Journal of Air Transport Management 40 (2014) 78–85

Contents lists available at ScienceDirect

Journal of Air Transport Management

journal homepage: www.elsevier.com/locate/jairtraman

Improving the IATA delay data coding system for enhanced data analytics

Cheng-Lung Wu*, Tiffany Truong

School of Aviation, UNSW Australia, Kensington, NSW 2052, Australia

A R T I C L E I N F O

Article history:
Available online

Keywords:
Delay codes
IATA delay coding scheme
Delay propagation
Propagation tree
Data analytics
Airline delays

A B S T R A C T

Aviation delays inconvenience travelers and result in financial losses for stakeholders. Without complex data pre-processing, delay data collected by the existing IATA delay coding system are inadequate to support advanced delay analytics, e.g. large-scale delay propagation tracing in an airline network. Consequently, we developed three new coding schemes aiming at improving the current IATA system. These schemes were tested with specific analysis tasks using simulated delay data and were benchmarked against the IATA system. It was found that a coding scheme with a well-designed reporting style can facilitate automated data analytics and data mining, and an improved grouping of delay codes can minimise potential confusion at the data entry and recording stages.

1. Background

Flight delays have always been costly. The extent and severity of their presence vary depending on the stages at which the delays appear. Despite the growing global demand for air transport, increases in space capacity have been slow at major airports in most countries. Managing delays in aviation has emerged as an immediate and sustainable option to relieve congestion at busy airports while waiting for air transport infrastructure (airport capacity in particular) to grow.

In the aviation industry, delays have long been a complicated and multidimensional issue because of numerous factors, such as the many causes of delays, the complex chain reaction of delays in airline networks, and, most notably, the consequences of delays such as passengers left stranded and flight cancelations (Sarseshiki et al., 2010). Delays may occur at any point of time in airline operations and are seldom the result of a single factor (Niehues et al., 2001; Wu, 2010). Since delays can occur at any stage, some are more manageable than others. For example, en-route (airborne) delays are mostly out of the airline management's reach and are dealt with mainly by air traffic control authorities as well as the flight crew. Hence because of the costs and complexities surrounding delays, Sarseshiki et al. (2010) noted that understanding delays requires more research in the field of delays to determine and evaluate the main (root) factors of delay, also considering other conditions contributing to incidences of delay.

High quality delay data are essential in post-operation analysis to guide future operational and scheduling improvements. The current practice in the airline industry is to adopt the International Air Transport Association (IATA) delay coding system, a standard recommended in the Airport Handling Manual (AHM 730) published by IATA (IATA, 2012). Delay recording is typically performed by ground handling agents or airlines' own ground handling units, although air traffic control authorities and airports may also record their own data for airspace/airport capacity monitoring purposes. The data collected by different parties is then used to serve their own individual purposes. For airlines, this data is used mostly for operational procedures and schedule improvements to reduce operating costs.

2. Current practice of delay data recording and analysis

The current data collected by most carriers follow the IATA coding system for meeting the requirements of delay data reporting to aviation authorities in various countries. As summarised in Table 1, the format usually includes twelve “columns” in a tabular form.

The first six columns record the basic flight information that is also available from an airline schedule. The actual departure and arrival times (the seventh and eighth columns) are either manually recorded by ground handling agents/airlines or automatically

* Corresponding author. Tel.: +61 2 9385 6767.
E-mail address: C.L.Wu@unsw.edu.au (C.-L. Wu).

http://dx.doi.org/10.1016/j.jairtraman.2014.06.001
0969-6997/

Table 1
Typical format of delay records using the IATA system.

Column no.	Data field	Column no.	Data field
1	Date of flight	7	Actual departure time
2	Aircraft registration no.	8	Actual arrival time
3	Origin–destination airports	9	Delay code/s
4	Flight number	10	Delay minutes
5	Scheduled departure time	11	Delay reason/s
6	Scheduled arrival time	12	Delay comments

transmitted via the Aircraft Communications Addressing and Reporting System (ACARS) equipment aboard an aircraft (Wu, 2010). The ninth column records the delay code or codes (often following the IATA system) with the corresponding delay minutes recorded in the tenth column for each individual delay code (if more than one is recorded). Column eleven also corresponds to column nine, providing a textual explanation of the reason for the delay. Further delay-related information is then manually noted in the twelfth column, *Delay Comments*. Some airlines allow the use of multiple delay codes for multiple causes to a single delay incident. When there are multiple delay codes, columns nine through twelve are repeated for each individual delay code.

The present layout, as seen in Table 1, has provided most airlines with sufficient data to investigate flight delays. Statistical analyses, such as delay frequency and average delay minutes by flight, can be easily conducted with these data. Given the networked nature of airline schedules, flight delays are often caused by resource connections and airline scheduling, such as delay propagation (Jetzki, 2009; Wu, 2010). For instance, departure delays for one flight may cause arrival delays for another flight (or flights) at the destination airport. Since the same aircraft is scheduled to serve a following flight at the arriving airport, inbound delays due to late aircraft or late crew boarding may lead to departure delays of other flights to which the passengers, crew, and aircraft connect. To combat the phenomenon of delay propagation and its impacts on airline operations and revenue losses, airlines have recently paid more attention to the study of delay propagation and robust scheduling practices (AhmadBeygi et al., 2008; Diana, 2009; Dunbar et al., 2012; Lan et al., 2006).

