

## Research Article

# Air-Ground Task Collaboration Safety Analysis Based on Real-Valued Differential Biclustering for Midterm Conflict in Single-Pilot Operations of Commercial Aircraft

Miao Wang <sup>1</sup>, Yue Luo <sup>1</sup>, Yong Chen <sup>1,2</sup>, Kelin Zhong <sup>3</sup>, and Guoqing Wang <sup>1</sup>

<sup>1</sup>School of Aeronautics and Astronautics, Shanghai Jiao Tong University, Shanghai 200240, China

<sup>2</sup>Commercial Aircraft Corporation of China, Ltd., Shanghai 200436, China

<sup>3</sup>China COMAC Shanghai Aircraft Design and Research Institute, Shanghai 201210, China

Correspondence should be addressed to Miao Wang; miaowang@sjtu.edu.cn

Received 16 January 2023; Revised 31 March 2023; Accepted 4 May 2023; Published 27 May 2023

Academic Editor: Jinchao Chen

Copyright © 2023 Miao Wang et al. This is an open access article distributed under the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

The International Civil Aviation Organization (ICAO) has mapped out Single-Pilot Operations (SPO) as the core development direction for the next generation of commercial aircraft operations in 2030. Safety is a key airworthiness factor in commercial aircraft design. Due to the higher degree of air-ground task collaboration and complexity in the SPO mode, the traditional safety analysis methods applied in two-pilot mode cannot effectively identify the potential hazard patterns in the system. To address the above problems, a safety analysis method that combines model-based safety analysis (MBSA) with hazard pattern mining is introduced, and a differential bicluster mining algorithm named *TFCluster* is proposed to identify maximum differential biclusters from real-valued function-resource matrices without candidate maintenance. Experiment studies on public datasets indicate that *TFCluster* is efficient and scalable, and outperform the existing differential bicluster algorithms. Taking the typical operating scenario—midterm conflict resolution in the SPO mode as an example—safety analysis of air-ground task collaboration for flight conflict in the SPO air-ground collaborative architecture is carried out. It is found that the proposed method can effectively identify potential hazard patterns, feedback to the system architecture design, and assist safety analysis.

## 1. Introduction

With the enhanced capability of the airborne systems, the crew size of the commercial aircraft is correspondingly reduced. As an important direction for future aviation technology, single-pilot operations (SPO) of commercial aircraft have completely revolutionized the traditional dual-pilot operating mode, which has drawn widespread attention from main manufacturers, airlines, and scholars in the world.

In the SPO mode, the original two pilots are reduced to one. In case of a flight conflict, the single pilot may not be able to respond promptly, posing a threat to flight safety. In order to avoid accidents, on the one hand, from the system architecture design perspective, we refer to the Class G Concept System proposed by NASA [1], which introduces

a ground operator to support air-ground collaborative decision-making in nonnominal flight. On the other hand, from the viewpoint of enhancing airborne system capability, high level engaged system integration is required to support air-ground task collaboration for enhancing airborne system capability. For example, the integration has the potential of enhancing air and ground surveillance capabilities for full flight schedule and avoid conflicts.

To ensure the safety of a flight task, it is necessary to collaborate with multiple subtasks, that is, task collaboration. In the SPO mode, flight tasks are accomplished through the collaboration between one pilot on board and one ground operator at the ground station. Therefore, air-ground task collaboration is a key feature of SPO. However, while improving system capability, air-ground task collaboration also brings safety issues in terms of fault mixing. Therefore,

safety analysis of air-ground task collaboration for flight conflict in SPO will be carried out in this paper.

Recently, research on SPO of commercial aircraft has focused on operating architecture design [2, 3], human-machine function allocation [4], single-pilot workload [5, 6], air-ground communication links [7], etc. Lachter et al. [8] and Schmid and Korn [9] summarize the existing conceptual design of SPO, which is mainly divided into alternative design [10, 11] and displacement design [12]. The current mainstream adoption is the advanced airborne automation system along with ground station assistance, which is proposed by NASA [1]. Faulhaber and Friedrich and Sprengart et al. [5, 13] conduct studies on the assessment of single-pilot workload. Foreign researchers have developed intelligent cognitive human-machine interfaces [4, 14, 15] based on artificial intelligence technology to sense and analyze pilot workload in real time, and conduct real-time management of workload to ensure flight safety [2, 16]. Min et al. [17] studied SPO human-machine task allocation by comparing and analyzing the current task allocation in two-crew operations. Carloni and Manica [7] studied the air-ground communication link requirements in SPO mode. Literature rarely focused on SPO safety. Yong et al. [18] studied task synthesis safety analysis in the SPO mode based on hazard pattern mining, but the algorithm is based on discrete data. Yue et al. [19] proposed a real-valued mining algorithm for SPO safety analysis; however, there is a lack of detailed description of the algorithm mechanism and no safety verification in conjunction with SPO scenarios, which will be compensated in this paper.

The avionics system in SPO mode has a high degree of integration and complexity, which makes traditional safety analysis methods unable to effectively identify the potential hazard elements. In view of the sharp rise in the number and complexity of system failures under SPO air-ground task collaboration, in this paper, a safety analysis method that combines model-based safety analysis (MBSA) with hazard pattern mining is proposed. Through mining a large number of simulation scenarios and historical test data in the SPO mode, the hidden hazard patterns and propagation mechanisms behind the model data could be uncovered and transformed into knowledge for safety analysis. Thus, comprehensive safety management of SPO air-ground collaborative architecture could be achieved, ultimately improving the overall safety of the aircraft.

The contributions of this paper mainly include the following aspects:

- (1) In response to the significant increase in the number and complexity of system failures under SPO air-ground task collaboration, this paper proposes a safety analysis method based on the combination of model-based analysis and hazard pattern mining, which can be used to guide safety analysis and system architecture design by mining and analyzing a large amount of model operating data
- (2) A real-valued differential bicluster mining algorithm—*TFCluster* algorithm is proposed for the analysis

of the function-resource allocation under different tasks. Compared with the conference paper [19], this paper provides a more detailed description of the mechanism of *TFCluster* and tests its efficiency on both public and SPO experimental datasets

- (3) A safety verification framework is constructed in this paper, which combines model-based analysis with hazard pattern mining. Taking the midconflict handling process in the SPO mode as an example, the model-based safety verification of SPO air-ground task collaboration is carried out. The mining results obtained from the *TFCluster* algorithm can effectively identify potential hazard patterns, feed back to the system architecture design and assist safety analysis

The remaining sections of this paper are organized as follows. Section 2 firstly introduces conflict detection in SPO, then analyzes safety issues, and finally proposes the safety analysis method combining MBSA and hazard pattern mining. Section 3 describes our proposed *TFCluster* algorithm in detail. Section 4 compares the proposed algorithm with the existing algorithms in mining efficiency, and takes a typical operating scenario—midterm conflict resolution in SPO for air-ground task collaboration safety analysis. Conclusions are drawn in the final section.

## 2. SPO Safety Analysis Method

**2.1. Conflict Detection in SPO.** There is a loss of visual information due to the absence of the copilot observing the situation outside the aircraft in the SPO mode, so surveillance capabilities need to be further enhanced to avoid the consequent conflicts. The next-generation surveillance system in SPO involves airborne automation system, onboard single pilot, ground station and air traffic control (ATC), and achieves coordination through air-ground data link and ground network.

In order to avoid collision accidents, it is necessary to predict flight trajectory and calculate probability of conflict in advance. For different predicted time, conflict detection is divided into long-term conflict detection, midterm conflict detection and near-term conflict detection, as shown in Figure 1. Among them, near-term conflict refers to potential conflicts within the next five minutes, and is addressed through the traffic collision avoidance system (TCAS), which alerts the onboard pilot and provides guidance for resolving the conflict. TCAS also supports emergency manoeuvring, such as emergency collision avoidance. Long-term conflict detection aims at potential conflicts that may occur in the future, beyond several hours. Aircraft in the airspace report their status to ATC, and ATC identifies the potential conflict through analysis. Coordinated resolution of the long-term conflict will be the responsibility of ATC. If adjustments by this aircraft are required, ATC will negotiate with the ground operator in advance to avoid conflicts by adjusting flight routes or modifying the required time of arrival (RTA). Midterm conflict detection is aimed at

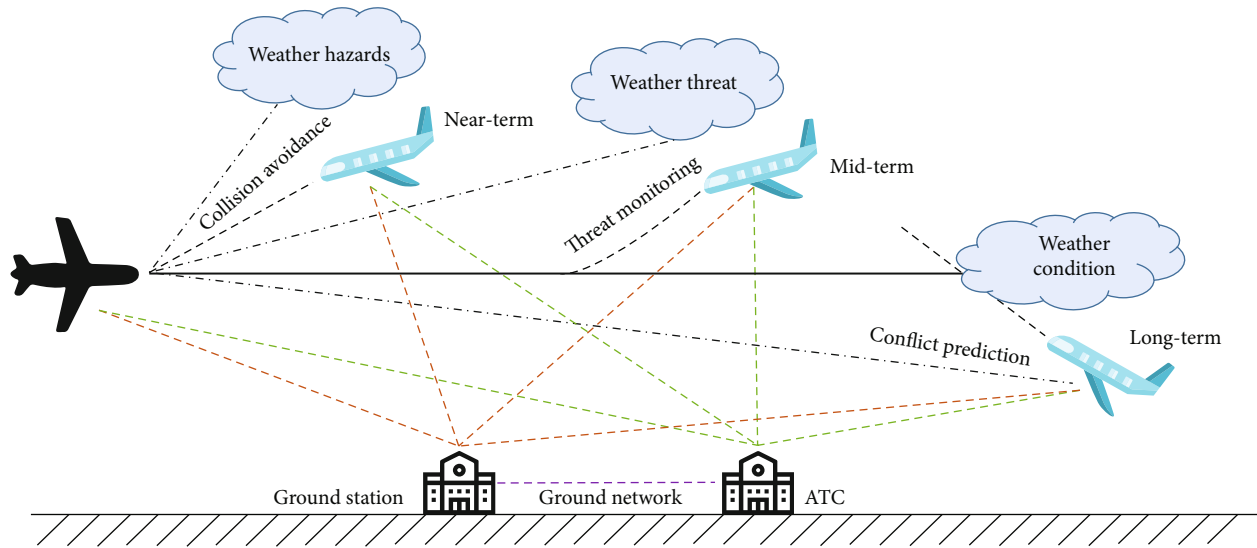


FIGURE 1: Different conflict detection modes in SPO.

potential conflicts in the next 10 to 30 minutes. The airborne surveillance system senses and identifies traffic threats and meteorological constraints, and then transmits them to the ground station through air-ground data link. The ground operator makes collaborative decisions with the onboard pilot and interacts with ATC to organize the maneuvering process and adjust the flight status, such as autonomous crossing, and altitude adjustments. Midterm conflict detection is accomplished through the air-ground collaboration among airborne automation system, onboard pilot, ground operator. Therefore, we choose midterm conflict for the follow-up study because it best reflects the characteristics of SPO air-ground coordination.

Midterm conflict does not require an immediate response from the pilot, but need to be handled by the pilot on board, the ground operator and ATC together. Several relevant flight tasks are included as follows.

- (1) Establish a collaborative surveillance among the pilot on board, the ground operator and ATC. The airborne data fusion and intelligent analysis system supports the onboard pilot to analyze the surveillance data and transmit the processed data to the ground station. The ground operator receives the surveillance data through air-ground data link, and receives airspace traffic/meteorological information from ATC through ground network. The data will be visualized through ground simulation software to facilitate the ground operator in making collaborative decisions. All the three parties simultaneously monitor the operating status of the aircraft, identify potential hazards and provide threat warnings
- (2) Determine collaborative handling measures among the onboard pilot, the ground operator and ATC. The onboard pilot senses and identifies the airspace traffic and route meteorology with the assistance of airborne automatic systems, ATC monitors the oper-

ating status of aircraft in the airspace and completes flight constraint management, and the ground operator is responsible for communicating and negotiating with ATC to finalize the conflict resolution, such as horizontal crossing and interval maintenance. The criteria and constraints for handling should be established, such as minimum safety separation and horizontal offset distance, to form a three-way collaborative handling method

**2.2. Safety Issues in SPO Mode.** In the absence of the complementary capabilities, interactive decision-making and status confirmation of another captain, safety issue is the first problem faced by SPO. For example, in midterm conflict detection, the lack of visual information from the copilot and the lack of face-to-face decision making may lead to safety hazards. To make up for the absence of another captain's ability, a ground operator is introduced to provide flight decision assistance and flight emergency response. Besides, the more integrated and advanced airborne automation system is necessary to assist the onboard pilot. Hence, air-ground task collaboration can be achieved through task synthesis, function fusion, and resource integration.

Among them, resource integration is achieved through the process of resource organization, operation and management, which enables resource capability sharing, resource operation reuse, and resource status management, thereby improving device resource utilization, operating efficiency, and availability. With the improvement of hardware computing capabilities, more and more functions are implemented through embedded software. Based on modularized function design, resource integration enables the same hardware platform to load multiple function modules, such as data processing function and display executing function, ultimately achieving process-based resource reuse and goal-based resource sharing. Function fusion enhances system task execution through dynamic collaboration between different functions. For example, the surveillance data of Automatic Dependent Surveillance-Broadcast

(ADS-B) and Traffic Collision Avoidance System (TCAS) can be fused and displayed comprehensively to achieve more accurate flight situation awareness [20] to support conflict detection. Task synthesis support the operation of nonnominal flight process and part of automated decision-making based on the current system capability status and the detected external environmental parameters, thereby reducing the workload of the single pilot.

