CONCORDIA UNIVERSITY
Dept. of Computer Science and Software Engineering

COMP 691 – Online Algorithms and Competitive Analysis
Winter 2024

$\mathcal{P}$roject

1 Phase 3: experiments and report

Deadline: submission on Moodle by 23:55PM on April 19, 2024

What to submit: a single .zip file with source code + readme.txt + report.pdf

Weight: 4% of the overall course grade

Description of the phase: in this phase, you will use the codebase you have developed in phases 1 and 2 to evaluate performance of the algorithms.

More specifically, you should perform several experiments, collect the data, generate plots, and write a short report discussing what you observe in the experiments.

Each experiment will consist of several batches of trials. Given values of $(k, N, n, \epsilon, \tau, w)$, a single trial consists of the following steps:

1. generate random sequence seq using k, N, n, ϵ,
2. generate the sequence of h-values,
3. generate the sequence of $\hat{h}$-values by adding noise to the h-values based on parameters τ and w,
4. compute the number of page faults of OPT on seq by running $BlindOracle$ on seq with true h-values,
5. compute the number of page faults of $BlindOracle$ when it is executed on seq with $\hat{h}$ values instead of true values of h,
6. compute the number of page faults of LRU on seq,
7. compute the number of page faults of $Combined$.

Given values of $(k, N, n, \epsilon, \tau, w)$, the above trial should be executed 100 times to obtain: average number of page faults of OPT, $BlindOracle$ with $\hat{h}$ values, LRU, and $Combined$. These 100 trials are referred to as a batch of trials.

In what follows, try to use $n = 10,000$ (if it is taking too long, you can either optimize your code or use a smaller value of n such as $n = 5,000$ or maybe even $n = 1,000$, but not less than $1,000$). As a first step, identify values of $(k, N, \epsilon, \tau, w)$ such that

Regime 1 where OPT is significantly better than $BlindOracle$ with $\hat{h}$ values which is significantly better than LRU, and

Regime 2 where OPT is significantly better than LRU which is significantly better than $BlindOracle$ with $\hat{h}$ values.

Having identified values for two regimes, you are ready to observe some trends.

Trend 1 (dependence on k): vary the value of k (adjusting the value of N accordingly – you can use a fixed relationship such as $N = 10k$, for example). For each value of k and each regime, run a batch of trials. Plot average page faults of all 4 algorithms (OPT, LRU, $BlindOracle$ with $\hat{h}$ values, $Combined$) versus k as 4 time series lines (on the same plot). Thus, you should obtain 2 plots – one for each regime.

Trend 2 (dependence on w): vary the value of w and plot average page faults of all 4 algorithms versus w.

Trend 3 (dependence on ϵ): vary the value of ϵ and plot average page faults of all 4 algorithms versus ϵ.

Trend 4 (dependence on τ): vary the value of τ and plot average page faults of all 4 algorithms versus τ.

Then write a report, which should have 5 sections (one section per each of the above trends + conclusion). Each of the first 4 sections should start out with two plots associated with the corresponding trend. You should use proper scientific presentation - figures should have labels on the axis, units of measurement, legends for trend-lines, and a title. Do not make plots too large (you can have two of them side-by-side, occupying not more than 1/3 of a page). These plots should be followed by a short discussion: what can you conclude from the plots? are there any surprises, or are the plots as you expected them to be? Why? If some lines intersect and overtake each other, or reach a peak – what is the meaning of this? If you cannot explain certain patterns, propose additional experiments that may help you understand what is going on (you don't have to do these additional experiments, just propose them). Lastly, in the conclusion section you should discuss what you have learnt from the project and highlight the most interesting things you encountered in the prject. The report should not exceed 4 pages, but it should contain at least 2 full pages of pure text (these 2 full pages won't happen consecutively, of course). Thus, if I remove all figures, whitespace, title, section names, etc, I should be left with at least 2 pages of pure text. Use regular US letter size of a page, single-spaced, font size 11, and margins no larger than 1 inch.

You should upload the updated code that contains experiments and plot generation (you are allowed to use 3rd party libraries of generating figures, such as matplotlib, for example. You can also produce csv files with results of experiments and then use other software to generate plots from csv files. Then you should document in the readme file how the plots were generated), updated readme file, and report.pdf.