

SERVER AND METHOD FOR FACILITATING PROCESSING A PLURALITY OF ORDERS FOR ON-DEMAND SERVICE

TECHNICAL FIELD

[0001] Various embodiments relate to a server and a method for facilitating processing a plurality of orders for an on-demand service.

BACKGROUND

[0002] Due to development of information technology, a user (who in some contexts herein may also be referred to as a “consumer” or an “eater”) may request an on-demand service using a computing device. The on-demand service may allow the user to fulfil the user’s demand via an immediate access to goods and/or services. The user may request the on-demand service, for example, a delivery service or a transport service, using a user interface presented on the computing device. After a server for the on-demand service receives an order (also referred to as a “booking”) for the on-demand service from the user, the server may allocate a delivery service provider (who in some contexts herein may also be referred to as a “driver”, a “delivery partner” or a “delivery agent”) to the received order.

[0003] FIG. 1 illustrates exemplary diagrams showing a concept of a normal allocation model and a concept of a JITA (just-in-time allocation) model according to conventional technologies. As shown in (A) of FIG. 1, conventionally, the normal allocation model may have been used. According to the normal allocation model, once the server receives the order from the user, the server may start finding the delivery service provider (referred to as a “DAX” in figures). After the delivery service provider is found, an item provider (referred to as a “MEX” in figures)

may be notified to prepare an item, for example, food, for the received order. In the meantime, the delivery service provider may travel to a location of the item provider to collect the food. After the food is ready for collection, the delivery service provider may collect the food and deliver the food to the user.

[0004] To improve efficiency in providing the on-demand service, the JITA (just-in-time allocation) model has been introduced. The JITA model may delay allocating orders to the delivery service provider for a predetermined time. The JITA model may serve to minimise a driver wait time (DWT) at the location of the item provider. For example, the JITA model may serve to minimise the driver wait time (DWT), by delaying a driver allocation start point and notifying the item provider to prepare the item before the delivery service provider arrives at the location of the item provider.

[0005] As shown in (B) of FIG. 1, according to the JITA model that has been introduced, once the server receives the order from the user, the item provider may be notified to prepare the food. After the item provider is notified to prepare the food, the server may start finding the delivery service provider. After the delivery service provider is found, the delivery service provider may travel to the location of the item provider to collect the food. After the food is ready for collection, the delivery service provider may collect the food and deliver the food to the user.

[0006] There may be a need to optimise the delay in allocation of the order to the delivery service provider by predicting the delay more accurately for right orders and item providers while keeping waiting time low. The JITA model that has been introduced may include a basic JITA model and a smart JITA model. The basic JITA model may use a confirmed allocation rate (CAR) with respect to a confirmed allocation rate (CAR) threshold which may be configured at a city or a geohash5 level. For example, if the confirmed allocation rate (CAR) falls below the CAR threshold, the basic JITA model may be deactivated. The smart JITA

model may use a CRWM (Calculated Risk per Waiting Minute saved) value of each order with respect to a CRWM threshold which may be configured at a city level. The CRWM value may be calculated using the compensation cost divided by the estimated wait time saved, to represent the risk of not allocating the order. For example, if the CRWM value of the order exceeds the CRWM threshold, the smart JITA model may be deactivated for the order. In this manner, the JITA model may simply be activated based on the confirmed allocation rate (CAR), which may refer to the number of allocated (confirmed) orders divided by the number of outstanding orders, as an approximation of how likely the order will be allocated.

[0007] However, the confirmed allocation rate (CAR) may not accurately reflect allocation probability of the order, as the confirmed allocation rate (CAR) may be reactive, consider insufficient features, and be a blanket signal calculated at an item provider, a geohash and/or a city level instead of at an order level. Specifically, the confirmed allocation rate (CAR) may be reactive, as the confirmed allocation rate (CAR) may be the confirmed allocation rate for a certain time period in the past (for example, in the past 10 minutes). In light of rapidly changing conditions, the confirmed allocation rate (CAR) may not fully represent current situations (for example, start of peak hours). This may create a delay in a decision of the JITA model, which may result in higher unallocated orders and compensation costs. In addition, the confirmed allocation rate (CAR) may purely be based on a ratio of the allocated orders with respect to the orders placed, and may not factor in underlying features such as a state of supply or allocation/batching levers. Further, the confirmed allocation rate (CAR) may not reflect an actual allocation probability based on each order's characteristics (attributes) which may affect a filtration and prioritisation. For example, large orders with long R2C (Restaurant to Consumer) may prefer a four-wheeled car as the delivery service provider, while the high confirmed allocation rate (CAR) in the past 10 minutes may be due to an abundance of two-wheeled motorcycles as the delivery service providers.

[0008] Therefore, there is a need to provide a solution for facilitating processing a plurality of orders for on-demand services.

SUMMARY

[0009] According to various embodiments, there is a server for facilitating processing a plurality of orders for an on-demand service, the server comprising: a memory configured to store instructions; a communication interface configured to receive the plurality of orders for the on-demand service from a plurality of computing devices; and a processor for executing the stored instructions and configured to: for each order of the plurality of orders, receive information about one or more characteristics of the order; obtain an indicator relating to a likelihood that the order will get allocated to a delivery service provider, based on the information about the one or more characteristics of the order; compare the indicator with a designated threshold; and determine whether to delay allocating the order to the delivery service provider for a certain time, based on the comparison of the indicator with the designated threshold.

[0010] In some embodiments, the processor is further configured to: for each order of the plurality of orders, check whether an item provider relating to the order has allowed a delay in allocating one or more received orders to a respective delivery service provider; and determine whether to assign the order to a first candidate group of orders, based on whether the item provider relating to the order has allowed the delay.

[0011] In some embodiments, the processor is further configured to: for each order assigned to the first candidate group of orders, compare the indicator with the designated threshold; and determine whether to assign the order to a second candidate group of orders, based on the comparison of the indicator with the designated threshold.

[0012] In some embodiments, the processor is further configured to: for each order assigned to the first candidate group of orders, determine that the indicator is greater than the designated threshold; and determine to assign the order to the second candidate group of orders.

[0013] In some embodiments, the processor is further configured to: for each order assigned to the second candidate group of orders, calculate the certain time for the delay, based on at least one of an item preparation time, an elapsed time, a pick-up buffer, a pick-up ETA (estimated time of arrival) and an allocation buffer; and determine whether to delay allocating the order to the delivery service provider, based on the calculated certain time.

[0014] In some embodiments, the processor is further configured to: for each order assigned to the second candidate group of orders, check if an upfront batching for the order is enabled; determine that the upfront batching for the order is enabled; and determine whether to delay allocating the order to the delivery service provider, based on the calculated certain time.

[0015] In some embodiments, the processor is further configured to: for each order assigned to the second candidate group of orders, obtain a maximum pooling time for the upfront batching; compare the calculated certain time with the maximum pooling time; and determine whether to delay allocating the order to the delivery service provider, based on the comparison of the calculated certain time with the maximum pooling time.

[0016] In some embodiments, the processor is further configured to: for each order assigned to the second candidate group of orders, determine that the calculated certain time is longer than the maximum pooling time; and determine to delay allocating the order to the delivery service provider for the calculated certain time.

[0017] In some embodiments, the processor is further configured to: for each order assigned to the second candidate group of orders, determine that the upfront batching for the order is disabled; and determine to delay allocating the order to the delivery service provider for the calculated certain time.

[0018] In some embodiments, the one or more characteristics of the order include at least one of geographical characteristics of the order, temporal characteristics of the order, a category of the order, and characteristics of orders geographically relevant to the order.

