ECE/CS 4434/6434
Homework 8
Due Date: Friday, November 15, 4:59 PM

Reading: Lectures 19, 20, 21 + Chapter 6 of I. Koren + Chapter 8 of R. K. Iyer
Note: Problem 1 is almost the same as Class Activity 6.

Problem 1 (40 pts) **–** Assume that you are using Google's TensorFlow to train a complex machine learning model (a deep neural network model with several layers) for speech recognition (e.g., see DeepSpeech https://github.com/mozilla/DeepSpeech). During training of the model interruptions due to unexpected failures might happen that lead to losing hours of training time. You use a check-pointing and recovery technique (e.g., see TensorFlow's Training Checkpoints https://www.tensorflow.org/guide/checkpoint) where regular checkpoints of the trained model are stored in separate directories on the disk. Resuming from the last checkpoint happens automatically by just (re)starting the training process with the same checkpoint of the former run. The time interval at which the checkpoints are taken can be configured by you based on the type of training experiment and the overhead of checkpointing.

Part A (10 pts) – What information (data and parameters) from each training experiment should be saved in a checkpoint so that a future training experiment can be successfully resumed from that checkpoint? **Hint:** Assume that the source code for your model can be re-executed without any changes/problems in each training experiment, During the training phase of a deep neural network model, the weights and biases for each layer of the model are learned from the training data using an iterative optimization algorithm with the goal of minimizing a loss function. Each training experiment is configured with hyperparameters, including the batch size, learning rate, optimization function, which are further tuned to obtain the optimal performance. You may answer this question based on your experience with machine learning or by consulting any online documents on checkpointing for deep learning models. Please provide any assumptions you make for your answer.

1. **Model Weights and Biases**: These represent the learned parameters for each layer of the neural network and are essential for resuming the model from the exact state it was in when the checkpoint was saved.
2. **Optimizer State**: The optimizer's state includes moment estimates, velocity terms, or accumulated gradients.
3. **Hyperparameters**: Key hyperparameters such as batch size, learning rate, and optimizer settings should be saved to ensure consistent conditions when resuming training.
4. **Training Epoch and Iteration**: Keeping track of the epoch and iteration number allows resumption from the precise point in the training process.
5. **Random Seed**: Saving the random seed ensures that any can be reproduced, maintaining consistency in the resumed training process.

If the whole training process for a given dataset is 5 hours (T) and you configure the checkpoints to be taken uniformly at the beginning of each hour (**Hint:** no checkpoints taken at times 0 or 5hrs), calculate the expected value of the time interval between completion of two checkpoints

Names: _____

Email IDs: _____

(E_{int}) and use that to estimate the total time it would take for the training process to be done in each of the following cases.

Assume the expected interval between the start and the end of checkpoint or checkpoint latency (T_{lt}) is 2 minutes, the checkpoint overhead (T_{ov}) is 1 minute, the failure happens in one of the one-hour intervals, half an hour after the last checkpoint (τ = 30 mins), and restarting of the training process (T_r) takes about 0.5 minute.

Part B (10 pts) – No failure occurs.

$T_{ov} = 1\ Minute$
$T = 300\ Minutes\ (5\ Hours)$
$T_{ex} = \frac{5}{4+1} = 1\ Hour$
$E_{int} = 60(T_{ex}) + 1(T_{ov}) = 60 + 1 = 61\ Minutes$
$Total\ Time = 300\ (T) + 4\big(1(T_{ov})\big) = 304\ Minutes$

Part C (20 pts) – One failure occurs.

