We Claim:

1. A processor implemented method (200) for minimizing passenger misconnects in airline operations using learning, the method comprising:

identifying (202), by one or more hardware processors, a plurality of Reinforcement Learning (RL) agents implemented by the one or more hardware processors, wherein each of the plurality of RL agents corresponds to a flight among a plurality of flights operated by an airline in a training environment, wherein the training environment is one of a simulated airline network generated for the airline operations and a real world airline operations network, and wherein the plurality of RL agents work in parallel in the training environment on a common simulation timeline;

initializing (204), by the one or more hardware processors, each of the plurality of RL agents with a random seed, and a learning rate (ε) with a gradual decay, wherein the random seed creates initial conditions for learned knowledge representation and the learning rate (ε) controls rate of exploration versus exploitation of each of the plurality of RL agents; and

training (206), by the one or more hardware processors, each of the plurality of RL agents, wherein the training is an iterative process and iterations are terminated when each of the plurality of RL agents consistently reaches an optimal value of a Quality (Q) metric indicating quality of the training, wherein the optimal value of Q is identified by a non-changing convergence among the iterations, and wherein steps in each iteration comprise:

a) setting (206a) random values for weights of each of the plurality of RL agents in accordance with the initial random seed;

b) determining (206b), by a Machine Learning (ML) based context engine implemented by the one or more hardware processors, a state of each of the plurality of RL agents defining a current state of the flight corresponding to each of the plurality of RL agents, wherein a plurality

of features defining the state of each of the plurality of RL agents comprise:

a local passenger utility vector ($\vec{P_L}$) representing a local utility, experienced by passengers of the flight, for every hold time identified for the flight;

a local airline utility vector ($\vec{A_L}$) representing the local utility, experienced by the airline operating the flight, for every hold time identified for the flight;

an average passenger global utility scalar ($P_G$) representing an average global utility experienced by the passengers of all the flights for a past predefined time window;

an average global airline utility scalar ($A_G$) representing an average global utility experienced by the airline for the past predefined time window; and

a hold time guidance value $\tau^*$ indicating a guideline for computing a hold time $\tau$ for the flight, wherein $\tau^*$ is derived from flight level data which consists of the local passenger utility vector ($\vec{P_L}$), the local airline utility vector ($\vec{A_L}$) and a business constant ($\alpha$) computed specific to the airline, and wherein α enables adjusting impact $\vec{P_L}$ and $\vec{A_L}$, on a local reward ($R_L$);

c) recommending (206c), by each of the plurality of RL agents the hold time $\tau$ for the flight corresponding to each of the plurality of the RL agents, wherein the hold time $\tau$ is computed by processing values of the plurality of features of the current state generated for each of the plurality of RL agents in accordance to the weights of each RL agent and underlying neural schema of each of the plurality of RL agents;

d) applying (206d), in the training environment, the hold time $\tau$ estimated by each of the plurality of RL agents to the flight corresponding to each of the plurality of RL agents to run a first round of the airline operations in the training environment for the plurality of flights in accordance to the hold time $\tau$ estimated by each of the plurality of RL agents;

e) regenerating (206e), by the ML based context engine a revised state to be communicated to each of the plurality of RL agents for a successive round of the airline operations for the flight corresponding to each of the plurality of RL agents;

f) receiving (206f) a reward ($R_{t-1}$) by each of the plurality of RL agents from a reward engine for the hold time $\tau$ derived from a local reward $R_L$ and a global reward $R_G$ and a business constant $\beta$ precomputed specific to the airline;

g) revising (206g) the values of the weights of each of the plurality of RL agents based on the received reward ($R_{t-1}$); and

h) recommending (206h), by each of the plurality of RL agents a revised hold time for the flight corresponding to each of the plurality of RL agents for the successive round of the airline operations for minimizing the passenger misconnects of one or more flights, wherein the revised hold time is computed by processing the revised values of the plurality of features of the revised state regenerated for each of the plurality of RL agents in accordance with the revised values of the weights of each of the plurality of RL agents,

