of the privileges and testing abilities - you should operate within these limitations.
All software based tools used through the labs should be available 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).
We have added a page on the wiki with pointers to various data centre measurement work [1]. This should provide you with ideas on what can be measured and how. Reproducing some of these experiments is certainly encouraged. Some of these papers also released the code they used for testing, and you can build upon such packages.
One useful package for PTP-based latency measurements is described in [2] and is available at [3].
For software based traffic generation ideas, you can refer to [4].

6 Lab report

A lab report summarising your work is required.

Submission Deadline: 29/11/2017, 12:00
Submission: submit through Moodle. Three 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)

The reproduction environment and the results files should include a README file, explaining the organization 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 and relevant metadata.
- Platform information and environment.
- Topology experiments.
- 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.
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.

References

[1] "Data centre measurements - related work," https://github.com/cucl-srg/L50/wiki/Data-Centre-Measurements.

[2] D. A. Popescu and A. W. Moore, "PTPmesh: Data center network latency measurements using PTP," in *IEEE MASCOTS*, 2017.

[3] ——, "PTPmesh: Data center network latency measurements using PTP - reproduction environment," 2017, http://www.cl.cam.ac.uk/research/srg/netos/projects/latency/mascots2017/.

[4] P. Emmerich, S. Gallenmüller, G. Antichi, A. W. Moore, and G. Carle, "Mind the gap: A comparison of software packet generators," in *Proceedings of the Symposium on Architectures for Networking and Communications Systems*. IEEE Press, 2017, pp. 191–203.