Even though air-ground task collaboration brings the above benefits, it also increases system complexity, such as resource sharing, function cross-linking, and software and hardware interaction, making system faults spread and mix in the integration, fusion and synthesis, which has a great impact on system safety. In particular, multiple tasks may occur during midterm conflict detection in SPO, where air-ground task collaboration not only brings efficiency improvements, but also makes the resources lose their closeness. Resources are interconnected in terms of professional capabilities, resource operations, and organizational aspects, which results in safety issues where the system state is difficult to determine and fault composition is challenging to diagnose.

**2.3. Safety Analysis Method Combining MBSA and Hazard Pattern Mining.** For the abovementioned safety issues resulted from air-ground task collaboration in SPO, the traditional safety methods are flawed. Traditional safety analysis methods such as FTA and FMEA mainly rely on engineering experience, which are generally based on static logical reasoning, and are not carried out simultaneously with system design. In the SPO mode, as the complexity of the avionics system increases, it is difficult to list all the system failure modes and effects. In addition, due to the iteration of the system design, the failure mode may not match the system architecture. It is difficult to carry out the collaborative design of system safety and functional performance. To address the inconsistency between the system architecture design model and the safety analysis model, model-based safety analysis (MBSA) [21] is applied in this paper. The system design process is combined with the safety analysis process to ensure data consistency at the model level, and then the complex integrated system will be modeled through layered modeling [22].

To cope with the increasing number and complexity of system failure modes in the context of air-ground task collaboration in SPO, a new safety analysis method combining model-based safety analysis (MBSA) and hazard pattern mining is proposed, as shown in Figure 2. In accordance with the top-down system design process, various system models are built in different design stages, such as the flight scenario model in the requirements definition stage, the organization structure model in the system architecture design stage, the operating process model in the system operating analysis stage, and the function decomposition model and resource allocation model in the architecture decomposition stage. Then, model data will be extracted and the hazard pattern mining algorithms will be designed for analysis. Through the mining and analysis of a large number of model operating data, poten-

tial hazard patterns or more reasonable function-resource allocation scheme can be obtained to assist safety analysis and system architecture design.

The new safety analysis method combining MBSA and hazard pattern mining does not abandon the traditional safety analysis method, but serves as an auxiliary means to deal with the significant increase in the number and complexity of system failure modes in SPO task collaboration. The complex integrated system model is constructed through hierarchical modeling [22], which realizes the synchronization of the system design and the safety analysis. Different safety analysts could use the same research object, and adopt different analysis methods, such as FTA and FMEA, based on different analysis focuses. As an aid to traditional safety analysis, the hazard pattern mining method can make up for the lack of manual analysis and make safety analysis and verification more reliable.

### 3. TFCluster Algorithm

Based on the proposed safety analysis method, this section presents a hazard pattern mining algorithm called *TFCluster*, which can facilitate the analysis of function-resource organization during the execution of different tasks, thereby establishing a task-function-resource safety relationship and enabling the safety analysis of SPO air-ground task collaboration.

**3.1. Problem Description.** When dealing with the midterm conflict in SPO, there may be situations of multiple concurrent tasks. For example, while the onboard pilot collaborates with the ground operator and ATC to resolve conflicts, he also needs to monitor the airspace traffic and meteorological conditions to prevent new conflicts. Hence, this section will analyze whether there are potential hazards in the specific function-resource allocation scheme under the requirements of multiple concurrent tasks through hazard pattern mining.

The execution process of a task is the organizational and operational process of system functions, while the execution process of a function is the organizational and operational process of system resources. Therefore, the completion of the top-level tasks cannot be inseparable from the support of resources and functions. Here focuses on the air-ground task collaboration in SPO. The execution of each task can be abstracted as an  $R \times F$  matrix called function-resource matrix, where each row represents a resource, and each column represents a function. The values in the matrix indicate the extent to which a function uses a resource, and it can be represented as discrete data or real-valued data. In *DFCluster* [18], the values in the function-resource matrix are discretized as 1, -1, and 0, which reflects whether a function uses a resource, similar to a switch, and lacks a representation of the degree of use. However, in real systems, the extent to which a function uses a resource varies from case to case. Hence, this section targets the real-valued function-resource matrices for mining. The values in the real-valued function-resource matrix range from 0 to 1.0 means that the resource does not need to be called when a certain function is executed, 1 means that the resource must be called when a



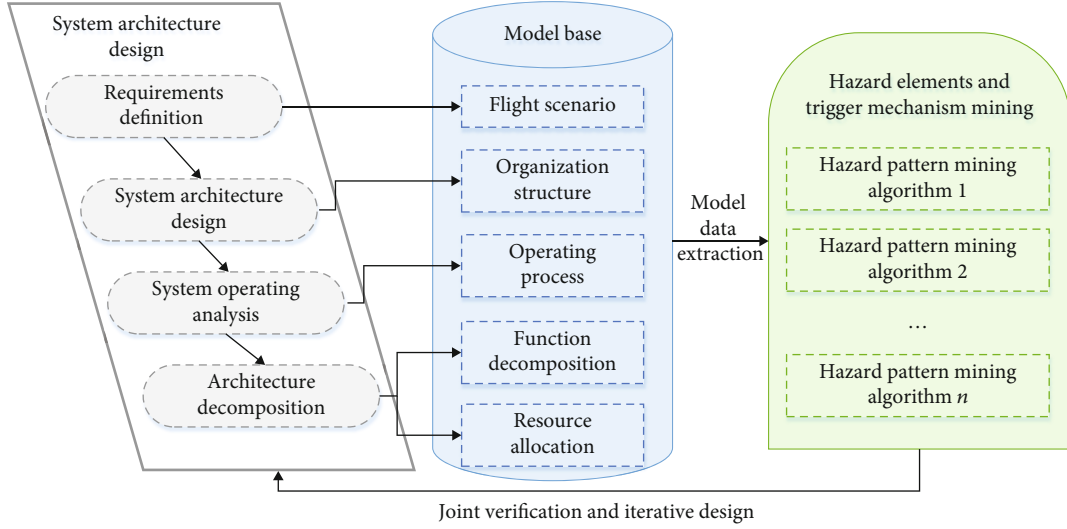


FIGURE 2: Safety analysis method combining MBSA and hazard pattern mining.

certain function is executed. For example, for a communication resource  $R_1$  with a bandwidth of 20 kHz, function  $F_1$  needs to occupy a bandwidth of 5 kHz to transmit data, so the occupation degree of the communication resource  $R_1$  by the function  $F_1$  is 0.25. Tables 1 and 2 show the function-resource matrices when tasks  $T_1$  and  $T_2$  are executing, respectively.

A differential bicluster mining algorithm named *TFCluster*, is proposed in this section. Maximum differential-used and low-used biclusters could be identified without candidate maintenance, and several pruning strategies are applied to improve mining efficiency. The mining process of *TFCluster* is shown in Figure 3. First, the original function-resource matrices are scanned to construct function-function weighted graph. Then, all maximum biclusters are identified using function expansion.

Through bicluster mining algorithm-*TFCluster* algorithm, differential biclusters could be identified from function-resource matrices of different tasks. The identified biclusters imply the relationship between multiple functions and resources, which can facilitate the analysis of function-resource organization during the execution of different tasks, thereby the task-function-resource safety relationship could be established, and SPO air-ground task collaboration could be performed. Specifically speaking, through analyzing the identified differential-used and low-used biclusters, it is possible to avoid task execution conflicts caused by simultaneous calls to the same resource under different functions. The mining results can be used to allocate resources reasonably. On the one hand, it can improve the efficiency of resources, on the other, it can avoid safety problems caused by air-ground task collaboration.

**3.2. Preliminaries and Definitions.** In order to enhance the differential bicluster algorithm for air-ground task collaboration safety analysis, three parameters are introduced to identify the differential-used and low-used biclusters from the

TABLE 1: Function-resource matrix of task  $T_1$ .

	$F_1$	$F_2$	$F_3$	$F_4$	$F_5$
$R_1$	0.15	0.1	0.01	0.01	0.9
$R_2$	0.1	0.8	0.2	0.1	0.2
$R_3$	0.2	0.1	0.1	0.2	0.9
$R_4$	0.1	0.2	0.1	0.1	0.1
$R_5$	0.2	0.1	0.1	0.9	0.1

TABLE 2: Function-resource matrix of task  $T_2$ .

	$F_1$	$F_2$	$F_3$	$F_4$	$F_5$
$R_1$	0.2	0.1	0.01	0.01	0.11
$R_2$	0.2	0.1	0.2	0.8	0.1
$R_3$	0.13	0.2	0.1	0.9	0.25
$R_4$	0.1	0.15	0.1	0.2	0.8
$R_5$	0.1	0.9	0.2	0.2	0.1

real-valued function-resource matrices. They are: (1) parameter  $\sigma$  used to measure the relevance between functions. (2) parameter  $\delta$  that restricts low usage of a resource. (3) parameter  $\mu$  used to measure the differential usage of a resource.

**Definition 1.** Low-used function-resource bicluster is defined as a bicluster where all resources and functions meet the Equation (1) and Equation (2):

$$\frac{\max_{f \in \{F_1, F_2\}} \{N_1, N_2\}_{r,f} - \min_{f \in \{F_1, F_2\}} \{N_1, N_2\}_{r,f}}{\min_{f \in \{F_1, F_2\}} \{N_1, N_2\}_{r,f}} \leq \sigma, \quad (1)$$

$$\max_{f \in \{F_1, F_2\}} \{N_1, N_2\}_{r,f} < \delta, \quad (2)$$

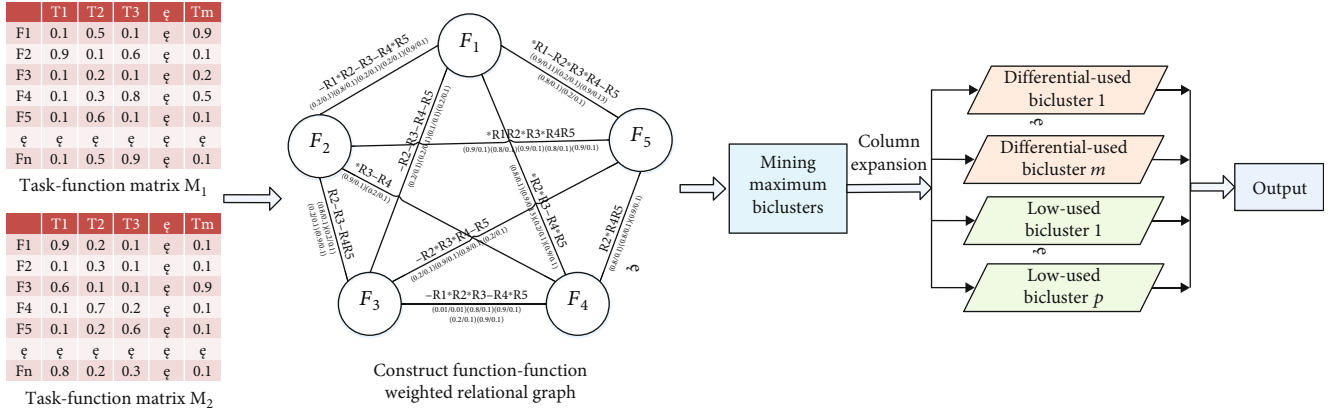


FIGURE 3: The mining process of TFCluster algorithm.

where  $N_1$  is the function-resource matrix of task  $T_1$ ,  $N_2$  is the function-resource matrix of task  $T_2$ ,  $\sigma$  is a user-defined parameter used to measure the relevance between functions,  $\delta$  is a user-defined parameter that restricts low usage of a resource,  $r$  is any resource in the function-resource matrices  $N_1$  and  $N_2$ ,  $F$  is the set of functions in  $N_1$  and  $N_2$ ,  $\max$  represents the maximum value of the resource under a set of functions, and  $\min$  represents the minimum value. Then, the safety index of low-used function-resource bicluster can be expressed as follows:

$$SI = \frac{\delta}{\max_{f \in \{F_1, F_2\}} \{N_1, N_2\}_{r,f}}. \quad (3)$$

**Definition 2.** Differential-used function-resource bicluster is defined as a bicluster where there is at least one resource that meets the Equations (4)–(7) under certain two functions, and the resource meets the constraints in Definition 1 under other functions.

$$\max_{f \in F} \{N_1, N_2\}_{r,f} > 1 - \delta, \quad (4)$$

$$\max_2 \{N_1, N_2\}_{r,f} < \delta, \quad (5)$$

$$\frac{\max_{f \in F} \{N_1, N_2\}_{r,f}}{\max_2 \{N_1, N_2\}_{r,f}} > \mu, \quad (6)$$

$$\frac{\max_2 \{N_1, N_2\}_{r,f} - \min_{f \in \{F_1, F_2\}} \{N_1, N_2\}_{r,f}}{\min_{f \in \{F_1, F_2\}} \{N_1, N_2\}_{r,f}} \leq \sigma, \quad (7)$$

where  $N_1$  is the function-resource matrix of task  $T_1$ ,  $N_2$  is the function-resource matrix of task  $T_2$ ,  $\sigma$  is a user-defined parameter used to measure the relevance between functions,  $\mu$  is a user-defined parameter used to measure the differential usage of a resource,  $\delta$  is a user-defined parameter that restricts low usage of a resource,  $r$  is any resource in the function-resource matrices  $N_1$  and  $N_2$ ,  $F$

is the set of functions in  $N_1$  and  $N_2$ ,  $\max$  represents the maximum value of the resource under a set of functions, and  $\max_2$  represents the second maximum value. Then, the safety index of differential-used function-resource bicluster can be expressed as follows:

$$SI = \frac{\max_{f \in F} \{N_1, N_2\}_{r,f} / \max_2 \{N_1, N_2\}_{r,f}}{\mu}. \quad (8)$$

**Definition 3.** In order to facilitate description of bicluster with differential usage rate and low usage rate, suppose the use values of resource  $R_1$  under the functions  $F_1$  and  $F_2$  of task  $T_1$  are  $V_{11}$  and  $V_{12}$ , and the use values of resource  $R_1$  under the functions  $F_1$  and  $F_2$  of task  $T_2$  are  $V_{21}$  and  $V_{22}$ . There are four representations for  $R_1$  under  $F_1$  and  $F_2$ : (1) If  $V_{11}, V_{12}, V_{21}, V_{22}$  meet Definition 2, and  $\max\{V_{11}, V_{12}, V_{21}, V_{22}\} = V_{11}$  or  $V_{21}$ , then the contribution rate of  $R_1$  to  $F_1$  and  $F_2$  meets differential-used requirement, expressed as “ $R_1$ ”. (2) If  $V_{11}, V_{12}, V_{21}, V_{22}$  meet Definition 2 and  $\max\{V_{11}, V_{12}, V_{21}, V_{22}\} = V_{12}$  or  $V_{22}$ , then the contribution rate of  $R_1$  to  $F_1$  and  $F_2$  meets differential-used requirement, expressed as “ $*R_1$ ”. (3) If  $V_{11}, V_{12}, V_{21}, V_{22}$  meet Definition 1, then the contribution rate of  $R_1$  to  $F_1$  and  $F_2$  meets low-used requirement, expressed as “ $-R_1$ ”. (4) If  $V_{11}, V_{12}, V_{21}, V_{22}$  does not meet Definition 1 or Definition 2, then no record is given.

