International Journal of Security and Its Applications
Vol.9, No.10 (2015), pp.1-20
http://dx.doi.org/10.14257/ijsia.2015.9.10.01

Efficient Mining Maximal Constant Row Bicluster in Function-resource Matrix for IMA Safety Analysis

Miao Wang[1,2,*], Zhiyong Xiong[1,2], Liang Xu[2], Lihua Zhang[1,2], Qingfan Gu[2] and Guoqing Wang[1,2]

[1] *Science and Technology on Avionics Integration Laboratory, Shanghai, China, 200233*

[2] *China National Aeronautical Radio Electronics Research Institute, Shanghai, China, 200233*

**wang_miao@careri.com*

Abstract

Integrated Modular Avionics (IMA) uses task synthesis, function fusion and resource integration to achieve the goal of low cost, high efficiency, high efficacy, high performance and high reliability. However, safety issue is caused in the system integration process. In this paper, firstly, we use data mining technology to describe resource-layer safety model, function-layer safety model and task-layer safety model; secondly, we proposed an efficient bicluster mining algorithm: LowCluster, to effectively mine all the maximal constant row biclusters with low usage rate in real-valued function-resource matrix for IMA safety analysis. In LowCluster algorithm, a sample weighted graph is constructed firstly, it includes all resource collections between both samples which meet the definition of low usage rate; then, all the maximal constant row biclusters with low usage rate are mined using sample-growth and depth-first method in the sample weighted graph. In order to improve the mining efficiency, LowCluster algorithm uses pruning strategy to ensure the mining of maximal bicluster without candidate maintenance. The experimental results show that LowCluster algorithm is more efficient than traditional constant row biclustering algorithm, and using our proposed LowCluster algorithm can find the error reason when executing more functions, which will help to improve system safety analysis.

Keywords: *IMA; Bicluster; Safety; Constant row; Function; resource*

1. Introduction

Integrated Modular Avionics (IMA) integrates the information, resource, ability and process of avionics system to form unified information collection, united resource allocation, unified capacity organization, unified process synergy, unified function integration and unified system management through sharing, integration, synergy and fusion so as to achieve the goal of low cost, high efficiency, high efficacy, high performance and high reliability. However, safety issue is caused in the system integration process: (1) resource integration gives rise to failure spread, forming system capability operation safety problem; (2) function information fusion causes failure implication and chaos, forming system information processing safety problem; (3) task synthesis leads to the difficulty in diagnosing the failure and expansion of failure damages, forming system result application safety problem. Therefore, it is necessary to analyze IMA safety. In order to effectively analyze safety of integrated modular avionics state, large quantities of state monitoring and data processing modules are used in the design phase. Effective analysis of these data is the foundation for constructing IMA

ISSN: 1738-9976 IJSIA

safety model. However, how to efficiently analyze experimental data is the challenge for traditional safety approaches.

System safety [1] is currently the most important development direction of avionics system. Especially for civil aircraft, no matter what Boeing B787, Airbus A380 or China C919, regards avionics system safety as the core character of system. Avionics system integration is currently important development character of avionics system. The avionics system integrations consist of physical resource integration, information function fusion and task synthesis. However, physical integration gives rise to the problems of failure correlation, transmission and spread; information fusion causes the problems of failure implication, latent and chaos; task synthesis leads to the problems of failure composition, condition and significant diffusion. Above three issues forms the serious challenge of avionics system safety.

The safety, economy, comfort, flight management and environmental protection are five essential attributes of civil aircraft. Safety Directed Development Concept (ARP4754) is the most important feature in civil development. The emphasis of airworthiness certification has changed from eliminating system error to relating potential system hazard. Aiming at design process and system full life circle, basing on organization architecture, System safety is management process which identifies, controls, eliminates, and mitigates potential hazards to acceptable risk ranks. For aircraft, the final object of avionics system organization is to maximize the system capability. The system safety design uses system hazard control to prevent system risk happening or eliminate system injure.

1.1 System Safety Design

The concept of system safety design needs to contain the following three parts.