Delay data recorded in the format displayed in Table 1 pose several challenges for airlines when conducting advanced delay analytics. The first challenge involves resources tracing for both aircraft and crew in post-operations analysis. The current delay data format contains only information about delayed flights, and often only departure delays and delay codes are recorded; en-route delay information is not logged. Therefore, an analyst often needs to combine flight and crew schedule data to trace resource connections in actual operations. Flight schedule data provide essential aircraft routing information that allows an analyst to connect flights and rebuild a "connection network" to trace delay propagation due to aircraft routing. Crew schedule data are used for a similar purpose, building a crewing network to trace delays due to crewing issues in operations. Although aircraft and crew schedules are readily available within an airline, it is not a trivial task to combine and match the two sources with delay records, because both schedules can be changed during operations. Moreover, delays classified under codes such as "aircraft rotation" (IATA code 93) may actually contain more information than is represented by a single delay code; further information is therefore recorded in the *Delay Comments* column for reference purposes.

The second challenge occurs when an analyst tries to explore delay propagation on a network scale. Once the connection network of resources is built, one only has information about the relationship between two connected flights via recorded delay codes (e.g., code 93, aircraft rotation). To include further information, manual processing of the records held in the *Delay Comments* column is necessary (see Table 2 for examples). Comprehensive and rich information is often recorded under this column in a textual format, and an analyst has to analyse this text manually to find out the true causes of delays. The current challenge with the IATA coding system is that this comments column is generally relied upon heavily in both data entry and after-operations delay analysis to attain better explanations of the causes of delays. This also means that crucial information is currently stored in a textual form without a standard format or style, making the process of extracting relevant information for advanced delay analysis difficult and time-consuming. Several excerpts from actual delay data from an airline are provided in Table 2.

The example data in Table 2 demonstrate that rich operational information not included elsewhere in the coding was noted by the airline within the *Delay Comments* column. Although delay codes were recorded for each delay incident, delay codes alone could not explain clearly the true delay causes occurring in actual operations. Due to time constraints and attempts to retain all necessary delay-relevant information, the comments were abbreviated in such a way that, in many cases, only the reporter could understand them. For those who analyse this post-operation data, it consequently becomes confusing and tedious to extrapolate, as well as understand the rich information held in the *Delay Comments* column. For a large network with a great deal of delay data (on the order of hundreds of thousands of records), it becomes impossible for an analyst to examine and analyse the mass quantity of information held within the comments section. This has been the biggest challenge for airlines seeking to explore operational bottlenecks via post-operational data mining exercises. Given the amount of information in the comments section, improvements in the data delay recording format or advanced "text-parsing" pre-processing techniques applied to delay data before analysis are necessary to achieve automated data mining and reporting.

In conjunction with the delay reporting style, the third challenge involves the limited quality of recorded delay data due to the current form of data entry and heavy reliance on the *Delay Comments* column to supplement delay codes (Niehues et al., 2001; Speri, 2006). The ultimate goal of delay data recording and reporting is to investigate and determine the sources of delays in past operations so that improvements can be made for the future. Given the heavy reliance on the comments column, analysis results are drawn mostly based on analysts' "understanding" of the information noted in this column. In the current method of data recording, the notes taken by ground agents are inconsistent in

Table 2
Excerpt of delay data of Airline X (domestic operations by a narrow body aircraft).

Route	Flight no.	Delay code	Delay min	Delay reason	Delay comments
C–D	4	94	5	Cabin crew connection	CC CONX EX 10 LATE BRD FOR SSRS AT 1352
C–D	4	70	5	Late cabin preparation	CC CONX EX 10 LATE BRD FOR SSRS AT 1353
D–E	5	93	14	Aircraft rotation/ delays upline	UPLINE DELAY EX 4 ATA 1525
E–D	6	93	6	Aircraft rotation/ delays upline	UPLINE DELAY EX 5 ATA 1526
D–C	7	94	3	Guest convenience	1XINF OFF 30, SSR PERM-173XDPL ON VERY SLOW TO BRD, ALL REQ. CHAIRS TO SEATFREIGHT ADDED AFT. LIR COMP. NOT NCL. ON P/WORK

style across data entries. This is particularly true when aircraft ground operations are conducted by different agents and/or in different countries. As a consequence, many industry managers tend to show a "reserved" position toward the quality of delay data (Speri, 2006).

The fourth challenge concerns the delay codes that are used in the current IATA system. It has been argued by both industry practitioners and researchers that the current set of IATA delay codes no longer effectively describes the delays encountered in modern airline operations. Some current codes also overlap, making it hard for the reporter to distinguish between the codes and use the appropriate code or codes to describe a delay incident for statistical analysis (Jetzki, 2009). Although the current IATA coding system has recently been augmented by the use of sub-codes on air traffic flow management delays (IATA, 2012), this change has not addressed the issues of code overlapping, sometimes confusing codes, and the inadequacy of the current set of delay codes in explaining complex delay incidents.