wherein the ML based context engine is pretrained to generate the state for the flight corresponding to each of the plurality of RL agents using a plurality of flight parameters affecting the hold time τ of the flight to minimize the passenger misconnects, wherein the pretraining comprises:

applying regression techniques on the plurality of flight parameters comprising passenger (Pax) specific Minimum Connection Time (Pax specific MCT) based on Gate to Gate (G2G) score prediction, crew availability, and tail health denoting health of a physical flight along with dimensionality reduction technique;

applying clustering techniques for Pax profiling; and

applying transformations to:

real world data concerning airline operations and passenger itineraries; and

current location and status of the passenger between gates in an airport derived from passenger history.

2. The method as claimed in claim 1, wherein for the reward $(R_{t-1})$ the business constant *(β)* enables adjusting balance of the reward towards local or global optima, and wherein

the local reward ($R_L$) comprises an actual local passenger utility ($\hat{P}_L$) component and an actual local airline utility ($\hat{A}_L$) combined using a business constant (α); and

the global reward ($R_G$) comprises an average of all the local rewards ($R_{Ls}$) of the plurality of flights operated in last round.

3. The method as claimed in claim 1, wherein learnt knowledge of the plurality of trained RL agents is transferred to a new RL agent associated with a new flight entering the airline as the state defined for any RL agent is portable across the plurality of flights, and wherein training of the new RL agent starts from the learnt knowledge.

4. The method as claimed in claim 1, wherein a single RL agent is used for the plurality of flights as the state defined for any RL agent is portable across the plurality of flights.

5. The method as claimed in claim 1 further comprising utilizing the trained plurality of RL agents in real time airline operations to determine the hold time $\tau$ of the flight corresponding to each of the plurality of RL agents in real time.

6. A system (100) for minimizing passenger misconnects in airline operations using learning, the system (100) comprising:

a memory (102) storing instructions;

one or more Input/Output (I/O) interfaces (106); and

one or more hardware processors (104) coupled to the memory (102) via the one or more I/O interfaces (106), wherein the one or more hardware processors (104) are configured by the instructions to:

identify a plurality of Reinforcement Learning (RL) agents implemented by the one or more hardware processors (104), wherein each of the plurality of RL agents corresponds to a flight among a plurality of flights operated by an airline in a training environment, wherein the training environment is one of a simulated airline network generated for the airline operations and a real world airline operations network, and wherein the plurality of RL agents work in parallel in the training environment on a common simulation timeline;

initialize each of the plurality of RL agents with a random seed, and a learning rate (ε) with a gradual decay, wherein the random seed creates initial conditions for learned knowledge representation and the learning rate (ε) controls rate of exploration versus exploitation of each of the plurality of RL agents; and

train each of the plurality of RL agents, wherein the training is an iterative process and iterations are terminated when each of the plurality of RL agents consistently reaches an optimal value of a Quality (Q) metric indicating quality of the training, wherein the optimal value of Q is identified by a non-

changing convergence among the iterations, and wherein steps in each iteration comprise:

a) setting random values for weights of each of the plurality of RL agents in accordance with the initial random seed;

b) determining by a Machine Learning (ML) based context engine implemented by the one or more hardware processors (104), a state of each of the plurality of RL agents defining a current state of the flight corresponding to each of the plurality of RL agents, wherein a plurality of features defining the state of each of the plurality of RL agents comprising:

a local passenger utility vector ($\vec{P_L}$) representing a local utility, experienced by passengers of the flight, for every hold time identified for the flight

a local airline utility vector ($\vec{A_L}$) representing the local utility, experienced by the airline operating the flight, for every hold time identified for the flight;

an average passenger global utility scalar($P_G$) representing an average global utility experienced by the passengers of all the flights for a past predefined time window;

an average global airline utility scalar ($A_G$) representing average global utility experienced by the airline for the past predefined time window; and

a hold time guidance value $\tau^*$ indicating a guideline for computing a hold time $\tau$ for the flight, wherein $\tau^*$ is derived from flight level data which consists of the local passenger utility vector ($\vec{P_L}$), the local airline utility vector ($\vec{A_L}$) and a business constant ($\alpha$) computed specific to the airline, and