[0019] In accordance with various embodiments, there is a method for facilitating processing a plurality of orders for an on-demand service, the method comprising: receiving the plurality of orders for the on-demand service from a plurality of computing devices; for each order of the plurality of orders, receiving information about one or more characteristics of the order; obtaining an indicator relating to a likelihood that the order will get allocated to a delivery service provider, based on the information about the one or more characteristics of the order; comparing the indicator with a designated threshold; and determining whether to delay allocating the order to the delivery service provider for a certain time, based on the comparison of the indicator with the designated threshold.

[0020] In some embodiments, the method further comprises: for each order of the plurality of orders, checking whether an item provider relating to the order has allowed a delay in allocating one or more received orders to a respective delivery service provider; and determining whether to assign the order to a first candidate group of orders, based on whether the item provider relating to the order has allowed the delay.

[0021] In some embodiments, the method further comprises: for each order assigned to the first candidate group of orders, comparing the indicator with the designated threshold; and determining whether to assign the order to a second candidate group of orders, based on the comparison of the indicator with the designated threshold.

[0022] In some embodiments, the method further comprises: for each order assigned to the first candidate group of orders, determining that the indicator is greater than the designated threshold; and determining to assign the order to the second candidate group of orders.

[0023] In some embodiments, the method further comprises: for each order assigned to the second candidate group of orders, calculating the certain time for the delay, based on at least one of an item preparation time, an elapsed time, a pick-up buffer, a pick-up ETA (estimated time of arrival) and an allocation buffer; and determining whether to delay allocating the order to the delivery service provider, based on the calculated certain time.

[0024] In some embodiments, the method further comprises: for each order assigned to the second candidate group of orders, checking if an upfront batching for the order is enabled; determining that the upfront batching for the order is enabled; and determining whether to delay allocating the order to the delivery service provider, based on the calculated certain time.

[0025] In some embodiments, the method further comprises: for each order assigned to the second candidate group of orders, obtaining a maximum pooling time for the upfront batching; comparing the calculated certain time with the maximum pooling time; and determining whether to delay allocating the order to the delivery service provider, based on the comparison of the calculated certain time with the maximum pooling time.

[0026] In some embodiments, the method further comprises: for each order assigned to the second candidate group of orders, determining that the calculated certain time is longer than the maximum pooling time; and determining to delay allocating the order to the delivery service provider for the calculated certain time.

[0027] In some embodiments, the method further comprises: for each order assigned to the second candidate group of orders, determining that the upfront batching for the order is disabled; and determining to delay allocating the order to the delivery service provider for the calculated certain time.

[0028] In some embodiments, the one or more characteristics of the order include at least one of geographical characteristics of the order, temporal characteristics of the order, a category of the order, and characteristics of orders geographically relevant to the order.

[0029] According to various embodiments, a data processing apparatus configured to perform the method of any one of the above embodiments is provided.

[0030] According to various embodiments, a computer program element comprising program instructions, which, when executed by one or more processors, cause the one or more processors to perform the method of any one of the above embodiments is provided.

[0031] According to various embodiments, a computer-readable medium comprising program instructions, which, when executed by one or more processors, cause the one or more processors to perform the method of any one of the above embodiments is provided. The computer-readable medium may include a non-transitory computer-readable medium.

BRIEF DESCRIPTION OF THE DRAWINGS

[0032] The invention will be better understood with reference to the detailed description when considered in conjunction with the non-limiting examples and the accompanying drawings, in which:

- FIG. 1 illustrates exemplary diagrams showing a concept of a normal allocation model and a concept of a JITA (just-in-time allocation) model according to conventional technologies.
- FIG. 2 illustrates an infrastructure of a system including a server for facilitating processing a plurality of orders for an on-demand service according to various embodiments.
- FIG. 3 illustrates a block diagram of a server for facilitating processing a plurality of orders for an on-demand service according to various embodiments.
- FIG. 4 illustrates an exemplary diagram showing an application of a JITA model according to various embodiments.

- FIG. 5 illustrates a flowchart for a method for facilitating processing a plurality of orders for an on-demand service according to various embodiments.
- FIG. 6 illustrates an exemplary diagram showing a JITA enabled orders group, a JITA active orders group, and a JITA effective orders group according to various embodiments.
- FIG. 7 illustrates a data flow diagram of an allocation probability model (APM) according to various embodiments.
- FIGS. 8, 9 and 10 illustrate exemplary diagrams showing a user interface of a computing device of an item provider according to various embodiments.

DETAILED DESCRIPTION

[0033] The following detailed description refers to the accompanying drawings that show, by way of illustration, specific details and embodiments in which the disclosure may be practiced. These embodiments are described in sufficient detail to enable those skilled in the art to practice the disclosure. Other embodiments may be utilized and structural, and logical changes may be made without departing from the scope of the disclosure. The various embodiments are not necessarily mutually exclusive, as some embodiments can be combined with one or more other embodiments to form new embodiments.

[0034] Embodiments described in the context of one of a server and a method are analogously valid for the other server and method. Similarly, embodiments described in the context of a server are analogously valid for a method, and vice-versa.

[0035] Features that are described in the context of an embodiment may correspondingly be applicable to the same or similar features in the other embodiments. Features that are described in the context of an embodiment may correspondingly be applicable to the other embodiments,

even if not explicitly described in these other embodiments. Furthermore, additions and/or combinations and/or alternatives as described for a feature in the context of an embodiment may correspondingly be applicable to the same or similar feature in the other embodiments.

[0036] In the context of various embodiments, the articles “a”, “an” and “the” as used with regard to a feature or element include a reference to one or more of the features or elements.

[0037] As used herein, the term “and/or” includes any and all combinations of one or more of the associated listed items.

[0038] Throughout the description, the term “module” may be understood as an application specific integrated circuit (ASIC), an electronic circuit, a combinational logic circuit, a field programmable gate array (FPGA), a processor which executes code, other suitable hardware components which provide the described functionality, or any combination thereof. The term of “module” may include a memory which stores code executed by the processor.

[0039] In the following, embodiments will be described in detail.

[0040] FIG. 2 illustrates an infrastructure of a system 200 including a server 100 for facilitating processing a plurality of orders for an on-demand service according to various embodiments.

[0041] As shown in FIG. 2, the system 200 may include, but is not limited to, the server 100, a database system 140, a network 150, a computing device 160, and one or more external devices 170, 180 (not shown).

[0042] In some embodiments, the on-demand service may be a service allowing a user 161 (who in some contexts herein may also be referred to as a “consumer” or an “eater”) to fulfil the user’s demand via an immediate access to items and/or services. The user 161 may request the on-demand service, for example, a transport service or an item delivery service, using a user interface presented on the computing device 160. The user 161 may make an order for the on-demand service.

[0043] In some embodiments, the network 150 may include, but is not limited to, a Local Area Network (LAN), a Wide Area Network (WAN), a Global Area Network (GAN), or any combination thereof. The network 150 may provide a wireline communication, a wireless communication, or a combination of the wireline and wireless communication between the server 100 and the computing device 160, and between the server 100 and the one or more external devices 170, 180, for example, one or more delivery service provider devices 170, and one or more item provider devices 180.

[0044] In some embodiments, the computing device 160 may be connectable to the server 100 via the network 150. In some embodiments, the computing device 160 may be arranged in data or signal communication with the server 100 via the network 150. In some embodiments, the computing device 160 may include, but is not limited to, at least one of the following: a mobile phone, a tablet computer, a laptop computer, a desktop computer, a head-mounted display and a smart watch. In some embodiments, the computing device 160 may be associated with the user 161. For example, the computing device 160 may belong to the user 161. Although not shown, it may be appreciated that the system 200 may further include a plurality of computing devices each associated with, for example, belonging to, a plurality of users.