In 2001, a federal register notice by the U.S. Department of Transportation (DoT) proposed improving the airline delay and cancelation data reporting style by reorganising existing IATA delay codes into seven major codes with the assistance of a comprehensive data reporting style (U.S. DoT, 2001). For instance, the proposed system replaced IATA delay codes with seven major codes such as Code A – Carrier Caused, Code B – Extreme Weather, Code D – Late Arriving Aircraft, and so on. This format was proposed to provide the DoT with statistical data that can be accessed by interested parties. However, the information gathered using this reporting style would not allow for in-depth network and operational analyses, because the proposed delay codes were broadly defined and not specific enough to identify delay causes for operational improvements.

Improvements are required to the existing IATA system such that the recorded delay data quality can meet the needs of both basic and advanced delay analysis and the choice of delay codes in practice can be standardised across the industry. Changing the reporting style and delay codes also raises the question of whether more or less delay information needs to be collected to ensure that all aspects of delays are well recorded without causing overwhelming data entry workloads in daily airline operations. At present, most airlines collect and examine information only on departure delays, whereas en-route and arrival delays are not well recorded. There are no coding standards or reporting schemes specifically allocated to collect data on en-route and arrival delays (Wu, 2010). Consequently, airlines and researchers primarily investigate the nature of departure delays, creating a gap in the understanding of all delays within flight operations (Jetzki, 2009).

The objective of this paper is to improve the existing IATA delay coding system by creating a more effective delay coding scheme and reporting framework that can be easily adopted universally by the industry. Specifically, this paper aims to produce a delay coding scheme that will enable airlines to record data with quality and uniformity that also allow delay data owners to conduct automated diagnoses of network problems via computer scripts without the pre-processing of delay data or manual intervention of analysts.

3. Methodology

Most airlines collect delay data for analysis and/or reporting purposes. The styles of delay coding and reporting schemes adopted by airlines vary according to individual needs, but the similarities are based mostly on the IATA coding system. Airlines may use a single delay code or multiple codes for a delay incident (i.e., a delayed flight) with the original IATA system. Others develop in-house "sub-codes" (i.e., a "sub-layer" of codes under the IATA system), reorganise codes into different categories based on the IATA system, develop new codes to augment the IATA system, or combine several of the above methods (see Lee and Moore, 2003 for the coding scheme developed by Air New Zealand). In the following analysis, we develop three new delay coding and reporting schemes (referred to as the "coding schemes" hereafter in the paper) based on the IATA system and the in-house coding schemes provided by two airlines in Asia. We then use a simulated schedule and simulated delay data to assess the performance of each coding scheme by conducting specific delay analysis tasks commonly performed by airline analysts.

3.1. Rationale and new coding schemes

The format of the current IATA standard coding system has been previously provided in Table 1. Comparing the set of sample delay data in Table 2 with Table 1, we can see that essential operational and delay-related data are stored within four columns: *Delay Code*, *Delay Minutes*, *Delay Reason*, and *Delay Comments*. Given the previously discussed drawbacks of using the current IATA coding system (and some of its variants), the development rationale of the new coding schemes focuses on two main areas: (1) improving the effectiveness (clarity) of delay codes and (2) improving the reporting style to enhance data readiness for automated analysis. Delay codes are improved by creating new codes and/or reorganising code groups to minimise the confusion of similar delay codes while maintaining a standard in the usage of delay codes (i.e., improving effectiveness/clarity of delay codes). The reporting style is improved by changing the way delay data are recorded. More specifically, the reporting style is changed to a structure that systematically absorbs the "texts" that are currently stored in the *Delay Comments* column. Based on this rationale, three sets of coding schemes were developed as follows. Set 1 adopts the original IATA codes with an improved coding and reporting style. Set 2 consolidates the existing IATA codes with improved clarity for delay coding and minimum deviation from the existing IATA data entry style. Set 3 retains the structure of the existing IATA system and adds "depth" to delay codes by using sub-codes (under certain delay codes) to eliminate the possibility of code entry confusion by ground agents.

3.1.1. Set 1: improved reporting style with the original IATA coding system

Set 1 is developed on the basis of minimal changes to the current IATA coding system, with a strong focus on the actual reporting style. The objective of Set 1 is to minimise the use of the *Delay Comments* column by creating four extra columns. The new reporting format of Set 1 has the style shown in Table 3. Columns 1–11 (please refer to Table 1 for column numbering) from the original IATA system are preserved in the Set 1 coding scheme. The *Delay Comments* column is now moved to column 16, where further comments can still be recorded for reference purposes.

The following four columns (columns 12–15 in Table 3) are added in Set 1:

- *Pre-flight Number* – This insertion option will appear in the reporting format only when a delay code from 91 to 96

Table 3
Set 1 reporting format.

Column 1–11 (same as IATA scheme)	12	13	14	15	16 (Same as IATA scheme)
	Pre-flight no	Rotation, connection	Aircraft Rego	Pax	Delay comments

(reactionary) is entered in Column 9. The flight number of the previous flight (or flights) directly affecting the flight being reported is recorded in this column. This column applies to all delay codes in the "Reactionary" category of the IATA system.