(1) Safety design object carrier. System safety design confirms the carrier firstly, which is avionics system organization mode. The system safety design contains confirming application requirement, organization mode and technology process, and forms the main object and character of system safety concept. System safety is the system application function requirement. Pure system safety does not exist. For the system, safety is relative. System function requirement conceals the existing hazard, perception, control and acceptable relevant risk rank in the system. System safety is based on system architecture organization. Different architecture organization determines different target, ability and way in the system. Therefore, architecture organization is the basic platform of system safety. Hazard is an existing or latent gather of environment and activity. This gather is an activity which is based on system architecture. System safety accepts different integrated mode influence in the system. Different integrated mode has different relationship with application mode, system mode and physical mode, which forms different respond to related system hazard mode. Through different integrated mode can confirm different hazard influence and form the ability, rank and probability of system safety.

(2) Safety design target. Safety target is based on the mode of safety design object. It confirms the object and content of system safety and forms the organization mode of object safety. Safety target is object-oriented system application mode. According to system application object requirement and system environment condition, based on system application requirement, organization architecture and technology character, safety target requirement confirms system safety object, self-ability and minimum acceptable safety rank of system application function. Safety hazard is based on system safety object and object-oriented latent hazard concept of system lifecycle. According to hazard analysis, safety hazard identifies system hazard mode, system environment, hazard cause, stimulation condition, hazard

influence and risk state; it also builds the method of hazard reduction or elimination. System safety control is based on system safety object and system hazard analysis organization. According to adopt meeting system application function safety minimum requirement and relevant identification, reduction, elimination and control influence system safety technology, safety control mode make sure system setting and control satisfy system application safety rank.

(3) Safety design task. According to system application and organization mode, the task of system safety design is to confirm system hazard identification, warning, isolation, monitoring, elimination or control mode, and build system safety mechanism. Firstly, it can use hazard identification mode for identifying potential hazard form and character; analyzing hazard influence and direction; building hazard controlling and eliminating method. Secondly, it uses hazard control mode for confirming the key element of system acceptable risk; evaluating the cost, probability, injure probability, hazard element and happen frequency of the system high risk and safety; forming controlling or reducing system risk mode. Finally, it uses hazard prompting and warning mode for building hazard and influence warning equipment; modifying hazard operation process; building hazard influence control; and perfecting safety evaluation train.

1.2 IMA System Safety Problems

The IMA system safety contains three problems, which are integrated avionics system task synthesis safety problem, integrated avionics system function fusion safety problem, and integrated avionics system architecture safety organization problem. The task synthesis safety problems contains multiple ability organization conditions problem, multiple process conditions organization problem, and multiple task conditions organization problem. Its main research contents include multiple failure element, condition and hazard mode, and mechanism, safety organization theory under situation ability synthesis, task mode synthesis and system condition synthesis in the task synthesis process. The key and important technologies include task ability of environment situation fault latent synthesis, task organization of function capability error differential synthesis, task management of task failure condition relation synthesis, and the next task of system condition failure transition synthesis.

The function fusion safety problems contain multiple resources sensor capability information fault, multiple resources element performance information fault, and multiple resources function specialty information fault. Its mains research include multiple resources failure implication, latent and chaos mode, and mechanism and safety organization theory under function information fusion, process information fusion and sensor input fusion in the information fusion process. The research centers and difficulties include different function specialty information fusion error chaos, different process reuse information fusion error relation, different environment element information fusion fault latent, and different sensor input data fusion performance fault.

The architecture safety organization problems contain multiple application task condition failure problem, multiple ability function mode error problem, multiple operation resource capacity fault problem. Its main research contents include the system architecture organization method of task organization mode, ability organization mode and physical organization mode, and forms the safety organization of system task architecture, the safety organization of function architecture and the safety organization of physical architecture. The key and important Technologies include multiple system task organization condition and failure mode, multiple system function organization mode and error mode, multiple system resource capability and fault mode, and system safety mode architecture and management mode.

