Management and Content Delivery for Smart Networks: Algorithms and Modeling

Academic Year 2020-2021

L2: Internet of Things (IoT) system simulation: expanding the baseline scenario - Versions A and B

****
Please note that, for this Lab activity, tasks come in two slightly different versions, A and B. Groups that have been assigned Version A must follow the corresponding requirements, whose peculiarities are highlighted in blue fonts, whereas Groups that have been assigned Version B must follow the requirement specifications that are highlighted in red fonts.
****

Summary

Refer to the same Internet of Things system considered in Lab. 1, that is reported below for your convenience in Fig. 1. In this Lab you are required to expand your scenario analysis including the Cloud Data Center operation in your system.
Two types of data can be generated by the sensor nodes and arrive at the Micro Data Center:

Packets of type A: these packets require simple processing tasks, that can be locally performed by the fog nodes;

Packets of type B: these packets require more complex computational tasks, hence, after a local pre-processing, they must necessarily be forwarded to the Cloud Data Center to complete the processing.
Denote f the fraction of packets of **type B**, i.e. the fraction of all the data packets arriving at the Micro Data Center from the sensors that must be further forwarded to the Cloud Data Center after local pre-processing by the fog node(s), due to more complex required computational tasks. In this case, assume a reasonably increased average service time on the Cloud server(s). Note that, although the required tasks are more complex, the Cloud Servers are likely to be better performing than the fog nodes.
Furthermore, when data cannot be locally pre-processed due to busy fog nodes and full buffer, any incoming data (either of type A or B) are forwarded to the Cloud without any local pre-processing. Since these forwarded data are fully processed in the Cloud, they experience an average service time that reflect the fact that both the simple pre-processing tasks (for type A and type B data) and the more complex computational tasks (only for type B data)

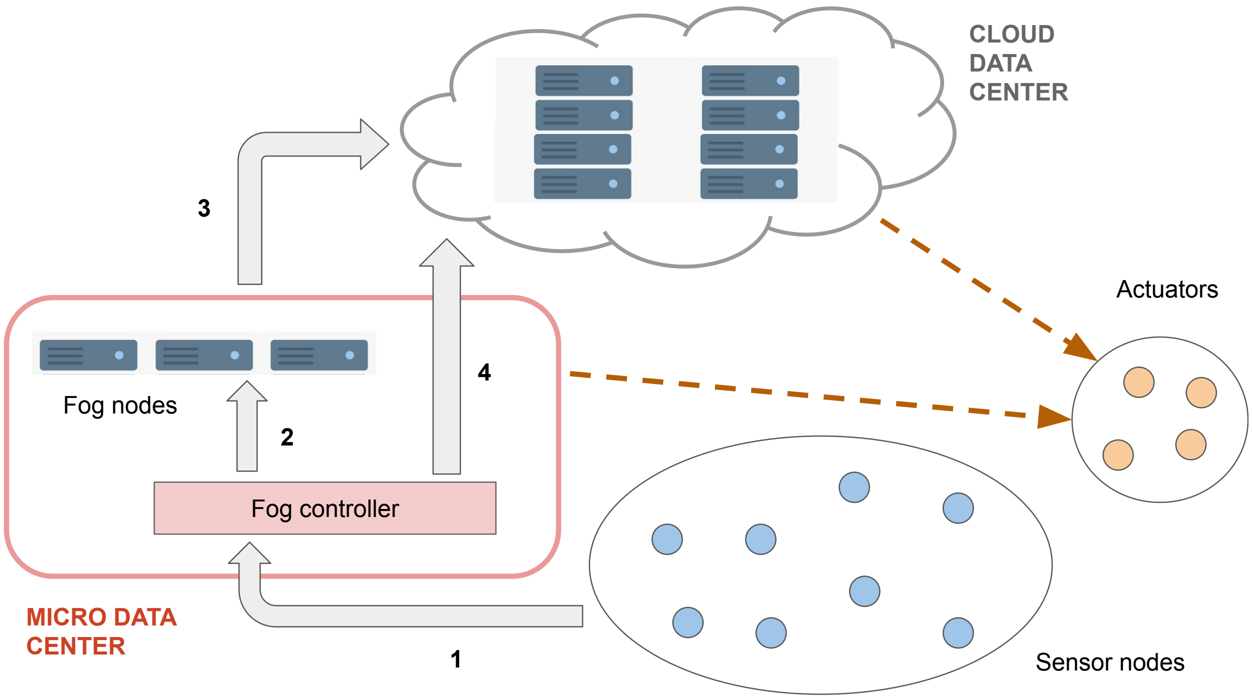

Figure 1: IoT scenario.

are performed in the Cloud Data center itself.
You can include the propagation delay in your analysis, assuming a reasonable additional delay due to the data transmission from the Micro Data Center to the Cloud Data Center and for the transmission of the commands from Cloud Data Center to the actuators. The other propagation delays can be neglected.