- *Rotation/Connection* – This column is concerned with delay code 91. Only when this code is chosen in Column 9, this optional column will appear, allowing the reporter to choose between crew (cabin, flight, entire, or technical), aircraft, loads, and passenger/baggage to provide further explanation of rotation- and connection-related causes of delay.
- *Aircraft Rego* – This column will appear only when code 46 is recorded in Column 9. Code 46 denotes aircraft swaps (aircraft change) due to service issues. Once the originally planned aircraft is changed, a new aircraft with a different registration will serve the flight, and the new aircraft's registration is recorded here for delay tracing purposes. Aircraft swap information is often stored in the *Delay Comments* column, but the registration number of the replacing aircraft is not always recorded. Without this specific information, delay propagation analysis is often compromised, because after-ops delay tracing analysis may not be able to "rebuild" a complete delay propagation tree by linking flights that are operated by different aircraft due to swapping.
- *Pax* – Only when codes 11–19 or code 91 is entered in Column 9, this *Passenger* column will become available for data entry. The number of passengers affected by delays on a particular flight can be entered here. This will facilitate delay analysis in which an analyst wishes to assess the "impact" of flight delays due to passenger handling and connection issues.

To facilitate the adoption of the Set 1 coding scheme, the delay codes in the "Reactionary" and "Miscellaneous" categories of the IATA system were consolidated. A comparison between the original IATA system and the Set 1 coding scheme for delay codes 91–99 is provided in Table 4. Code 91 in the new Set 1 scheme consolidates the original codes 91, 93, 94, and 95 by providing a new reporting format that can clearly distinguish between these codes during data entry.

Table 4
IATA and Set 1 coding scheme comparison (codes 91–99).

IATA coding scheme	Set 1 coding scheme
Reactionary	**Reactionary**
91: Passenger or load connection	91: Rotation/connection • Cabin crew (ex-code 94) • Flight crew (ex-code 95) • Technical crew (ex-code 95) • Entire crew (ex-code 95) • Aircraft (ex-code 93) • Load (ex-code 91) • Passenger/baggage (ex-code 91)
92: Through check-in error (passenger and baggage)	92: Through check-in error (ex-code 92)
93: Aircraft rotation	
94: Cabin crew rotation	
95: Crew rotation (flight or entire)	
96: Operations control, rerouting, diversion, consolidation, aircraft change for reasons other than technical	93: Operations control (ex-code 96)
Miscellaneous	**Miscellaneous**
97: Industrial action within own airline	94: Industrial action within own airline
98: Industrial action outside own airline	95: Industrial action outside own airline
99: Miscellaneous	96: Miscellaneous

3.1.2. *Set 2: revamped IATA coding system with improved reporting style*

Set 2 is built based on the identification of the codes most commonly used by airlines in actual practice. Two sets of delay data from two different airlines in Asia were used to identify the frequency of delay code usage. The rationale adopted for revamping the IATA coding system in Set 2 is based on two observations of industry practices in the analysis of these delay data. First, certain delay causes have significantly higher financial implications than others and should be collected with more detailed information (e.g., aircraft technical defects – code 41 and aircraft rotation – code 93). Other codes with low usage or low financial/operational implications are seen as trivial in data collection and analysis. Second, confusion occurs in delay data entry primarily owing to unclear and inadequate definitions of certain codes. For instance, code 96 (operations control) overlaps with code 46 (aircraft change). In an incident of delay due to aircraft change, either code 46 or code 96 could be used to explain the incident. However, the use of code 96 would require noting further information in the *Delay Comments* column, because code 96 represents a wide coverage of delay causes.

After processing delay datasets from the two samples, the frequency of code usage was calculated separately for each airline. These calculations were later compared across airlines to establish a new coding scheme. The new code grouping of Set 2 is presented along with the original IATA code grouping in Table 5. The consolidated structure of code groupings in Set 2 is created with the existing IATA structure in mind, maintaining as many similar groups as possible while eliminating unnecessary and confusing codes. This consideration also helps ease the efforts of staff members retraining in the future deployment of Set 2 in the industry.

To improve Set 2 data quality, the reporting style developed for Set 1 was slightly altered. The new reporting style of Set 2 retains column 12 (*Pre-flight No*) and column 16 (*Delay Comments*), and eliminates columns 13–15 (see Table 3 for column numbering). The simple reporting style of Set 2 is based on the fact that the new coding groups of Set 2 have already minimised code duplication and confusion during data entry, so the codes in Set 2 can now effectively represent delay causes in sufficient detail by adding column 12, *Pre-flight No*; this facilitates delay tracing in post-operations analyses. Set 2 represents a compromise that takes under consideration the efforts that would be required to retrain staff to use the new scheme and the ease of on-site data entry by ground handling agents. The adoption of the reporting style developed for Set 1 is also possible for Set 2.

3.1.3. *Set 3: extending the IATA system with sub-codes*

The focus of Set 3 is to retain the structure of the IATA coding system while adding depth and eliminating repetitive codes to

Table 5
IATA and Set 2 delay code grouping comparison.

