

Received November 19, 2019, accepted November 30, 2019, date of publication December 5, 2019, date of current version December 26, 2019.

Digital Object Identifier 10.1109/ACCESS.2019.2957874

# A Multi-Agent Approach for Reactionary Delay Prediction of Flights

YASH GULERIA<sup>1</sup>, QING CAI<sup>1</sup>, SAMEER ALAM<sup>1</sup>, AND LISHUAI LI<sup>2</sup>

<sup>1</sup>School of Mechanical and Aerospace Engineering, Air Traffic Management Research Institute, Nanyang Technological University, Singapore 639798

<sup>2</sup>Department of Systems Engineering and Engineering Management, City University of Hong Kong, Hong Kong

Corresponding author: Sameer Alam (sameeralam@ntu.edu.sg)

This work was supported by the Nanyang Technological University-Civil Aviation Authority of Singapore (NTU-CAAS) Research through the School of MAE, Air Traffic Management Research Institute, Nanyang Technological University, Singapore, under Grant M4062429.052.

**ABSTRACT** Flight schedules are highly sensitive to delays and witness these events on a very frequent basis. In an interconnected and interdependent air transportation system, these delays can magnify and cascade as the flight itineraries progress, causing reactionary delays. The airlines, passengers and airports bear the negative economic implications of such phenomenon. The current research draws motivation from this behavior and develops a multi-agent based method to predict the reactionary delays of flights, given the magnitude of primary delay that the flights witness at the beginning of the itinerary. Every flight is modeled as an agent which functions in a dynamic airport environment, receives information about other agents and updates its own arrival and departure schedule. To evaluate the performance of the method, this paper carries out a case study on the flights in Southeast Asia, which covers eleven countries. The model is tested on a six-month ADS-B dataset that is collected for the calendar year 2016. Through the reactionary delay values predicted by the multi-agent based method, the flights are first classified as delayed or un-delayed in terms of departure. The classification results show an average accuracy of 80.7%, with a delay classification MAE threshold of 15 minutes. Further, a delay multiplier index is evaluated, which is a ratio of the total delays (primary+reactionary delays) and the primary delays for each aircraft. The majority of delay multiplier values range between 1-1.5, which signifies that for except a few outliers, the primary delays do not significantly cascade into reactionary delays for the flights in Southeast Asia. The outliers represent scenarios where primary delays magnify and propagate as reactionary delays over subsequent flight legs. Therefore, the proposed method can assist in better flight scheduling by identifying itineraries which experience higher reactionary delays.

**INDEX TERMS** Air traffic management, agent-based method, ground delay analysis, Southeast Asian airports, reactionary delays.

## I. INTRODUCTION

### A. BACKGROUND AND LITERATURE REVIEW

The commercial civil aviation sector is a complex, distributed network, with a vast number of elements interacting with each other to meet a common objective of safe and in-time transport of passengers [1]. This sector has witnessed immense growth in terms of infrastructure, territorial coverage and customers, over the past few decades. The International Air Transport Association (IATA) forecast predicts 8.2 billion passengers by the year 2037, at an annual compounded growth rate of 3.5% [2]. With the ever-increasing air traffic

and limited CNS (communication, navigation and surveillance) and airspace capacities, the flight schedule planners try to reduce the buffers between flight arrivals and departures in order to maximize aircraft utilization and meet the increasing demand. Similarly, the airport management aims to service and prepare the aircraft for departure faster, to increase the airport's gate availability. Such plans and practices increase the likelihood of delay generation [3], [4].

The direct and indirect costs associated with delays [5], its effect on passengers and an increasingly competitive commercial aviation market have driven research to predict and minimize air traffic delays [6]–[8]. Strategies such as the Ground Delay Programs (GDP) [9], [10], Collaborative Decision Making (CDM) [11] and Air Traffic Flow

The associate editor coordinating the review of this manuscript and approving it for publication was Weiguo Xia<sup>1</sup>.

Management (ATFM) [12], [13] have been heavily researched and implemented in order to improve the information flow between the participating airports and contain the overall flight delays. It has been reported in the literature that the major portion of all the flight delays occurs on ground at the airports [14]. These delays originating at one location propagate downstream to other airports, leading to the emergent behaviour which is called delay propagation. Over the past few decades, research methodologies to characterize and predict delay propagation have evolved from pure statistical methods and heuristics [15], [16], to queuing models, network theory approaches, machine learning and agent based techniques [14], [17]–[19].

With regard to delay propagation, Pyrgiotis *et al.* [20] considered an analytical queuing and network decomposition model to characterize delay propagation. This study, however, was only able to give a macroscopic analysis of daily and hourly delay at 30 main airports of US. Rebello and Balakrishnan [21] developed a two-step approach to first classify the delay above or below a threshold and then, estimate the departure delay propagation along a link at some time in the future. Their delay classification results for a day's forecast with a threshold of 60 minutes yielded an accuracy of 72.8%. On similar lines, Belcastro *et al.* [22] attempted to classify and predict flight delays due to weather conditions. They reported an accuracy of 74.2% while classifying delays with a delay threshold of 15 minutes. Choi *et al.* [18] proposed a binary classification model to predict weather induced arrival delays of flights, which was tested on multiple algorithm like random forest, AdaBoost, kNearest-Neighbors and Decision Tree. Researchers have also explored Bayesian network models to understand the effects of delay propagation [23], [24]. In one of the recent research efforts, Chen and Li [25] introduced an optimum feature selection process to improve the prediction performance of their machine learning based model and claimed to overcome the problem of feature selection, prevalent in most machine learning works in air traffic management.

Agent based modelling is a different paradigm, which has also been used in transportation studies. In the past, agent based methods have been implemented in research involving aircraft collision avoidance [26], air traffic flow management [27] and aircraft sequencing and scheduling [28]. Bouarfa *et al.* [19] discussed the emergent phenomenon shown by air transportation systems and used an agent based model to assess the safety risk of an active runway crossing; due to the dependency of safety on the joint functions of many parameters like pilots, controllers, ground staff, technology, etc. Agent based models have also been implemented in delay propagation analysis. Fleurquin *et al.* [7] developed an agent based model to assess the effect of three factors namely, airport congestion, aircraft rotation and flight connectivity on delay propagation by evaluating the size of the delayed airport clusters. The basis of selection of only these parameters is although not clear. In their subsequent paper [8], they extended their work

to analyze the effect of meteorological disruptions (weather) in the US air traffic network. For both of the papers, aircraft were modeled as agents.

## B. RESEARCH MOTIVATION

According to the Central Office for Delay Analysis (CODA), Eurocontrol, reactionary delay remained the largest cause of departure delay of the flights for July 2019 (see Fig. 1). Similar trend has been documented for the months of April, May, June, etc for 2019 and 2018 as well. In order to contain the problem of delay propagation, identifying the individual influence of major sources of delay is believed to be critical, so that remedial actions to curb delay are focused on reducing these critical parameters.

Departure Delay Causes - July 2019

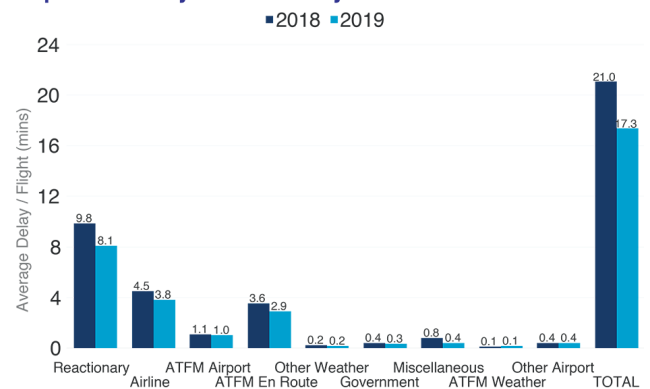


FIGURE 1. Major causes of departure delay for July 2019 for the flights in Europe. The picture is obtained from the Central Office for Delay Analysis (CODA), Eurocontrol [29].

It can be seen from the documented literature that the problem of flight delay has been studied with a view to predict delays due to the aggregated effect of multiple factors. Research efforts which focus on the effect of individual factors on the overall delays are mainly concentrated to predicting weather related effects, with no significant literature on predicting reactionary delays over an aircraft's flight itinerary and its influence on the total delays. With an aim to bridge this research gap, the current research provides a unique yet simple approach to predict the reactionary delays and identify cases where primary delays show a magnified and cascading effect leading to higher reactionary delays, with a case study of the Southeast Asian (SEA) airport network. This research can assist in better flight schedule planning through prior knowledge of flights expecting higher delay. We develop a multi-agent based method to evaluate the reactionary delays on various flight legs of a daily aircraft itinerary, wherein each flight is regarded as an agent. In order to maximize flight services, an aircraft normally has to undertake multiple flight legs. Given that each aircraft has a primary departure delay in the first flight leg, we let the agents (flight) travel among the airports based on their daily itineraries and predict the reactionary delays on each of the subsequent flight leg. The functionality of the proposed method is tested through a case

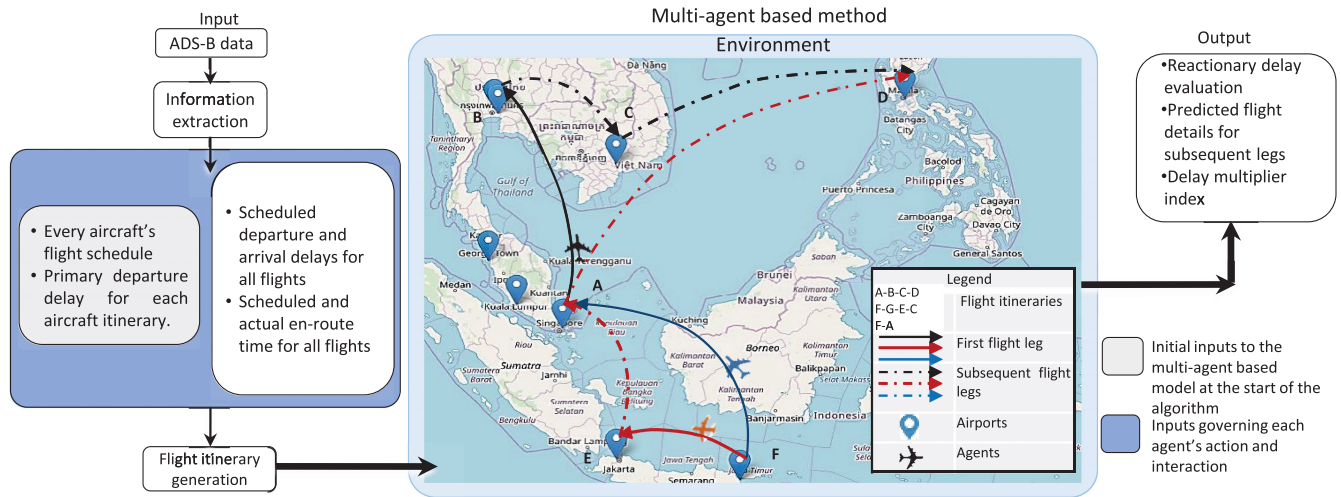


FIGURE 2. Proposed multi-agent based method for reactionary delay prediction of flights.

study by analyzing the Automatic Dependent Surveillance Broadcast (ADS-B) data for SEA Airport network, for a period of six months.