[0045] In some embodiments, the computing device 160 may include a location sensor. In some embodiments, the location sensor may communicate with at least one of a global positioning satellite (GPS) server, a network server, and a Wi-Fi server, to detect a location of the computing device 160. In some embodiments, the computing device 160 may generate information about the location of the computing device 160.

[0046] In some embodiments, the server 100, for example, implemented by a server computer, may include a communication interface 110, a processor 120, and a memory 130 (as will be described with reference to FIG. 3).

[0047] In some embodiments, the server 100 may communicate with the computing device 160 via the network 150. In some embodiments, the computing device 160 may receive a request (hereinafter, referred to as an “order” or a “booking”) from the user 161 for the on-demand service. The computing device 160 may send the order to the server 100 via the network 150. In some embodiments, the computing device 160 may send the information about the location of the computing device 160 to the server 100 via the network 150. The location of the computing device 160 may be considered as a location of the user 161.

[0048] In some embodiments, the system 200 may further include a database 141. In some embodiments, the database 141 may be a part of the database system 140 which may be external to the server 100. The server 100 may communicate with the database 141. In some other embodiments, although not shown, the database 141 may be implemented locally in the memory 130 of the server 100.

[0049] In some embodiments, the server 100 may communicate with the one or more external devices 170, for example, the one or more delivery service provider devices 170, via the network 150. The one or more delivery service provider devices 170 may be associated with one or more delivery service providers 171 (who in some contexts herein may also be referred to as a “drivers”, “delivery partners” or “delivery agents”) respectively. For example, the one or more delivery service provider devices 170 may belong to the one or more delivery service providers 171 respectively. The one or more delivery service providers 171 may provide a delivery service from a first location, for example, a location of a selected item provider 181a, to a second location, for example, a delivery location such as a location of the user 161.

[0050] In some embodiments, the server 100 may communicate with the one or more external devices 180, for example, the one or more item provider devices 180, via the network 150. The one or more item provider devices 180 may be associated with one or more item providers 181 respectively. For example, the one or more item provider devices 180 may belong to the one

or more item providers 181 respectively. The one or more item providers 181 may include, but are not limited to, a food provider and a goods provider, that can manufacture and/or provide items, for example, foods or goods. For example, the food provider may include, but is not limited to, a restaurant and a café. As an example, the goods provider may include, but is not limited to, a store, a market, and a supermarket. In some embodiments, the server 100 may receive the order for the item delivery service with the information about the location of the user 161, and then produce a list of items, associated with at least one item provider, which can be prepared and delivered to the user 161. In some embodiments, the server 100 may communicate with the one or more item provider devices 180 to check the one or more item providers' 181 availability. In some embodiments, the server 100 may communicate with the one or more item provider devices 180 to aggregate information including, but not limited to, a list of available items, an estimated time of preparation of each item, and an estimated price of each item, in order to produce the list of items. In some embodiments, the computing device 160 may display the list of items with the aggregated information of the at least one of the one or more item providers 181 on the user interface. In some embodiments, after the user 161 makes selections on the user interface for the order for the item delivery service, for example, by selecting the item provider 181a and the item, the server 100 may communicate with an item provider device 180a (not shown) of the selected item provider 181a to prepare the selected item.

[0051] In some embodiments, the server 100 may receive the order from the computing device 160. After the server 100 receives the order from the computing device 160, the server 100 may allocate (assign) the order to a suitable delivery service provider 171a. To allocate the order to the delivery service provider 171a, the server 100 may use a normal allocation model or a JITA (just-in-time) allocation model. According to the normal allocation model, once the server 100 receives the order from the computing device 160 of the user 161, the server 100

may start finding the suitable delivery service provider 171a. After the server 100 selects the delivery service provider 171a and allocates the order to the selected delivery service provider 171a, the server 100 may send a notification to the item provider device 180a, so that the item provider 181a may prepare an item, for example, food, for the order. In the meantime, the selected delivery service provider 171a may travel to a location of the item provider 181a to collect the food. After the food is ready for collection, the selected delivery service provider 171a may collect the food and deliver the food to the user 161.

[0052] The JITA model may delay allocating orders to the delivery service provider 171a for a certain time. According to the JITA model, once the server 100 receives the order from the computing device 160 of the user 161, the server 100 may send a notification to the item provider device 180a, so that the item provider 181a may prepare an item, for example, food, for the order. After the item provider 181a is notified to prepare the food, the server 100 may start finding the suitable delivery service provider 171a. After the server 100 selects the delivery service provider 171a and allocates the order to the selected delivery service provider 171a, the selected delivery service provider 171a may travel to a location of the item provider 181a to collect the food. After the food is ready for collection, the selected delivery service provider 171a may collect the food and deliver the food to the user 161.

[0053] FIG. 3 illustrates a block diagram of a server 100 for facilitating processing a plurality of orders for an on-demand service according to various embodiments. FIG. 4 illustrates an exemplary diagram showing an application of a JITA model according to various embodiments.

[0054] As shown in FIG. 3, the server 100, for example, implemented by a server computer, may include a communication interface 110, a processor 120, and a memory 130.

[0055] In some embodiments, the memory 130 (also referred to as a “database”) may store input data and/or output data temporarily or permanently. In some embodiments, the memory

130 may store program code which allows the server 100 to perform a method S300 (as will be described with reference to FIG. 5). In some embodiments, the program code may be embedded in a Software Development Kit (SDK). The memory 130 may include an internal memory of the server 100 and/or an external memory. The external memory may include, but is not limited to, an external storage medium, for example, a memory card, a flash drive, and a web storage.

[0056] In some embodiments, the communication interface 110 may allow one or more computing devices, including a computing device 160, to communicate with the processor 120 of the server 100 via a network 150, as shown in FIG. 2. In some embodiments, as shown in FIG. 2, the computing device 160 may belong to a user 161 who wants to make an order for an on-demand service. In some embodiments, the communication interface 110 may transmit signals to the computing device 160, and/or receive signals from the computing device 160 via the network 150.

[0057] In some embodiments, the communication interface 110 may allow one or more external devices 170, 180, for example, one or more delivery service provider devices 170, and one or more item provider devices 180, to communicate with the processor 120 of the server 100 via the network 150, as shown in FIG. 2. In some embodiments, the communication interface 110 may transmit signals to the one or more external devices 170, 180, and/or receive signals from the one or more external devices 170, 180, via the network 150.

[0058] In some embodiments, the communication interface 110 may receive an order for an on-demand service, for example, an item delivery service to be provided to the user 161, from the computing device 160 associated with the user 161 via the network 150. In some embodiments, the order for the item delivery service may include information about at least one selected item and at least one selected item provider 181a providing the at least one selected

item. The communication interface 110 may then send the order for the item delivery service to the processor 120.

[0059] In some embodiments, the communication interface 110 may further receive information about a location of the computing device 160 from the computing device 160 via the network 150. The communication interface 110 may then send the information about the location of the computing device 160 to the processor 120.

[0060] In some embodiments, the communication interface 110 may receive concurrently the order for the item delivery service and the information about the location of the computing device 160 from the computing device 160 via the network 150. In some other embodiments, the communication interface 110 may receive the order for the item delivery service first, and subsequently receive the information about the location of the computing device 160, from the computing device 160 via the network 150.

[0061] The processor 120 may include, but is not limited to, a microprocessor, an analogue circuit, a digital circuit, a mixed-signal circuit, a logic circuit, an integrated circuit, a Central Processing Unit (CPU), a Graphics Processing Unit (GPU), a Digital Signal Processor (DSP), a Field Programmable Gate Array (FPGA), an Application Specific Integrated Circuit (ASIC), or any combination thereof. Any other kind of implementation of the respective functions, which will be described below in further detail, may also be understood as the processor 120.