IATA code groups	Set 2 code groups
0x: Internal	0x: Reactionary
1x: Passenger/baggage	1x: Passenger-boarding and services
2x: Cargo/mail	2x: Cargo
3x: Aircraft and ramp handling	3x: Crew
4x: Technical and aircraft equipment	4x: Shortages
5x: Damage/equipment failure	5x: Aircraft and equipment
6x: Flight operations and crewing	6x: Operations
7x: Weather	7x: Air traffic services
8x: Air traffic flow management and airport/authorities restrictions	8x: Airport and government authorities
9x: Reactionary	

reduce confusion for the user. Adding depth to the IATA codes relates to the specificity and clarification of descriptions of existing IATA codes. In the current system, certain IATA codes are general in nature and cover various delay causes. The use of these codes in data entry requires manual identification of specific causes of delay using notes recorded in the *Delay Comments* column. Through adding extra levels of codes (sub-codes), a delay cause can be well explained without resorting to notes in the *Delay Comments* column. This new coding scheme reduces the need for noting extra delay information in textual format and facilitates automated delay data analysis on a large scale. A comparison of code groupings (for codes 51–58, Damage to Aircraft and Equipment Failure) in the IATA and Set 3 schemes is provided in Table 6.

An added advantage of bringing in extra levels of information without changing the overarching IATA framework is that it allows airlines to simplify mandatory reporting for government authorities. With the Set 3 coding scheme, an airline can produce reports with the minimum amount of delay information aggregated to the top code group level without manually adjusting the data or exposing all their data to authorities.

Similarly to Set 2, Set 3 requires the use of a new reporting style to record key delay information for advanced data analytics. Without a new reporting style, Set 3 codes alone will serve to explain delays in further detail with sub-codes but will not facilitate in-depth investigations, because certain data would still be kept in the *Delay Comments* section. The reporting style of Set 3 is identical to that of Set 2 and is based on a similar rationale. This reporting style facilitates advanced analysis of delay tracing without having to extract reactionary delay information that would otherwise be found in textual format in the *Delay Comments* section.

3.2. Delay propagation tracing and coding scheme implementation

3.2.1. Study schedule, delay data, and model testing

In this proof-of-concept study, it is practically challenging to test the new coding schemes in an actual airline operating environment without costly ground staff data entry retraining and without affecting airline operations. As an alternative, we built a test schedule and used simulated delay data for the testing and validation of the coding schemes. The test schedule was built for domestic operation in Australia among six capital cities (including Sydney, Melbourne, Gold Coast, Brisbane, Cairns, and Perth), and was designed to operate by six narrow-body aircraft. A total of 30 flights were incorporated in the design of the test schedule, with sector lengths ranging from one and a half hours (e.g. Sydney–-Brisbane) to 5 h for cross-continent flights such as the Sydney–Perth sector. A short turnaround time (between 30 and 40 min) was planned for most ground operations.

The design of this study schedule allowed us to simulate tight flight connections and potential delay propagation impacts on an airline network with complex resource connections among aircraft, crew, and passengers. Delay scenarios and delay data were generated on an ad hoc basis for this case study and "recorded" by the respective requirements of the three new coding schemes as well as the original IATA system. Table 7 lists an example subset of the simulated delay data in Set 1 reporting style. Columns 12–15 are specifically designed in Set 1, while other columns follow the original IATA style.

For coding scheme comparison and benchmarking, we conducted six delay analysis tasks using the same test schedule and delay data. The six tasks were to obtain the following information:

1. Identification of flights affected by delays in the database
2. Number of flights affected by code X and Identification of number of minutes these flights were delayed by this particular code
3. Determination of whether any passengers were affected (and how many) by a specific delay cause (code X)
4. Determination of whether any flights were delayed due to delay propagation in the network
5. Number of flights delayed because of propagation due to delay code X
6. Identification of all connected delayed flights (affected by delay propagation) and sorting of these flights to build delay propagation trees for each incident

Delay data are commonly analysed for two major purposes in the industry: (1) developing key performance indicators for various operating units of an airline to benchmark operational performance and (2) exploring operational bottlenecks resulting from schedule planning (including aircraft routing and crewing) and passenger demands (in the form of passenger flight connections). The selected tasks 1–3 above can achieve the first purpose by generating all types of delay statistics for benchmarking purposes, including mean delay minutes and delay probabilities. Tasks 4–6 can achieve the second objective aiming to explore the network effects of an airline schedule from both planning and operational perspectives. Tasks 4 and 5 explore delay propagation and the scale of impact in actual operations. Task 6 further advances delay propagation analysis by rebuilding flight connections and exploring how schedule planning can mitigate delay propagation (Dunbar et al., 2012).

3.2.2. Delay tracing algorithms and implementation

To improve the efficiency of the proposed new coding schemes, one objective in developing the new coding schemes was to develop a scheme that allows automated data mining by computer

Table 6
IATA and Set 3 delay code grouping comparison (codes 51–58).

IATA scheme-group 5x codes	Set 3 group 5x codes
51: Damage during flight operations (bird, lightning, turbulence etc.)	51: Damage 511: During flight 512: On ground
52: Damage during ground operations (collisions other than during taxiing etc.)	
55: Departure control (check-in weight and balance (load control), computer system error, baggage sorting	52: System failure/error 521: Cargo 522: Check-in
56: Cargo preparation/documentation system	523: Load control 524: Others
57: Flight plans	53: Flight plans
58: Other computer systems	

Table 7
An example subset of the simulated delay data in Set 1 reporting style.