Therefore, each resource in the biclusters identified by TFCluster algorithm satisfies the first three cases in Definition 3 for all functions. In order to improve the mining efficiency, the mining process takes the form of function expansion without candidate maintenance.

**3.3. Mining Process.** The mining procedure of the TFCluster algorithm consists of two main steps. Firstly, the original function-resource matrices are scanned, and a weighted function-function relational graph is constructed based on Definition 3. Secondly, maximum differential-used and low-used function-resource biclusters are identified through function expansion. Moreover, some pruning strategies are designed to improve the mining efficiency.

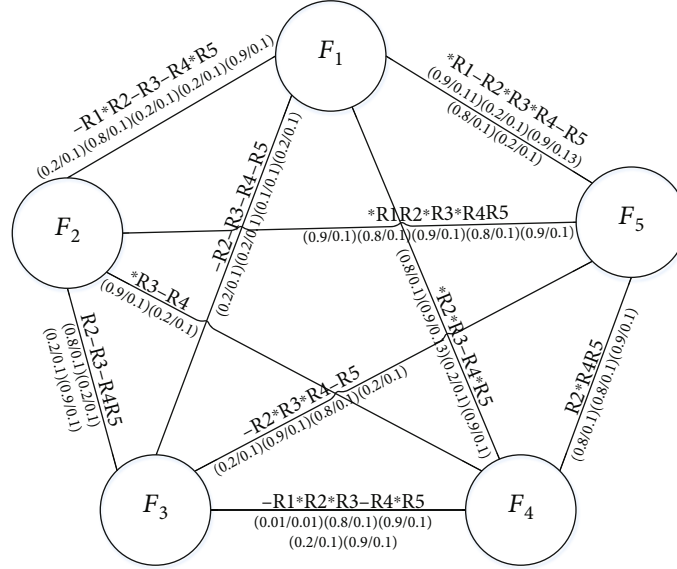


FIGURE 4: Weighted function-function relational graph based on Tables 1 and 2.

### 3.3.1. Weighted Function-Function Relational Graph

**Definition 4.** Weighted function-function relational graph is defined by the set  $L = \{S, P, V\}$ . Each node in the vertex set  $P$  in the graph represents a function. If there is an edge between a pair of vertices, it means that the resource set under above vertices must satisfy Definition 3, and the set of edges is denoted as  $S$ . The weights of each edge are the resource set satisfying Definition 3 under the two functions connected to this edge, and the set of the weights is denoted as  $V$ .

Different from *DFCluster* algorithm [18], the weight set  $V$  not only records the expression and symbol of the resource, but also records the maximum value ( $\max$ ) and the second maximum value ( $\max_2$ ) of each resource under these two functions. Taking the data in Tables 1 and 2 as an example, the weighted function-function relational graph as shown in Figure 4 can be constructed.

Through storing the original function-resource matrices in the weighted function-function relational graph, repeated access to the original data can be avoided, thereby improving the efficiency of the algorithm.

**3.3.2. Maximum Biclusters Mining.** Based on the above definitions, the extended bicluster satisfies anti-monotonicity. If the bicluster extended by  $F_1F_2, \dots, F_m$  does not satisfy the constraints, then its arbitrary superset  $F_1F_2, \dots, F_mF_n$  also does not satisfy. Hence, the larger-scale bicluster can be obtained through intersecting the weights on each edge in the weighted function-function relational graph. Since the weighted graph of *TFCluster* also records the maximum value and the second maximum value of each resource, then the bicluster expansion will be different from that in the *DFCluster* algorithm [18]. In *DFCluster*, if  $F_3$  is extended from  $F_1F_2$ , the weights of edges  $F_1F_2$ ,  $F_1F_3$ , and  $F_2F_3$  must be intersected to obtain the resource

set under  $F_1F_2F_3$ . But in the *TFCluster* algorithm, the resource set under  $F_1F_2F_3$  can be obtained only by calculating the intersection of the weights under  $F_1F_2$  and  $F_1F_3$ , as well as satisfying the following constraints.

Assuming that  $P$  is the current extended differential-used bicluster,  $ER.\max$  is the maximum value of resource  $R$  under function  $E$  to be extended,  $CR.\min$  is the minimum value. Based on Definition 3, the intersection of the resources under the current extended function and the candidate function can only be obtained in the following four cases: (1) The resource under the current extended function is “ $R$ ”, and the resource under the candidate function is also “ $R$ ”. (2) The resource under the current extended function is “ $*R$ ”, and the resource under the candidate function is “ $-R$ ”. (3) The resource under the current extended function is “ $-R$ ”, and the resource under the candidate function is “ $*R$ ”. (4) The resource under the current extended function is “ $-R$ ”, and the resource under the candidate function is also “ $-R$ ”.

**Theorem 5.** Assume that the maximum value of the resource  $R$  under the function  $E$  to be extended is  $ER.\max$  and the minimum value is  $ER.\min$ . The maximum value of resource  $R$  under candidate function  $C$  is  $CR.\max$ , and the minimum value is  $CR.\min$ . The form of the resource under function  $E$  to be extended is “ $R$ ”, and the form of the resource under candidate function  $C$  is also “ $R$ ”. Then, the following conditions must be met to make the resource  $R$  exist in the intersection of function  $E$  and function  $C$ :

$$\frac{\max \{ER.\min, CR.\min\} - \min \{ER.\min, CR.\min\}}{\min \{ER.\min, CR.\min\}} \leq \sigma. \quad (9)$$

It can be seen from the resource forms under current extended function  $E$  and candidate function  $C$  that the resource

with a utilization rate greater than  $1 - \delta$  must be under  $P$ . The resource utilization rate under function  $E$  and function  $C$  are both lower than  $\delta$ , then the extended PEC must be a differential-used bicluster, whose difference has been satisfied by the differential-used biclusters  $PE$  and  $PC$ . The resource  $R$  will exist in the intersection of function  $E$  and function  $C$  as long as the minimum values of the resource under the  $PE$  and  $PC$  satisfy the correlation.

*Proof of Theorem 1. Proof:*

$$R \in (PE \cap PC) \text{ if } \frac{\max \{ER.\min, CR.\min\} - \min \{ER.\min, CR.\min\}}{\min \{ER.\min, CR.\min\}} \leq \sigma, \quad (10)$$

if

$$\exists \left\{ \frac{\max \{ER.\min, CR.\min\} - \min \{ER.\min, CR.\min\}}{\min \{ER.\min, CR.\min\}} \leq \sigma \mid "R" \in PE, "R" \in PC \right\}, \quad (11)$$

then

$$\text{type}(PEC) = \text{differential-used} \&\& PE.\min \cap PC.\min = \text{correlation}, \quad (12)$$

then

$$R \in (PE \cap PC). \quad (13)$$

□

**Theorem 6.** Assume that the maximum value of resource  $R$  under function  $E$  to be extended is  $ER.\max$  and the minimum value is  $ER.\min$ . The maximum value of resource  $R$  under the candidate function  $C$  is  $CR.\max$ , and the minimum value is  $CR.\min$ . The form of the resource under function  $E$  to be extended is " $*R$ ", and the form of the resource under candidate function  $C$  is " $-R$ ". Then, the following conditions must be met to make the resource  $R$  exist in the intersection of function  $E$  and function  $C$ :

$$\frac{\max \{ER.\min, CR.\max\} - \min \{ER.\min, CR.\min\}}{\min \{ER.\min, CR.\min\}} \leq \sigma, \quad (14)$$

$$\frac{ER.\max}{\max \{ER.\min, CR.\max\}} > \mu.$$

It can be seen from the resource forms under current extended function  $E$  and candidate function  $C$  that the resource whose utilization rate is greater than  $1 - \delta$  must be under  $E$ , and the utilization rates of the resource under  $P$  and  $C$  are both lower than  $\delta$ . Then, the extended PEC must be a differential-used bicluster. The resource  $R$  will exist in the intersection of function  $E$  and function  $C$ , as long as the maximum value under  $PE$ , the minimum value under  $PE$ , and the maximum value under  $PC$  satisfy the difference, and the minimum values under  $PE$  and  $PC$  satisfy the correlation.

*Proof of Theorem 2. Proof:*

$$R \in (PE \cap PC) \text{ if } \frac{\max \{ER.\min, CR.\max\} - \min \{ER.\min, CR.\min\}}{\min \{ER.\min, CR.\min\}} \leq \sigma \&\& \frac{ER.\max}{\max \{ER.\min, CR.\max\}} > \mu, \quad (15)$$

if

$$\exists \left\{ \frac{\max \{ER.\min, CR.\max\} - \min \{ER.\min, CR.\min\}}{\min \{ER.\min, CR.\min\}} \leq \sigma \&\& \frac{ER.\max}{\max \{ER.\min, CR.\max\}} > \mu \mid "R" \in PE, "-R" \in PC \right\}, \quad (16)$$

then

$$\text{type}(PEC) = \text{differential-used} \&\& PE.\max \cap PE.\min \cap PC.\max = \text{difference} \&\& PE.\min \cap PC.\min = \text{correlation}, \quad (17)$$

then

$$R \in (PE \cap PC). \quad (18)$$

□

**Theorem 7.** Assume that the maximum value of resource  $R$  under function  $E$  to be extended is  $ER.\max$  and the minimum value is  $ER.\min$ . The maximum value of the resource  $R$  under candidate function  $C$  is  $CR.\max$ , and the minimum value is  $CR.\min$ . The form of the resource under the function  $E$  to be extended is " $-R$ ", and the form of the resource under candidate function  $C$  is " $*R$ ". Then, the following conditions must be met to make the resource  $R$  exist in the intersection



of function  $E$  and function  $C$ :

$$\frac{\max \{ER.\max, CR.\min\} - \min \{ER.\min, CR.\min\}}{\min \{ER.\min, CR.\min\}} \leq \sigma,$$

$$\frac{CR.\max}{\max \{ER.\max, CR.\min\}} > \mu. \quad (19)$$

It can be seen from the resource forms under current extended function  $E$  and candidate function  $C$  that the resource whose utilization rate is greater than  $1 - \delta$  must be under  $C$ , and the utilization rates of the resource under  $P$  and  $E$  are both lower than  $\delta$ . Then, the extended PEC must be a differential-used bicluster. The resource  $R$  will exist in the intersection of function  $E$  and function  $C$ , as long as the maximum value under  $PC$ , the minimum value under  $PC$ , and the maximum value under  $PE$  satisfy the difference, and the minimum values under  $PE$  and  $PC$  satisfy the correlation.