[0062] In some embodiments, the processor 120 may be connectable to the communication interface 110. In some embodiments, the processor 120 may be arranged in data or signal communication with the communication interface 110 to receive the order for the item delivery service and the information about the location of the computing device 160.

[0063] In some embodiments, the processor 120 may receive concurrently the order for the item delivery service and the information about the location of the computing device 160 from the communication interface 110. In some other embodiments, the processor 120 may receive

the order for the item delivery service first, and subsequently receive the information about the location of the computing device 160 from the communication interface 110.

[0064] In some embodiments, the processor 120 may receive a request for a search for the item delivery service from the computing device 160. In some embodiments, the processor 120 may produce a list of items, associated with at least one item provider 181, which can be prepared and delivered to the user 161. The processor 120 may then provide the list of items to the computing device 160, so that the user 161 may select the item provider 181a and the item for the item delivery service.

[0065] In some embodiments, the user 161 may select the item provider 181a and the item from the list of items. The computing device 160 may generate information about the selected item provider 181a and the selected item, based on the user's 161 input. The processor 120 may receive the information about the selected item provider 181a and the selected item from the computing device 160, via the communication interface 110. In some embodiments, the processor 120 may then provide the information about the selected item to the selected item provider 181a for preparing the selected item, via the communication interface 110.

[0066] In some embodiments, the processor 120 may select a delivery service provider 171a from one or more delivery service providers 171 based on a geographical location of the one or more delivery service providers 171. In some embodiments, the processor 120 may request the selected delivery service provider 171a to pick up the selected item at a geographical location of the selected item provider 181a and deliver the selected item to the user 161.

[0067] In some embodiments, after the processor 120 receives the order from the computing device 160 via the communication interface 110, the processor 120 may select the delivery service provider 171a and allocate the order to the selected delivery service provider 171a. To allocate the order to the delivery service provider 171a, the processor 120 may use a normal allocation model or a JITA (just-in-time) allocation model. According to the normal allocation

model, once the processor 120 receives the order from the computing device 160 of the user 161 via the communication interface 110, the processor 120 may start finding the suitable delivery service provider 171a. After the processor 120 selects the delivery service provider 171a and allocates the order to the selected delivery service provider 171a, the processor 120 may send a notification to the item provider device 180a via the communication interface 110, so that the item provider 181a may prepare an item, for example, food, for the order. In the meantime, the selected delivery service provider 171a may travel to a location of the item provider 181a to collect the food. After the food is ready for collection, the selected delivery service provider 171a may collect the food and deliver the food to the user 161.

[0068] According to the JITA model, once the processor 120 receives the order from the computing device 160 of the user 161 via the communication interface 110, the processor 120 may send a notification to the item provider device 180a via the communication interface 110, so that the item provider 181a may prepare an item, for example, food, for the order. After the item provider 181a is notified to prepare the food, the processor 120 may start finding the suitable delivery service provider 171a. After the processor 120 selects the delivery service provider 171a and allocates the order to the selected delivery service provider 171a, the selected delivery service provider 171a may travel to the location of the item provider 181a to collect the food. After the food is ready for collection, the selected delivery service provider 171a may collect the food and deliver the food to the user 161.

[0069] In some embodiments, the communication interface 110 may receive the plurality of orders for the on-demand service from a plurality of computing devices each associated with a plurality of users. In some embodiments, the processor 120 may receive the plurality of orders from the communication interface 110.

[0070] In some embodiments, for each order of the plurality of orders, the communication interface 110 may receive information about one or more characteristics of the order. In some

embodiments, the processor 120 may receive the information about the one or more characteristics of the order from the communication interface 110. In some embodiments, the one or more characteristics of the order may include, but are not limited to, at least one of geographical characteristics of the order, temporal characteristics of the order, a category of the order, and characteristics of orders geographically relevant to the order.

[0071] In some embodiments, the geographical characteristics of the order may include a location relating to the order. For example, the location relating to the order may include, but is not limited to, a latitude and a longitude of a pick-up point and a drop-off point for the order, and an estimated distance of a trip for the order, and a delivery radius for the order. As an example, the distance of the trip may be estimated by the processor 120 before the delivery service provider 171a is allocated.

[0072] In some embodiments, the temporal characteristics of the order may include time of the order. For example, the time of the order may include, but is not limited to, at least one of a day of a week of the order, an hour of a day of the order, and a current delay time. As an example, the day of the week may be indicated as a range of 1 to 7, each representing Monday to Sunday. As an example, the hour of the day may be indicated as a range of 0 to 23 representing each hour. As an example, the current delay time (also referred to as a “calculated certain time”) may be a duration from a current time to a time of starting allocation of the delivery service provider 171a. For example, a total delay time may be predetermined in a service level agreement (SLA), and the current delay time may be updated every time after the order is created. For each order, in order to obtain/calculate the current delay time for the order, the processor 120 may look up a row from a predetermined delay time table to obtain a pre-computed threshold (also referred to as a “pre-computed score”). In some embodiments, the pre-computed threshold may be a value used to compare against the order’s allocation probability score (also referred to as an “APM (allocation probability model) score”), where

the order's allocation probability is below the pre-computed threshold, it may be deemed that the order is unlikely to get fulfilled and hence the JITA model is not activated for this order. This look-up may be done using HoD (hour of the day), DoW (day of the week) and a location of the order. In some embodiments, the current delay time (delayTime) may be calculated using at least one of an item preparation time (for example, a food preparation time), an elapsed time, a pick-up buffer, a pick-up ETA (estimated time of arrival) and an allocation buffer. The food preparation time (foodPrepTime) may be a time period for preparing the food. The elapsed time (elapsedTime) may be a time difference between a current time and a time for creating an order. The pick-up buffer (pickupBuffer) may be a buffer time for pick-up of the food by the delivery service provider. The pick-up ETA (pickupETA) may be an ETA of the pick-up of the food by the delivery service provider. The allocation buffer (allocationBuffer) may be a buffer time for allocation of the order to the delivery service provider. In some embodiments, the current delay time may be calculated as follows:

- $\text{delayTime} = \text{foodPrepTime} - \text{elapsedTime} - \text{pickupBuffer} - (\text{pickupETA} + \text{allocationBuffer})$
- foodPrepTime = by calling food ETA API
- elapsedTime = current time – created order time
- pickupBuffer = either using static value or Dynamic PickUp Buffer
- pickupETA = by calling food ETA API GetAllocationBufferTime
- allocationBuffer = by calling food ETA API GetPreBookingPickUpETA

In some embodiments, the pre-computed threshold may be used to decide whether to activate the JITA model for the order, while the calculation of the current delay time may be used to decide how long the order will be delayed if it is activated for the JITA model.

[0073] In some embodiments, the category of the order may include a type of the order. For example, the type of the order may include, but is not limited to, at least one of information

about whether the order is batched with at least one other order, and information about the user 161. For example, the processor 120 may prioritise the user 161 whose membership grade with the on-demand service provider platform is higher than a certain grade, who has made at least a certain number of orders (for example, 22 completed orders) in the last few days (for example, last 29 days), and/or who has spent at least a certain amount (for example, USD 220) for the on-demand service.

[0074] In some embodiments, the characteristics of orders geographically relevant to the order may include information about allocation health. In some embodiments, the allocation health may include, but is not limited to, at least one of real-time supply data, real-time demand data, past average supply data, and past average demand data.

[0075] In some embodiments, the real-time supply data may be a unified supply signal, which may indicate all delivery service providers 171 who are available at least once and accessible to a current geohash. In some embodiments, the real-time demand data may include a total demand in a last few minutes (for example, last 5 minutes) within the current geohash. For example, the real-time supply data and the real-time demand data may be updated periodically (for example, every 1 minute). As another example, the real-time supply data and the real-time demand data may be updated upon receipt of an input and/or instructions.

