Assignment for the PhD position:
READY: Rethinking Monitoring for Large Distributed Systems

Romaric Duvignau[1]

[1]*Assistant Professor, Computer and Network Systems,* ***Computer Science and Engineering Department,***
`duvignau@chalmers.se`

September 11, 2023

1 Part 1: Small coding assignment

1. Download the dataset `cpu.csv` for the assignment: click here[1] (Password: chalmers).

   The dataset contains CPU-usage values (in %) for 720 nodes during 2022 monitoring rounds (1 round = 1 second, so the data of this experiment lasts for about 35min). In the csv file, Column i indicates the 2022 values observed by node i.

2. If every node needs 1 message to send its current CPU-value to a certain coordinator C, how many messages would be needed in total?

3. Assume now that nodes only share the current CPU-value with C if it is different from the last value that was read by the node. How many messages would be needed in this case?

4. Assume now that nodes only share the current CPU-value v_t with C if it is at least ε-far from the last value $v_{t'}$ that was shared with C, i.e. if
$$\frac{|v_t - v_{t'}|}{|v_{t'}|} > \varepsilon.$$
   How many messages would be needed in this case if (a) $\varepsilon = 0.05$, (b) $\varepsilon = 0.10$ and (c) $\varepsilon = 0.25$?

5. The problem with the strategy used in the previous question is that C might remember a stale value instead of the current value and hence the monitoring is not optimal but *approximate* (that is, there is an *error* $|v_t - v_{t'}|$ between the value $v_{t'}$ remembered at C and the current value v_t read at time t). For each of the three cases (i.e., $\varepsilon = 0.05$, $\varepsilon = 0.10$ and $\varepsilon = 0.25$), calculate the Mean Absolute Error (MAE) over all nodes and monitoring rounds?

6. Plot the number of messages sent and the MAE at C for the full experiment (with time as X-axis) and for each of the three values considered for ε (for better readability, use a moving average over the last 30 seconds).

7. Gather your source code[2] and share it along a brief report with your answers. Alternatively, submit a jupyter notebook containing both your code, the answers and the generated figure(s).

Disclaimer: *The assignment is only meant to check that the candidate has all necessary basic skills to start working on the PhD project. If you find the assignment to be too easy, it is rather a good sign!*

[1] `https://tinyurl.com/4yka2ba9` (Password: chalmers).

[2] Written preferably in Python.

2 Part 2: Small writing assignment

1. Read the following short paper:

    - Duvignau, R., Papatriantafilou, M., Peratinos, K., Nordström, E., & Nyman, P. (2019, June). Continuous distributed monitoring in the evolved packet core. In Proceedings of the 13th ACM International Conference on Distributed and Event-based Systems (pp. 187-192). `https://doi.org/10.1145/3328905.3329761`

2. Write 2-3 paragraphs (400-500 words maximum) explaining:

    - Why decreasing monitoring communication costs is essential in high-performance packet processing infrastructures.
    - Some initial ideas on how you would extend the results of the paper.

3 Note

Be ready to present, explain and discuss your answers to this assignment during the interview taking place at the second step of the interview process. You are thus encouraged to prepare a short presentation (of 10min maximum) including your answers to the assignment (e.g. with some of your plots and research ideas), that can be followed by a free discussion of the topic.

Also, please share your answers to the assignment (including any code) by the submission given date (October 6, 2023) as per the provided instructions.