A higher level abstract diagram of the central idea of research in this paper is shown in Fig. 2. The ADS-B data is filtered to extract information of departure delays, arrival delays and actual en-route time for every flight and flight itineraries are obtained. Every agent (flight) receives information regarding its initial departure delay and its schedule. The agents then, function in the airport environment according to this information and the rules discussed in Section 3. The output of this analysis is the reactionary delays for every aircraft’s subsequent flight legs, predicted flight arrival and departures details for subsequent legs and the evaluation of delay multiplier index.

The document is further organized as follows - Section II documents the preliminaries of the paper. Section III describes the research problem and research methodology for reactionary delay prediction. In section IV we discuss the case study used to test the methodology, data reprocessing and discussions on the results obtained. Section V accumulates the concluding comments and the future work scopes of this research.

II. RESEARCH PRELIMINARIES

A. AIRCRAFT, FLIGHT AND ITINERARY

An aircraft can be identified by its registration number, which is unique. An aircraft is tagged as a flight when it travels between different airports. As shown in Fig. 3, the aircraft with registration number 9MAQC is a different flight for each origin-destination pair. These two terminologies have been cautiously used in the paper.

An aircraft itinerary is a sequence of flights that an aircraft takes from origin to final destination. An itinerary may consist of a single or multiple flight legs, where one flight leg is an origin destination pair (see Fig. 3).

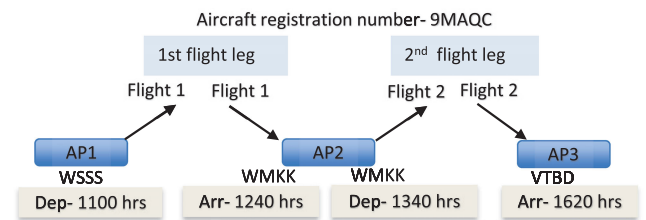


FIGURE 3. A sample itinerary of a given aircraft with two flight legs.

B. MULTI-AGENT BASED METHOD

Air traffic management (ATM) operations show an emergent behavior due to the combined dynamic interactions between various individual components such as human resources, technical facilities, passengers, airlines, airports, different ATM centres and other stakeholders. With the continuously increasing complexity of such ATM systems, agent based methods (ABMs) provide a suitable tool to model such systems with the desired level of abstraction [19]. ABMs are suitable and can enhance the analysis of problem domains which are geographically distributed, contain subsystems in a dynamic environment and where subsystems are required to interact with the environment [30]–[32] and hence, are very suitable for applications in a complex airport network. Individual components of an dynamic, distributed system communicate with each other through simple rules. It is the interaction guided by the cumulation of these simple rules, that gives rise to complex behavior at the global level. On a similar note, agents in this method are guided by simplistic rules to interact with the environment and update their current state. The key components of the multi-agent based method used in the current research are described as below-

- Agent - The term ‘agent’ can be defined as a system situated in a certain environment, which is capable of performing autonomous actions to meet its design objectives [33], [34]. In the current method, every flight is

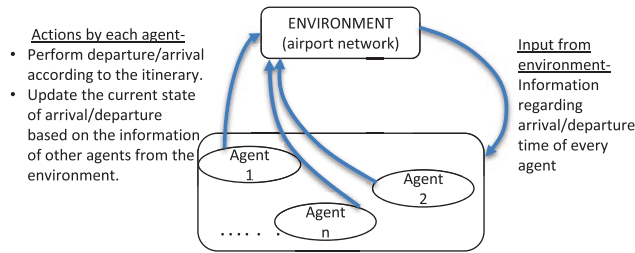


FIGURE 4. Agent-environment abstract diagram.

modeled as an agent whose objective is to depart from the origin and arrive at the destination at the stipulated time and receive information from the environment regarding simultaneous arrivals or departures of multiple agents. The information of the primary departure delays of the first agent and a daily schedule are provided from the ADS-B data.

- Agent environment - For every agent, the environment corresponds to its previous departure airport and its current destination airport. All the agents in unison, function in an environment incorporating all the airports.
- Agent decisions (action and interaction) - an agent’s actions involve utilizing information of scheduled departure time, scheduled en-route time, the agent’s origin and destination, from the data, to complete its assigned task (travel from its origin to destination). Agents interact with the environment when there is a commonality in their origin or destination, i.e if multiple agents depart from or arrive at the same airport at the same time. In this case the agent decide their sequence of arrival and departure based on pre-defined rules. For instance, in case of simultaneous arrivals at an airport, the sequence of arrival is decided on the basis of departure delays that these agents witnessed on the flight leg. Similar rules apply for the agents departing from a common airport.

III. PROBLEM DESCRIPTION AND METHODOLOGY

A. PROBLEM DESCRIPTION

An aircraft’s schedule for the entire day may consist of multiple flight legs and it is prone to witnessing delay on one or all of these flight legs. The current research problem focuses on the prediction of reactionary departure delays that the aircraft witnesses on subsequent flight legs based on its delay in the primary flight leg and also, the inputs from the environment regarding the schedules of other flights. The departure delays that are obtained from the ADS-B data are ‘observed delays’, which can be a cumulation of multiple factors other than just the propagated delay, such as delays due to gate unavailability, ground holding due to weather or even crew unavailability at some instances, to name a few. In this research, a multi-agent based method has been used to predict the reactionary departure delays, i.e. the delays that have been solely propagated from the primary flight leg due to the initial departure delays, to the subsequent flight legs,

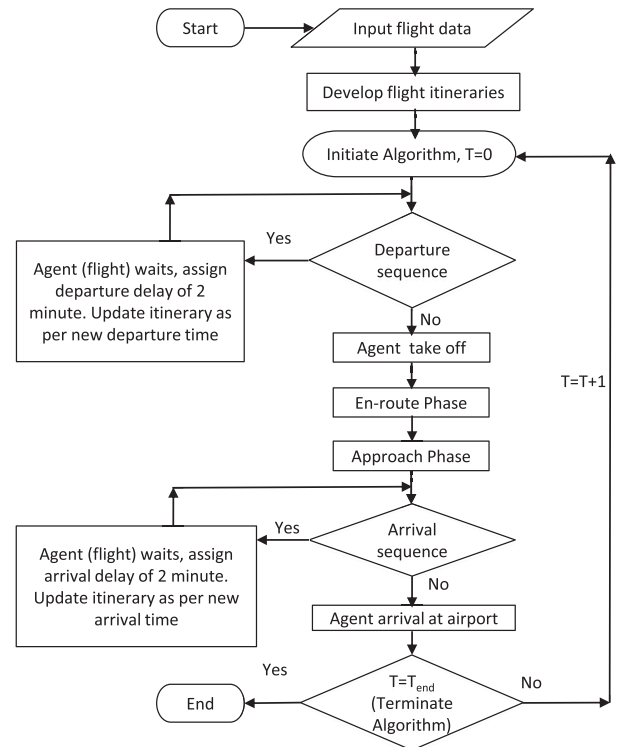


FIGURE 5. Flowchart of the proposed method for reactionary delay prediction.

by making use of the information of the flight’s primary delay in the first flight leg and the scheduled arrival time, scheduled departure time, scheduled en-route time over the subsequent flight legs and the input from the environment regarding the schedules of other flights.

The research focus i.e the SEA airport network possesses a unique characteristic of fairly short haul flights, with flight time of most flights within 2 hours (also discussed in Section IV-C, Fig. 7). With such short flight times the flights that incur delay during departures are unable to recover the delay en-route and most of the times, have a delayed arrival at the destination. This claim is also supported in Fig. 7 by the plot between departure delays and the subsequent arrival delays, which is linear, signifying the inability of the flight to recover delays.

B. METHOD OVERVIEW

In the proposed methodology, every flight acts as an agent which carries out its own flight schedule in an environment of multiple airports, receives information regarding arrivals and departures of other agents from the environment and updates its schedule accordingly. In this process generation of reactionary delays takes place. An overview of the methodology is shown by a flowchart in Fig. 5.

C. RESEARCH ASSUMPTIONS AND REASONS

In order to solve the research problem given in the above section, the following assumptions together with their corresponding reasons are provided below.

**TABLE 1. Definitions of parameters used in the proposed method.**

Parameters	Definition
$N_f$	number of flights in a day's schedule
$T_{SDf_i}$	the scheduled departure time of flight $f_i$
$T_{ADf_i}$	the actual departure time of flight $f_i$
$T_{SAf_i}$	the scheduled arrival time of flight $f_i$
$T_{AAf_i}$	the actual arrival time of flight $f_i$
$T_{DDf_i}$	the departure delay of flight $f_i$
$T_{ADf_i}$	the arrival delay of flight $f_i$
$TAT_{f_i}$	the turn-around time of flight $f_i$
$T_{enf_i}$	the flight time of flight $f_i$
$T_{begin}$	the starting time point for the delay analysis
$T_{end}$	the ending time point for the delay analysis

- 1) One of the fundamental assumptions in the current work is that the flights do not witness or absorb delays enroute their destination. Similar assumption has been made for the delay analysis in existing works [35]. Further, for the SEA region, the majority of flight are short haul (see Fig. 7 for reference). It can be argued that due to smaller enroute times, recovering delays enroute is not feasible.
- 2) Further, flight arrival and departure take place at separate runways at every airports considered in the analysis. This has been done to relatively simplify the model and keep it generic in nature.
- 3) A flight is classified as delayed if the arrival or departure takes place 15 minutes after the scheduled time. This is in reference to the delay definition used by the Federal Aviation Administration (FAA) [36].
- 4) It has also been hypothesized that no flight can depart more than 5 minutes prior to its actual departure. If a flight has an early departure of more than 5 minutes, the actual departure time is set to be equal to the scheduled departure time.
- 5) The minimum turn around time for every flight has been fixed at 70 minutes. This is based on the data analysis in section IV-C.

The abbreviation used in the algorithms have been mentioned in Table 1.

Based on the above assumptions, we then predict  $T_{ADf_i}$  and  $T_{AAf_i}$  for all  $f_i \in [1, N_f]$  in the following way.

1) If  $f_i$  is the first flight leg of an aircraft's itinerary, then we estimate  $T_{ADf_i}$  as

$$T_{ADf_i} = T_{SDf_i} + T_{DDf_i}$$

2) If  $f_i$  is an intermediate flight leg of an aircraft's itinerary, then we calculate  $T_{ADf_i}$  in the following ways.

When  $T_{SDf_i} - T_{ED} < T_{AAf_{i-1}} + TAT_{f_i} < T_{SDf_i}$  or  $T_{AAf_{i-1}} + TAT_{f_i} \geq T_{SDf_i}$ , then we set

$$T_{ADf_i} = T_{AAf_{i-1}} + TAT_{f_i}$$

The term  $T_{ED}$  is the maximum early departure time which is set to be 5 minutes, and the term  $TAT$  is the minimum turnaround time which is set to be 70 minutes.