Tasks

1. For the first task, assume a single server with finite buffer for both the Micro Data Center and the Cloud Data Center. Focusing on the Cloud Data Center sub-system, analyze the average waiting delay (version A) or the packet drop probability (version B) of those data packets that are forwarded to the Cloud for any reason:

   (a) observe the system behavior during the warm-up transient period and identify the transition to the steady state. Try to apply a method to remove the warm-up transient in your simulations;

   (b) observe how the confidence interval of the estimated value of the considered performance metric is affected by the input settings. How can the confidence be improved? For the considered performance metric, properly configure the system settings to obtain a desired confidence interval corresponding to a target confidence level.

2. Consider now the overall system operation, including both the Micro Data Center and

the Cloud Data Center.

(a) How does the size of the buffer in the micro Data Center impact on the overall system performance?

(b) Does the size of the buffer in the Cloud Data Center show a similar impact on the system performance?

(c) How do the characteristics of input data (i.e. different values of f) affect the packet drop probability (version A) or the overall average queuing delay (version B)?

3. Now define a desired value for the maximum average queuing time, denoted T_q.

   (a) Assuming again a single-server Micro Data Center, replace the fog node with a progressively faster server. Which is the minimum value of the average service time that allows to reduce the queuing time below the threshold T_q?

   (b) Try to increase the number of fog nodes, assuming the same fixed average service time for each fog node: which is the minimum number of servers required to reduce the queuing time below the threshold T_q?

4. **Version A:** Consider a multi-server system, assuming various servers both in the Micro and in the Cloud Data Centers. Simulate the system operation over a period of 24 hours. **In order to avoid issues related to the long time required to run your simulations, you can properly reduce the simulation duration, considering for example few hours instead of the entire day.** Assign an operational cost to each fog node and to the Cloud servers, with the operational cost for a Cloud server being different with respect to the one for a fog node.

   (a) Vary the packet arrival rate or the value of f over time, depending on the different periods of the day. Investigate how these configuration settings affect the overall system performance.

   (b) Now assume at least 4 Cloud servers. Furthermore, consider at least three different types of Cloud servers, each featuring a different service speed and a specific operational cost (that may depend, for example, on the service speed). Test various combinations of server types and investigate the trade off between the overall cost, the queuing delays and the packet drop probability.

   (c) Identify the best combination allowing to either reduce the cost or queuing delay below a desired threshold. Now, assume to install half the number of Cloud servers. In this case, which is the proper configuration of server types allowing to reduce the considered performance indicator below the same desired threshold? Compare the obtained queueing delay and cost under these two scenarios, also highlighting how packets of type A and packets of type B are differently affected in terms of delay and packet drop probability.

   Version B: Consider a multi-server system, assuming various servers both in the Micro and in the Cloud Data Centers. Simulate the system operation over a period of 24 hours, varying the average packet arrival rate over time. **In order to avoid issues related to the long time required to run your simulations, you can properly reduce the simulation duration, considering for example few hours instead of the entire day.** Assign an operational cost to each fog node and to the Cloud servers, with the operational cost for a Cloud server being different with respect to the one for a fog node. In addition, define two different priority classes for packets of type B, featuring different delay constraints. For example, for each class set a different value of the maximum threshold for the queuing delay. The fraction of packets of type B belonging to the high priority class is denoted f_1, while f_2 represents the fraction of packets of type B having low priority, with $f = f_1 + f_2$.

(a) Design a strategy to assign all the incoming packets to the different Cloud servers aiming at satisfying the delay constraints of the different priority classes of packets of type B. How are cost and delays affected?

(b) How could the designed strategy be improved in order to further reduce cost, still respecting the delay constraints? How would the packet drop probability be affected?

(c) Try to vary the proportion between the values of f_1 and f_2. How does the variation of these configuration parameters affect the system performance under the operation of the designed strategy?

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 15 pages. You need to delivery both the written report and your source code before the end of exam session in September.