

CLAIMS

1. A server for facilitating batching orders of an on-demand service, the server comprising:
 - a memory for storing instructions;
 - a communication interface configured to receive an order for the on-demand service from a computing device; and
 - a processor for executing the stored instructions and configured to:
 - receive the order from the communication interface;
 - predict an OTE (overall time efficiency) for a predetermined time slot at a predetermined area, based on a market condition signal and a batching configuration parameter; compare the predicted OTE with a predetermined OTE threshold;
 - Determine whether the predicted OTE is greater than the predetermined OTE threshold;
 - if it is determined that the predicted OTE is greater than the predetermined OTE threshold, select a first batching strategy, and attempt upfront batching upon a receipt of the order based on the selected first batching strategy; and
 - if it is determined that the predicted OTE is equal to or less than the predetermined OTE threshold, select a second batching strategy, and attempt in-transit batching upon the receipt of the order based on the selected second batching strategy.
2. The server according to claim 1, wherein, based on the first batching strategy, the processor is further configured to:
 - determine that the upfront batching is attempted upon the receipt of the order;
 - determine if a predetermined upfront pooling duration expires; and

if it is determined that the predetermined upfront pooling duration expires, switch to an alternation mode that alternates between the in-transit batching or an idle allocation.

3. The server according to claim 1, wherein, based on the second batching strategy, the processor is further configured to switch to an alternation mode that alternates between the in-transit batching or an idle allocation.

4. The server according to claim 3, wherein the processor is further configured to:
start the alternation mode with tightest constraints; and
gradually loosen the constraints with attempts.

5. The server according to any one of claims 1 to 4, wherein the processor is further configured to adjust the batching configuration parameter using the predicted OTE.

6. The server according to any one of claims 1 to 5, wherein the batching configuration parameter includes at least one of a batching interval, an order delivery delay, and a maximum amount of time the order can be kept.

7. The server according to any one of claims 1 to 6, wherein the processor is further configured to:

predict an IAR (inter-arrival rate) for the predetermined time slot at the predetermined area; and

use the predicted IAR as the market condition signal.

8. The server according to claim 7, wherein the processor is configured to use an OTE predictor model to predict the OTE, and

the processor is further configured to train the OTE predictor model using at least one of the predicted IAR, the predicted OTE, and the batching configuration parameter.

9. The server according to any one of claims 1 to 8, wherein the processor is further configured to:

predict a reference OTE based on a first market condition signal and a first batching configuration parameter in a reference state, and a target OTE based on a second market condition signal and a second batching configuration parameter in a target state; and

obtain a predicted OTE gain based on the predicted reference OTE and the predicted target OTE.

10. The server according to claim 9, wherein the processor is further configured to determine the OTE threshold based on the predicted OTE gain.

11. A method for facilitating batching orders of an on-demand service, the method comprising:

receiving an order for the on-demand service;

predicting an OTE (overall time efficiency) for a predetermined time slot at a predetermined area, based on a market condition signal and a batching configuration parameter;

comparing the predicted OTE with a predetermined OTE threshold;

determining whether the predicted OTE is greater than the predetermined OTE threshold;

if it is determined that the predicted OTE is greater than the predetermined OTE threshold, selecting a first batching strategy, and attempting upfront batching upon a receipt of the order based on the selected first batching strategy; and

if it is determined that the predicted OTE is equal to or less than the predetermined OTE threshold, selecting a second batching strategy, and attempting in-transit batching upon the receipt of the order based on the selected second batching strategy.

12. The method according to claim 11 further comprising: based on the first batching strategy,

determining that the upfront batching is attempted upon the receipt of the order;

determining if a predetermined upfront pooling duration expires; and

if it is determined that the predetermined upfront pooling duration expires, switching to an alternation mode that alternates between the in-transit batching or an idle allocation.

13. The method according to claim 11 further comprising: based on the second batching strategy, switching to an alternation mode that alternates between the in-transit batching or an idle allocation.

14. The method according to claim 13 further comprising:

starting the alternation mode with tightest constraints; and

gradually loosening the constraints with attempts.

15. The method according to any one of claims 11 to 14 further comprising adjusting the batching configuration parameter using the predicted OTE.

16. The method according to any one of claims 11 to 15, wherein the batching configuration parameter includes at least one of a batching interval, an order delivery delay, and a maximum amount of time the order can be kept.
17. The method according to any one of claims 11 to 16 further comprising:
predicting an IAR (inter-arrival rate) for the predetermined time slot at the predetermined area; and
using the predicted IAR as the market condition signal.
18. The method according to claim 17, wherein the predicting an OTE comprises using an OTE predictor model to predict the OTE, and
the method further comprises training the OTE predictor model using at least one of the predicted IAR, the predicted OTE, and the batching configuration parameter.