wherein α enables adjusting impact $\vec{P}_L$ and $\vec{A}_L$, on a local reward ($R_L$);

c) recommending by each of the plurality of RL agents the hold time $\tau$ for the flight corresponding to each of the plurality of RL agents, wherein the hold time $\tau$ is computed by processing values of the plurality of features of the current state generated for each of the plurality of RL agents in accordance to the weights of each RL agent and underlying neural schema of each of the plurality of RL agents;

d) applying in the training environment, the hold time $\tau$ estimated by each of the plurality of RL agent to the flight corresponding to each of the plurality of RL agents to run a first round of the airline operations in the training environment for the plurality of flights in accordance to the hold time $\tau$ estimated by each of the plurality of RL agents;

e) regenerating by the ML based context engine a revised state to be communicated to each of the plurality of RL agents for a successive round of the airline operations for the flight corresponding to each of the plurality of RL agents;

f) receiving a reward ($R_{t-1}$) by each of the plurality of RL agents from a reward engine for the hold time $\tau$ derived from a local reward $R_L$and a global reward $R_G$ and a business constant $\beta$ precomputed specific to the airline;

g) revising the values of the weights of each of the plurality of RL agents based on the received reward ($R_{t-1}$); and

h) recommending by each of the plurality of RL agents a revised hold time for the flight corresponding to each of the plurality of RL agents for the successive round of the airline operations for minimizing the passenger misconnects of one or more flights, wherein the revised hold time is computed by processing the revised values of the plurality of features of

the revised state regenerated for each of the plurality of RL agents in accordance with the revised values of the weights of each of the plurality of RL agents,

wherein the ML based context engine is pretrained to generate the state for the flight corresponding to each of the plurality of RL agents using a plurality of flight parameters affecting the hold time τ of the flight to minimize the passenger misconnects, wherein the pretraining comprises:

applying regression techniques on the plurality of flight parameters comprising passenger (Pax) specific Minimum Connection Time (Pax specific MCT) based on Gate to Gate (G2G) score prediction, crew availability, and tail health denoting health of a physical flight along with dimensionality reduction technique;

applying clustering techniques for Pax profiling; and

applying transformations to:

real world data concerning airline operations and passenger itineraries; and

current location and status of the passenger between gates in an airport derived from passenger history.

7. The system (100) as claimed in claim 6, wherein for the reward $(R_{t-1})$, the business constant *(β)* enables adjusting balance of the reward towards local or global optima, and wherein

the local reward ($R_L$) comprises an actual local passenger utility($\hat{P}_L$) component and an actual local airline utility ($\hat{A}_L$) combined using a business constant (α); and

the global reward ($R_G$) comprises an average of all the local rewards ($R_{Ls}$) of the plurality of flights operated in last round.

8. The system (100) as claimed in claim 6, wherein learnt knowledge of the plurality of trained RL agents is transferred to a new RL agent associated with

a new flight entering the airline as the state defined for any RL agent is portable across the plurality of flights, and wherein training of the new RL agent starts from the pre-acquired learnt knowledge.

9. The system (100) as claimed in claim 6, wherein a single RL agent is used for the plurality of flights as the state defined for any RL agent is portable across the plurality of flights.

10. The system as claimed in claim 6 further comprising utilizing the trained plurality of RL agents in real time airline operations to determine the hold time $\tau$ of the flight, corresponding to each of the plurality of RL agents, in real time.