When  $T_{AAf_{i-1}} + TAT_{f_i} < T_{SDf_i} - T_{ED}$ , then we set

$$T_{ADf_i} = T_{SDf_i}$$

3) Based on  $T_{ADf_i}$ , we then update  $T_{AAf_i}$  as

$$T_{AAf_i} = T_{ADf_i} + T_{enf_i}$$

#### D. DEPARTURE SEQUENCE MANAGEMENT PRINCIPLE

The algorithm starts with a daily itinerary generated from the ADS-B data. For an aircraft's flight itinerary, the first agent (flight) has its specified actual departure time and the primary departure delay and it departs from the origin at that specified time. In case of multiple departures at the same time from an airport, we propose a departure sequence management principle, which is explained in Algorithm 1. A list of all origin and destination airports is obtained for all agents (flights) from the itinerary and is filtered to obtain the list of agents departing at time  $T$ . A list of unique origins is the obtained and the agents departing from common airports are identified. Step 5 of the algorithm sequences the flights in case multiple agents depart from the same airport. The agents are sequenced in the increasing order of magnitude of delays, for the case when departure delays are different. If the departure delays are a combination of negative and positive delays, the agents with positive delays are sequenced first, followed by the agents with negative delays, all on the basis of the previous rules. For the rare case of same departure delays, random sequencing has been adopted. After an agent departs for the airport, the other agents are assigned a delay  $T_\delta$ . For every departing agent, its arrival time at the destination is updated.

#### E. ARRIVAL SEQUENCE MANAGEMENT PRINCIPLE

The algorithm for arriving agents (flights) follows the same logic as the departure sequence assignment mechanism. For the flights arriving at an airport at the same time, Algorithm 2 is followed to sequence them for arrival. Similar to the case of departure sequence management, Step 4 of the algorithm defines the rules for sequencing the arriving flights at an airport. When the arrival delays are different, the arriving flight are sequenced in the increasing order of their magnitude of delays. If the arrival delays are a combination of negative and positive delays, the flights with positive delays are sequenced first, followed by the flights with negative delays. For the case of same arrival delays for all flights arriving at an airport, random sequencing has been adopted.

The sequencing of the arriving and the departing flights is fundamentally based on 'first come first served' principle but this is applicable if the arrivals and departures are at different times. Due to delays, if multiple departures are scheduled at the same time at a given airport, we assign a departure delay of 2 minute to the next sequenced flight, to compensate for

**Algorithm 1** Departure Sequence Management Mechanism

- 1) Define the origin and destination sequences  $Orgs := (O_1, O_2, \dots, O_{N_f})$  and  $Dsts := (D_1, D_2, \dots, D_{N_f})$  with  $O_i$  and  $D_i$  being the origin and destination of flight  $f_i$ ;
- 2) Given a time point  $T$ ,  $\forall i \in [1, N_f]$ , get the flight sequence  $F := (f_i, f_j, \dots, f_k)$  with  $t_{AD_{f_i}} = T$ ;
- 3) Get the subset  $Org = (O_i^F, O_j^F, \dots, O_k^F)$  with  $O_i^F$  being the origin of flight  $f_i \in F$ ;
- 4)  $UniqOrg = unique(Org)$ ; //identify the unique origins
- 5) **For**  $i = 1$  to  $length(UniqOrg)$ , **do**
  - a)  $F' = find(Org == UniqOrg(i))$ ; //  $F' \subset F$
  - b)  $T_{DD_{F'}} = T - T_{SD_{F'}}$ ; //delays for all  $f'_i \in F'$
  - c) If  $size(unique(T_{DD_{F'}})) == 1$ , i.e., all  $f'_i \in F'$  have the same departure delay, then randomly select a  $f'_i$  from  $F'$ ;
  - d) If  $max(T_{DD_{F'}}) \leq 0$ , then select the  $f'_i$  from  $F'$  with the largest departure delay;
  - e) If  $min(T_{DD_{F'}}) \geq 0$ , then select the  $f'_i$  from  $F'$  with the smallest departure delay;
  - f) If  $min(T_{DD_{F'}}) < 0 \wedge max(T_{DD_{F'}}) > 0$ , then from the flights with nonnegative delays select the  $f'_i$  with the smallest departure delay;
  - g) Set  $T_{AD_{f'_i}} = T$  and set  $T_{AD_{f'_j}} = T + T_\delta, \forall f'_j \in F' \setminus f'_i$ .
- 6) **End**
- 7) **For** all  $f_i \in F$ , **do**
  - a) Set  $T_{AA_{f_i}} = T_{AD_{f_i}} + T_{enf_i}$ ; //update the arrival time
  - b) If  $D(i) == O(i+1)$ , then set  $T_{AD_{f_{i+1}}} = T_{AA_{f_i}} + TAT_{f_i}$ .
- 8) **End**

the vortex effects. The Manual of Standards-Air Traffic Services, Civil Aviation Authority of Singapore (CAAS), [37] mentions a 2 minute separation while take-off and landing when a medium category aircraft follows a heavy category aircraft and a 3 minutes separation when a medium category aircraft follows a super heavy category aircraft. This value has been standardized to 2 minutes for the current research considering that the majority of aircraft used for short haul flight are medium and heavy aircraft. For flights arriving at a common airport, if the final approach fix is the same due to the delays associated with the flights (which we refer to as simultaneous arrival for the current research), we assign a 2-minute delay to the arriving aircraft which, on a higher abstraction level represents the ATC actions such as holding the aircraft in order to maintain the separation standards.

**IV. CASE STUDY****A. ADS-B DATA**

The findings represented in the current paper have been obtained using the ADS-B data for the months of June, July, September, October, November and December 2016,

**Algorithm 2** Arrival Sequence Management Mechanism

- 1) Given a time point  $T$ ,  $\forall i \in [1, N_f]$ , get the flight sequence  $F := (f_i, f_j, \dots, f_k)$  with  $T_{AA_{f_i}} = T$ ;
- 2) Get the subset  $Dst = (D_i^F, D_j^F, \dots, D_k^F)$  with  $D_i^F$  being the destination of flight  $f_i \in F$ ;
- 3)  $UniqDst = unique(Dst)$ ; //the unique destination
- 4) **For**  $i = 1$  to  $length(UniqDst)$ , **do**
  - a)  $F' = find(Dst == UniqDst(i))$ ; //  $F' \subset F$
  - b)  $T_{AD_{F'}} = T - T_{SA_{F'}}$ ; //delays for all  $f'_i \in F'$
  - c) If  $size(unique(T_{AD_{F'}})) == 1$ , i.e., all  $f'_i \in F'$  have the same arrival delay, then randomly select a  $f'_i$  from  $F'$ ;
  - d) If  $max(T_{AD_{F'}}) \leq 0$ , then select the  $f'_i$  from  $F'$  with the largest arrival delay;
  - e) If  $min(T_{AD_{F'}}) \geq 0$ , then select the  $f'_i$  from  $F'$  with the smallest arrival delay;
  - f) If  $min(T_{AD_{F'}}) < 0 \wedge max(T_{AD_{F'}}) > 0$ , then from the flights with nonnegative delays select the  $f'_i$  with the smallest arrival delay;
  - g) Set  $T_{AA_{f'_i}} = t$  and set  $T_{AA_{f'_j}} = T + T_\delta, \forall f'_j \in F' \setminus f'_i$ .
- 5) **End**
- 6) For all  $f_i \in F$ , if  $D(i) == O(i+1)$ , then set  $T_{AD_{f_{i+1}}} = T_{AA_{f_i}} + TAT_{f_i}$ .

**TABLE 2.** Related variables and their definitions.

Variables	Definition
$TAT_{sch}$	scheduled turn around time
$TAT_{act}$	actual turn around time
$T_{S.BlkOut}$	scheduled block out time
$T_{S.BlkIn}$	scheduled block in time
$T_{A.BlkOut}$	actual block out time
$T_{A.BlkIn}$	actual block in time
$D_D$	departure delay
$D_A$	arrival delay

procured from Flight Aware. The ADS-B data, out of the many parameters, provides information regarding the flight ID, registration number, origin, destination, actual departure time, actual arrival time, en-route time, scheduled and actual block in time and block out time.

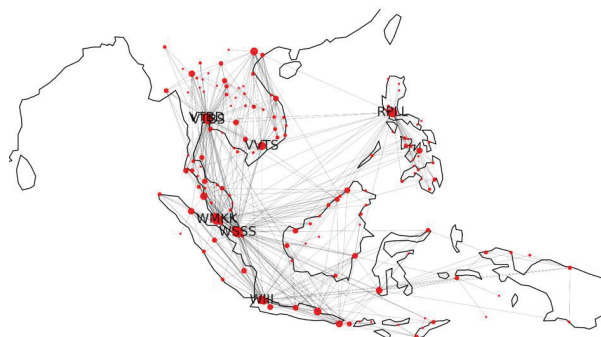
Through the ADS-B data, the following information for each flight leg can be derived-

$$\begin{aligned}
 TAT_{sch} &= T_{S.BlkOut} - T_{S.BlkIn} \\
 TAT_{act} &= T_{A.BlkOut} - T_{A.BlkIn} \\
 D_D &= T_{S.BlkOut} - T_{A.BlkOut} \\
 D_A &= T_{S.BlkIn} - T_{A.BlkIn}
 \end{aligned}$$

The terminologies used in these equations are defined in Table 2.

**B. SOUTHEAST ASIAN AIRPORT NETWORK**

An air transport network is usually described and visualized as a graph with nodes indicating the airports and the edges



**FIGURE 6.** Southeast Asian Airport Network. The airports are weighted according to the degree centrality, where bigger dots imply higher connectivity.

indicating the flight movements between the airports [38], [39]. The Southeast Asian Airport Network can be visualized as an undirected network where the nodes represent the airports and the edges highlight the flights between the airports, as shown in Fig. 6. Contrasting geographical and weather conditions, relatively smaller flight duration in the SEA as compared to the US, and the lack of research inclination to this region has been the driving factor to analyze the SEA airport network.

The concept of centrality is used to identify the most important or central nodes in this network [40]. Out of a multiple centrality measures like the degree centrality, between centrality, eign-vector centrality and closeness centrality, we choose degree centrality as the most apt for our analysis. Degree centrality is a measure of the number of connection or the number of edges connected to a node. The airports with highest degree centrality are shown in Table 3. Through this analysis we can identify the 20 most important airports from a total of 136 airports (ICAO codes) in the Southeast Asian Airport network.

**C. STATISTICS FOR FLIGHTS IN THE SEA**

ADS-B data for six months in 2016 is analyzed to draw initial inferences regarding the active airports, number of flights and delay patterns of the Southeast Asian airport network. The basic statistics for the flight data are provided in Fig. 7.

The monthly active number of airports can be seen in the first subfigure of Fig. 7. There are on average, 130 airports in operation in each of the studied months. The number of active airports in December is the highest. Also, the number of flights are the highest in July and December.