[0076] In some embodiments, the real-time supply data and the real-time demand data may include an allocation rate about orders which are geographically relevant to the order and received within a certain time slot. In some embodiments, the allocation rate about the orders which are geographically relevant to the order and received within the certain time slot may include a confirmed allocation rate (CAR) for the orders which are received within the certain time slot and made in a region same as that of the order. In some embodiments, the confirmed allocation rate (CAR) may give a real-time overview of allocation health (fulfilment) of a given space and time, by evaluating allocated orders with respect to outstanding orders. For example,

the confirmed allocation rate (CAR) may be calculated by dividing the number of the allocated orders into the number of all the orders which are geographically relevant to the order and received within the certain time slot.

[0077] In some embodiments, to obtain the confirmed allocation rate (CAR), the orders may be aggregated in real time, in the certain time slot, for example, in a 10-minute window. In some embodiments, the certain time slot may be changed (updated), for example, depending on an external environment, to improve accuracy of the confirmed allocation rate (CAR). In some embodiments, in a region of high population density with high order volume, the certain time slot may be lower, to reflect real-time data, in order to improve the accuracy of the confirmed allocation rate. In some other embodiments, in a region of low population density with low order volume, the certain time slot may be higher, to collect more order data, in order to improve the accuracy of the confirmed allocation rate (CAR).

[0078] In some embodiments, to obtain the confirmed allocation rate (CAR), the orders which are geographically relevant to the order may be aggregated. In some embodiments, a certain size-level area, for example, a city, a county or a province, may be divided into a plurality of regions depending on a predetermined granularity, for example, geohash5, geohash6, etc. In some embodiments, the orders made in the region same as that of the order may be aggregated. In some embodiments, the granularity may be changed (updated), for example, depending on an external environment, to improve accuracy of the confirmed allocation rate (CAR). In some embodiments, in a city of high population density with high order volume, the granularity may be higher, to collect more geographically relevant order data, in order to improve the accuracy of the confirmed allocation rate (CAR). In some other embodiments, in a county of low population density with low order volume, the granularity may be lower, to collect more order data, in order to improve the accuracy of the confirmed allocation rate (CAR). For example, in some areas of higher volume density, geohash6 may be used to obtain accurate confirmed

allocation rate (CAR), while in some areas geohash6 may not be good enough to obtain the accurate confirmed allocation rate (CAR) and geohash5 may be used (for example, in a combination of multiple geohash6).

[0079] In some embodiments, the past average supply data may be obtained from past supply data derived from the unified supply signal received during a certain time slot in the past (for example, 2 weeks). In some embodiments, the past supply data may be partitioned by the date of the week and/or the hour of the day. For example, the processor 120 may obtain a date in different weeks (for example, in 2 weeks). As an example, the processor 120 may fetch the past supply data from the unified supply signal received 7 days ago and 14 days ago respectively. For each date, the processor 120 may query the past supply data at a corresponding hour of the day (for example, 10 A.M.), and then calculate an average 1-minute supply data of 10 A.M. The processor 120 may then average all the 1-minute supply data in the different weeks. In some embodiments, the past average demand data may be obtained from past demand data derived from the number of orders (order counts) received during a certain time slot in the past (for example, 2 weeks). In some embodiments, the past demand data may be partitioned by the date of the week and/or the hour of the day. For example, the processor 120 may obtain a date in different weeks (for example, in 2 weeks). As an example, the processor 120 may fetch the past demand data from the unified supply signal received 7 days ago and 14 days ago respectively. For each date, the processor 120 may query the past demand data at a corresponding hour of the day (for example, 10 A.M.), and then calculate an average 1-minute demand data of 10 A.M. The processor 120 may then average all the 1-minute demand data in the different weeks.

[0080] In some embodiments, for each order of the plurality of orders, the processor 120 may obtain an indicator that indicates how allocable the order is, for example, using an allocation probability model (APM). In some embodiments, the indicator may be in the form of a score

(also referred to as the “APM (allocation probability model) score” or the “allocation probability score”). In some embodiments, the indicator may relate to a likelihood that the order will get allocated to a delivery service provider 171a (also referred to as an “allocability”). In some embodiments, the indicator for the order may be determined based on the information about the one or more characteristics of the order. In some embodiments, the processor 120 may input the information about the one or more characteristics of the order into the allocation probability model (APM), and obtain the indicator for the order, for example, in the form of the APM score for the order, as an output from the allocation probability model (APM). In some other embodiments, the allocation probability model (APM) may collect the information about the one or more characteristics of the order from one or more data sources, for example, the server 100, an external server, etc., and output the indicator for the order, for example, in the form of the APM score for the order.

[0081] In some embodiments, for each order of the plurality of orders, the processor 120 may compare the indicator for the order with a designated threshold (also referred to as a “designated APM threshold”). In some embodiments, the APM threshold may be determined by historical data where a threshold that leads to an acceptable trade-off between the number of delayed orders and expected compensation costs (i.e. failed orders) is selected. In some embodiments, the processor 120 may determine whether to delay allocating the order to the delivery service provider 171a for a certain time (for example, whether to apply the JITA model), based on the comparison of the indicator with the predetermined threshold.

[0082] In some embodiments, within each region, orders having the APM score which is greater than the designated APM threshold may be selected to apply the JITA model. In some embodiments, the processor 120 may use the designated APM threshold to select the at least one order to delay the allocation to the delivery service provider 171a (for example, to apply the JITA model). In some embodiments, the processor 120 may check if the indicator for an

order of the plurality of orders is greater than the designated APM threshold. In some embodiments, if the processor 120 determines that the indicator for the order is greater than the designated APM threshold, the processor 120 may select the order to delay the allocation to the delivery service provider 171a (for example, to apply the JITA model).

[0083] In some embodiments, for each order of the plurality of orders, the processor 120 may assign the order to a JITA enabled orders group (also referred to as a “first candidate group of orders”) if a certain criterion for the JITA enabled orders group is met. In some embodiments, for each order assigned to the first candidate group of orders, the processor 120 may assign the order to a JITA active orders group (also referred to as a “second candidate group of orders”) if a certain criterion for the JITA active orders group is met. In some embodiments, for each order assigned to the second candidate group of orders, the processor 120 may assign the order to a JITA effective orders group if a certain criterion for the JITA effective orders group is met. In some embodiments, the JITA model may be applied to the order assigned to the JITA effective orders group.

[0084] In some embodiments, for each order of the plurality of orders, the processor 120 may check whether an item provider 181a relating to the order has allowed a delay in allocating one or more received orders to a respective delivery service provider. In some embodiments, the processor 120 may determine whether to assign the order to the first candidate group of orders, based on whether the item provider 181a relating to the order has allowed the delay. For example, the processor 120 may check if the item provider 181a relating to the order has been whitelisted for the JITA model. In other words, the processor 120 may check if the item provider 181a relating to the order may be a JITA-enabled item provider. The processor 120 may assign all orders relating to JITA-enabled item providers to the first candidate group of orders. In some embodiments, the processor 120 may consider one or more factors, including, but not limited to, at least one of a ratio of item providers that may prepare the item ahead, and

identified item providers that may prepare the item ahead. For example, one of the consideration may be that the item provider (MEX) prepares the order ahead of time.

