Management and Content Delivery for Smart Networks: Algorithms and Modeling

Academic Year 2020-2021

L1: Internet of Things (IoT) system simulation

Summary

The purpose of this lab is to simulate the operation of a portion of an Internet of Things (IoT) network that is modeled as a queuing system and investigate its performance under variable configurations, to understand how different parameter settings may affect the system behaviour and how modifying some parameters of the network system may impact on performance metrics.

The scenario is depicted in Fig. 1. Consider an IoT system, consisting of sensors and actua-

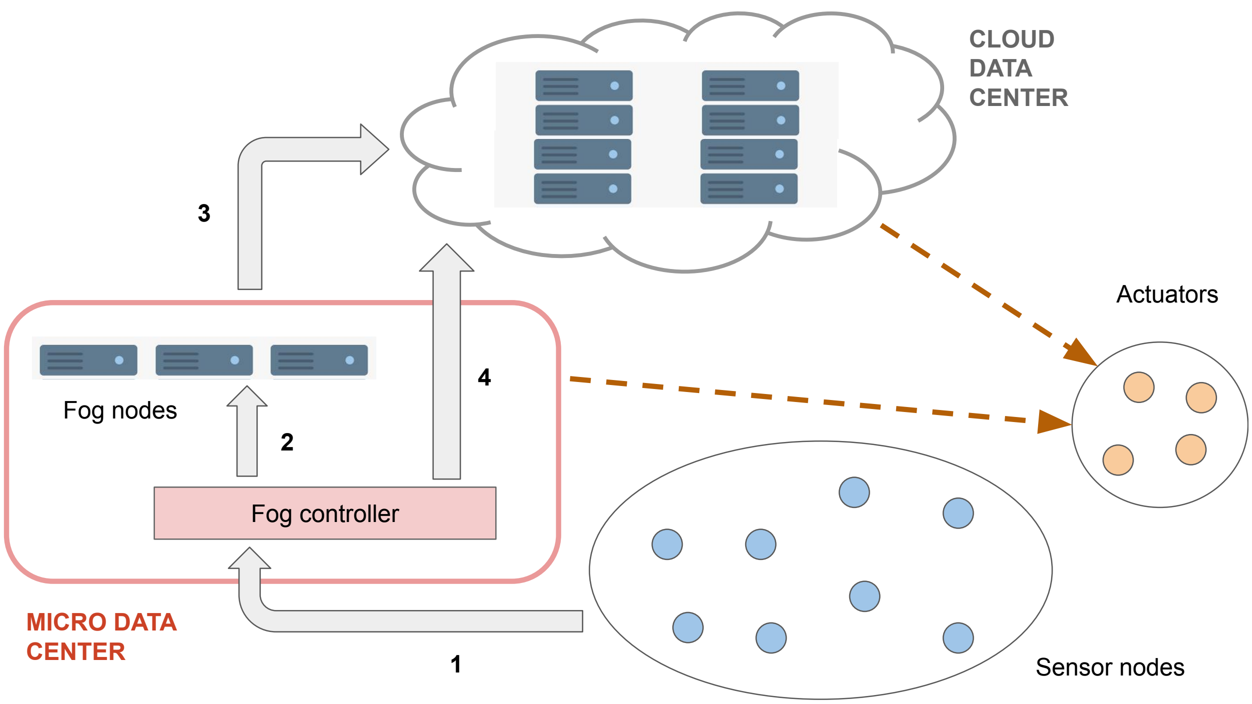

Figure 1: IoT scenario.

tors. A number of sensor devices collect real time monitoring data at the edge of the network. Collected data are sent to a local Micro Data Center (arrow 1), that provides computational capacity close to edge of IoT network by means of fog nodes, which are managed by a fog controller. Fog nodes pre-process data (arrow 2), so that most requests can be fulfilled locally, whereas only computationally intensive requests are forwarded to a Cloud data center (arrow 3), thus saving bandwidth and energy. Finally, based on the data processing output,

operative commands are generated and sent to the actuators (dashed arrows). In case all the fog nodes are busy, incoming data can be buffered, if a buffer is present. However, if the buffer is full or it is not envisioned at all, data packets are directly forwarded to the Cloud Data Center for performing the entire processing tasks (arrow 4).
We focus on the Micro Data Center. In the considered queuing system, the customers represent the IoT data packets that arrive at that Micro Data Center. The service represents the local processing of the data by fog nodes before generating the actuation message packets or before forwarding data to the Cloud for further computational processing. Finally, the waiting line represents the buffer where packets are stored before performing the computational tasks.