The third subfigure demonstrates the average departure delays as well as overall delays for the flights operated in the studied six months. We can see from the figure that the departure delays and the overall average delays are around 22 minutes. Since in July and December the flight activities are the highest, the average delays in this two months are also the largest, above 25 minutes as can be seen from the figure.

The second row of Fig. 7 further visualizes the probability distributions of arrival and departure delays of the flights

**TABLE 3.** Top twenty airports in the Southeast Asia. The un-weighted degree represents the number of airports to which the current airport is connected. Weighted degree represents total arrivals to and departures from the current airport.

ID	Airport IATA Code	Location	Unweighted Degree	Weighted Degree
1	WMKK	Malaysia	51	71582
2	WSSS	Singapore	42	45271
3	VTBD	Thailand	40	27583
4	RPLL	Philippines	36	27015
5	WIII	Indonesia	30	60677
6	VVTS	Vietnam	24	19178
7	VVNB	Vietnam	24	13301
8	VTBS	Thailand	23	34392
9	WADD	Indonesia	16	15202
10	WMKP	Malaysia	15	9713
11	VTCC	Vietnam	15	8284
12	RPVM	Philippines	15	7706
13	WARR	Indonesia	14	14171
14	WAAA	Indonesia	13	7285
15	WBKK	Malaysia	12	8911
16	VDSR	Cambodia	12	3405
17	VTSP	Thailand	11	12058
18	WIMM	Indonesia	11	7994
19	VVDN	Vietnam	11	5728
20	WMKJ	Malaysia	10	3458

operated in the SEA. The plots show that arrival and the departure delays show similar trends over the months. In the literature, it has been reported that flight delays normally follow Weibull distribution. To verify whether this phenomenon also applies to flights in the SEA, we use three probability distribution functions, i.e., Weibull (1), Lognormal (2), and Exponential (3) [41], [42] to fit the departure delay distributions. The PDF of those three distributions are given below.

$$f(x, k, \lambda) = \frac{k}{\lambda} \left(\frac{x}{\lambda}\right)^{k-1} e^{-(x/\lambda)^k} \tag{1}$$

$$f(x, u, \sigma) = \frac{1}{x\sqrt{2\pi\sigma^2}} e^{-\frac{(\ln x - u)^2}{2\sigma^2}} \tag{2}$$

$$f(x, \alpha, \beta) = \alpha e^{-\beta x} \tag{3}$$

In the above equation,  $k$  is the shape and  $\lambda$  is the scale parameter,  $\mu$  is the expected value (mean) and  $\sigma$  is the standard deviation. Table 4 lists out the statistics for the curve fittings of the departure delays. For each curve fitting we record the  $R^2$  and RMSE values. We also record the confidence intervals for the parameters in the corresponding probabilistic distribution function at the significance level of 95%.

The  $R^2$  test and the RMSE show the maximum value of 0.8533 and minimum value of 0.0034, respectively, for the Weibull distribution. The curve fitting results shown in Table 4 indicate that the delays of the flights operated in the SEA also follow the Weibull distribution.

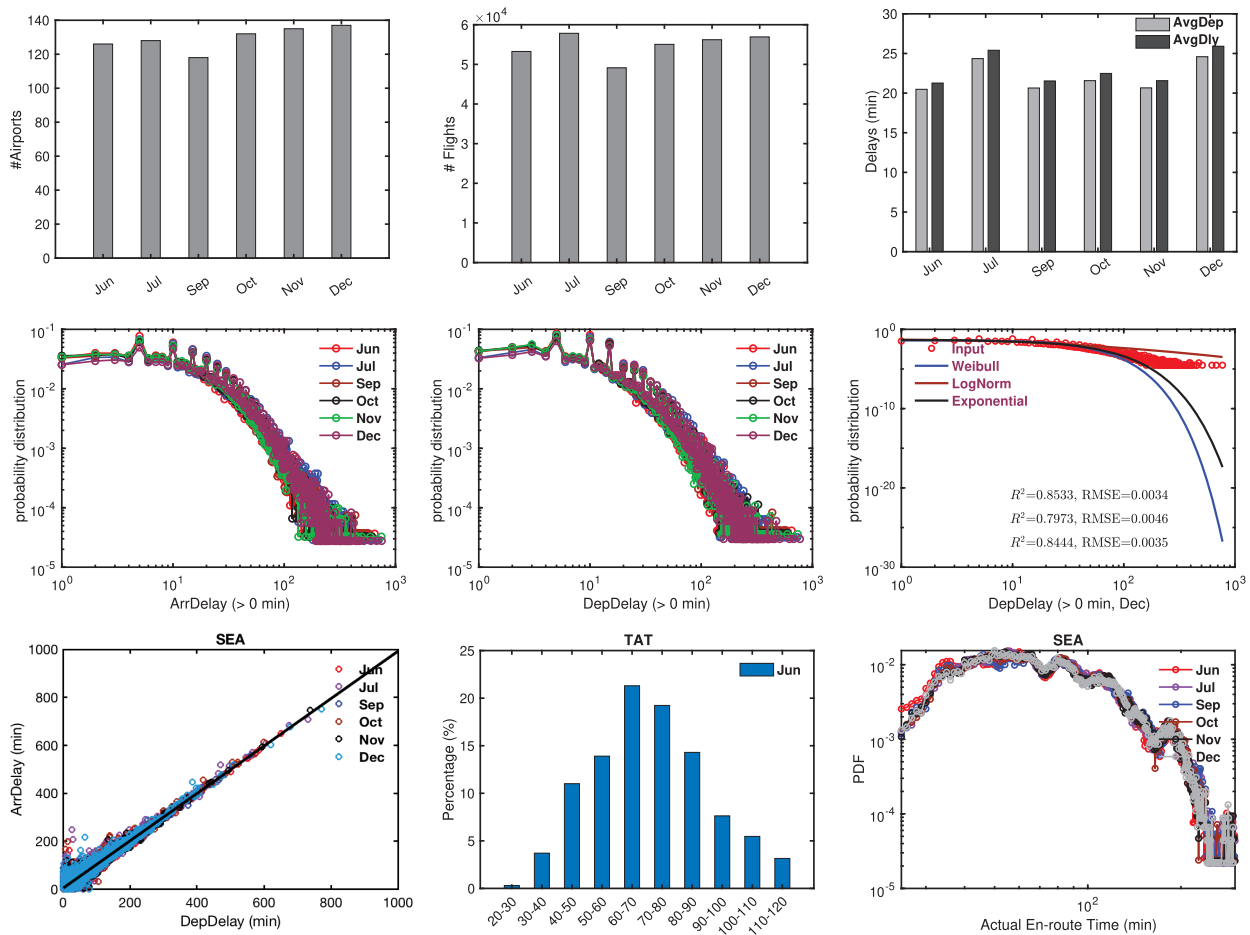


FIGURE 7. Statistics of flight over a period of 6 months (June, July, September, October, November and December) obtained from the ADS-B data.

TABLE 4. Statistical values for the curve fittings of the departure delays. SL – Significance Level; CI – Confidence Interval.

PDF	Weibull	Lognormal	Exponential
SL	95%	95%	95%
Variable	$k = 1.1095$ $\lambda = 19.5092$	$u = -0.9185$ $\sigma = 1.9245$	$\alpha = 0.0483$ $\beta = -0.0479$
CI	$k \in [1.0550, 1.1640]$ $\lambda \in [18.3477, 20.6708]$	$u \in [-0.9744, -0.8628]$ $\sigma \in [1.8596, 1.9893]$	$\alpha \in [0.0452, 0.0515]$ $\beta \in [-0.0523, 0.0436]$
R <sup>2</sup>	0.8533	0.7973	0.8444
RMSE	0.0034	0.0046	0.0035

Capturing the correlation between departure delays from the previous airport and the arrival delay to the destination airport reveals that except for a few outliers, most of the flights that depart late tend to arrive late at the destination i.e a flight that witnesses positive departure delay will have a positive arrival delay. This is also the reason that the subsequent analysis focuses on departure situations. Plotting the distribution of the turn around time (TAT) reveals that majority of flight have a turn around time of 60 to 70 minutes. Hence, a TAT of 70 minutes is chosen for the analysis.

The distribution of the flight en-route time have also been analyzed it can be see from the last subfigure that majority

of flight are short haul, with en-route time between 40 to 120 minutes.

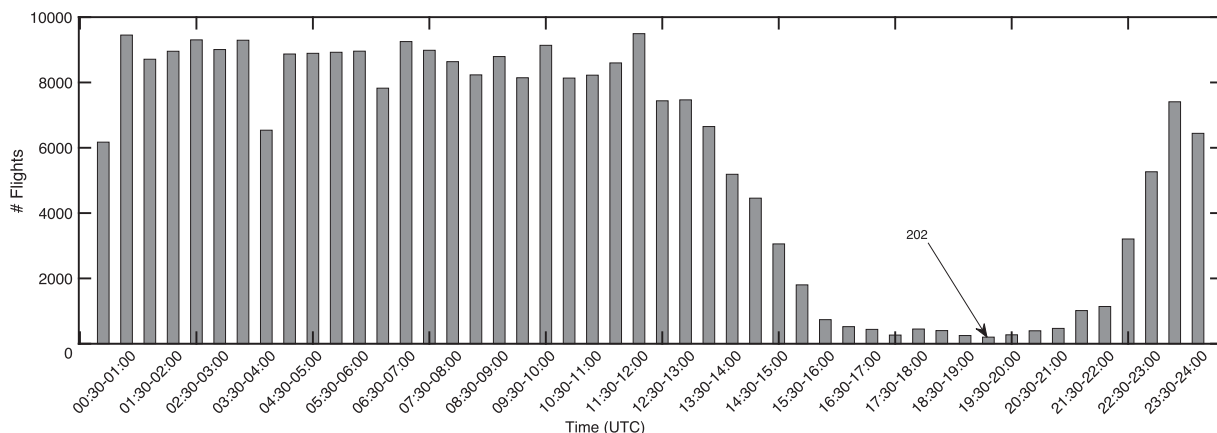
#### D. FLIGHT ITINERARY EXTRACTION

The experimental design focuses on the airport network of 11 Southeast Asian countries namely Indonesia, Vietnam, Cambodia, Malaysia, Singapore, Laos, Myanmar, Thailand, Philippines, Brunie and Timor-Leste. The time horizon for itinerary generation has been kept as one day for the analysis. In order to choose a starting time to obtain primary delays and generate flight itineraries, the number of hourly departures of the flights are plotted. It is visible from Fig. 8 that the time period between 16:00 hours (UTC) and 20:00 hours (UTC) has the minimum departure activities. Hence, the starting time point of the itinerary is chosen from 19:00 hours (UTC) since this time is a relatively low activity period [7].