C-4[*]	C-9	C-10	C-12[a]	C-13	C-14	C-15	C-16
Flight no.	Delay (code)	Delay (mins)	Pre-Flt no.	Rotation/ connection	Aircraft Rego	Pax	Delay comments
1	41	40					Oil leak on right wing
1	62	20					Waiting for final fuel figure from tech crew
2	91	40	1	Aircraft			LATAC
2	32	20	1				Late loading due ground ops disrupted
3	91	50	2	Aircraft			LATAC
4	91	40	3	Aircraft			LATAC
5	91	25	4	Aircraft			LATAC
6	91	10	5	Aircraft			LATAC

Note: *– 'C-4' denotes column 4 of Set 1 coding scheme.
[a] Shaded cells (columns) are specific for Set 1 coding scheme.

scripts so that delay reporting can be generated without manual intervention from analysts. Therefore, we implemented the selected analysis tasks via Visual Basic scripting in Excel and used these Visual Basic scripts to demonstrate the advantages of the new schemes, and to understand how automated data mining and reporting can be carried out in practice with the improved coding schemes.

The current methodology of delay tracing in post-operational analysis involves either sifting through the *Delay Comments* column to identify which flight causes a delay to a particular flight or tracing aircraft tail numbers (Jetzki, 2009; Wu, 2010). To date, tracing aircraft routes through tail numbers has been the most popular method, as it can be easily analysed using software tools such as Excel or SQL database queries. Using the simple function of "Search" or "Find," one can identify all flights operated by the same aircraft that are affected by a specific delay code/incident (i.e., with the same tail number). From this step, delay propagation trees can be built to display the links between flights and can be used for further analysis (see e.g., AhmadBeygi et al., 2008 for a demonstration on delay propagation trees). However, this method of delay tracing limits the amount of data that can be analysed, because large quantities of data require more time for semi-automated processes or are otherwise too complicated to analyse because of the heavy reliance on text-based information contained in the *Delay Comments* column in the current IATA coding system.

With the improved reporting styles of the proposed coding schemes, we developed a series of delay tracing algorithms to conduct the test analysis tasks automatically. The pseudo-codes of the delay tracing algorithm for building delay propagation trees (Task 6) based on the proposed new coding schemes are provided in Table 8.

4. Delay coding scheme testing and results

4.1. Delay propagation tracing

The delay tracing algorithm in Table 8 is implemented by analysing the same set of delay data in four different coding schemes. As expected, the delay tracing algorithm could not produce delay propagation trees without manual intervention to sift through the *Delay Comments* column with the data coded in the existing IATA system. The three new coding schemes produce similar results with different levels of detail; Sets 2 and 3 can only produce delay propagation trees, whereas Set 1 can generate other statistics for impact assessment, such as passengers affected by delays. Fig. 1 illustrates a delay propagation tree generated with Set 1 delay data.

Table 8
Task 6 – identification of all connected delayed flights (affected by delay propagation) and sorting of these flights to build delay propagation trees for each incident.

1. From the delay data (Excel) sheet, start searching from the first row of data by the entry in the *Pre-flight* column.
2. Assign the (first) value found from the *Pre-flight* column to **pf** and note the current row number (denoted by **r**) of the found flight. Copy and paste this row into a target sheet.
3. While **pf** is not null (i.e. no entry in the *Pre-flight* column), do
   a. search the delay data sheet by the entry of *Flight Number* column and select the entire row of the flight whose flight number matches **pf** value; copy and paste this newly found row into the target sheet by appending the new row to the dataset,
   b. find the **pf** value from the *Pre-flight* column of the newly found row and update the **pf** parameter.
4. Advance the row number **r** by 1 and repeat Steps 2–3 until the end of the delay data sheet.
5. Use the selected delay data from the target sheet and build delay propagation trees.

4.2. Performance benchmarking

A comparison of test results from the four coding schemes is presented in Table 9. Test results of the existing IATA system show that Tasks 3, 5, and 6 cannot be fully achieved without manual processing of delay data, whereas the other tasks are fully fulfilled using only automated analysis. This result is consistent with our current observations in the industry: (a) detailed delay impact analyses such as the number of passengers affected by a specific delay incident (e.g. Task 3) cannot be easily achieved without manual (or semi-manual) processing of the delay-related information contained in the *Delay Comments* column (this is the case for Task 3) and (b) delay propagation tracing cannot be done without manual (or semi-manual) processing of the delay-related information stored in the *Delay Comments* column (this is the case for Tasks 5 and 6). Although Tasks 3, 5, and 6 could be conducted with the IATA system in our test, fully automated processing of delay data was not possible in this case; certain tasks are "partially" fulfilled by the current IATA system (denoted by "P" in Table 9).

Compared with the existing IATA system, the Set 1 coding scheme allows all six tasks to be automatically conducted without manual intervention. This outcome is expected, because the new coding scheme contains an improved reporting style that is designed to record key delay-related information in specific columns for automated analysis, leaving only trivial information in the *Delay Comments* column.

Sets 2 and 3 deliver similar results to the IATA system: Tasks 3 and 5 could not be conducted on an automated basis, because certain key delay-related information (e.g., the number of passengers affected by a delay incident) is still recorded in the *Delay Comments* column. In contrast to the IATA system, Task 6 (delay propagation tree building) could be conducted automatically using computer scripts because of the improved reporting style that recorded the "pre-flight" data in a separate column.