Currently, China’s IMA study is still in the initial stage. There are no special materials about IMA system safety in domestic public literatures and topics. So, it is badly necessary to establish basic theory on IMA system safety, safety guarantee, measurement and evaluation system for IMA system to lay the foundation for studying new generation of IMA system. One of the most challenging issues in system development today is to take into consideration, during development, all possible failures modes of a system and to ensure safe operation of a system under all conditions. Current informal methodologies, like manual fault tree analysis (FTA) [2] and failure mode and effect analysis (FMEA) [3] that rely on the ability of the safety engineer to understand and to foresee the system behavior are not ideal when dealing with highly complex systems, due to the difficulty in understanding the system under development and in anticipating all its possible behaviors.

In order to effectively analyze safety of airplane operation state, large quantities of state monitoring and data processing modules are used in the airplane. According to incomplete statistics, condition monitoring sensors installed in A380 exceed 15000, and these sensors are collecting data all the time when the airplane is flying. Due to numerous sensors, a large quantity of information based on time series will be produced. Effective analysis of these data is the foundation for constructing IMA safety model. These are exactly the research contents of data mining technology. Data mining technology can mine rich information hidden behind these data and transform them to the knowledge which can be used by airplane avionics system so as to achieve safety management of airplane system and flight tasks. Due to the influences of environmental factor, there are many noises in the data collected. The efficiency is low when traditional data mining algorithm is used to mine data. So, it is not applicable to efficient and multi-task safety system. Therefore, we need to apply more efficient data mining algorithm to mine the knowledge of sense data and finally improve system safety rapidly and effectively.

1.3 Our Contributes

Therefore, analysis of the relation among functions and resources can mine the health relations so as to complete the functions through using healthy resources and improve the health degree of functions. The call relation of functions and resources can be denoted as a matrix, in which each row means a resource and each column means a function, the value in the matrix is the usage degree of a function to a resource. This value is defined during functional design, i.e. resource dependence degree of this function in aircraft system in order to complete a function. Through function-resource matrix mining, in order to achieve a group of functions, the resources which can meet all functional demands simultaneously and the resources which can satisfy all functional demands through multiple accesses can be mined. Therefore, using data mining technology can construct the call relation between function and resource, as so to construct the model which can compute the health degree of functions using the health degree of resources.

The above mining concept complies with the mining concept of bicluster in data mining field. Biclustering [4] is a methodology allowing for condition set and item set points clustering simultaneously. It finds clusters of items possessing similar characteristics together with conditions creating these similarities. The main advantage of biclustering is the simultaneous mining module on items and conditions, another advantage is its applicability on original data instead of discretized data. However, mining biclustering approach presents the following four challenges [5]. First, the computing of biclustering method is NP-hard [4]. Second, biclustering method deals with original data, it should adapt to the noise-sensitive character of microarray dataset. Third, the biclustering method should allow overlapping biclusters which share some items or conditions, which would increase the complex of biclustering algorithm. Finally, the biclustering method should be flexible enough to handle different types of biclusters. [6] classified biclusters into four categories.

Currently, large quantities of algorithms based on greedy strategy or exploratory strategy are applied in mining bicluster. Cheng and Church proposed an algorithm based on greedy strategy [2]. This algorithm adopts a low square root residue to delete redundant nodes step by step. After that, many algorithms based on greedy strategy were raised [7-17]. All the above algorithms' efficiency is not high. Thus, to design a high-efficiency bicluster mining algorithm is current research hotspot. So, Wang et al. came up with the mining algorithm to mine the maximal bicluster from discretized data [18, 19]. However, this algorithm cannot effectively mine bicluster with low usage rate meeting difference restraint from function-resource matrix.

