L50 - Lab 4, System Characterisation

Dr Noa Zilberman, Prof Andrew Moore

Michaelmas, 2019/20

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 cloud based (Azure). Each student is assigned two "clusters" 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 during the labs the instructors will the review the plans with you to make sure you are on the right track. You should bring the plan to Lab 5, but it can also be discussed during Lab 4. 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 reproducibilty 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

Microsoft Azure is used as a cloud environment. By now you should have received an invitation to Lab 4 on Azure. Please contact the course's team if you haven't.
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.
Login to Microsoft Azure using Cambridge CRSid:

1. To login, go to `https://portal.azure.com` and use your CRSid to login.
2. Go to "Resource Groups" on the top menu (or under "all services" on the top-left menu button).
3. Make sure that the selected subscription is "Lab4 - System Characterization ARTIFACT $<$ *number* $>$"
4. Make sure that you see "Cluster1" and "Cluster2", these are your artifacts. Ignore "NetworkWatcherRG".
5. Click on a cluster resource group to see all the resources within in.
6. From the main menu, select "Virtual Machines".
7. Select all your VM instances and start them by pressing "Start" at the menu on the top.
8. Selecting a specific VM will display information about this VM.

9. The VMs don't have a public IP address. You can use either a serial console or Bastion. The username and password will be provided in class).

5.1 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 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: 06/12/2019, 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 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(s) and all relevant metadata.
- Platform information and environment.
- Latency experiments.
- Throughput experiments.
- Comparison of Cluster1 and Cluster2.
- 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.
The comparison of Cluster1 and Cluster2 can be embedded within the earlier sections, or it may be a stand alone section.
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.