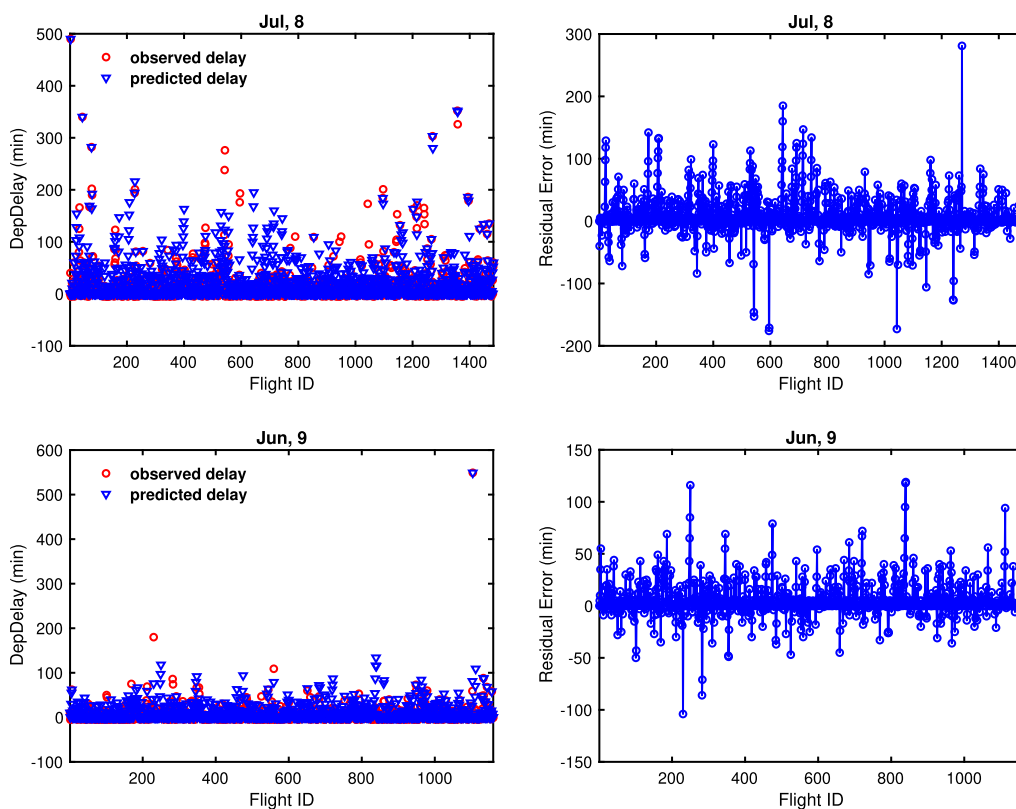
#### E. ANALYSIS OF REACTIONARY DELAYS

The method is tested on 2 days with the highest average delay and 2 days with lowest average delay from each of the six months, for the analysis of reactionary delays. Table 5





**FIGURE 8.** Hourly flight departures plotted for a period of 6 month to select starting time point and primary delay on the basis of minimum flight activity period. It shows a total of 202 flights departing between the time period of 19:00 - 19:30 hours (UTC).



**FIGURE 9.** Comparisons of departure delay values for each flight obtained from the ADS-B data (red) and the predicted departure delays by the multi-agent based method (blue). Figures in the second column show the residuals of the two values for every flight.

summarizes the delay information of the studied 24 days. The values of average delays are in minutes.

For each studied day, we extract a one-day flight itinerary from the ADS-B data with the starting time point being chosen in the way described in subsection IV-D. We then apply the proposed agent-based method to the one-day flight itinerary to evaluate the reactionary delays for the flights contained in the itinerary.

Fig. 9 demonstrates the two-day comparisons between the predicted departure delay values (blue dots) yielded by the

proposed agent-based method and the original flight departure delays (red dots) as recorded in the ADS-B data. The first set of images (row 1) are for 8<sup>th</sup> July 2016 and the second set is for 9<sup>th</sup> June, 2016. It can be seen from the figure that the predicted results for departure delay (blue markers) overlap the departure delay obtained from the ADS-B data (red markers). The RMSE for these two days are 16.21 minutes and 29.18 minutes, respectively. From the residual plots, few instances of outliers can be detected where the residuals are abnormally positive and negative, which may be a reason

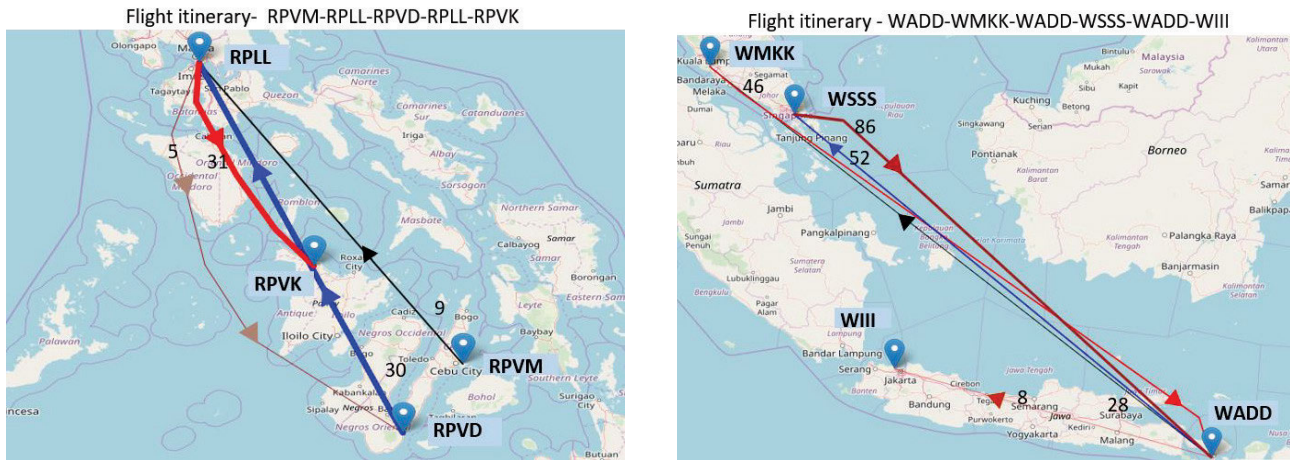


FIGURE 10. Departure delay visualization on subsequent flight legs of two flight itineraries for 8<sup>th</sup> July 2016, highlighting the propagation of primary delays.

TABLE 5. Test days selected on the basis of monthly maximum and minimum average delays.

M	Highest Delay Days			Lowest Delay Days		
	Day	$N_f$	AvgDly	Day	$N_f$	AvgDly
Jun	17	1515	23.745	9	1318	15.635
	29	1597	23.362	12	1021	17.311
Jul	8	1742	31.265	25	1656	19.182
	18	1644	31.071	6	1499	19.708
Sept	9	1450	28.680	13	1253	16.894
	29	1318	27.478	24	1262	17.009
Oct	14	1621	27.835	10	1391	17.025
	30	1497	26.469	3	1360	17.27
Nov	2	1546	26.723	12	1509	16.794
	1	1466	25.876	19	1568	16.795
Dec	16	1863	30.451	14	1581	17.890
	23	1697	29.553	31	1314	18.346

of higher RMSE values. Similar plots can be generated for all the critical days. An interesting conclusion can be drawn from these plots. The agent based method predicts reactionary delay only on the basis of primary delays and the information of other agents during arrivals and departures. The close proximity of these results to the actual delays suggests that most of the flights suffer form delays which originate from the first flight leg. For the 24 tested days, we further calculate the delay classification accuracy based on the reactionary delays. During the calculation, a flight is classified as delayed if the delay is 15 minutes or more than the scheduled time of departure.

Table 6 records the delay classification accuracy for the 24 studied days. In addition to this, average classification accuracy for all the days of the dataset, for a delay threshold of 15 minutes is obtained, which is 80.7%. Here, accuracy implies the flights which were delayed according to the data and were correctly classified as delayed by the model. The average value of recall (true positive rate) is 78.1%. By comparing these values with the existing literature, [21], [22] it

is visible that the current multi-agent based method is able to match the accuracy with significantly less input data. Their delay classification results by Rebello and Balakrishnan [21] for a day’s forecast with a threshold of 60 minutes yielded an accuracy of 72.8%. On similar lines, the machine learning algorithm proposed by Belcastro *et al.* [22] classifies flight delays due to weather conditions with an accuracy of 74.2%, with a delay threshold of 15 minutes.

The predicted departure delays on each aircraft’s itinerary can be visualized as in Fig. 10. The departure delays on the subsequent flight legs constitute the reactionary component of departure delays. The figure on the left shows the predicted departure delays in the itinerary of a aircraft with 4 flight legs. The itinerary starts with an initial departure delays of 9 minutes from RPVM to RPLL. The Subsequent departure delays from RPLL to RPVD, RPVD to RPLL and RPLL to RPVK are 5, 30 and 31 minutes respectively. The lines are weighted according to the departure delay witnessed at the airport of the flight leg. Hence, here the relative weights are 1.8, 1, 6 and 6.2. The lines do not represent the actual trajectory of flight. Some lines have been curved to provide better visibility.

The figure on the right side has an itinerary of 5 flight legs, starting at WADD. From the predicted departures delays, it is visible that the initial departure delay of 28 minutes escalates to 46, 52 and 86 minutes for the departure in 2nd, 3rd and 4th flight leg. Thus, the weights on the lines are 1, 1.6, 1.8, 3.07, which are proportional to the predicted departure delays. From the ADS-B data, it is observable that for the first three flight legs of this itinerary, the scheduled turn around time are relatively shorter, which leads to the cumulation of delays over subsequent flight leg. The turn around time before the start of the 5th flight leg is 2hours and 28 minutes, which helps in absorbing the majority of the previously propagating departure delay. These two cases highlight the instances where primary departure delays visible magnify over the subsequent flight legs.

TABLE 6. Delay classification evaluation for the selected days. Acc – accuracy; Prec – precision; F1-S – F1 Score.