*Proof of Theorem 3. Proof:*

$$R \in (PE \cap PC) \text{ if } \frac{\max \{ER.\max, CR.\min\} - \min \{ER.\min, CR.\min\}}{\min \{ER.\min, CR.\min\}} \leq \sigma \ \&\& \ \frac{CR.\max}{\max \{ER.\max, CR.\min\}} > \mu, \quad (20)$$

if

$$\exists \left\{ \frac{\max \{ER.\max, CR.\min\} - \min \{ER.\min, CR.\min\}}{\min \{ER.\min, CR.\min\}} \leq \sigma \ \&\& \ \frac{CR.\max}{\max \{ER.\max, CR.\min\}} > \mu \mid \text{"-R"} \in PE, \text{"*R"} \in PC \right\}, \quad (21)$$

then

$$\begin{aligned} \text{type}(PEC) &= \text{differential-used} \ \&\& \ PC.\max \cap PC.\min \cap PE.\max \\ &= \text{difference} \ \&\& \ PE.\min \cap PC.\min = \text{correlation}, \end{aligned} \quad (22)$$

then

$$R \in (PE \cap PC). \quad (23)$$

Even if the resource forms under current extended function  $E$  and candidate function  $C$  are both “-R”, “R” or “\*R” may still appear under the intermediate functions, which is because the definition of the resource symbol only depends on the functions of the beginning and the end. Different cases can be distinguished by comparing the relationship between the maximum value of the resource and the parameter  $\delta$ .  $\square$

**Theorem 8.** Assume that the maximum value of resource  $R$  under function  $E$  to be extended is  $ER.\max$  and the minimum value is  $ER.\min$ . The maximum value of resource  $R$  under the candidate function  $C$  is  $CR.\max$ , and the minimum value is  $CR.\min$ . The form of the resource under function  $E$  to be extended is “-R”, and the form of the resource under candidate function  $C$  is also “-R”. Then, the following conditions must be met to make the resource  $R$  exist in the intersection of function  $E$  and function  $C$ :

$$CR.\max > 1 - \delta,$$

$$\frac{CR.\max}{\max \{ER.\max, CR.\min\}} > \mu,$$

$$\frac{\max \{ER.\max, CR.\min\} - \min \{ER.\min, CR.\min\}}{\min \{ER.\min, CR.\min\}} \leq \sigma. \quad (24)$$

Based on the resource forms under current extended function  $E$  and candidate function  $C$ , and the minimum value under  $PC$  is greater than  $1 - \delta$ , then the extended PEC must be a differential-used bicluster. The resource  $R$  will exist in the intersection of function  $E$  and function  $C$ , as long as the maximum value under  $PC$  and the larger between the minimum value under  $PC$  and the maximum value under  $PE$  satisfy the difference, and the larger between the maximum value under  $PE$  and the minimum value under  $PC$ , and the smaller between the minimum values under  $PE$  and  $PC$  satisfy the correlation.

*Proof of Theorem 4. Proof:*

$$R \in (PE \cap PC) \text{ if } CR.\max > 1 - \delta \ \&\& \ \frac{CR.\max}{\max \{ER.\max, CR.\min\}} > \mu \ \&\& \ \frac{\max \{ER.\max, CR.\min\} - \min \{ER.\min, CR.\min\}}{\min \{ER.\min, CR.\min\}} \leq \sigma, \quad (25)$$

if

$$\exists \left\{ CR.\max > 1 - \delta \ \&\& \ \frac{CR.\max}{\max \{ER.\max, CR.\min\}} > \mu \ \&\& \ \frac{\max \{ER.\max, CR.\min\} - \min \{ER.\min, CR.\min\}}{\min \{ER.\min, CR.\min\}} \leq \sigma \mid \text{"-R"} \in PE, \text{"-R"} \in PC \right\}, \quad (26)$$

then

$$\begin{aligned} \text{type}(PEC) &= \text{differential-used} \ \&\& \ PC.\text{max} \cap \max \{PC.\text{min}, PE.\text{max}\} \\ &= \text{difference} \ \&\& \ \max \{PE.\text{max}, PC.\text{min}\} \cap \min \\ &\quad \cdot \{PE.\text{min}, PC.\text{min}\} = \text{correlation}, \end{aligned} \quad (27)$$

then

$$R \in (PE \cap PC). \quad (28)$$

□

**Theorem 9.** Assume that the maximum value of resource  $R$  under function  $E$  to be extended is  $ER.\text{max}$  and the minimum value is  $ER.\text{min}$ . The maximum value of resource  $R$  under candidate function  $C$  is  $CR.\text{max}$ , and the minimum value is  $CR.\text{min}$ . The form of the resource under function  $E$  to be extended is “ $-R$ ”, and the form of the resource under candidate function  $C$  is also “ $-R$ ”. Then, the following conditions

must be met to make the resource  $R$  exist in the intersection of function  $E$  and function  $C$ :

$$\begin{aligned} &CR.\text{max} < \delta, \\ &\frac{\max \{ER.\text{max}, CR.\text{max}\} - \min \{ER.\text{min}, CR.\text{min}\}}{\min \{ER.\text{min}, CR.\text{min}\}} \leq \sigma. \end{aligned} \quad (29)$$

Based on the resource forms under current extended function  $E$  and candidate function  $C$ , and the minimum values under  $PE$  and  $PC$  are both lower than  $\delta$ , then the extended  $PEC$  must be a low-used bicluster. The resource  $R$  will exist in the intersection of function  $E$  and function  $C$ , as long as the larger between the maximum values under  $PE$  and  $PC$ , and the smaller between the minimum values under  $PE$  and  $PC$  satisfy the correlation.

Proof of Theorem 5. Proof:

$$R \in (PE \cap PC) \text{ if } CR.\text{max} < \delta \ \&\& \ \frac{\max \{ER.\text{max}, CR.\text{max}\} - \min \{ER.\text{min}, CR.\text{min}\}}{\min \{ER.\text{min}, CR.\text{min}\}} \leq \sigma, \quad (30)$$

if

$$\exists \left\{ CR.\text{max} < \delta \ \&\& \ \frac{\max \{ER.\text{max}, CR.\text{max}\} - \min \{ER.\text{min}, CR.\text{min}\}}{\min \{ER.\text{min}, CR.\text{min}\}} \leq \sigma \mid \text{“} - R \text{”} \in PE, \text{“} - R \text{”} \in PC \right\}, \quad (31)$$

then

$$\begin{aligned} \text{type}(PEC) &= \text{low-used} \ \&\& \ \max \{PE.\text{max}, PC.\text{max}\} \\ &\cap \min \{PE.\text{min}, PC.\text{min}\} = \text{correlation}, \end{aligned} \quad (32)$$

then

$$R \in (PE \cap PC). \quad (33)$$

□

**3.4. Pruning Strategy.** Predecessor detection [23] is applied in the *TFCluster* algorithm to identify maximum biclusters without candidate maintenance. If the resource set under the candidate function has an inclusive relationship with that under the prior function, then the biclusters extended by the candidate function have been extended by the prior function and are a subset thereof. Hence, the candidate function can be pruned to reduce the search space and avoid unpromising processes. When describing the representation of the resources, the resource  $R_1$  is represented as “ $R_1$ ”, “ $*R_1$ ” and “ $-R_1$ ”, respectively, which is to facilitate the design of the pruning strategy.

**Pruning 1.** Assume that  $E$  is currently extended bicluster,  $C$  is the candidate function set of  $E$ , and  $P$  is the prior function set of  $E$ . For any resource  $R_j$  under the candidate function  $C_m$  ( $C_m \in C$ ), if its representation is “ $R_j$ ”, and there is a prior function  $P_n$  ( $P_n \in P$ ) and a resource “ $R_j$ ” under  $P_n$ , then resource  $R_j$  under function  $C_m$  can be extended by the prior function  $P_n$ , thus  $C_m$  can be pruned.

Since  $R_j$  under function  $C_m$  is represented as “ $R_j$ ”, the resource whose utilization rate is higher than  $1 - \delta$  must exist under a certain function in  $E$ , and the utilization rate of  $R_j$  under function  $C_m$  and function  $P_n$  are both lower than  $\delta$ . Hence, the current extended bicluster is a differential-used bicluster. Since there can only be one resource usage rate higher than  $1 - \delta$  under all functions in differential-used bicluster, then the bicluster obtained by  $EC_m$  can also be obtained by  $EP_nC_m$ .

Proof of Pruning 1. Proof:

$$C_m \text{ can be pruned iff “} R_j \text{”} \in EC_m \ \&\& \ \text{“} R_j \text{”} \in EP_n, \quad (34)$$

if

$$"R_j" \in EC_m \ \&\& \ "R_j" \in EP_n, \quad (35)$$

then

$$\exists R \in E \text{ where } r(R) > 1 - \delta \ \&\& \ r(C_m.R) < \delta \ \&\& \ r(P_n.R) < \delta, \quad (36)$$

then

$$\text{type}(E) = \text{differential-used}, \quad (37)$$

then

the bicluster obtained by  $EC_m$  can also be obtained by  $EP_n C_m$ .

Hence,  $C_m$  can be pruned.

Pruning 2. Assume that  $E$  is currently extended bicluster,  $C$  is the candidate function set of  $E$ , and  $P$  is the prior function set of  $E$ . For any resource  $R_j$  under the candidate function  $C_m$  ( $C_m \in C$ ), if its representation is  $"*R_j"$ , and there is a prior function  $P_n$  ( $P_n \in P$ ) and a resource  $"-R_j"$  under  $P_n$ , then resource  $R_j$  under function  $C_m$  can be extended by the prior function  $P_n$ , thus  $C_m$  can be pruned.

Since  $R_j$  under function  $C_m$  is represented as  $"*R_j"$ , the resource whose utilization rate is higher than  $1 - \delta$  must exist under the candidate function  $C_m$ , and the utilization rate of  $R_j$  under all functions in  $E$  is lower than  $\delta$ . The resource  $R_j$  under prior function  $P_n$  is recorded as  $"-R_j"$ , so the resource usage rate of  $R_j$  under  $P_n$  is lower than  $\delta$ . Hence, the current extended bicluster is a differential-used bicluster. Since there can only be one resource whose usage rate is higher than  $1 - \delta$  under all functions in the differential-used bicluster, then the bicluster obtained by  $E C_m$  can also be obtained by  $EP_n C_m$ .

Proof of Pruning 2. Proof:

$$C_m \text{ can be pruned if } "*R_j" \in EC_m \ \&\& \ "-R_j" \in EP_n, \quad (38)$$

if

$$"*R_j" \in EC_m \ \&\& \ "-R_j" \in EP_n, \quad (39)$$

then

$$\forall R \in E \text{ where } r(R) < \delta \ \&\& \ r(C_m.R) > 1 - \delta \ \&\& \ r(P_n.R) < \delta, \quad (40)$$

then

$$\text{type}(E) = \text{differential-used}, \quad (41)$$

then

the bicluster obtained by  $EC_m$  can also be obtained by  $EP_n C_m$ .

Hence,  $C_m$  can be pruned.

Pruning 3. Assume that  $E$  is currently extended bicluster,  $C$  is the candidate function set of  $E$ , and  $P$  is the prior function set of  $E$ . For any resource  $R_j$  under the candidate function  $C_m$  ( $C_m \in C$ ), if its representation is  $"-R_j"$ , and there is a prior function  $P_n$  ( $P_n \in P$ ) and a resource  $"-R_j"$  under  $P_n$ , then resource  $R_j$  under function  $C_m$  can be extended by the prior function  $P_n$ , thus  $C_m$  can be pruned.

Since  $R_j$  under function  $C_m$  is represented as  $"-R_j"$ , then the utilization rate of resource  $R_j$  under candidate function  $C_m$  is lower than  $\delta$ , and the utilization rate of resource  $R_j$  under all functions in  $E$  may be lower than  $\delta$  or higher than  $1 - \delta$  under a certain function. The resource  $R_j$  under the prior function  $P_n$  is also recorded as  $"-R_j"$ , so the resource usage rate of  $R_j$  under  $P_n$  is lower than  $\delta$ . Hence, the differential-used bicluster or low-used bicluster  $EP_n C_m$  can be extended by  $EP_n$ .

Proof of Pruning 3. Proof:

$$C_m \text{ can be pruned if } "-R_j" \in EC_m \ \&\& \ "-R_j" \in EP_n, \quad (42)$$

if

$$"-R_j" \in EC_m \ \&\& \ "-R_j" \in EP_n, \quad (43)$$

then if

$$\forall R \in E \text{ where } r(R) < \delta \ \&\& \ r(C_m.R) < \delta \ \&\& \ r(P_n.R) < \delta, \quad (44)$$

then

$$\text{type}(E) = \text{low-used}, \quad (45)$$

then if

$$\exists R \in E \text{ where } r(R) > 1 - \delta \ \&\& \ r(C_m.R) < \delta \ \&\& \ r(P_n.R) < \delta, \quad (46)$$

then

$$\text{type}(E) = \text{differential-used}, \quad (47)$$

then

the bicluster obtained by  $EC_m$  can also be obtained by  $EP_n C_m$ .

Hence,  $C_m$  can be pruned.

The pseudocode of the *TFCluster* is shown in the following Algorithm 1.

The *TFCluster* algorithm is used in Tables 1 and 2, assuming that the correlation parameter  $\sigma$  is set to 10, the low usage parameter  $\delta$  is set to 0.3, differential usage parameter  $\mu$  is set to 10, the minimum thresholds for the number of functions and the number of resources are set to 3 and 1, respectively. The mining process is shown in Figure 5.

**Input:** Two real-valued function-resource datasets:  $M_1$  and  $M_2$ , the minimum number of resources in bicluster:  $R_{min}$ , the minimum number of functions in bicluster:  $F_{min}$ , low-used parameter  $\delta$ , differential-used parameter  $\mu$ , weighted function-function relational graph:  $L$ , the current extending differential function-resource bicluster:  $E$ .

**Output:** the maximal differential-used and low-used function-resource bicluster set;

**Initialization:**  $L = \emptyset$ ;  $E = \emptyset$ ;

**Method:**  $TFCluster(M_1, M_2, R_{min}, F_{min}, \delta, \mu, L, E)$

**if**  $L = \emptyset$  **then**  
 construct  $L$ ;  
**end if**  
 scan  $L$  and find all the candidate set  $C$  of  $E$ ;  
**for** each candidate  $C_m$  in  $C$   
**if** the differential function-resource bicluster  $EC_m$  does not satisfy **Pruning 1.** and **Pruning 2.** and **Pruning 3.**, and the number of resources in  $EC_m$  is greater than  $R_{min}$ , **then**  
 $E.Function = EC_m.Function$ ;  
 $E.EResources = E.EResources \cap EC_m.EResources$ ;  
 $E.PResources = E.PResources \cap EC_m.PResources$ ;  
 $TFCluster(M_1, M_2, R_{min}, F_{min}, \delta, \mu, L, E)$   
**else if**  $EC_m$  satisfies **Pruning 1.** or **Pruning 2.** or **Pruning 3.** **then**  
 delete  $C_m$   
**end if**  
**end for**  
**if**  $E$  is greater than any candidate bicluster of  $E$  and the number of functions in  $E$  is greater than  $F_{min}$  **then**  
 output( $P$ );  
**end if**  
**Return**

ALGORITHM 1: TFCluster algorithm.