4.3. Overall comparison and discussion

Based on the results and comparisons made among the three new coding/reporting schemes, it is clear that there are many positive attributes of each set. The basic requirement established for these new coding schemes is the ability to produce statistical results (Tasks 1–3) as good as or better than the existing IATA system. This goal was fulfilled by each of the new coding schemes. However, as was the case for the IATA system, Sets 2 and 3 could not complete Task 3 without manual intervention in data analysis (see Table 9). A desired feature of the new coding schemes is to allow for automated in-depth delay propagation analysis (Tasks 4–6). Set 1 fully achieves Tasks 4–6, whereas Sets 2 and 3 cannot achieve Task 5 without manual data mining. The bottom line is that these new coding schemes are at least as good as the existing IATA system in terms of producing statistics and advanced delay analysis.

An overall comparison of the three new coding schemes with the IATA system is presented in Table 10. The biggest advantage of Set 1 over Sets 2 and 3 is the added benefit of producing more specific statistical results without the complication of the comments column. Through the (dynamically) added columns, Set 1 provides the most detailed data, and it is easier to implement automated reporting by computer scripts under the Set 1 scheme. When implementing Set 1 in the future, staff members only need to be trained to be familiar with data entry in the newly added columns since the report style is almost the same as the IATA system. Hence, the implementation of Set 1 in actual industry practice can be seen as the easiest of the three schemes.

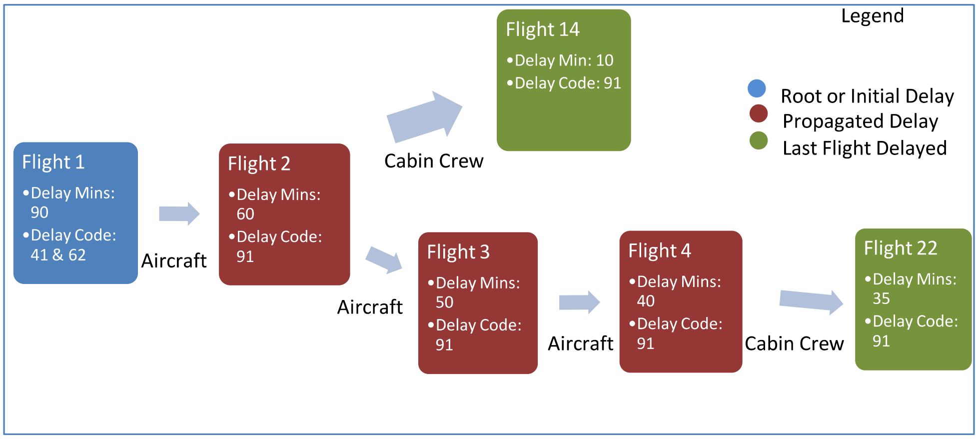

Fig. 1. Delay propagation tree generated by Set 1 coding scheme.

For Set 2, with its elimination of unnecessary and duplicated codes, the benefit is the reduction of delay codes, thereby reducing the possibility of confusion during data entry, analysis, and reporting. The new codes are straightforward (largely based on the existing IATA structure) and address the most commonly found delay causes in airline operations. Further, the complementary reporting style allows for advanced delay tracing analyses. However, of the three schemes, Set 2 would be the most difficult to implement by the industry because of further training required for ground staff in understanding the new code groups and data entry style. Existing in-house computer scripts would also need to be adjusted to the new coding and reporting scheme, although not all analysis tasks can be fully automated under this set.

Finally, the advantage of the implementation of Set 3 lies in its easy application due to the code groups largely remaining the same as the existing IATA framework. With the added depth in delay information through comprehensive sub-codes, reasons for delays are more accurately explained in this scheme. This allows for both basic and in-depth views of the delay data without compromising on delay analysis quality and depth. Like Set 2, the complementary reporting style also aids in delay tracing analysis. However, the reporting style adopted by Sets 2 and 3 is not as comprehensive as that of Set 1. Nevertheless, the reporting style of Set 1 could easily be adopted in Set 2 or 3 without any changes to the code groupings of these sets. This is a major advantage of these new delay coding schemes.

Based on the different advantages of each coding and reporting scheme, different parties within the industry can utilise the coding scheme that best suits their needs. Depending on the type and depth of data the user wishes to collect, the option of mixing and matching the three sets is also possible to further enhance data quality and facilitate the automated data analytics that are now more commonly seen in the aviation industry. When implementing a new coding scheme, it is suggested that the interested party implement a new scheme parallel to the existing in-house scheme at a medium-sized airport where a sufficient amount (and variety) of data can be recorded from actual operations. Comparisons could then be made between the two resulting datasets, and further improvements of the new coding scheme could be practically achieved. Following this trial, a full-scale roll-out of the new coding scheme can be deployed across the flight network, and more sophisticated computer scripts and algorithms (e.g., SQL/Java/C++ scripts) on larger platforms (e.g., SQL servers and data warehouses) can be designed to automatically produce in-depth network and operation diagnoses as well as routine delay data statistics reporting.

Table 9
New coding scheme testing and benchmarking against the IATA system.

	IATA	Set 1	Set 2	Set 3
Task 1	X[a]	X	X	X
Task 2	X	X	X	X
Task 3	P[b]	X	P	P
Task 4	X	X	X	X
Task 5	P	X	P	P
Task 6	P	X	X	X

[a] X denotes that the task is accomplished with automated analysis and reporting.
[b] P denotes that the task is accomplished only with manual data pre-processing.