Days with maximum delay					Days with minimum delay				
mm-dd	Acc	Prec	Recall	F1-S	mm-dd	Acc	Prec	Recall	F1-S
0629	0.8209	0.7802	0.7460	0.7627	0609	0.8325	0.6178	0.7433	0.6748
0617	0.7864	0.7313	0.7313	0.7313	0612	0.8643	0.7610	0.6994	0.7289
0718	0.7910	0.7227	0.7130	0.7178	0725	0.8050	0.6838	0.7745	0.7264
0708	0.8121	0.7171	0.8166	0.7636	0706	0.7988	0.6717	0.7893	0.7258
0929	0.8274	0.7819	0.7983	0.7900	0913	0.8142	0.5853	0.8750	0.7014
0909	0.8194	0.8251	0.8220	0.8235	0924	0.7968	0.6339	0.7695	0.6952
1030	0.8176	0.7731	0.7809	0.7769	1010	0.7863	0.5762	0.7683	0.6585
1014	0.7749	0.6810	0.7419	0.7102	1003	0.8018	0.6702	0.8182	0.7368
1101	0.8213	0.7622	0.8386	0.7986	1112	0.8028	0.6339	0.7891	0.7030
1102	0.7992	0.7648	0.7685	0.7667	1119	0.7922	0.6033	0.7631	0.6738
1223	0.7730	0.7966	0.7472	0.7711	1214	0.8039	0.6978	0.7634	0.7291
1216	0.8069	0.8163	0.7989	0.8075	1231	0.7984	0.6543	0.8149	0.7258

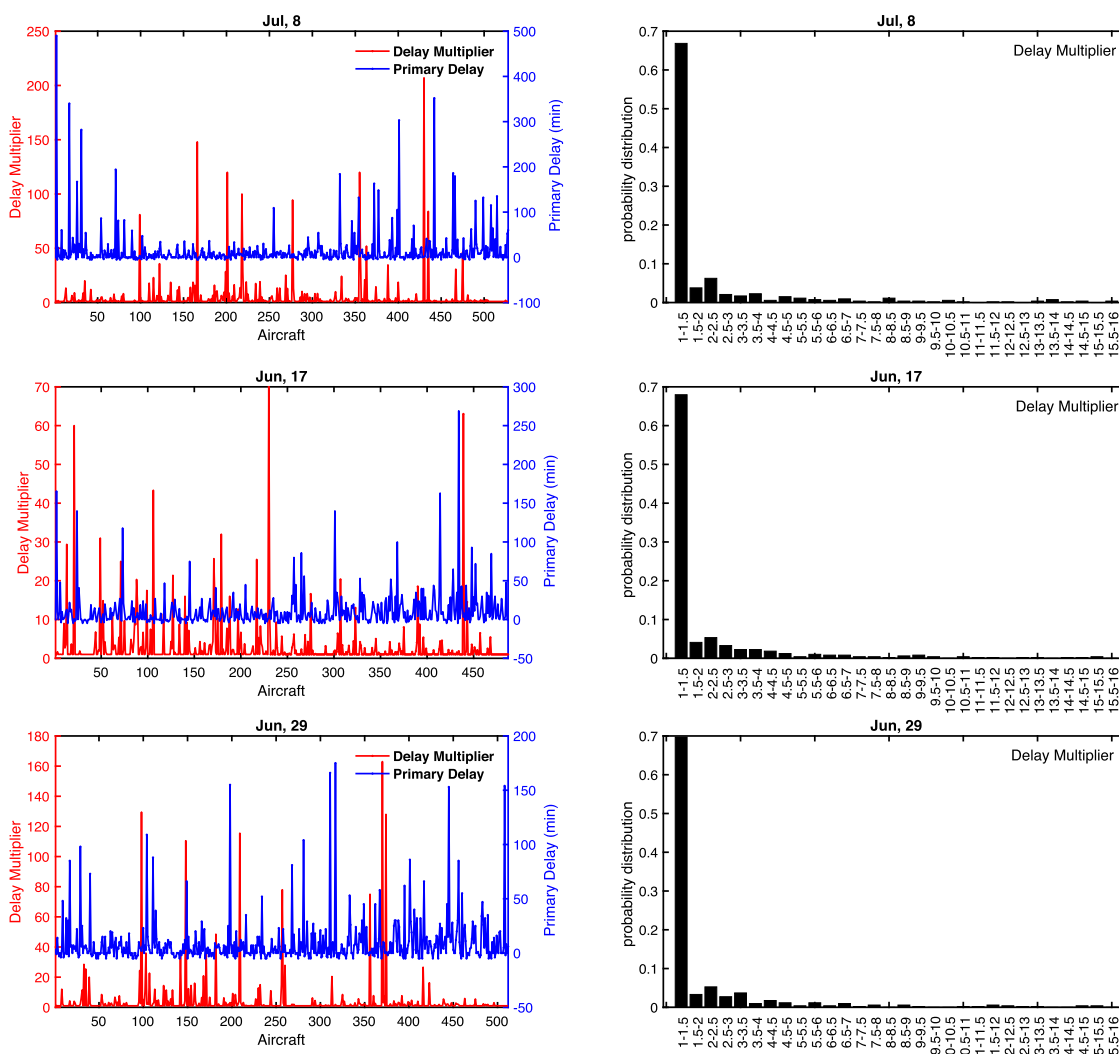


FIGURE 11. Delay multiplier calculated from the predicted reactionary delays for three of the chosen problematic days. Figures on the left show the delay multipliers for each aircraft. Probability distribution of the delay multiplier values are plotted on the right.

The results on delays witnessed by each agent can contribute to develop a global picture as to which aircraft itineraries in a daily schedule witness significant

magnification of the primary delays into higher reactionary delays Thus, the concept of delay multiplier [4] is invoked to evaluate a ratio, based on the downline reactionary

**TABLE 7.** Delay classification results for June and July 2016. Acc – accuracy; Prec – precision; F1-S – F1 Score.

mm-dd	AvgDly	Acc	Prec	Recall	F1-S	mm-dd	AvgDly	Acc	Prec	Recall	F1-S
0609	15.64	0.8325	0.6178	0.7433	0.6748	0725	19.18	0.8050	0.6838	0.7745	0.7264
0912	17.31	0.8643	0.7610	0.6994	0.7289	0706	19.71	0.7988	0.6717	0.7893	0.7258
0621	17.39	0.8149	0.6121	0.7511	0.6745	0731	20.00	0.7965	0.7089	0.7735	0.7398
0610	17.50	0.8221	0.6866	0.7566	0.7199	0712	20.53	0.8082	0.7235	0.7474	0.7353
0608	18.13	0.8276	0.6327	0.7803	0.6988	0726	20.64	0.8164	0.6836	0.7965	0.7357
0628	18.14	0.8401	0.6930	0.7889	0.7379	0713	20.67	0.8138	0.7413	0.7354	0.7384
0607	18.23	0.8352	0.6628	0.8028	0.7261	0710	20.85	0.7946	0.7173	0.7156	0.7165
0613	18.31	0.8351	0.6696	0.7476	0.7064	0729	20.91	0.8149	0.7241	0.7772	0.7497
0624	18.83	0.8426	0.7499	0.8281	0.7871	0730	21.10	0.7941	0.6625	0.7288	0.6941
0604	19.05	0.8163	0.6960	0.7960	0.7426	0707	21.82	0.7873	0.6765	0.7775	0.7235
0620	19.44	0.8420	0.7340	0.7803	0.7561	0723	21.83	0.8160	0.7015	0.7790	0.7382
0625	19.51	0.7953	0.6533	0.7427	0.6951	0702	22.05	0.8079	0.7726	0.7335	0.7525
0605	19.62	0.8096	0.6338	0.7059	0.6679	0705	22.16	0.8122	0.7384	0.7513	0.7448
0602	19.63	0.8026	0.7216	0.7257	0.7236	0709	22.34	0.7994	0.6461	0.8242	0.7244
0619	19.88	0.8319	0.7107	0.7661	0.7374	0701	22.77	0.7870	0.7343	0.7559	0.7449
0606	19.95	0.8324	0.6460	0.7763	0.7052	0704	23.04	0.7955	0.6998	0.7380	0.7184
0618	20.59	0.8239	0.7175	0.7826	0.7487	0714	23.18	0.8044	0.7768	0.6907	0.7312
0614	20.61	0.8471	0.7027	0.8089	0.7521	0727	23.97	0.7735	0.7220	0.6928	0.7071
0623	21.31	0.8430	0.7196	0.7970	0.7563	0721	24.42	0.7797	0.7405	0.7222	0.7313
0611	21.58	0.8278	0.6656	0.8103	0.7308	0724	24.58	0.7902	0.6947	0.7740	0.7322
0626	21.59	0.7936	0.7135	0.6862	0.6996	0717	24.61	0.7799	0.7294	0.7008	0.7147
0603	21.60	0.7879	0.7383	0.6802	0.7081	0716	26.07	0.8044	0.7368	0.7505	0.7436
0616	22.15	0.8244	0.7138	0.7340	0.7238	0715	27.11	0.7891	0.7897	0.7483	0.7684
0627	22.32	0.8374	0.7551	0.8043	0.7789	0728	27.22	0.8143	0.7546	0.7935	0.7736
0622	22.80	0.8295	0.7258	0.7706	0.7475	0720	27.45	0.7856	0.7827	0.7458	0.7638
0630	22.97	0.7964	0.7484	0.7420	0.7452	0711	27.88	0.8008	0.7344	0.8020	0.7667
0615	23.26	0.8165	0.7613	0.7671	0.7642	0703	29.32	0.7986	0.7692	0.7296	0.7489
0629	23.36	0.8209	0.7802	0.7460	0.7627	0722	29.51	0.7918	0.7401	0.7824	0.7607
0617	23.75	0.7864	0.7313	0.7313	0.7313	0719	30.23	0.7965	0.7526	0.7604	0.7565
						0718	31.07	0.7910	0.7227	0.7130	0.7178
						0708	31.27	0.8121	0.7171	0.8166	0.7636

departure delay over the subsequent flight legs of the aircraft and its primary departure delay at the start of the itinerary. The delay multiplier (D.M) can be described as -

$$D.M = (D_p + D_R)/D_p \quad (4)$$

Here,  $D_p$  is the primary departure delay and  $D_R$  denotes the reactionary departure delay. The D.M, here, measures the ratio of the total departure delay in a flight's itinerary to the primary departure delay that it witnessed. Fig. 11 shows the distribution of delay multipliers for 3 out of the 24 critical days. It is evident from the probability distribution of the delay multipliers that for these days, the majority of values lie between 1 and 1.5. This implies that for most of the aircraft, the primary departure delays do not significantly propagate and magnify into reactionary delays. Some minority cases are visible where D.M values are significantly higher, signifying higher delay propagation over the subsequent flight legs as reactionary delays.

## V. CONCLUSION

Delays are inherent in an air transportation network and are detrimental to this transportation system as a whole. Flight delays occur and propagate due to a multitude of factors and identifying the major sources of delay can help in containing delays, if not completely mitigating them. This paper presents a multi-agent based method to predict the reactionary delays in an airport network and adopts the SEA region a case study. Each flight is modeled as an agent which operates in airport environment.