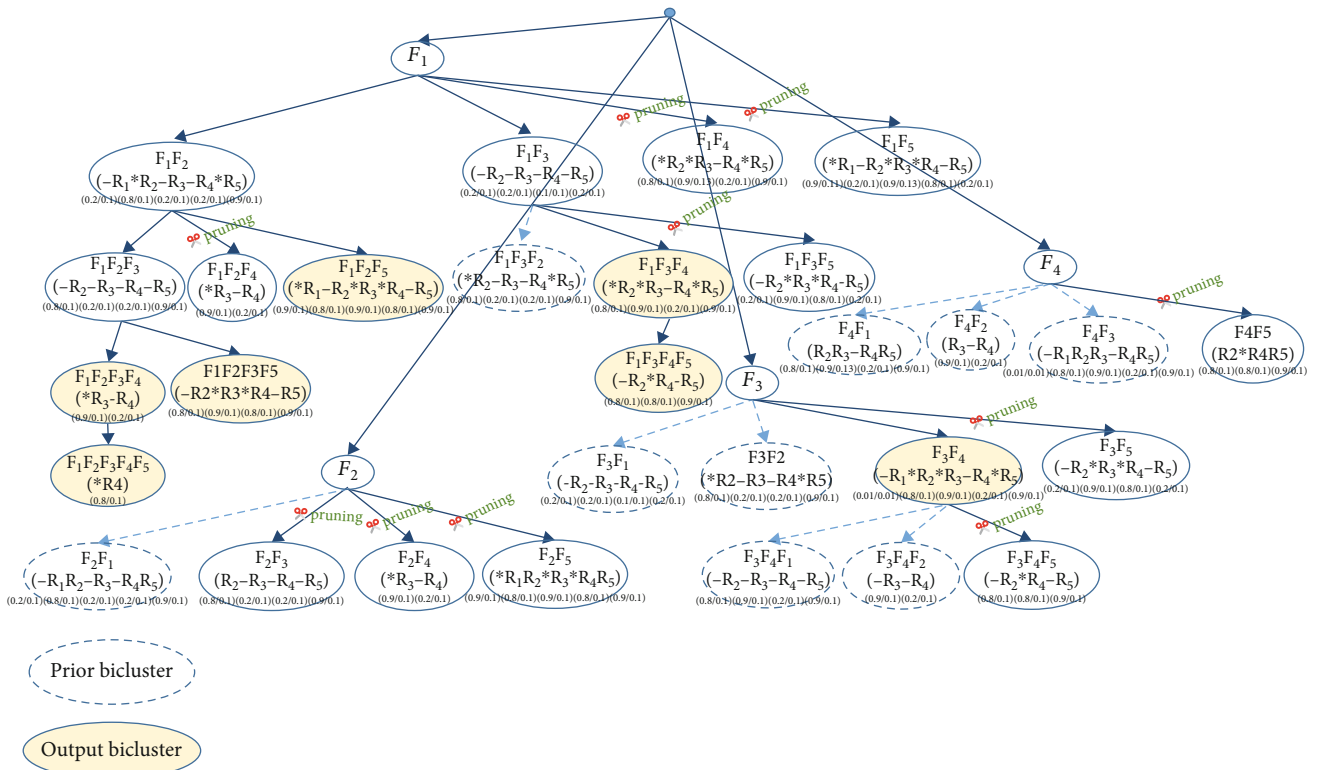


FIGURE 5: Example mining process of TFCluster algorithm.

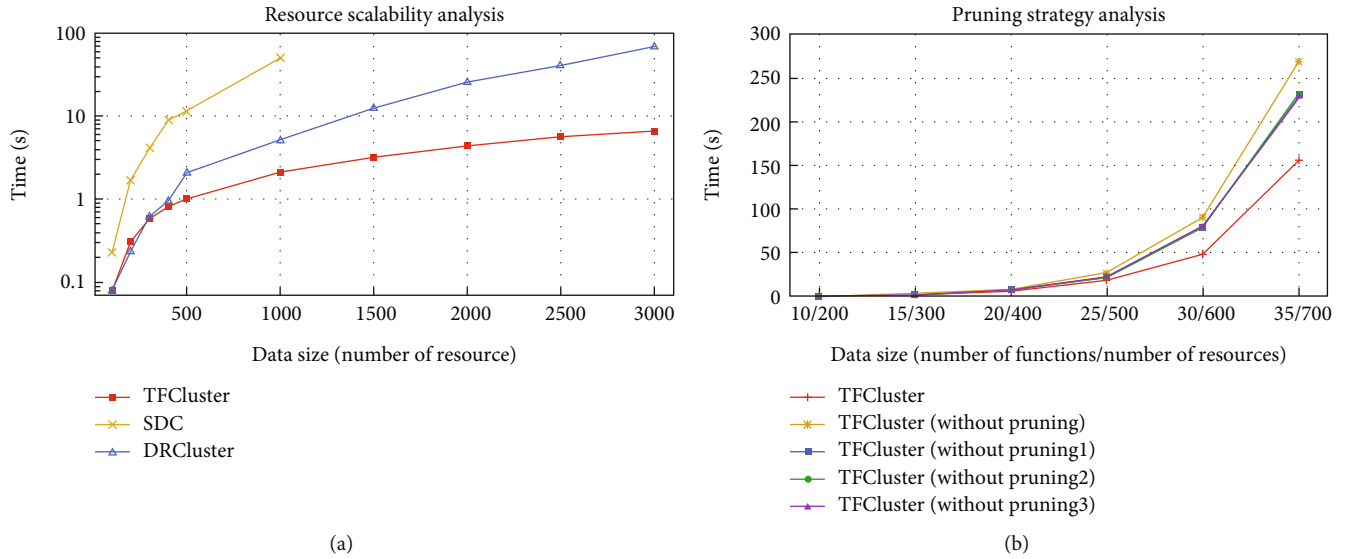


FIGURE 6: Running time comparison in AGEMAP database. (a) Runtime benchmark of different algorithms varying dataset size. (b) Pruning strategy analysis of TFCluster algorithm.

## 4. Experiment and Analysis

**4.1. Efficiency Comparison.** To evaluate the performance of *TFCluster*, two bicluster mining algorithms, *SDC* [24] and *DRCluster* [25] are compared with *TFCluster*. All algorithms were implemented in C language and tested on a laptop with Intel(R) Core(TM) CPU and 8G memory. Differential bicluster mining algorithms have been widely used in the field of bioinformatics for detecting disease-causing or growth-related genes. In this study, we selected *AGEMAP* [26], a gene database, for experimentation to test the efficiency and scalability of *TFCluster*. *AGEMAP* records gene expression data in mice as they age. The dataset includes samples from five groups of mice aged 1, 6, 16, and 24 months, respectively. Based on the quality of the original dataset, we chose the data from mice aged 6 and 16 months as the input for subsequent experiments.

**SDC algorithm:** The algorithm identifies all the differential coexpression biclusters from two real-valued datasets based on the idea of the Apriori algorithm. In the original SDC algorithm, only the differential coexpression relative constant column biclusters can be identified. In order to identify the differential coexpression relative constant row biclusters, original SDC algorithm is improved as follows: first, a set of association samples for each item in the two datasets are generated based on the association threshold  $\alpha$ . Then, all differential coexpression relative constant row biclusters that satisfy the differential support are identified using item expansion based on the *Apriori* principle. Finally, the results are judged to output only the maximum biclusters.

**DRCluster algorithm:** The algorithm identifies the maximum differential coexpression biclusters from two real-valued datasets. The mining process of *DRCluster* is divided into two main steps: first, construct sample weighted graph of differential expressions. Then, identify the maximum dif-

ferential expression biclusters from the constructed weighted graph using sample expansion.

Multiple datasets of different sizes are generated based on *AGEMAP*. The parameters of each algorithm are set in such a way that *SDC* and *DRCluster* could obtain the highest mining efficiency. (1) In *SDC*, the differential support is set to 1, and both the minimum thresholds for the number of items and the number of samples is set to 3. (2) In *DRCluster*, both the minimum threshold for the number of items and the number of samples are set to 3. (3) In *TFCluster*, the correlation parameter  $\sigma$  is set to 10, the low usage parameter  $\delta$  is set to 0.3, the differential usage parameter  $\mu$  is set to 10, and the minimum thresholds for the number of functions and the number of resources are set to 2 and 1, respectively.

**4.1.1. Resource Scalability.** All algorithms are benchmarked in terms of runtime by varying the number of resources, where the number of functions is fixed at 10, and the number of resources is set to 100, 200, 300, 400, 500, ..., 3000. Figure 6 shows the runtime (in seconds) to identify the biclusters for each algorithm. It can be seen that *TFCluster* is a clear winner for all datasets, and its advantage is more obvious when the number of resources is larger. At 3000 resources, *TFCluster* can still complete mining within 10s (6.623 s), it is more than 10 times faster than *DRCluster* (68.561 s). The *SDC* algorithm suffers from memory explosion when the number of resources is greater than 1500, while this is not the case with *TFCluster* as it does not store candidate sets.

**4.1.2. Pruning Strategy Analysis.** *TFCluster* is compared with *TFCluster with pruning 1*, *TFCluster with pruning 2*, *TFCluster with pruning 3*, *TFCluster without pruning* in terms of runtime at different sizes of datasets. It can be seen from Figure 6 that pruning strategies have a significant impact on the execution time. There is a lot of repetitive mining



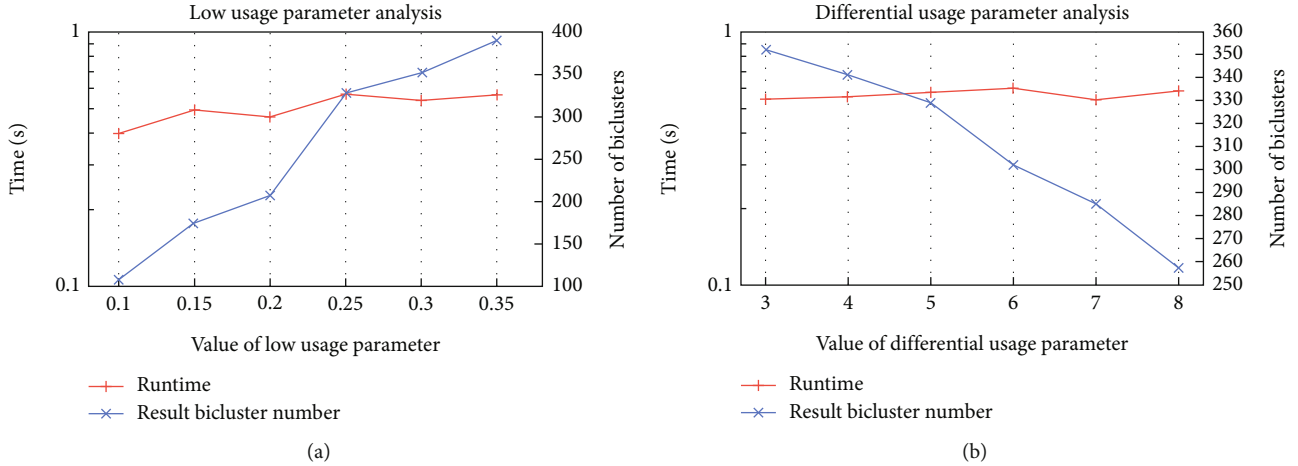


FIGURE 7: Comparison of the running time and the number of mining results under different parameters in AGEMAP database. (a) Varying low usage parameter  $\delta$ . (b) Varying differential usage parameter  $\mu$ .

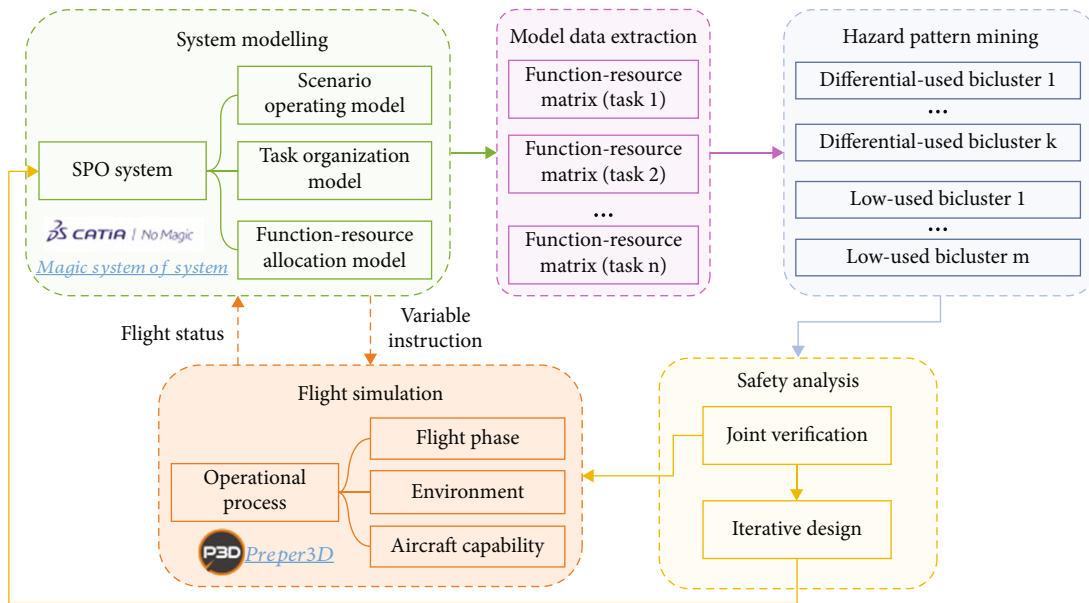


FIGURE 8: Model-based simulation framework.

in *TFCluster without pruning*, which results in low mining efficiency. *TFCluster with pruning 3* outperforms all algorithms except *TFCluster*, this is because *pruning 3* greatly simplifies the process of mining low-used biclusters as described in Section 3.4. When the number of function is up to 30, and the number of resource is up to 600, *TFCluster without pruning* (89.875 s) is about twice as slow as *TFCluster* (48.236 s), and the advantage of *TFCluster* becomes more obvious as the dataset increases, which shows that the designed pruning strategy greatly improves the mining efficiency of the algorithm.