[0085] In some embodiments, for each order assigned to the first candidate group of orders, the processor 120 may compare the indicator with the designated APM threshold. In some embodiments, the processor 120 may determine whether to assign the order to the second candidate group of orders, based on the comparison of the indicator with the designated APM threshold. In some embodiments, the processor 120 may determine that the indicator is greater than the designated APM threshold, and determine to assign the order to the second candidate group of orders. For example, where the indicator is greater than the designated APM threshold, the processor 120 may determine to assign the order to the second candidate group of orders. In some embodiments, the processor 120 may assign orders to the second candidate group of orders, based on allocation conditions. In some embodiments, the processor 120 may consider one or more factors, including, but not limited to, at least one of a non-allocation and cancellation policy, and a compensation policy shifts for the item provider 181a.

[0086] In some embodiments, for each order assigned to the first candidate group of orders, in addition to checking if the indicator in the form of the APM score is greater than the designated APM threshold, the processor 120 may check if at least one other indicator for the order is greater than a respective threshold. For example, the processor 120 may further check if a CAR (confirmed allocation rate) score and/or a CRWM score for the order is greater than a CAR threshold and/or the CRWM threshold, respectively. For example, if all the APM score, the CAR score, and the CRWM score for the order are greater than the designated APM threshold, the CAR threshold, and the CRWM threshold respectively, the processor 120 may determine to assign the order to the second candidate group of orders.

[0087] In some embodiments, for each order assigned to the second candidate group of orders, the processor 120 may calculate the certain time for the delay, and determine whether to delay

allocating the order to the delivery service provider 171a (for example, whether to apply the JITA model), based on the calculated certain time.

[0088] In some embodiments, the processor 120 may check if an upfront batching for the order is enabled. In some embodiments, the processor 120 may determine that the upfront batching for the order is disabled, and determine to delay allocating the order to the delivery service provider 171a for the calculated certain time. For example, where the upfront batching for the order is disabled, the processor 120 may determine to apply the JITA model to the order. In some embodiments, the processor 120 may determine that the upfront batching for the order is enabled, and determine whether to delay allocating the order to the delivery service provider 171a, based on the calculated certain time. For example, where the upfront batching for the order is enabled, the processor 120 may determine whether to apply the JITA model to the order. In some embodiments, the processor 120 may obtain a maximum pooling time (also referred to as a “max pooling time”) (for example, fub (food upfront batching) max pooling time) which may be a configuration for the upfront batching, compare the calculated certain time with the maximum pooling time, and determine whether to delay allocating the order to the delivery service provider 171a, based on the comparison of the calculated certain time with the maximum pooling time. In some embodiments, the processor 120 may determine that the calculated certain time is longer than the maximum pooling time, and determine to delay allocating the order to the delivery service provider 171a for the calculated certain time. For example, where the calculated certain time is longer than the maximum pooling time, the processor 120 may determine to apply the JITA model to the order for the calculated certain time.

[0089] In this manner, in case the upfront batching for the order is enabled, if the calculated certain time is longer than the maximum pooling time, the processor 120 may delay an allocation start time for the order to start the allocation of the delivery service provider 171a

later than the max pooling time. In case the upfront batching is disabled, the processor 120 may delay the allocation start time for the order, to start the allocation of the delivery service provider 171a later than the calculated certain time (which is longer than 0). In some embodiments, the processor 120 may consider one or more factors, including, but not limited to, at least one of FPT (food preparation time) inaccuracy, a JITA delay duration/accuracy, and a distance of the delivery service provider 171a.

[0090] In some embodiments, the processor 120 may, for the selected order to apply the JITA model, use the JITA model in allocation to the delivery service provider 171a. In some embodiments, the processor 120 may, for the selected order to apply the JITA model, delay allocation to the delivery service provider 171a for the calculated certain time. The delay of the allocation to the delivery service provider 171a may allow the item provider 181a to have more time to prepare an item for the selected order. As shown in FIG. 4, the processor 120 may delay allocating the selected order to the delivery service provider 171a for the calculated certain time t . Advantageously, the item provider 181a may start preparing the item before the delivery service provider 171a physically arrives at the location of the item provider 181a. The item provider 181a may have more time (i.e. the calculated certain time t) to prepare the item for the selected order which is highly likely to get allocated to the delivery service provider 171a.

[0091] In some embodiments, the processor 120 may, for non-selected orders from the plurality of orders, use the normal allocation model in the allocation to the delivery service provider 171a. In some embodiments, the processor 120 may, for non-selected orders from the plurality of orders, activate the normal allocation model in the allocation to the delivery service provider 171a. In some embodiments, the processor 120 may, for the non-selected orders, notify the item provider 181a to prepare the item after allocating the order to the delivery service provider 171a. Advantageously, the item provider 181a may prepare the item for the selected order

which is unlikely to get allocated to the delivery service provider 171a, only after the delivery service provider 171a is allocated.

[0092] As described above, the various embodiments may use the allocation probability model (APM) which may predict the allocability of each order and respond to market conditions real-time, instead of solely relying on the confirmed allocation rate (CAR) which is backward looking, to select orders for using the JITA model. The use of the allocation probability model (APM) may predict the allocation likelihood per order with higher accuracy, and thus allow to programmatically pick orders that will most likely be fulfilled. The various embodiments may allow to decide whether the JITA model is applied to the order, considering compensation costs and associated item providers' and users' experience downside if this order should not get allocation (i.e. false positive by APM prediction). Advantageously, the various embodiment may improve JITA active rate and reduce driver wait time (DWT), since the allocation probability model (APM) may allow the JITA model to be remaining active during periods of critical and/or low confirmed allocation rate (CAR) (for example, poor weather conditions, festive periods, peak hours, etc.). In addition, the various embodiments may reduce compensation costs due to a non-allocation of the delivery service provider 171a under the JITA model. Further, the various embodiments may reduce the disappointment to the item provider 181a (where they see the order they prepared are not delivered) and the user 161 (where after a long time their orders placed are cancelled).

[0093] FIG. 5 illustrates a flowchart for a method S300 for facilitating processing a plurality of orders for an on-demand service according to various embodiments. According to various embodiments, the method S300 for facilitating processing the plurality of orders for the on-demand service may be provided.

[0094] In some embodiments, the method S300 may include a step S301 of receiving a plurality of orders for the on-demand service from a plurality of computing devices.

[0095] In some embodiments, the method S300 may include a step S302 of, for each order of the plurality of orders, receiving information about one or more characteristics of the order.

[0096] In some embodiments, the method S300 may include a step S303 of, for each order of the plurality of orders, obtaining an indicator relating to a likelihood that the order will get allocated to a delivery service provider, based on the information about the one or more characteristics of the order.

[0097] In some embodiments, the method S300 may include a step S304 of, for each order of the plurality of orders, comparing the indicator with a designated threshold.

[0098] In some embodiments, the method S300 may include a step S305 of, for each order of the plurality of orders, determining whether to delay allocating the order to the delivery service provider for a certain time, based on the comparison of the indicator with the designated threshold.

[0099] FIG. 6 illustrates an exemplary diagram showing a JITA enabled orders group, a JITA active orders group, and a JITA effective orders group according to various embodiments.

[00100] In some embodiments, a processor 120 of a server 100 (as described with reference to FIG. 3) may assign the order to a JITA enabled orders group 401 (also referred to as a “first candidate group of orders”) if a certain criterion for the JITA enabled orders group 401 is met. In some embodiments, for each order assigned to the JITA enabled orders group 401, the processor 120 may assign the order to a JITA active orders group 402 (also referred to as a “second candidate group of orders”) if a certain criterion for the JITA active orders group 402 is met. In some embodiments, for each order assigned to the JITA active orders group 402, the processor 120 may assign the order to a JITA effective orders group 403 if a certain criterion for the JITA effective orders group 403 is met. In some embodiments, the JITA model may be applied to the order assigned to the JITA effective orders group 403.