Through this analysis, the reactionary departure delays in an aircraft's itinerary, emanating solely due to the primary departure delay witnessed by flights at the beginning of their first flight legs, are predicted. Each agent (flight), after experiencing an initial primary departure delay, operates in a dynamic airport environment consisting of other agents arriving and departing at stipulated times. Through the complex interaction of a multitude of agents with the airport environment, the primary delay of agent cascades further

**TABLE 8.** Delay classification results for September and October 2016. Acc – accuracy; Prec – precision; F1-S – F1 Score.

mm-dd	AvgDly	Acc	Prec	Recall	F1-S	mm-dd	AvgDly	Acc	Prec	Recall	F1-S
0913	16.85	0.8142	0.5853	0.8750	0.7014	1010	17.03	0.7863	0.5762	0.7683	0.6585
0924	17.10	0.7968	0.6339	0.7695	0.6952	1003	17.27	0.8018	0.6702	0.8182	0.7368
0911	17.18	0.7832	0.5934	0.8491	0.6986	1021	17.35	0.8028	0.6510	0.7851	0.7118
0912	17.33	0.8033	0.6269	0.9145	0.7439	1001	17.97	0.7827	0.6187	0.8084	0.7009
0914	17.41	0.7781	0.6511	0.7572	0.7001	1026	18.15	0.8122	0.7028	0.8112	0.7531
0917	17.55	0.7913	0.6684	0.8113	0.7330	1013	18.40	0.8136	0.7069	0.7482	0.7270
0903	17.61	0.7942	0.6331	0.8278	0.7175	1024	18.82	0.8502	0.7166	0.8462	0.7760
0902	18.40	0.7554	0.6044	0.7285	0.6607	1006	18.91	0.8333	0.7273	0.8553	0.7861
0916	18.83	0.7734	0.6429	0.7814	0.7054	1022	18.94	0.8345	0.7182	0.7799	0.7478
0921	18.91	0.8073	0.7009	0.7810	0.7388	1012	19.45	0.8235	0.7101	0.8167	0.7597
0905	18.98	0.7690	0.6322	0.7666	0.6929	1009	19.53	0.8311	0.7360	0.8415	0.7852
0922	19.08	0.7863	0.6632	0.7950	0.7232	1004	19.97	0.8185	0.7006	0.8464	0.7666
0906	19.19	0.7875	0.6462	0.8046	0.7167	1020	20.96	0.8128	0.6728	0.8185	0.7385
0925	19.63	0.7966	0.6717	0.7621	0.7141	1011	21.09	0.8269	0.6686	0.8582	0.7516
0918	19.88	0.8106	0.7273	0.8085	0.7657	1025	21.16	0.7998	0.6773	0.8309	0.7463
0915	19.95	0.8087	0.6771	0.8258	0.7441	1007	21.61	0.7867	0.7215	0.7703	0.7451
0908	20.49	0.8012	0.7002	0.8169	0.7541	1002	21.76	0.8073	0.7291	0.8106	0.7676
0910	20.54	0.7927	0.7242	0.7580	0.7407	1015	21.92	0.8061	0.6975	0.8017	0.7460
0926	20.97	0.8475	0.7414	0.8144	0.7762	1005	21.96	0.8162	0.7026	0.8164	0.7552
0919	21.12	0.7934	0.6585	0.8333	0.7357	1028	22.15	0.7860	0.6938	0.7500	0.7208
0907	21.40	0.8052	0.7392	0.7880	0.7628	1019	22.57	0.8157	0.7262	0.8112	0.7663
0923	21.85	0.7624	0.7047	0.7449	0.7243	1018	23.06	0.8193	0.7500	0.8425	0.7936
0927	21.97	0.8362	0.7108	0.8259	0.7640	1008	23.08	0.8028	0.7324	0.8005	0.7649
0920	23.79	0.8021	0.7368	0.8214	0.7768	1029	23.11	0.8110	0.7014	0.8127	0.7529
0904	24.08	0.8095	0.7634	0.7721	0.7677	1031	23.72	0.8215	0.7205	0.7962	0.7564
0928	24.80	0.8224	0.7820	0.7982	0.7900	1016	23.95	0.8209	0.7593	0.8025	0.7803
0930	25.87	0.7979	0.7920	0.7885	0.7903	1023	24.30	0.8165	0.7600	0.7308	0.7451
0929	27.48	0.8274	0.7819	0.7983	0.7900	1017	25.76	0.8019	0.6843	0.8462	0.7567
0909	28.68	0.8194	0.8251	0.8220	0.8235	1027	25.85	0.8147	0.7867	0.7925	0.7896
						1030	26.47	0.8176	0.7731	0.7809	0.7769
						1014	27.84	0.7749	0.6810	0.7419	0.7102

into reactionary delay. Such predictions can aid in better flight schedule planning through prior knowledge of potential delays on different flight legs, when the prediction of other events which may cause delays, such as en-route weather, airport services and equipment being unserviceable, are significantly uncertain.

The results on classification of flights as delayed or un-delayed show an average accuracy of 80.7% for a period of 6 months with a delay threshold of 15 minutes. This value is at par with the existing literature on delay propagation prediction. This method utilizes the information of the primary value of departure delay and the scheduled arrival and departure times of the agents (flights), unlike the other machine learning algorithms requiring multiple features, training and test sets. The evaluation of the delay multiplier for the SEA region highlights that the majority of values are in the range of 1-1.5. As the values of D.M increase, it suggests an increased contribution of the reactionary delays to the overall delay associated with an aircraft's itinerary. In other words, magnification and cascading of primary delays into reactionary components. The obtained

range of D.M (1-1.5) indicates that for the SEA region, the primary delays do not significantly cascade and magnify to reactionary delays for most aircraft itineraries. There are some cases of itineraries where the magnified cascading effect of primary delays are witnessed. This method also enables evaluation of individual flight legs where the departure delay propagation is significantly high and at what stage the delay is absorbed. For instance, in Fig. 10, for the itinerary WADD-WMCK-WSSS-WADD-WIII, the departure delays of 86 minutes in the 4<sup>th</sup> flight leg is absorbed during the departure of 5<sup>th</sup> flight leg due to the turn around time of 2 hours and 28 minutes.

The methodology developed in this research is simple and generic but is able to provide key insights on the case study of the SEA Airport network and the behavior of delays in this region. This research problem can further be fine tuned by modelling the flight crew, ground staff and passengers as agents affecting the delay and contributing to the delay propagation effect. An analysis of these factors on individual aircraft's primary and reactionary delays can be done. Also, with the information of passenger distribution to and from

**TABLE 9.** Delay classification results for November and December 2016. Acc – accuracy; Prec – precision; F1-S – F1 Score.

mm-dd	AvgDly	Acc	prec	Recall	F1-S	mm-dd	AvgDly	Acc	Prec	Recall	F1-S
1112	16.79	0.8028	0.6339	0.7891	0.7030	1214	17.89	0.8039	0.6978	0.7634	0.7291
1119	16.80	0.7922	0.6033	0.7631	0.6738	1231	18.35	0.7984	0.6543	0.8149	0.7258
1113	17.03	0.7823	0.6453	0.7446	0.6914	1225	19.39	0.7936	0.6793	0.8023	0.7357
1130	17.91	0.8143	0.6891	0.7649	0.7250	1213	20.40	0.8090	0.7263	0.7933	0.7583
1122	18.04	0.8206	0.6828	0.8194	0.7449	1201	20.64	0.8134	0.7632	0.7580	0.7606
1127	18.08	0.8147	0.6824	0.7448	0.7122	1211	20.85	0.8208	0.7677	0.7694	0.7685
1121	18.48	0.8228	0.6826	0.7834	0.7296	1206	21.69	0.8055	0.7275	0.7678	0.7471
1104	18.72	0.8081	0.6863	0.8174	0.7462	1208	21.72	0.7984	0.7285	0.8075	0.7660
1109	19.00	0.8223	0.6898	0.8058	0.7433	1212	21.96	0.8065	0.7047	0.7850	0.7427
1105	19.25	0.8299	0.6970	0.7830	0.7375	1219	22.09	0.8713	0.8491	0.8517	0.8504
1108	19.98	0.8107	0.6533	0.8133	0.7246	1229	22.47	0.8061	0.7473	0.7896	0.7679
1106	20.07	0.8237	0.7316	0.8043	0.7663	1202	22.87	0.8032	0.7533	0.7788	0.7658
1123	20.10	0.8320	0.7093	0.8110	0.7568	1230	23.57	0.8092	0.7741	0.8067	0.7901
1128	20.51	0.8122	0.6869	0.7946	0.7368	1205	24.00	0.8027	0.6955	0.7771	0.7341
1120	20.63	0.8320	0.7462	0.8179	0.7802	1209	24.06	0.7922	0.7790	0.7495	0.7640
1126	20.65	0.7864	0.6821	0.7762	0.7261	1224	24.38	0.7841	0.7628	0.7791	0.7709
1129	20.92	0.8170	0.6856	0.8309	0.7513	1215	24.77	0.7783	0.8025	0.7253	0.7619
1114	21.12	0.8546	0.8141	0.8249	0.8195	1204	24.93	0.7698	0.7348	0.7416	0.7382
1118	21.15	0.8285	0.7469	0.7815	0.7638	1210	25.22	0.8136	0.8043	0.7860	0.7950
1124	21.26	0.7953	0.7363	0.7260	0.7311	1227	25.62	0.7989	0.7490	0.8039	0.7755
1117	21.31	0.8362	0.7615	0.8137	0.7868	1218	26.15	0.7939	0.7818	0.7518	0.7665
1110	21.32	0.8346	0.7582	0.8177	0.7868	1217	26.23	0.7849	0.7852	0.7557	0.7702
1125	21.47	0.7899	0.7088	0.7468	0.7273	1221	26.51	0.8068	0.8401	0.7750	0.8062
1107	21.80	0.7875	0.6628	0.8383	0.7403	1228	26.91	0.7897	0.7761	0.7842	0.7801
1116	22.00	0.7779	0.6954	0.7466	0.7201	1226	27.10	0.7744	0.7002	0.7506	0.7246
1111	22.24	0.8104	0.7299	0.7586	0.7440	1207	28.28	0.8125	0.8051	0.7957	0.8004
1103	24.49	0.8175	0.7844	0.8224	0.8029	1220	28.36	0.7850	0.7793	0.7555	0.7672
1115	24.59	0.8115	0.7161	0.8258	0.7671	1203	28.53	0.7949	0.8012	0.7221	0.7596
1101	25.88	0.8213	0.7622	0.8386	0.7986	1222	29.01	0.7791	0.8012	0.7416	0.7702
1102	26.78	0.7992	0.7648	0.7685	0.7667	1223	29.53	0.7730	0.7966	0.7472	0.7711
						1216	30.45	0.8069	0.8163	0.7989	0.8075

flights, this research can be extended to evaluate delay propagation trees originating from individual flights, i.e the effect of delay of one flight on the other flights who have connecting passengers from it.

## ANNEX

The tables below show the classification accuracy for each day, based on a delay threshold of 15 minutes.