**4.1.3. Parameter Setting Analysis.** The *TFCluster* algorithm introduces three parameters: (1) parameter  $\sigma$  used to measure the relevance between functions, (2) parameter  $\delta$  that restricts low usage of a resource, and (3) parameter  $\mu$  used to measure the differential usage of a resource. Among them, parameter  $\sigma$  is used to constrain the identified biclusters to

be constant row. As the data in database *AGEMAP* all satisfy the constant row constraint, parameter analysis experiments here focus on the low usage parameter  $\delta$  and the differential usage parameter  $\mu$ .

*TFCluster* is benchmarked in terms of runtime and result number by varying the low usage parameter  $\delta$  and the differential usage parameter  $\mu$ . The number of functions is fixed at 10, the number of resources is fixed at 300, the differential usage parameter  $\mu$  is fixed at 3, and the low usage parameter  $\delta$  is set to 0.1, 0.15, 0.2, 0.25, ..., 0.35. From Figure 7, it can be seen that as the low usage parameter  $\delta$  increases, the smaller the constraint on the mining results, the more biclusters will be identified, but the running time of the algorithm is largely unaffected by  $\delta$ , remaining at around 0.5 s.

The number of functions is fixed at 10, the number of resources is fixed at 300, the low usage parameter  $\delta$  is fixed at 0.3, and the differential usage parameter  $\mu$  is set to 3, 4, 5, ..., 8. From Figure 7, it can be seen that as the differential

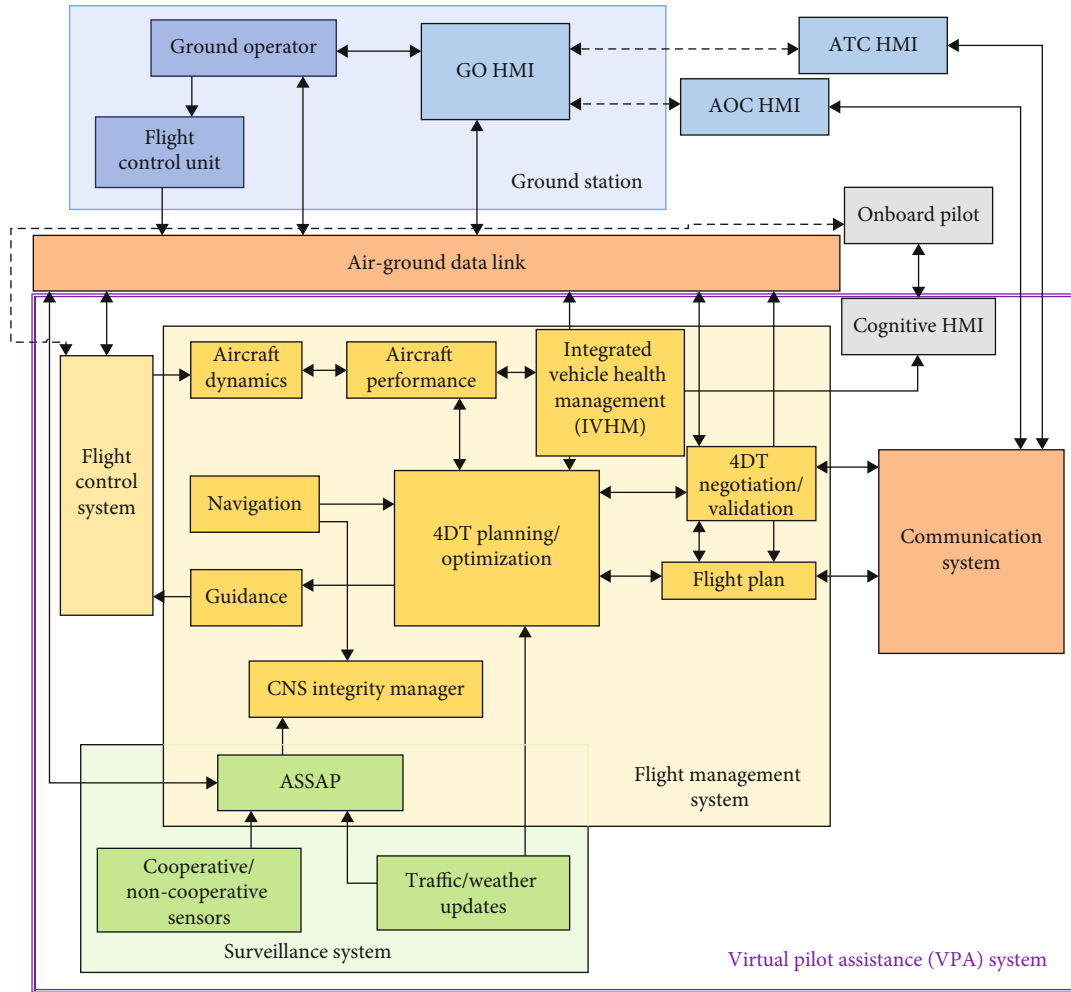


FIGURE 9: SPO system composition.



FIGURE 10: The modeling of the above flight scenario.

usage parameter  $\mu$  increases, the greater the constraint on the mining results, the less biclusters will be identified, but the running time of the algorithm is largely unaffected by  $\mu$ , remaining at around 0.5 s.

Overall, the experimental results show that *TFcluster* outperforms all the algorithms tested. The resource scalability experiments show that *TFcluster* has excellent scalability even with large datasets. The pruning experiments show that pruning strategies can improve mining effi-

ciency, particularly in large databases. The parameter experiments show that the low-used parameter  $\delta$  and the differential-used parameter  $\mu$  have little impact on mining efficiency, but have a significant impact on mining results.

**4.2. Model-Based Analysis.** Here takes the midterm conflict in the SPO mode as an example to carry out safety analysis. The model-based simulation framework is shown in Figure 8. Flight simulation is realized by the flight scenario

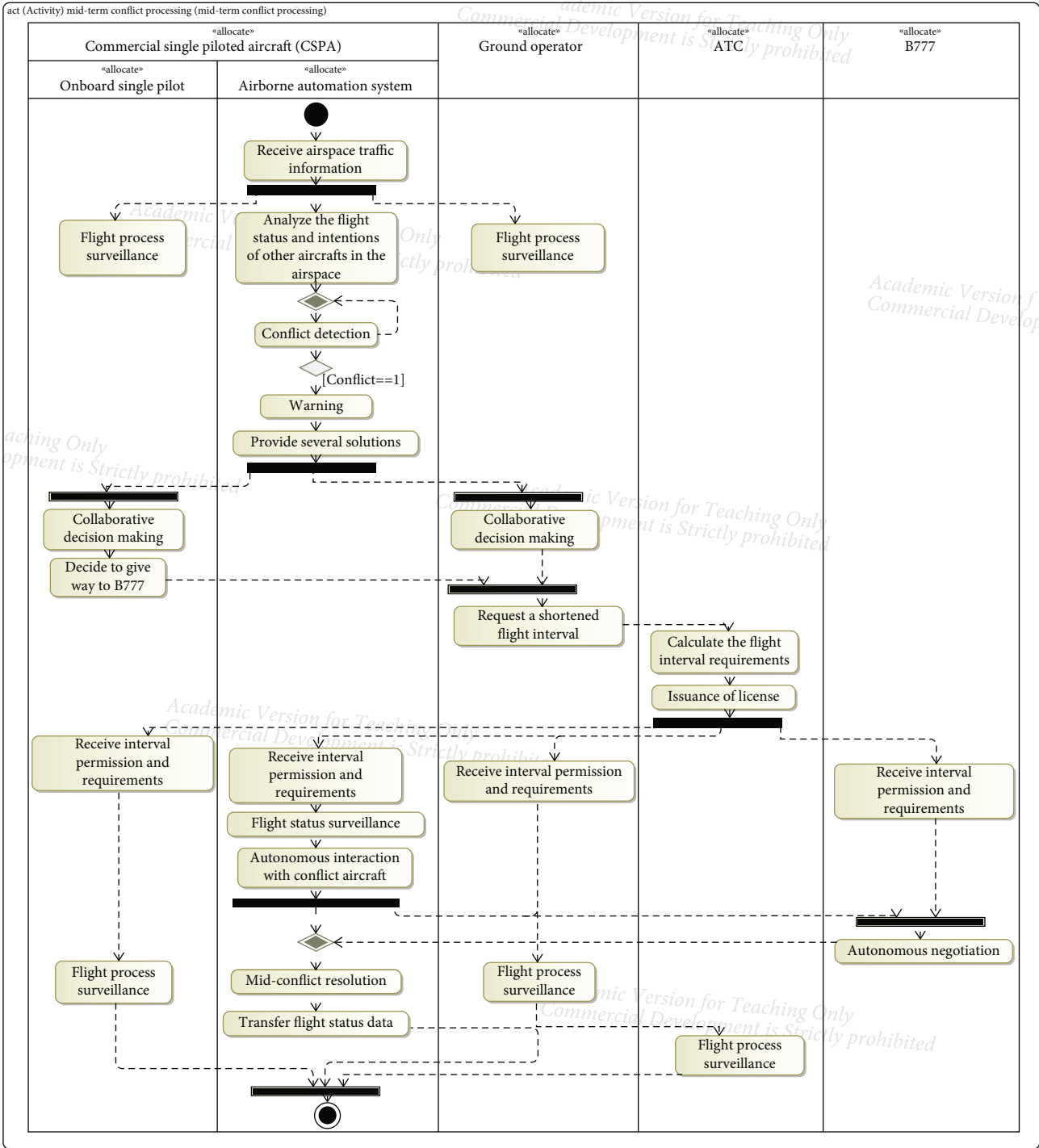


FIGURE 11: The activity diagram of mid-term conflict resolution.

simulation software *Prepar3D*, and SPO system architecture is modeled by *Magic System of System*, including scenario operating model, task organization model, function-resource allocation model, etc. *Prepar3D* passes flight state information (e.g., position, velocity, and attitude) to *Magic System of System* to trigger the execution of the internal system model, and *Magic System of System* feeds back key variables or flight instructions (e.g., velocity variable and ignition instruction) to drive the flight simulation. Then,

the real-valued function-resource matrices of different tasks are extracted from *Magic System of System* model. Next, potential hazard patterns could be identified using the proposed *TFCluster* algorithm, and based on this, safety analysis could be carried out. Finally, the SPO system model could be jointly verified and iteratively designed based on the mining results.

In the future SPO system, virtual pilot assistance (VPA) system [14] is introduced to reduce the workload of the pilot

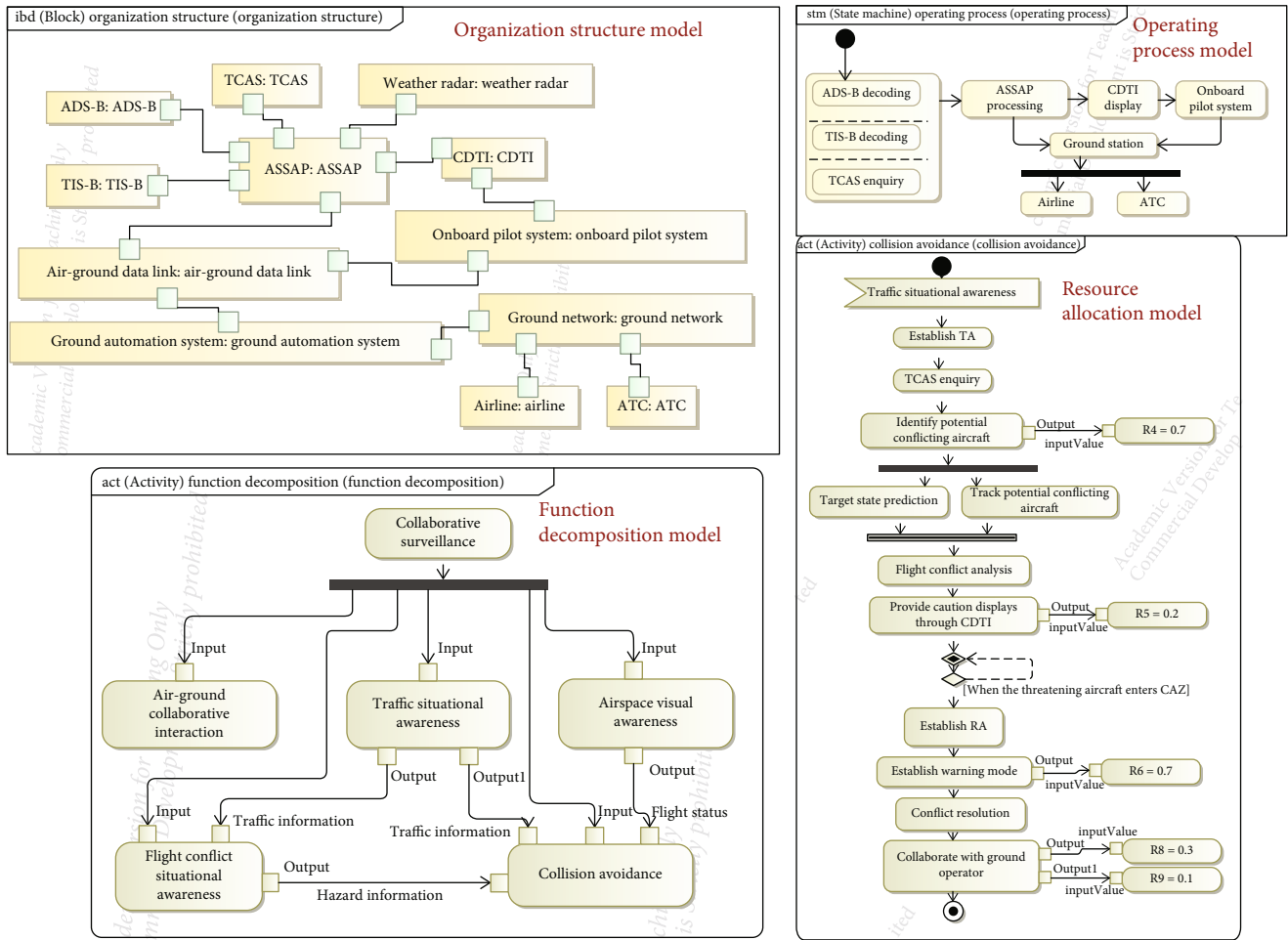


FIGURE 12: The model of midterm conflict resolution in SPO.