[00101] In some embodiments, for each order of the plurality of orders, the processor 120 may check if the item provider 181a relating to the order has been whitelisted for the JITA model. In other words, the processor 120 may check if the item provider 181a relating to the order may be a JITA-enabled item provider. The processor 120 may assign all orders relating to JITA-enabled item providers to the JITA enabled orders group 401. In some embodiments, the processor 120 may consider one or more factors, including, but not limited to, at least one of a ratio of item providers that may prepare the item ahead, and identified item providers that may prepare the item ahead.

[00102] In some embodiments, for each order assigned to the JITA enabled orders group 401, the processor 120 may determine to assign the order to the JITA active orders group 402, based on allocation conditions. For example, for each order assigned to the JITA enabled orders group 401, where an indicator for the order, for example, in the form of an APM score, is greater than a designated APM threshold, the processor 120 may determine to assign the order to the JITA active orders group 402. In some embodiments, the processor 120 may consider one or more factors, including, but not limited to, at least one of a non-allocation and cancellation policy, and a compensation policy shifts for the item provider 181a.

[00103] In some embodiments, for each order assigned to the JITA enabled orders group 401, in addition to checking if the indicator in the form of the APM score is greater than the designated APM threshold, the processor 120 may check if at least one other indicator for the order is greater than a respective threshold. For example, the processor 120 may further check if a CAR (confirmed allocation rate) score and/or a CRWM score for the order is greater than a CAR threshold and/or the CRWM threshold, respectively. For example, if all the APM score, the CAR score, and the CRWM score for the order are greater than the designated APM threshold, the CAR threshold, and the CRWM threshold respectively, the processor 120 may determine to assign the order to the JITA active orders group 402.

[00104] In some embodiments, for each order assigned to the JITA active orders group 402, the processor 120 may calculate the certain time for the delay, and determine whether to delay allocating the order to the delivery service provider 171a (for example, whether to apply the JITA model), based on the calculated certain time. In some embodiments, for each order assigned to the JITA active orders group 402, the processor 120 may further check if an upfront batching for the order is enabled.

[00105] In some embodiments, for each order assigned to the JITA active orders group 402, in case the upfront batching for the order is enabled, if the calculated certain time is longer than a maximum pooling time, the processor 120 may apply the JITA model, and the allocation of the delivery service provider 171a for the order may be delayed for the maximum pooling time. In case the upfront batching is disabled, the processor 120 may apply the JITA model, and the allocation of the delivery service provider 171a for the order may be delayed for the calculated certain time (which is longer than 0). In some embodiments, the processor 120 may consider one or more factors, including, but not limited to, at least one of FPT inaccuracy, a JITA delay duration or accuracy, and a distance of the delivery service provider 171a.

[00106] FIG. 7 illustrates a data flow diagram of an allocation probability model (APM) according to various embodiments.

[00107] As shown in FIG. 7, the allocation probability model (APM) may receive information about one or more characteristics of the order, as an input. In some embodiments, the one or more characteristics of the order may include, but are not limited to, at least one of geographical characteristics of the order, temporal characteristics of the order, a category of the order, and characteristics of orders geographically relevant to the order.

[00108] In some embodiments, the geographical characteristics of the order may include a location relating to the order. For example, the location relating to the order may include, but is not limited to, a latitude and a longitude of a pick-up point and a drop-off point for the order,

and an estimated distance of a trip for the order, and a delivery radius for the order. In some embodiments, information (features) about the latitude and the longitude of the pick-up point and the drop-off point for the order, and the estimated distance of the trip for the order, and the delivery radius for the order may be included as a part of an API (application program interface) request for the order sent to the allocation probability model (APM).

[00109] In some embodiments, the temporal characteristics of the order may include time of the order. For example, the time of the order may include, but is not limited to, a day of a week of the order, an hour of a day of the order, and a current delay time. As an example, the current delay time (also referred to as a “calculated certain time”) may be a duration from a current time to a time of starting allocation of the delivery service provider 171a. In some embodiments, information (features) about the day of the week of the order, the hour of the day of the order, and the current delay time may be included as a part of the API request for the order sent to the allocation probability model (APM).

[00110] In some embodiments, the category of the order may include a type of the order. For example, the type of the order may include, but is not limited to, at least one of information about whether the order is batched with at least one other order, and information about the user 161. In some embodiments, the information about whether the order is batched with at least one other order, and the information about the user may be included as a part of the API request for the order sent to the allocation probability model (APM).

[00111] In some embodiments, the characteristics of orders geographically relevant to the order may include information about allocation health. In some embodiments, the allocation health may include, but is not limited to, at least one of real-time supply data, real-time demand data, past average supply data, and past average demand data. In some embodiments, the past average supply data and the past average demand data may be computed in batch jobs (for example, a Spark module 501) and saved in Amphawa offline store 508. In some embodiments,

the real-time supply data and the real-time demand data may be computed in streams (for example, a Flink module 505 for streaming some online signal like real time supply and demand signal, and a Kafka module 506) and saved in Amphawa real-time store 508.

[00112] As shown in FIG. 7, in some embodiments, the Spark module 501 may receive historical data relating to orders, and compute the past average supply data and the past average demand data. In some embodiments, the past average supply data and the past average demand data may be saved in an Amphawa offline store 508 which is used to store the aggregated supply demand signal. In some embodiments, the past average supply data and the past average demand data may be saved in a datalake 502 which is used to store the aggregated supply demand signal. In some embodiments, a Catwalk module 509, which is a platform where a latest allocation probability model (APM) is run, may read the past average supply data and the past average demand data. In some embodiments, the Flink module 505 may receive raw real-time signal, and the Flink module 505 and the Kafka module 506 may compute the real-time supply data and the real-time demand data. In some embodiments, the real-time supply data and the real-time demand data may be saved in the Amphawa real-time store 508. In some embodiments, the Catwalk module 509 may read the real-time supply data and the real-time demand data. In some embodiments, the past average supply data and the past average demand data may be used to train the allocation probability model (APM) by a Tensorflow model training module 503, and the trained model may be saved in a Catwalk model bank 504. The latest model may be input to the Catwalk module 509. In some embodiments, an orchestrator 510 may be an integrated model serving service (including creating API, some simple data ETL etc.). For example, the simple data ETL may refer to simple data extraction, transformation and loading (ETL). In some embodiments, an engineering backend 511 may be a food backend. In some embodiments, an API Caller 512 may be an API created by the orchestrator 510.

[00113] FIGS. 8, 9 and 10 illustrate exemplary diagrams showing a user interface of a computing device 180a of an item provider 181a for an order according to various embodiments.

[00114] As shown in FIG. 8, in some embodiments, where an order (hereinafter, referred to as a “first order”) for an on-demand service is selected to use the JITA model, a processor 120 of a server 100 (as described with reference to FIG. 3) may notify an item provider 181a to prepare an item for the first order, before allocating the first order to a delivery service provider 171a. The processor 120 may provide a user interface of the computing device 180a of the item provider 181a with an option, for example, by displaying an image object, to check “order is ready” corresponding to the first order. Once the item is prepared and ready for collection, the item provider 181a may select the image object to check “order is ready”, to inform the processor 120 that the item is ready for collection. Thereafter, for example, the processor 120 may inform the allocated delivery service provider 171a to rush to a location of the item provider 181a to collect the item. As another example, if the delivery service provider 171a is still not allocated, the processor 120 may prioritise finding the delivery service provider 171a to allocate to the first order.