Based on the above analysis, in this paper, aiming at the characters of IMA, firstly, we use data mining technology to describe resource-layer safety model, function-layer safety model and task-layer safety model, which is shown in Fig. 1; secondly, we proposed a new bicluster mining algorithm: *LowCluster* algorithm to mine all maximal low usage rate constant row biclusters without candidate maintenance in real-valued function-resource matrix. Since the number of functions is far lower than that of resources in function-resource matrix, our algorithm uses sample-growth method for mining. First, a sample weighted graph is constructed, which includes all the pairs of resources between both samples that meet the definition of low usage rate; then, all maximal biclusters with low usage rate definition and constant row definition are mined using sample-growth and depth-first method in the constructed weighted graph. In order to improve the mining efficiency of the algorithm, *LowCluster* algorithm uses multiple pruning strategies to ensure the mining of maximal bicluster without candidate maintenance.

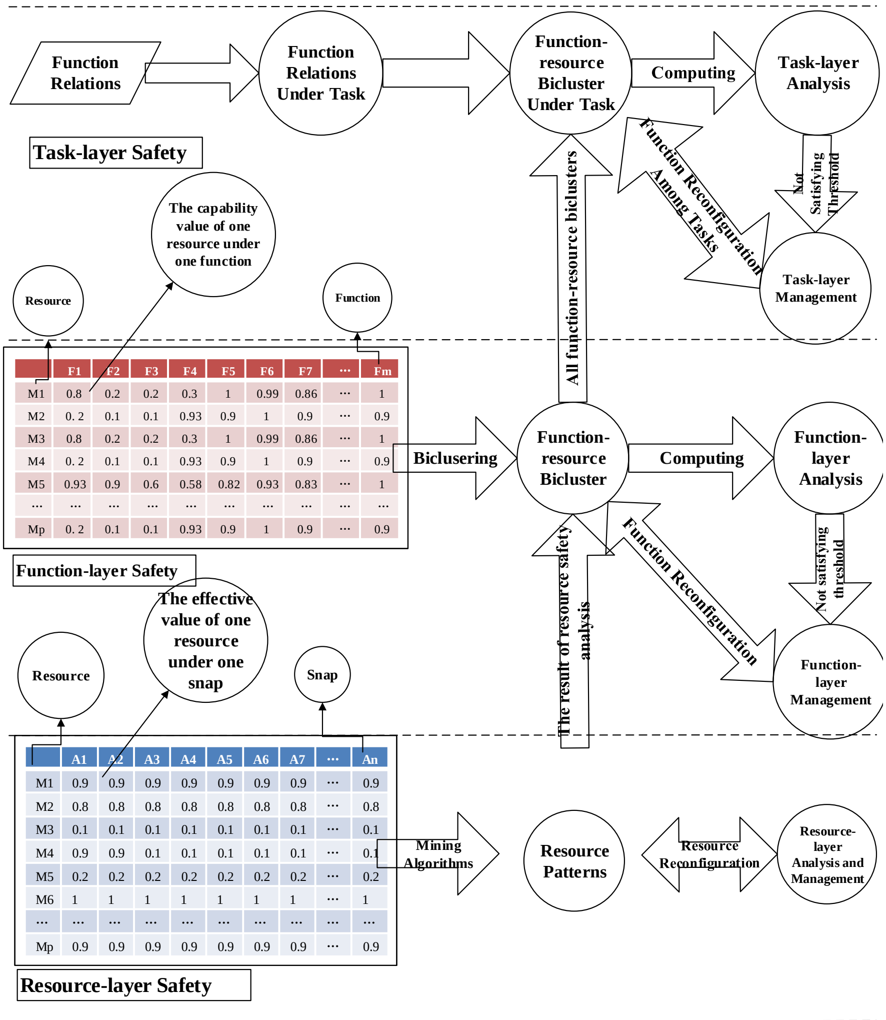

Figure 1. IMA System Safety Monitoring Process based on Data Mining Technology

2. IMA System Safety Analysis

2.1 Safety Problem of IMA-based Resource Sharing Integration

From the perspective of system design, every resource can be regarded as a subsystem with specific ability. On the one hand, physical integration boosts resource utilization rate through resource capacity sharing; on the other hand, it improves resource capacity efficiency through resource capacity organization and allocation. The above two aspects can reduce system resource allocation and decrease complexity influences. But resource integration makes the resource produce capacity relevance at the operation layer and organization layer. Thus, shared resource fault gives rise to failure spread and leads to mutual infection of different function errors, finally triggering greater safety problem. Aiming at the safety problem caused by failure speared resulted from IMA resource dynamic sharing, through studying the elements influencing resource integration, mutual relations among elements and resource capacity model, defining relevant rules, limiting or reducing unpredictable behaviors, resource uncertainty is transformed to system behavior determinacy so as to achieve resource dynamic sharing and reconfiguration based on safety level. To realize the above functions, resource-layer safety management is divided into three parts logically: safety information collection, safety information analysis and safety information modeling. The logic structure is shown in Figure 2.

Safety information collection unit is responsible for collecting all input information related to resource integration, mainly involving resource capacity information, environmental state information, resource state information and resource interaction information. Resource capacity information is the capacity set (resource space) which can be supplied by resource cooperation behaviors. Environmental state information includes ambient temperature, vibration and other information. Resource state information contains validity, fault, sharing, conflict, activation, idleness, current value, voltage value, performance parameter and resource redundancy. Resource interaction information involves the communication between processing resource and communication resource or the communication between processing resource and I/O resource.