onboard, enhance aircraft surveillance capability, and promote air-ground collaboration and information sharing, as shown in Figure 9. The communication system ensures information sharing between aircraft and ground station, ATC, and airlines to support air-ground collaborative decision-making. Airborne Surveillance and Separation Assurance Processing (ASSAP) subsystem is applied by the surveillance system, which is integrated into the flight management system (FMS) to provide automated SA&CA capabilities. FMS obtains flight environment information through cooperative or noncooperative sensors, and receives real-time traffic and weather information from ATC and airlines through data link to provide guidance, navigation and control, and supports 4DT route planning, optimization, negotiation and verification with the pilot through human-machine interface.

Then, taking the midterm conflict resolution during land-based cruise in SPO as an example, this paper conducts a safety analysis of SPO based on the combination of MBSA and hazard pattern mining. The specific flight scenario is designed as follows: during the flight, commercial single-piloted aircraft (CSPA) receives airspace traffic information through ADS-B and Traffic Information Services-Broadcast (TIS-B), and weather information through weather radar.

The flight management system (FMS) analyzes the surveillance information and finds that there will be a cross-route conflict with the B777 after 20 minutes, that is, their minimum separation does not meet the safety isolation requirements. FMS automatically warns and provides several route modifications to resolve the potential conflict. The onboard pilot and the ground operator decide to give way to the B777 after air-ground collaborative decision. Since the two aircraft have self-isolation capabilities, the ground operator of CSPA requests a shortened flight interval from ATC. ATC calculates the flight interval requirements based on the conflict location, meteorological conditions, navigation accuracy and airspace traffic, and sends the flight interval constraints to CSPA and B777. The medium-term conflict resolution of CSPA and B777 is carried out autonomously by themselves, and ATC only monitors the entire process. The above flight scenario was modeled and visualized by *Prepar3D*, as is shown in Figure 10, and the activity diagram of the above flight scenario is shown in Figure 11.

Midterm conflict does not require an immediate response from the pilot, but needs to be handled by the pilot on board, the ground operator and the ATC together. The system operating process model, organization structure model, function decomposition model and resource

TABLE 3: Mining results of TFCluster.

	Mining results	SI
1	$F_1 F_2 (*R_1 - R_3 - R_4 - R_6 - R_7)$	$R_1 : 0.4/R_3 : 1.5/R_4 : 3/R_6 : 3/R_7 : 3$
2	$F_1 F_2 F_3 F_4 F_5 (-R_1)$	$R_1 : 0.4$
3	$F_1 F_2 F_4 F_5 (-R_1 - R_3 - R_6)$	$R_1 : 0.4/R_3 : 0.35/R_6 : 0.35$
4	$F_1 F_2 F_5 (-R_1 - R_3 - R_6 - R_7)$	$R_1 : 0.4/R_3 : 1.5/R_6 : 1.5/R_7 : 0.35$
5	$F_1 F_3 (-R_1 - R_2 * R_3 * R_{10} - R_{11})$	$R_1 : 1.5/R_2 : 1.5/R_3 : 0.7/R_{10} : 0.4/R_{11} : 1.5$
6	$F_1 F_3 F_4 (-R_1 - R_2 - R_{10} - R_{11})$	$R_1 : 1.5/R_2 : 1.5/R_{10} : 0.4/R_{11} : 1.5$
7	$F_1 F_3 F_4 F_5 (-R_1 - R_2)$	$R_1 : 1.5/R_2 : 1.5$
8	$F_1 F_4 (-R_1 - R_2 * R_3 * R_6 - R_{10} - R_{11})$	$R_1 : 3/R_2 : 3/R_3 : 0.7/R_6 : 0.7/R_{10} : 1.5/R_{11} : 1.5$
9	$F_1 F_4 F_5 (-R_1 - R_2 - R_3 - R_6)$	$R_1 : 3/R_2 : 3/R_3 : 0.7/R_6 : 0.35$
10	$F_1 F_5 (-R_1 - R_2 - R_3 - R_6 - R_7)$	$R_1 : 3/R_2 : 3/R_3 : 3/R_6 : 1.5/R_7 : 3$
11	$F_3 F_4 (-R_1 - R_2 - R_5 - R_9 R_{10} - R_{11})$	$R_1 : 1.5/R_2 : 1.5/R_5 : 1.5/R_9 : 1.5/R_{10} : 0.4/R_{11} : 1.5$
12	$F_3 F_4 F_5 (-R_1 - R_2 - R_5)$	$R_1 : 1.5/R_2 : 1.5/R_5 : 1.5$
13	$F_3 F_5 (-R_1 - R_2 R_3 - R_5)$	$R_1 : 1.5/R_2 : 1.5/R_3 : 0.7/R_5 : 1.5$
14	$F_4 F_5 (-R_1 - R_2 R_3 - R_5 R_6)$	$R_1 : 3/R_2 : 3/R_3 : 0.7/R_5 : 1.5/R_6 : 0.35$

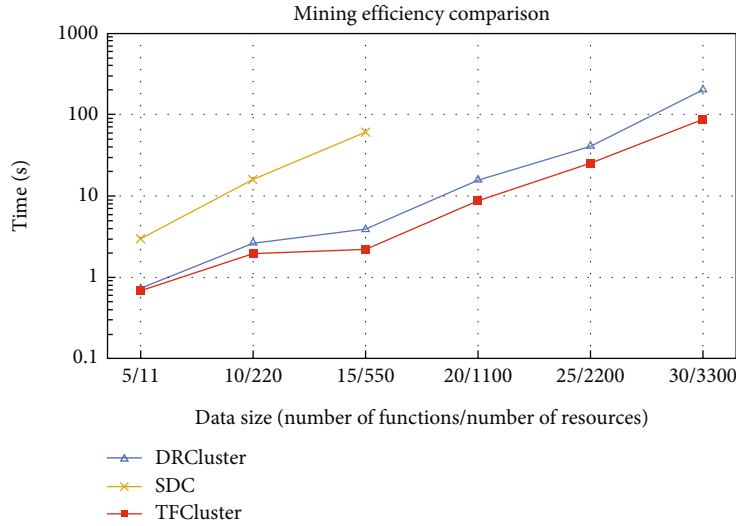


FIGURE 13: Mining efficiency comparison of different algorithms varying dataset size.

allocation model for the above scenario are constructed separately, as shown in Figure 12.

Taking the two flight tasks of collaborative surveillance ( $T_1$ ) and collaborative handling ( $T_2$ ) mentioned in Section 2.1 as an example, their respective real-valued function-resource matrices can be extracted, where  $F_1 - F_5$  denote airspace visual awareness, traffic situational awareness, flight conflict situational awareness, collision avoidance and air-ground collaborative interaction, respectively.  $R_1 - R_4$  denote computer resources: ADS-B decoding, TIS-B decoding, weather radar decoding, and hazard identification.  $R_5 - R_7$  denote execution resources: CDTI display, hazard alarm, TCAS enquiry.  $R_8 - R_9$  denote communication resources: air-ground data link and ground network.  $R_{10} - R_{11}$  denote awareness resources: meteorological awareness and traffic

awareness. The specific data is obtained through human assessment, which is not the focus of this paper.

The *TFCluster* algorithm is used to mine the two  $11 \times 5$  matrices extracted above. The correlation parameter  $\sigma$  is set to 10, the low usage parameter  $\delta$  is set to 0.3, the differential usage parameter  $\mu$  is set to 10. There are 14 items of mining results, as shown in Table 3, and the mining time is 0.669 s.

Then, the original data in the two  $11 \times 5$  matrices are replicated to expand the dataset. Then, algorithms are benchmarked in terms of runtime by varying the number of resources, where the number of functions is set to 5, 10, 15, 20, 25, and 30, and the number of resources is set to 11, 220, 550, 1100, 2200, and 3300. Figure 13 depicts the comparison result. When the number of resources reaches 1100 and



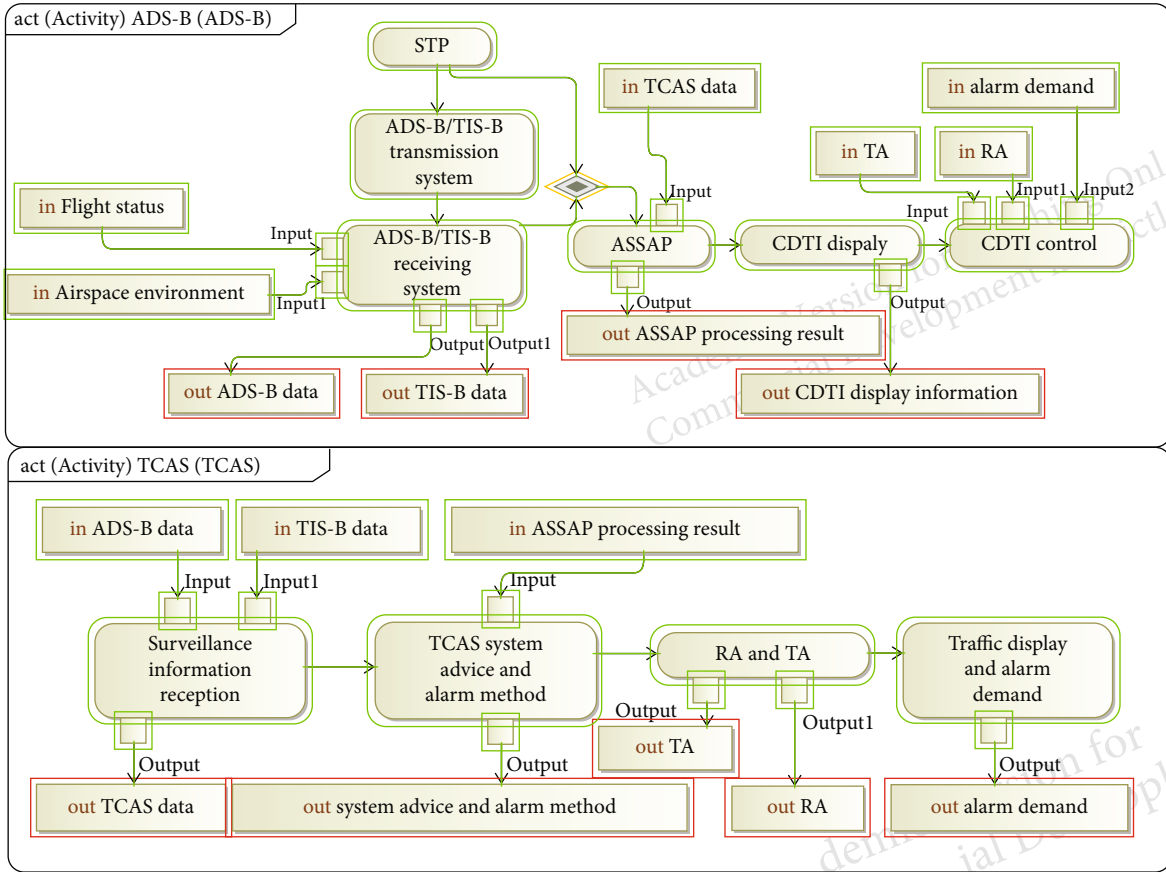


FIGURE 14: The newly constructed surveillance process model of midterm conflict resolution in SPO.

TABLE 4: Mining results of the newly constructed surveillance process.