## REFERENCES

- [1] A. Sternberg, J. Soares, D. Carvalho, and E. Ogasawara, "A review on flight delay prediction," 2017, *arXiv:1703.06118*. [Online]. Available: <https://arxiv.org/abs/1703.06118>
- [2] IATA. *2036 Forecast Reveals Air Passengers Will Nearly Double to 7.8 Billion*. Accessed: Dec. 2, 2019. [Online]. Available: <https://www.iata.org/pressroom/pr/pages/2017-10-24-01.aspx>
- [3] S. AhmadBeygi, A. Cohn, Y. Guan, and P. Belobaba, "Analysis of the potential for delay propagation in passenger airline networks," *J. Air Transp. Manage.*, vol. 14, no. 5, pp. 221–236, 2008.
- [4] R. Beatty, R. Hsu, L. Berry, and J. Rome, "Preliminary evaluation of flight delay propagation through an airline schedule," *Air Traffic Control Quart.*, vol. 7, no. 4, pp. 259–270, 1999.
- [5] M. O. Ball, C. Barnhart, M. Dresner, M. Hansen, K. Neels, A. Odoni, E. Peterson, L. Sherry, A. Trani, and B. Zou, "Total delay impact study: A comprehensive assessment of the costs and impacts of flight delay in the united states," Nat. Centre Excellence Aviation Oper. Res., Washington, DC, USA, Tech. Rep., 2010.
- [6] B. Campanelli, P. Fleurquin, A. Arranz, I. Etxebarria, C. Ciruelos, and V. M. Eguiluz, and J. J. Ramasco, "Comparing the modeling of delay propagation in the us and European air traffic networks," *J. Air Transp. Manage.*, vol. 56, pp. 12–18, Sep. 2016.
- [7] P. Fleurquin, J. J. Ramasco, and V. M. Eguiluz, "Systemic delay propagation in the us airport network," *Sci. Rep.*, vol. 3, p. 1159, Jan. 2013.
- [8] P. Fleurquin, J. J. Ramasco, and V. M. Eguiluz, "Data-driven modeling of systemic delay propagation under severe meteorological conditions," 2013, *arXiv:1308.0438*. [Online]. Available: <https://arxiv.org/abs/1308.0438>
- [9] M. O. Ball, R. Hoffman, and A. Mukherjee, "Ground delay program planning under uncertainty based on the ration-by-distance principle," *Transp. Sci.*, vol. 44, no. 1, pp. 1–14, 2010.
- [10] Y. Liu and M. Hansen, "Incorporating predictability into cost optimization for ground delay programs," *Transp. Sci.*, vol. 50, no. 1, pp. 132–149, 2015.
- [11] M. O. Ball, C.-Y. Chen, R. Hoffman, and T. Vossen, "Collaborative decision making in air traffic management: Current and future research directions," in *New Concepts and Methods in Air Traffic Management*. Berlin, Germany: Springer, 2001, pp. 17–30.
- [12] B. Sridhar, S. R. Grabbe, and A. Mukherjee, "Modeling and optimization in traffic flow management," *Proc. IEEE*, vol. 96, no. 12, pp. 2060–2080, Dec. 2008.
- [13] A. K. Agogino and K. Tumer, "A multiagent approach to managing air traffic flow," *Auto. Agents Multi-Agent Syst.*, vol. 24, no. 1, pp. 1–25, 2012.
- [14] E. Mueller and G. Chatterji, "Analysis of aircraft arrival and departure delay characteristics," in *Proc. AIAA's Aircr. Technol. Integr., Oper. (ATIO) Tech. Forum*, 2002, p. 5866.
- [15] S. Bratu and C. Barnhart, "An analysis of passenger delays using flight operations and passenger booking data," *Air Traffic Control Quart.*, vol. 13, no. 1, pp. 1–27, 2005.

- [16] H. Ali, Y. Guleria, S. Alam, V. N. Duong, and M. Schultz, "Impact of stochastic delays, turnaround time and connection time on missed connections at low cost airports," in *Proc. 13th USA/Eur. Air Traffic Manage. R&D Seminar*, Vienna, Austria, Jun. 2019.
- [17] Y. Tu, M. O. Ball, and W. S. Jank, "Estimating flight departure delay distributions a statistical approach with long-term trend and short-term pattern," *J. Amer. Stat. Assoc.*, vol. 103, no. 481, pp. 112–125, 2008.
- [18] S. Choi, Y. J. Kim, S. Briceno, and D. Mavris, "Prediction of weather-induced airline delays based on machine learning algorithms," in *Proc. IEEE/AIAA 35th Digit. Avionics Syst. Conf. (DASC)*, Sep. 2016, pp. 1–6.
- [19] S. Bouarfa, H. A. Blom, R. Curran, and M. H. Everdij, "Agent-based modeling and simulation of emergent behavior in air transportation," *Complex Adapt. Syst. Model.*, vol. 1, no. 1, p. 15, 2013.
- [20] N. Pyrgiotis, K. M. Malone, and A. Odoni, "Modelling delay propagation within an airport network," *Transp. Res. C, Emerg. Technol.*, vol. 27, pp. 60–75, Feb. 2013.
- [21] J. J. Rebollo and H. Balakrishnan, "Characterization and prediction of air traffic delays," *Transp. Res. C, Emerg. Technol.*, vol. 44, pp. 231–241, 2014.
- [22] L. Belcastro, F. Marozzo, D. Talia, and P. Trunfio, "Using scalable data mining for predicting flight delays," *ACM Trans. Intell. Syst. Technol. (TIST)*, vol. 8, no. 1, p. 5, 2016.
- [23] N. Xu, G. Donohue, K. B. Laskey, and C.-H. , "Estimation of delay propagation in the national aviation system using Bayesian networks," in *Proc. 6th USA/Europe Air Traffic Manage. Res. Develop. Seminar*, Baltimore, MD, USA, 2005.
- [24] G. Wu and K. Law, "Modelling the delay propagation effects of multiple resource connections in an airline network using a Bayesian network model," *Transp. Res. E, Logistics Transp. Rev.*, vol. 122, pp. 62–77, 2019.
- [25] J. Chen and M. Li, "Chained predictions of flight delay using machine learning," in *Proc. AIAA Scitech Forum*, 2019, p. 1661.
- [26] M. Pěchouček, D. Šišlák, D. Pavlíček, and M. Uller, "Autonomous agents for air-traffic deconfliction," in *Proc. 5th Int. Joint Conf. Auton. Agents Multiagent Syst.*, 2006, pp. 1498–1505.
- [27] K. Tumer and A. Agogino, "Distributed agent-based air traffic flow management," in *Proc. 6th Int. Joint Conf. Auto. Agents Multiagent Syst.*, 2007, p. 255.
- [28] M. Ljungberg and A. Lucas, "The oasis air traffic management system," Austral. Artif. Intell. Inst., Melbourne, VIC, Australia, Tech. Rep. 28, 1992.
- [29] CODA. (Oct. 2018). *All-Causes Delay To Air Transport In Europe*. Accessed: Dec. 02, 2019. [Online]. Available: <https://www.eurocontrol.int/publication/all-causes-delay-air-transport-europe-july-2019>
- [30] B. Chen and H. H. Cheng, "A review of the applications of agent technology in traffic and transportation systems," *IEEE Trans. Intell. Transp. Syst.*, vol. 11, no. 2, pp. 485–497, Jun. 2010.
- [31] M. R. Friesen and R. D. McLeod, "A Survey of Agent-Based Modeling of Hospital Environments," *IEEE Access*, vol. 2, pp. 227–233, 2014.
- [32] L. Mastroeni, P. Vellucci, and M. Naldi, "Agent-based models for opinion formation: A bibliographic survey," *IEEE Access*, vol. 7, pp. 58836–58848, 2019.
- [33] G. Weiss, *Multiagent Systems: A Modern Approach to Distributed Artificial Intelligence*. Cambridge, MA, USA: MIT Press, 1999.
- [34] S. Franklin and A. Graesser, "Is it an agent, or just a program?: A taxonomy for autonomous agents," in *Proc. Int. Workshop Agent Theories, Archit. Lang.* Berlin, Germany: Springer, 1996, pp. 21–35.
- [35] C. Ciruelos, A. Arranz, I. Etxebarria, S. Peces, B. Campanelli, P. Fleurquin, V. Eguluz, and J. Ramasco, "Modelling delay propagation trees for scheduled flights," in *Proc. 11th USA/EUROPE Air Traffic Manage. R&D Seminar; Lisbon, Portugal*, 2015, pp. 23–26.
- [36] FAA. *ASQP-Definition Of Variables*. Accessed: Dec. 02, 2019. [Online]. Available: [https://aspmhelp.faa.gov/index.php/asqp\\_definitions\\_of\\_variables](https://aspmhelp.faa.gov/index.php/asqp_definitions_of_variables)
- [37] *Manual of Standards—Air Traffic Services*, Singapore, Civil Aviation Authority of Singapore, 2014.
- [38] A. Barrat, M. Barthélemy, R. Pastor-Satorras, and A. Vespignani, "The architecture of complex weighted networks," *Proc. Nat. Acad. Sci. USA*, vol. 101, no. 11, pp. 3747–3752, Mar. 2004.
- [39] R. Guimera, S. Mossa, A. Turttschi, and L. N. Amaral, "The worldwide air transportation network: Anomalous centrality, community structure, and cities," *global roles*," *Proc. Nat. Acad. Sci. USA*, vol. 102, no. 22, pp. 7794–7799, 2005.
- [40] M. E. J. Newman, *Networks: An Introduction*. Oxford, U.K.: Oxford Univ. Press, 2010.

- [41] H. Fricke and M. Schultz, "Delay impacts onto turnaround performance," in *Proc. 8th USA/Europe Air Traffic Manage. R&D Seminar*, California, CA, USA, Jun./Jul. 2009.
- [42] M. Schultz, T. Kunze, B. Oreschko, and H. Fricke, "Microscopic process modelling for efficient aircraft turnaround management," in *Proc. 4th Int. Air Transp. Oper. Symp.*, Delft, The Netherlands, Jul. 2013.



**YASH GULERIA** received the B.S. degree in mechanical engineering from NIT Hamirpur, India, and the M.S. degree in aerospace engineering from Nanyang Technological University and TU Munich, in a joint program. He is currently a Research Associate with the Air Traffic Management Research Institute, Nanyang Technological University, Singapore. His research interests include delay propagation analysis, conflict detection and resolution, and air side operation in air traffic management.



**QING CAI** received the B.S. degree in electronic information engineering from Wuhan Textile University, Wuhan, China, in 2010, and the Ph.D. degree in pattern recognition and intelligent systems from the School of Electronic Engineering, Xidian University, Xi'an, China, in 2015. His current research interests are in the area of artificial intelligence, complex network analytics, recommender systems, air traffic management, and population ecology.



**SAMEER ALAM** received the M.S. degree in computer science from the Birla Institute of Technology, Mesra, India, in 1999, and the Ph.D. degree in computer science from the University of New South Wales, Canberra, Australia, in 2008. He is currently an Associate Professor with the School of Mechanical and Aerospace Engineering, Nanyang Technological University. His research interests include simulation, modeling, risk assessment, and the optimization of advanced air traffic management concepts. He is the Founding Member of the IEEE Computational Intelligence Society Chapter in Canberra. He is also the Chief Investigator of ICAOMIDRMA Collision Risk Project and served as a Co-Chief Investigator for several projects of Air Service Australia and EUROCONTROL. He was a recipient of the Australian National University Science Medal, in 2011, the Tall Poppy Science Award, in 2011, and the Fresh Science Award, in 2009.



**LISHUAI LI** received the B.Eng. degree in aircraft design and engineering from Fudan University and the M.Sc. and Ph.D. in air transportation systems from the Department of Aeronautics and Astronautics, Massachusetts Institute of Technology (MIT). She is currently an Assistant Professor with the Department of Systems Engineering and Engineering Management, City University of Hong Kong. Before joining CityU, she was a Consultant at McKinsey & Company, San Francisco. She is interested in innovative methods for the design, management, and operation of air transportation systems and drawing expertise in big data and information technology. She develops data analysis and decision support tools to improve aircraft operations, airspace efficiency, and system safety. Her current research focuses on the application of data mining techniques in airline safety management.

...