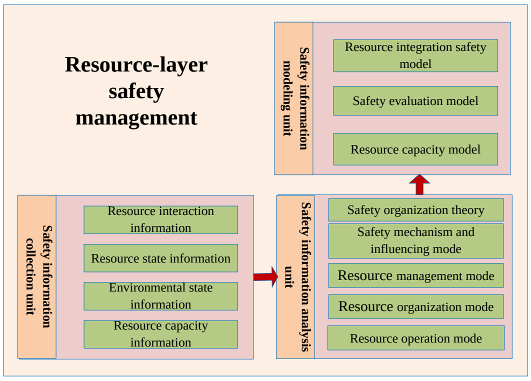

Figure 2. Structure Chart of Resource-layer Safety

Safety information analysis unit takes charge of analyzing and studying the information collected, mainly including resource operation mode, organization mode, management mode, system safety mechanism and influencing mode. On this basis, resource safety organization theory is formed. Resource organization mode refers to different resource allocations according to resource requirements. Different resource organizations have diverse capacity integration ways, while the influences of capacity relevance caused by different capacity integration ways on system safety and results are also different. Resource operation mode means due to different resource types, resource operation modes and results are different. Thus, heritage relevance safety mechanism and influencing modes are different. Resource management mode refers to resource allocation management based on resource demands. Through resource capacity organization and allocation, resource sharing based on time sharing and changes of system state can be realized. Since resource capacities are different, the changes of system state are also diverse. Thus, safety influences caused by state relevance are different, too. From the above, it is necessary to analyze and study safety mechanism and influencing mode and form integrated safety organization theory.

Safety information modeling unit contains modeling for resource capacity model, safety evaluation model and resource integration safety model. Resource capacity model is a process of mapping resource space, system state and demand space to resource capacity space. The following three aspects are mainly considered: 1) resource operation behavior (resource space); 2) organization behaviors among resources (demand space); 3) resource allocation management (system stage space). In resource space, the operation attribute of resource capacity model is formed through analysis of resource operation mode. In demand space, sharing attribute of resource capacity model is formed through analysis of resource organization mode. In system state space, management attribute of resource capacity model is formed through analysis of resource management mode. Safety evaluation model means to study the technique to assess the influence of resource integration on system safety to achieve safety evaluation of resource integration aiming at safety requirement of integrated system resource capacity sharing. Resource integration safety model considers inter-behaviors of components, constructs the relationship between safety elements influencing system safety behaviors and safety elements, forms resource equivalence class and adopts formalization method to set up system safety model based on event relevance and behavior restriction on the basis of analyzing system safety mechanism caused by resource capacity sharing, operation result sharing and resource management mode sharing.