	Mining results	SI
1	$F_1 F_2 F_3 F_4 F_5 (-R_1)$	$R_1 : 0.4$
2	$F_1 F_2 F_3 F_5 (-R_1 - R_3 - R_7)$	$R_1 : 0.4/R_3 : 0.35/R_7 : 3$
3	$F_1 F_2 F_4 F_5 (-R_1 - R_3 - R_6)$	$R_1 : 0.4/R_3 : 0.35/R_6 : 0.7$
4	$F_1 F_2 F_5 (-R_1 - R_3 - R_4 - R_6 - R_7)$	$R_1 : 0.4/R_3 : 1.5/R_6 : 3/R_7 : 3$
5	$F_1 F_3 (-R_1 - R_2^* R_3 - R_7^* R_{10} - R_{11})$	$R_1 : 1.5/R_2 : 1.5/R_3 : 0.7/R_7 : 3/R_{10} : 0.4/R_{11} : 1.5$
6	$F_1 F_3 F_4 (-R_1 - R_2 - R_{10} - R_{11})$	$R_1 : 1.5/R_2 : 1.5/R_{10} : 0.4/R_{11} : 1.5$
7	$F_1 F_3 F_4 F_5 (-R_1 - R_2)$	$R_1 : 1.5/R_2 : 1.5$
8	$F_1 F_3 F_5 (-R_1 - R_2 - R_3 - R_7)$	$R_1 : 1.5/R_2 : 1.5/R_3 : 0.7/R_7 : 3$
9	$F_1 F_4 (-R_1 - R_2^* R_3^* R_6 - R_{10} - R_{11})$	$R_1 : 3/R_2 : 3/R_3 : 0.7/R_6 : 0.7/R_{10} : 1.5/R_{11} : 1.5$
10	$F_1 F_4 F_5 (-R_1 - R_2 - R_3 - R_6)$	$R_1 : 3/R_2 : 3/R_3 : 0.7/R_6 : 0.7$
11	$F_1 F_5 (-R_1 - R_2 - R_3 - R_4 - R_6 - R_7)$	$R_1 : 3/R_2 : 3/R_3 : 3/R_4 : 1.5/R_6 : 3/R_7 : 3$
12	$F_3 F_4 (-R_1 - R_2 - R_5 - R_9 R_{10} - R_{11})$	$R_1 : 1.5/R_2 : 1.5/R_5 : 1.5/R_9 : 1.5/R_{10} : 0.4/R_{11} : 1.5$
13	$F_3 F_4 F_5 (-R_1 - R_2 - R_5)$	$R_1 : 1.5/R_2 : 1.5/R_5 : 1.5$
14	$F_3 F_5 (-R_1 - R_2 R_3 - R_5 - R_7)$	$R_1 : 1.5/R_2 : 1.5/R_3 : 0.7/R_5 : 1.5$
15	$F_4 F_5 (-R_1 - R_2 R_3 - R_5 R_6)$	$R_1 : 3/R_2 : 3/R_3 : 0.7/R_5 : 1.5/R_6 : 0.7$

the number of functions reaches 20, the SDC algorithm cannot output results. It can be seen that the TFCluster algorithm has high algorithm efficiency and resource scalability.

From the perspective of hazard pattern analysis, the mining results indicate that the concurrent execution of collaborative monitoring and collaborative handling under SPO

can lead to a shortage of air-ground communication resources. To address this issue, we recommend the implementation of dedicated links, such as 5G AeroMACS 2.0, 5G LDACS 2.0, 5G ATG, and 5G public networks [27], as a means of mitigating the pressure on air-ground communication. From the perspective of improving resource efficiency and effectiveness, ADS-B and TCAS can be considered for integrated surveillance to enhance surveillance capabilities in the face of midterm conflict. ADS-B system sends the current traffic situation and predicted trajectory of the aircraft in the airspace to cockpit display of traffic information (CDTI) in real time to establish flight environment situational awareness. ADS-B system determines the distance-based or time-based threat ranking and establishes the conflict avoidance situational awareness, which will be sent to TCAS. TCAS provides the conflict avoidance process organization based on conflict avoidance situational awareness, that is, TCAS flight conflict situational awareness is based on ADS-B traffic situational awareness. The newly constructed surveillance process model is shown in Figure 14.

Then, the newly constructed surveillance process model is mined using the *TFCluster* algorithm and the results are shown in Table 4, where the safety index enhancements have been bolded. Taking the biclusters  $F_1 F_5(-R_1 - R_2 - R_3 - R_6 - R_7)$  in Table 3 and  $F_1 F_5(-R_1 - R_2 - R_3 - R_4 - R_6 - R_7)$  in Table 4 as an example, the safety index of resource  $R_6$  TCAS enquiry is improved when function  $F_1$  airspace visual awareness and function  $F_5$  air-ground collaborative interaction are executed simultaneously, and resource  $R_4$  hazard identification is changed from being unsafe to having a safety index of 1.5. It can be seen that after integrated surveillance by ADS-B and TCAS, the safety index of some resources are improved, such as the computer resource  $R_4$  hazard identification and the execution resource  $R_7$  TCAS enquiry, and finally the enhancement of the surveillance capabilities in case of midterm conflicts is achieved.

## 5. Conclusions

Due to the higher degree of air-ground task collaboration and complexity in the SPO mode, the traditional safety analysis methods applied in two-pilot mode cannot effectively identify the potential hazard patterns in the system. Hence, air-ground task collaboration safety analysis based on real-valued differential bicluster mining is carried out in this paper.

- (1) A new safety analysis method combining model-based safety analysis (MBSA) and hazard pattern mining is proposed, which combines the system design process with the safety analysis process, ensuring data consistency at model level
- (2) A real-valued differential bicluster mining algorithm—*TFCluster* algorithm is proposed to analyze whether hazards will occur in specific function-resource allocation scheme under multiple concurrent tasks. Taking the midconflict handling process

in SPO as an example, and the model-based safety verification of SPO air-ground task collaboration is carried out. The newly constructed surveillance process model based on the mining results finally enhances surveillance capabilities in the face of midterm conflict

- (3) In addition, further investigation of more accurate model construction and weighted sequence pattern mining is needed. First, multidimensional matrices can be introduced to study the more complex redundancy design, monitoring management and fault tolerant strategy in SPO. Secondly, in real-world applications, the invocation of functions and resources is based on time intervals, hence adopting weighted sequence pattern mining would be more meaningful for analyzing SPO ground-air task collaboration

## Data Availability

Data used to support the findings of this study are available from the corresponding author upon request.

## Disclosure

Portions of this work were presented at the Digital Avionics Systems Conference (DASC) in 2021, *TFCluster: An Efficient Algorithm to Mine Maximal Differential Function-resource Biclusters for Single-Pilot Operations Safety Analysis*.

## Conflicts of Interest

The authors declare that there is no conflict of interest regarding the publication of this paper.

## Acknowledgments

This research was funded by the Natural Science Foundation of Shanghai (20ZR1427800) and the New Young Teachers Launch Program of Shanghai Jiao Tong University (20X100040036).

## References

- [1] M. Matessa, T. Strybel, K. Vu, V. Battiste, and T. Schnell, *Concept of Operations for RCO/SPO*, 2017, No. ARC-E-DAA-TN44254.
- [2] P. L. Myers and A. W. Starr, "Single pilot operations IN commercial cockpits: background, challenges, and options," *Journal of Intelligent & Robotic Systems*, vol. 102, no. 1, pp. 1–15, 2021.
- [3] D. Schmid and N. A. Stanton, "Progressing toward airliners' reduced-crew operations: a systematic literature review," *The International Journal of Aerospace Psychology*, vol. 30, no. 1–2, pp. 1–24, 2020.
- [4] Y. Shuo, Y. Zhong, W. Yagang, Z. Wei, W. Yongxiong, and Z. Jianhua, "Assessing cognitive mental workload via EEG signals and an ensemble deep learning classifier based on denoising autoencoders," *Computers in Biology and Medicine*, vol. 109, pp. 159–170, 2019.

- [5] A. K. Faulhaber and M. Friedrich, "Eye-tracking metrics as an indicator of workload in commercial single-pilot operations," in *International Symposium on Human Mental Workload: Models and Applications*, pp. 213–225, Springer, Cham, 2019.
- [6] G. Tokadli, M. C. Dorneich, and M. Matessa, "Evaluation of playbook delegation approach in human-autonomy teaming for single pilot operations," *International Journal of Human-Computer Interaction*, vol. 37, no. 7, pp. 703–716, 2021.
- [7] M. Carloni and L. Manica, "Wireless channel modeling analysis and evaluation for single pilot operation," in *Proceedings of the International Conference on Wireless Communication and Sensor Networks*, pp. 1–6, Warsaw Poland, 2020.
- [8] J. Lachter, S. L. Brandt, V. Battiste, S. V. Ligda, M. Matessa, and W. W. Johnson, "Toward single pilot operations: developing a ground station," in *Proceedings of the International Conference on Human-Computer Interaction in Aerospace*, pp. 1–8, New York, NY, USA, 2014.
- [9] D. Schmid and B. Korn, "A tripartite concept of a remote-copilot center for commercial single-pilot operations," *AIAA Information Systems-AIAA Infotech@ Aerospace*, p. 0064, 2017.
- [10] N. A. Stanton, D. Harris, and A. Starr, "The future flight deck: modelling dual, single and distributed crewing options," *Applied Ergonomics*, vol. 53, pp. 331–342, 2016.
- [11] N. A. Stanton, D. Harris, and A. Starr, "Modelling and analysis of single pilot operations in commercial aviation," in *Proceedings of the International Conference on Human-Computer Interaction in Aerospace*, pp. 1–8, Santa Clara California, 2014.
- [12] D. Harris, "A human-centred design agenda for the development of single crew operated commercial aircraft," *Aircraft Engineering and Aerospace Technology*, vol. 79, no. 5, pp. 518–526, 2007.
- [13] S. M. Sprengart, S. M. Neis, and J. Schiefele, "Role of the human operator in future commercial reduced crew operations," in *IEEE/AIAA 37th Digital Avionics Systems Conference (DASC)*, pp. 1–10, London, UK, 2018.
- [14] L. Yixiang, B. Vincent, R. Subramanian, L. Jing, and S. Roberto, "Commercial airline single-pilot operations: system design and pathways to certification," *IEEE Aerospace and Electronic Systems Magazine*, vol. 32, no. 7, pp. 4–21, 2017.
- [15] Y. Zhong and Z. Jianhua, "Cross-session classification of mental workload levels using EEG and an adaptive deep learning model," *Biomedical Signal Processing and Control*, vol. 33, pp. 30–47, 2017.
- [16] M. Narek, P. Alain, D. Charles-Alban, A. Jean-Marc, and S. Didier, "A study of human-machine teaming for single pilot operation with augmented reality," in *IEEE International Symposium on Mixed and Augmented Reality Adjunct (ISMAR-Adjunct)*, pp. 397–402, Bari, Italy, 2021.
- [17] L. Min, D. Dongjin, W. Miao, W. Guoqing, X. Gang, and Y. Xiaofan, "Going SPO: hierarchical task analysis of pilot flying and pilot monitoring in two-crew operations," in *IEEE/AIAA 40th Digital Avionics Systems Conference (DASC)*, pp. 1–5, San Antonio, TX, USA, 2021.
- [18] C. Yong, L. Yue, W. Miao, Z. Kelin, X. Gang, and W. Guoqing, "DFCluster: an efficient algorithm to mine maximal differential biclusters for single pilot operations task synthesis safety analysis," *Chinese Journal of Aeronautics*, vol. 35, no. 5, pp. 400–418, 2022.
- [19] L. Yue, W. Miao, C. Yong, X. Gang, and W. Guoqing, "TFCluster: an efficient algorithm to mine maximal differential function-resource biclusters for single pilot operations safety analysis," in *2021 IEEE/AIAA 40th Digital Avionics Systems Conference (DASC)*, pp. 1–6, San Antonio, TX, USA, 2021.
- [20] X. Yajun, "TCAS/ADS-B integrated surveillance and collision avoidance system," in *Conference of the 2nd International Conference on Computer Science and Electronics Engineering (ICCSEE 2013)*, pp. 666–669, Hangzhou, P.R. China, 2013.
- [21] G. D. Pasa and V. A. de Santiago Júnior, "Aircraft navigation systems safety assessment via probabilistic model checking," in *Computational Science and Its Applications-ICCSA 2021: 21st International Conference*, pp. 465–480, Cagliari, Italy, 2021.
- [22] Y. Fang, X. Chenyang, D. Lei, and W. Peng, "Fault propagation behavior analysis and evaluation of aircraft all-electric brake system under DIMA architecture," *Acta Aeronautica et Astronautica Sinica*, vol. 42, no. 6, article 324674, 2021.
- [23] J. Wang and J. Han, "BIDE: efficient mining of frequent closed sequences," in *Proceedings. 20th International Conference on Data Engineering*, pp. 79–90, Boston, MA, USA, 2004.
- [24] F. Gang, K. Rui, P. Gaurav, S. Michael, L. M. Chan, and K. Vipin, "Subspace differential coexpression analysis: problem definition and a general approach," *Biocomputing 2010*, pp. 145–156, 2010.
- [25] W. Miao, S. Xuequn, L. Xiaoyuan, L. Zhanhuai, and L. Wenbin, "Efficient mining differential co-expression constant row bicluster in real-valued gene expression datasets," *Applied Mathematics & Information Sciences*, vol. 7, no. 2, pp. 587–598, 2013.
- [26] J. M. Zahn, S. Poosala, A. B. Owen et al., "AGEMAP: a gene expression database for aging in mice," *PLoS Genetics*, vol. 3, no. 11, pp. 2326–2337, 2007.
- [27] C. Yong, Z. Kelin, W. Miao, and L. Yue, "The key technology and future development of regional airliner," *Acta Aeronautica et Astronautica*, vol. 44, no. 2, pp. 156–180, 2023, <http://kns.cnki.net/kcms/detail/11.1929.V.20220311.1142.024.html>.