Tasks

In order to investigate the performance of the Micro Data Center, you should perform the following tasks:

1. Assuming that the Micro Data Center consists of a single fog node, derive the metrics that describe the system performance in the baseline scenario. Consider both the following cases:

   (a) the incoming data packets are either immediately processed or, if the fog node is already busy, instead of being dropped they are forwarded to the Cloud Data Center for processing. They are hence added to the fraction of data that, after completing local pre-processing, still need to be forwarded to the Cloud Data Center, since they require more complex computational tasks;

   Baseline (b) the incoming data can be buffered whenever they cannot be immediately processed.

2. Modify the baseline scenario varying the size of the buffer and considering a multi-server system, assuming that two fog nodes are present in the system. Test the system behaviour under different buffer sizes (infinite waiting line, medium to short finite waiting line, no buffer at all) and compare its performance with the single server scenario.

You can start from the network simulator code available in the course material, that represents the operation of an M/M/1 queuing system with infinite waiting line, either choosing an event scheduling approach or a process interaction based approach. The simulator is provided in two versions:

- *queueMM1-ES.py*: this version is based on an event scheduling approach;
- *queueMM1-PI.py*: this version is based on process interaction.

You can make your modifications and integration to the code, in order to investigate the system performance via simulation. Different metrics can be useful to understand the system operation and evaluate its performance, like:

- **number of locally pre-processed packets**, i.e. the packets that are pre-processed in the local Micro Data Center
- **number of pre-processing forwarded packets**, i.e. the number of packets that are immediately forwarded to the Cloud Data Center before undergonig any pre-processing, since the fog node(s) is/are busy. This number does not include packets that are forwarded to the Cloud Data Center after the local pre-processing, since they require more intensive computational tasks
- **average number of packets** in the system
- **average queuing delay**, i.e. the average time spent in the system
- **distribution of queuing delay**
- **average waiting delay** per packet, assuming that the packet waiting delay is the time elapsed from the moment at which a packet enters the system to the instant at which the service begins.
  It is useful to compute both of the following types of waiting delay:
  - the average waiting delay experienced by any packet (**averaging over all packets**);
  - the average waiting delay **considering only those packets that actually experience some delay**, since they enter the system while the server(s) is (are) busy and are hence buffered.
- **average buffer occupancy**, i.e. the average number of packets in the buffer
- **pre-processing forward probability**, i.e. the fraction of packets that cannot be pre-processed since the fog node(s) is (are) currently busy and no buffer is present in the system or, in case of finite buffer, it is full, and are hence immediately forwarded to the Cloud Data Center
- **busy time**, i.e. the cumulative time that during the simulation each server spends in a busy state serving packets requests. In a single-server system, you can observe how it varies with the buffer size. In a multi-server system, based on the busy time analysis per each server, you can examine the distribution of the load among servers.

You can think of other metrics that may be helpful in your analysis to highlight specific aspects or critical issues in the system operation. When running your simulations, pay attention on setting appropriate values of the simulation duration.
In your analysis of the network operation and performance, focus on the following aspects:

1. assuming that the local Micro Data Center consists of a single fog node and no buffer is envisioned, investigate the system performance under different service rates, **keeping a fixed value for the average arrival rate**. Highlight your findings by showing the variation of relevant performance metrics;

2. compare the simulated performance metrics with the expected theoretical values, when available;

3. consider a Micro Data Center consisting of two fog nodes and assume a finite buffer:

   - compare its performance against the case with infinite buffer size;
   - compare its performance with the single server case;

4. keep focusing on the Micro Data Center system, considering a scenario with multiple fog nodes:

   - test different algorithms to assign each new processing request to an available fog node (e.g. random assignment; round robin; assignment to the least costly node, assuming for example that each node is assigned a different score to represent its operational cost...) and compare the load distribution in the various cases, analysing the busy time per node and the overall operational cost;
   - assume that each fog node features a different service rate: how does this affect the system performance? Investigate the impact in terms of load distribution, cost and delay, assuming that the faster the fog node the higher its operational costs. Optionally, you can design and test a load distribution algorithm to better balance cost and delay.

5. try to vary the distribution of the service time, i.e. considering the case of M/G/1, and observe how the system performance changes, assuming one or more different distribution types for the service time instead of exponential distribution;

.
Discuss in your report the main findings emerged from the various investigated scenarios. Support your claims and observations by plotting graphs that report the most significant performance metrics from different cases and help to highlight the relevant findings under variable configuration settings. Remember to always specify the unit of measure for the parameters and metrics represented in the graphs. For each graph, check that the main configuration settings that have been adopted to obtain the corresponding simulation results are reported.

Group and Final Reporting

You are expected to work in groups of up to three students. Each group is required to prepare a **single** report describing results of **all labs in the course**. This report must not exceed 10 pages. You need to delivery both the written report and your source code before the end of exam session in September.