11. One or more non-transitory machine-readable information storage mediums comprising one or more instructions, which when executed by one or more hardware processors causes:

identifying a plurality of Reinforcement Learning (RL) agents implemented by the one or more hardware processors, wherein each of the plurality of RL agents corresponds to a flight among a plurality of flights operated by an airline in a training environment, wherein the training environment is one of a simulated airline network generated for the airline operations and a real world airline operations network, and wherein the plurality of RL agents work in parallel in the training environment on a common simulation timeline;

initializing each of the plurality of RL agents with a random seed, and a learning rate (ε) with a gradual decay, wherein the random seed creates initial conditions for learned knowledge representation and the learning rate (ε) controls rate of exploration versus exploitation of each of the plurality of RL agents; and

training each of the plurality of RL agents, wherein the training is an iterative process and iterations are terminated when each of the plurality of RL

agents consistently reaches an optimal value of a Quality (Q) metric indicating quality of the training, wherein the optimal value of Q is identified by a non-changing convergence among the iterations, and wherein steps in each iteration comprise:

a) setting random values for weights of each of the plurality of RL agents in accordance with the initial random seed;

b) determining by a Machine Learning (ML) based context engine implemented by the one or more hardware processors, a state of each of the plurality of RL agents defining a current state of the flight corresponding to each of the plurality of RL agents, wherein a plurality of features defining the state of each of the plurality of RL agents comprise:

a local passenger utility vector ($\vec{P}_L$) representing a local utility, experienced by passengers of the flight, for every hold time identified for the flight;

a local airline utility vector ($\vec{A}_L$) representing the local utility, experienced by the airline operating the flight, for every hold time identified for the flight;

an average passenger global utility scalar ($P_G$) representing an average global utility experienced by the passengers of all the flights for a past predefined time window;

an average global airline utility scalar ($A_G$) representing an average global utility experienced by the airline for the past predefined time window; and

a hold time guidance value $\tau^*$ indicating a guideline for computing a hold time $\tau$ for the flight, wherein $\tau^*$ is derived from flight level data which consists of the local passenger utility vector ($\vec{P}_L$), the local airline utility vector ($\vec{A}_L$) and a

business constant ($\alpha$) computed specific to the airline, and wherein α enables adjusting impact $\vec{P}_L$ and $\vec{A}_L$, on a local reward ($R_L$);

c) recommending by each of the plurality of RL agents the hold time $\tau$ for the flight corresponding to each of the plurality of the RL agents, wherein the hold time $\tau$ is computed by processing values of the plurality of features of the current state generated for each of the plurality of RL agents in accordance to the weights of each RL agent and underlying neural schema of each of the plurality of RL agents;

d) applying in the training environment, the hold time $\tau$ estimated by each of the plurality of RL agents to the flight corresponding to each of the plurality of RL agents to run a first round of the airline operations in the training environment for the plurality of flights in accordance to the hold time $\tau$ estimated by each of the plurality of RL agents;

e) regenerating, by the ML based context engine a revised state to be communicated to each of the plurality of RL agents for a successive round of the airline operations for the flight corresponding to each of the plurality of RL agents;

f) receiving a reward ($R_{t-1}$) by each of the plurality of RL agents from a reward engine for the hold time $\tau$ derived from a local reward $R_L$ and a global reward $R_G$ and a business constant $\beta$ precomputed specific to the airline;

g) revising the values of the weights of each of the plurality of RL agents based on the received reward ($R_{t-1}$); and

h) recommending by each of the plurality of RL agents a revised hold time for the flight corresponding to each of the plurality of RL agents for the successive round of the airline operations for minimizing the passenger misconnects of one or more flights, wherein the revised hold time is

computed by processing the revised values of the plurality of features of the revised state regenerated for each of the plurality of RL agents in accordance with the revised values of the weights of each of the plurality of RL agents,

wherein the ML based context engine is pretrained to generate the state for the flight corresponding to each of the plurality of RL agents using a plurality of flight parameters affecting the hold time τ of the flight to minimize the passenger misconnects, wherein the pretraining comprises:

applying regression techniques on the plurality of flight parameters comprising passenger (Pax) specific Minimum Connection Time (Pax specific MCT) based on Gate to Gate (G2G) score prediction, crew availability, and tail health denoting health of a physical flight along with dimensionality reduction technique;

applying clustering techniques for Pax profiling; and

applying transformations to:

real world data concerning airline operations and passenger itineraries; and

current location and status of the passenger between gates in an airport derived from passenger history.