Table 10
Overall comparison of the new coding schemes with the IATA system.

	Set 1	Set 2	Set 3
Statistical reporting	Allow full automation[a]	Automation/manual data processing[b]	Automation/manual data processing
Advanced delay analysis on a network scale	Allow full automation	Automation/manual data processing	Automation/manual data processing
Data recording style (data entry style)	Same as IATA with a few extra columns	Same as IATA with one extra column	Same as IATA with one extra column
Delay codes alternation from IATA scheme	Minor changes	Major consolidation of codes and grouping	Minor changes with adoption of sub-codes for depth
Implementation efforts	Low; only require data entry training	Highest; require new code grouping and data entry training	Modest; require sub-codes and data entry training
Major benefits over IATA scheme	- eliminate Comments column - use Pre-flight column to trace delays - data recorded for analysis ease	- less codes/groups; less confusion in using codes - same recording style as IATA - use Pre-flight column to trace delays	- allow code depth - same recording style as IATA - use Pre-flight column to trace delays

[a] "Full automation" means data analysis could be accomplished automatically using computer scripts without manual data pre-processing.
[b] "Automation/manual data processing" means some data analysis tasks could be automated, but others would require manual data pre-processing.

5. Conclusion

To conduct advanced delay propagation analysis, there is an emerging need to improve the quality of data currently collected by the IATA delay coding system. Three new coding schemes were developed, tested, and validated by conducting a set of analysis tasks with simulated delay data coded in four different schemes. Results show that these three new coding schemes are able to record delay data with better quality than the existing IATA system. Amongst the three new coding schemes, Set 1 provides the reporting style that is the most suitable for conducting automated analysis. Sets 2 and 3 utilise the same reporting style that records the "pre-flight" data to facilitate automated delay propagation tracing. However, because of a simpler reporting style in Sets 2 and 3, the depth of analysis that can be automatically conducted with data coded in these schemes is not as deep as that of Set 1. Additionally, Set 2 presents certain challenges to future implementation, mostly stemming from the restructuring of delay code groups through the consolidation and regrouping of some commonly and rarely used codes into groups that better reflect the true needs of delay data recording and analysis.

The three proposed delay coding schemes achieved the objective of this study: to develop a coding scheme that can record data of high quality and allow for delay analysis using automated computer scripts without manual intervention. Without the improvement in delay coding schemes, one would need to use complex "text mining" tools to extract delay-related information and then rebuild the delay database to reflect resource connections for each delay incident. The current industry practice of doing this is both costly and time-consuming. The outcome of this study represents a substantial advance over the existing IATA delay coding system in that we are able to improve both the quality of delay data and the depth of delay analysis that can be conducted with such improved data. With the assistance of improved data quality and advanced delay analysis on a network scale, airlines will be equipped to pinpoint specific operational bottlenecks after operations and develop strategies to improve schedules and operation management tactics. This will lead to significant cost savings for the airline industry as well as travelers.

References

AhmadBeygi, S., Cohn, A., Guan, Y., Belobaba, P., 2008. Analysis of the potential for delay propagation in passenger aviation flight networks. J. Air Transp. Manag. 14, 221–236.
Diana, T., 2009. Do market-concentrated airports propagate more delays than less concentrated ones? A case study of selected U.S. airports. J. Air Transp. Manag. 15 (6), 280–286.
Dunbar, M., Froyland, G., Wu, C.-L., 2012. Robust airline schedule planning: minimizing propagated delay in an integrated routing and crewing framework. Transp. Sci. 46 (2), 1–13.
International Air Transport Association (IATA), 2012. Airport Handling Manual, 32nd ed.
Jetzki, M., 2009. The Propagation of Air Transport Delays in Europe (Master's thesis). Department of Airport and Air Transportation Research, RWTH Aachen University, Aachen, The Netherlands.
Lan, S., Clarke, J., Barnhart, C., 2006. Planning for robust airline operations optimizing aircraft routings and flight departure times to minimize passenger disruptions. Transp. Sci. 40 (1), 15–28.
Lee, K., Moore, J., 2003. On-time performance. In: AGIFORS Airline Operations Conference. Auckland, New Zealand. Available from: http://www.agifors.org.
Niehues, A., Belin, S., Hansson, T., Hauser, R., Mostajo, M., Richter, J., 2001. Punctuality: How Airlines Can Improve On-Time Performance. Available from: http://www.aviation.go.th/rbm/Punctuality.pdf.
Sarseshiki, F., Arjroody, A.R., Zamani, K., 2010. Determination of delay factors in domestic passenger flights using AHP technique. In: 12th World Conference for Transportation Research (WCTR).
Speri, J., 2006. Potential On-Time Performance Improvements at the Lufthansa Station at Frankfurt Airport – a Human Factors Approach (MSc thesis). Massey University, Auckland, New Zealand.
U. S. Department of Transportation (U.S. DoT), 2001. Reporting the Causes of Airline Delays and Cancellations. Federal Register Notice 66(248).
Wu, C.L., 2010. Airline Operations and Delay Management. Ashgate Publishing Limited, Surrey.