[00115] As shown in FIG. 9, where an order (hereinafter, referred to as a “second order”) for an on-demand service is selected to use the JITA model, the processor 120 may notify the item provider 181a to prepare an item for the second order, before allocating the second order to a delivery service provider 171a. In some embodiments, once the processor 120 receives the second order which is selected to use the JITA model, details of the second order may be displayed under a “preparing” tab, so that the item provider 181a may start preparing the item for the second order. In some embodiments, thereafter, when an estimated time of arrival (ETA) of the allocated delivery service provider 171a is earlier than a predetermined time, for

example, 3 minutes, identity information, for example, a name, of the allocated delivery service provider 171a may be displayed on the computing device 180a of the item provider 181a.

[00116] As shown in FIG. 10, where an order (hereinafter, referred to as a “third order”) for an on-demand service is not selected to use the JITA model, the processor 120 may notify the item provider 181a to prepare an item for the third order, after allocating the third order to a delivery service provider 171a. In some embodiments, once the processor 120 receives the third order which is selected not to use the JITA model, details of the third order may be displayed under a “upcoming” tab. After the delivery service provider 171a is allocated, the details of the third order may be displayed under the “preparing” tab, so that the item provider 181a may start preparing the item for the third order.

[00117] As described above, according to various embodiments, rather than using only allocation health (for example, a confirmed allocation rate (CAR)) as is the current practice, a signal to indicate the likelihood that an order will get allocated may be used. Therefore, the various embodiments may allow to avoid or minimise delaying a delivery for orders that have a high likelihood of allocation and do not exceed network-level capacity as indicated by the allocation health. Advantageously, the various embodiment may improve JITA active rate and reduce driver wait time (DWT), since the allocation probability model (APM) may allow the JITA model to be remaining active during periods of critical and/or low confirmed allocation rate (CAR). In addition, the various embodiments may result in less compensation costs (i.e. less false positives) due to a non-allocation of the delivery service provider under the JITA model.

[00118] While the disclosure has been particularly shown and described with reference to specific embodiments, it should be understood by those skilled in the art that various changes in form and detail may be made therein without departing from the spirit and scope of the invention as defined by the appended claims. The scope of the invention is thus indicated by

the appended claims and all changes which come within the meaning and range of equivalency of the claims are therefore intended to be embraced.

CLAIMS

1. A server for facilitating processing a plurality of orders for an on-demand service, the server comprising:

a memory configured to store instructions;

a communication interface configured to receive the plurality of orders for the on-demand service from a plurality of computing devices; and

a processor for executing the stored instructions and configured to: for each order of the plurality of orders,

receive information about one or more characteristics of the order;

obtain an indicator relating to a likelihood that the order will get allocated to a delivery service provider, based on the information about the one or more characteristics of the order;

compare the indicator with a designated threshold; and

determine whether to delay allocating the order to the delivery service provider for a certain time, based on the comparison of the indicator with the designated threshold.

2. The server according to claim 1, wherein the processor is further configured to: for each order of the plurality of orders,

check whether an item provider relating to the order has allowed a delay in allocating one or more received orders to a respective delivery service provider; and

determine whether to assign the order to a first candidate group of orders, based on whether the item provider relating to the order has allowed the delay.

3. The server according to claim 2, wherein the processor is further configured to: for each order assigned to the first candidate group of orders,

compare the indicator with the designated threshold; and

determine whether to assign the order to a second candidate group of orders, based on the comparison of the indicator with the designated threshold.

4. The server according to claim 3, wherein the processor is further configured to: for each order assigned to the first candidate group of orders,

determine that the indicator is greater than the designated threshold; and

determine to assign the order to the second candidate group of orders.

5. The server according to claim 3 or claim 4, wherein the processor is further configured to: for each order assigned to the second candidate group of orders,

calculate the certain time for the delay, based on at least one of an item preparation time, an elapsed time, a pick-up buffer, a pick-up ETA (estimated time of arrival) and an allocation buffer; and

determine whether to delay allocating the order to the delivery service provider, based on the calculated certain time.

6. The server according to claim 5, wherein the processor is further configured to: for each order assigned to the second candidate group of orders,

check if an upfront batching for the order is enabled;

determine that the upfront batching for the order is enabled; and

determine whether to delay allocating the order to the delivery service provider, based on the calculated certain time.

7. The server according to claim 6, wherein the processor is further configured to: for each order assigned to the second candidate group of orders,
- obtain a maximum pooling time for the upfront batching;
 - compare the calculated certain time with the maximum pooling time; and
 - determine whether to delay allocating the order to the delivery service provider, based on the comparison of the calculated certain time with the maximum pooling time.
8. The server according to claim 7, wherein the processor is further configured to: for each order assigned to the second candidate group of orders,
- determine that the calculated certain time is longer than the maximum pooling time;
- and
- determine to delay allocating the order to the delivery service provider for the calculated certain time.
9. The server according to any one of claims 6 to 8, wherein the processor is further configured to: for each order assigned to the second candidate group of orders,
- determine that the upfront batching for the order is disabled; and
 - determine to delay allocating the order to the delivery service provider for the calculated certain time.
10. The server according to any one of claims 1 to 9, wherein the one or more characteristics of the order include at least one of geographical characteristics of the order, temporal characteristics of the order, a category of the order, and characteristics of orders geographically relevant to the order.

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

receiving the plurality of orders for the on-demand service from a plurality of computing devices;

for each order of the plurality of orders,

receiving information about one or more characteristics of the order;

obtaining an indicator relating to a likelihood that the order will get allocated to a delivery service provider, based on the information about the one or more characteristics of the order;

comparing the indicator with a designated threshold; and

determining whether to delay allocating the order to the delivery service provider for a certain time, based on the comparison of the indicator with the designated threshold.

12. The method according to claim 11 further comprising: for each order of the plurality of orders,

checking whether an item provider relating to the order has allowed a delay in allocating one or more received orders to a respective delivery service provider; and

determining whether to assign the order to a first candidate group of orders, based on whether the item provider relating to the order has allowed the delay.

13. The method according to claim 12 further comprising: for each order assigned to the first candidate group of orders,

comparing the indicator with the designated threshold; and

determining whether to assign the order to a second candidate group of orders, based on the comparison of the indicator with the designated threshold.

14. The method according to claim 13 further comprising: for each order assigned to the first candidate group of orders,

determining that the indicator is greater than the designated threshold; and

determining to assign the order to the second candidate group of orders.

15. The method according to claim 13 or claim 14 further comprising: for each order assigned to the second candidate group of orders,

calculating the certain time for the delay, based on at least one of an item preparation time, an elapsed time, a pick-up buffer, a pick-up ETA (estimated time of arrival) and an allocation buffer; and

determining whether to delay allocating the order to the delivery service provider, based on the calculated certain time.

16. The method according to claim 15 further comprising: for each order assigned to the second candidate group of orders,

checking if an upfront batching for the order is enabled;

determining that the upfront batching for the order is enabled; and

determining whether to delay allocating the order to the delivery service provider, based on the calculated certain time.

17. The method according to claim 16 further comprising: for each order assigned to the second candidate group of orders,

obtaining a maximum pooling time for the upfront batching;

comparing the calculated certain time with the maximum pooling time; and

determining whether to delay allocating the order to the delivery service provider, based on the comparison of the calculated certain time with the maximum pooling time.

18. The method according to claim 17 further comprising: for each order assigned to the second candidate group of orders,

determining that the calculated certain time is longer than the maximum pooling time;
and

determining to delay allocating the order to the delivery service provider for the calculated certain time.

19. The method according to any one of claims 16 to 18 further comprising: for each order assigned to the second candidate group of orders,

determining that the upfront batching for the order is disabled; and

determining to delay allocating the order to the delivery service provider for the calculated certain time.

20. The method according to any one of claims 11 to 19, wherein the one or more characteristics of the order include at least one of geographical characteristics of the order, temporal characteristics of the order, a category of the order, and characteristics of orders geographically relevant to the order.