$T = 300\ Minutes$
$T_{ov} = 1\ Minute$
$T_{lt} = 2\ Minutes$
$T_{ex} = 60\ Minutes$
$\tau = 30$ Minutes
$T_r = 0.5\ Minutes$
$E_{int} = 60(T_{ex}) + 1(T_{ov}) = 61\ Minutes$
$Total\ Time = 300(T) + (30(\tau) + 0.5(T_r) + 2(T_{lt}) - 1(T_{ov})) + 4(1(T_{ov}) = 335.5\ Minutes$

Problem 2 (30 pts) – You have a task with execution time, T. You take N checkpoints, equally spaced through the lifetime of the task. The overhead for each checkpoint is T_{ov} and $T_{lt} = T_{ov}$. Given that during execution, the task is affected by a total of k point failures (i.e., failures from which the processor recovers in negligible time), answer the following:

Part A (15 pts) What is the maximum execution time of the task?

$Maximum\ Checkpoint\ Overhead = N * T_{ov}$
$T_{ex} = \frac{T}{N+1}$
$Maximum\ Lost\ Time = k * \frac{T}{N+1}$
$T_{Max} \leq T + N * T_{ov} + k * \frac{T}{N+1}$

Names: _____

Email IDs: _____

Part B (15 pts) Find N such that this maximum execution time is minimized. It is fine to get a non-integer answer (say x): in practice, this will mean that you will pick the better of $\lfloor x \rfloor$ and $\lceil x \rceil$.

$$T_{MAX}(N) = T + N * T_{ov} + k * \frac{T}{N+1}$$

$$\frac{dT_{MAX}(N)}{dN} = T_{ov} - \frac{kt}{(N+1)^2} = 0$$

$$T_{ov} = \frac{kt}{(N+1)^2}$$

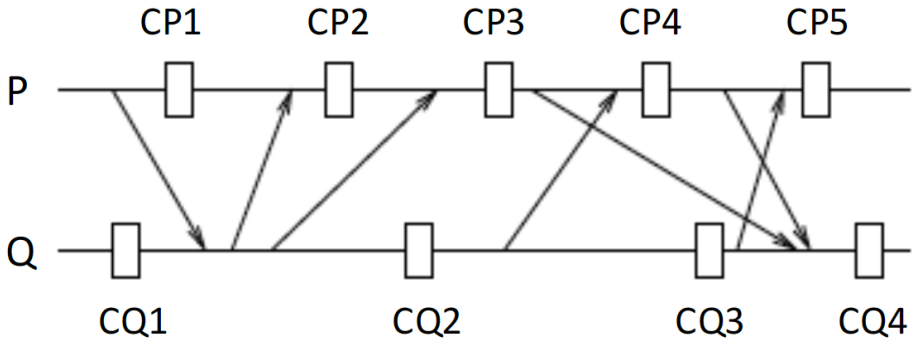
$$(N+1)^2 = \frac{kt}{T_{ov}}$$

$$N \geq \sqrt{\frac{kt}{T_{ov}}} - 1$$

Problem 3 (30 pts) – Remember the distributed check-pointing scheme from **Lecture 20**. The state of a distributed system is said to be consistent if, for every message delivery recorded in the state, there is a corresponding message-sending event.

Part A (20 pts) Identify all the consistent states in the following execution of two concurrent processes P and Q. For example, {CP1; CQ2} is a consistent state because the delivery of first message is recorded in CQ2 and its sending event is recorded in CP1. However, {CP2; CQ1} is not consistent, because the delivery of the second message is recorded in CP2, but its corresponding sending event is not recorded in CQ1.

1. {CP1;CQ2}
2. {CP2;CQ2}
3. {CP3;CQ2}
4. {CP4;CQ3}
5. {CP4;CQ4}
6. {CP5;CQ4}



Names: _____

Email IDs: _____

Part B (10 pts) A strongly consistent set of checkpoints or "recovery line" corresponds to a strongly consistent global state, where:

- there is one recovery point for each process in the set during the interval spanned by checkpoints,
- there is no information flow between any pair of processes in the set and a process in the set and any process outside the set

Identify all the consistent recovery lines in the execution of the two processes.

1. {CP3;CQ2}
2. {CP5;CQ4}