2.2 Safety Problem of IMA-based Function Information Fusion

Function information fusion means to establish function specialty, process reuse and input difference processing modes and form system function information fusion capability according to function structure organization. Function information fusion optimizes system function organization, boosts the quality of mission system function capability and specifies system validity capability. However, although information quality improves, sensor input, processing process reuse and processing capability fusion arouse failure fuzziness, implication and chaos, which gives rise to difficulties for failure diagnosis and prognostic, thus leading to safety problems.

Aiming at safety problems resulted from function information fusion, this paper studies the influence mechanism and influencing mode of function information fusion on safety from three attributes (professional complementation of capability model, process reuse and input difference) through setting up function capability model, analyzes dependency relationship of each function in the system, failure spread mode, sequential relationship of function invalidation as well as time-dependent dynamics and random characteristics of system state in the failure state and finally achieves function reconfiguration based on safety level. In a bid to realize the above functions, function-layer safety management is similar to resource-layer safety management and also divided into three parts logically. The logic structure is shown in Figure 3.

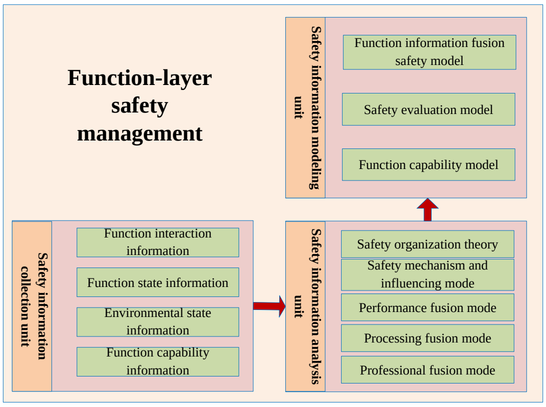

Figure 3. Structure Chart of Function-layer Safety

Safety information collection unit is responsible for collecting function capability information, environmental state information, function state information and function interaction information. Function capability information is the capability set which is supplied by the function itself (individual space). Environmental state information includes ambient temperature, vibration and other information. Analysis of environmental state information aims to the effects of environmental changes on function performance. Function status information contains basic ability and track information of the function, interior structure and top-grade configuration information of the function as well as function state parameter information. Function interaction information involves topological relation and cross-linking relation information among functions.

Safety information analysis unit takes charge of analyzing and studying the information collected, mainly including analysis and research of professional fusion mode, processing fusion mode, performance fusion mode, system safety mechanism and influencing mode. Safety organization theory is formed in this basis. Function specialty refers to the classification of function execution results based on system resource capacity type. Professional fusion mode means to study the influencing mode of function fusion on safety from the perspective of professional complementation. Processing fusion mode means to study the influencing mode of function fusion on safety from the perspective of input difference. Since safety influencing factors related to function capability caused by function specialty fusion, safety influencing factors related to function quality caused by function processing fusion and safety influencing factors related to sensors caused by function performance fusion are different, it is necessary to analyze capability integration mode of each function. In addition, it is required to analyze corresponding safety mechanisms and finally form safety organization theory of function information fusion.

Safety information modeling unit contains modeling for function capability model, safety evaluation model and function information fusion safety model. Function capability model describes function capability from three aspects: result capability, processing capability and input capability. Since function specialties are different, the processing process is different, too. Thus, the influences on result quality are diverse. Therefore, professional capability attribute of capability model is formed through analysis of professional capability fusion. As the function environments are diverse, the influences of processing process on result ability are different. Thus, processing capacity attribute of capability mode is formed through analysis of processing fusion mode. As the input of function information processing, the usability, integrity and reliability of resources directly influence system information processing quality. So, input capacity attribute of function capability model is formed through analysis of sensor information fusion of function performance. Safety evaluation model means aiming at safety demand of integrated system information processing to study the technique to assess the influences of function information fusion on system safety, achieve evaluation of the safety of function information fusion. The operation of integrated avionics system is based on information processing, so the precondition of achieving system safety operation is to study system function fusion safety model and explore safety influencing factors in operation. Function fusion safety model establishes information processing logic equivalence model, analyzes interaction process of function logic, defines system safety state, studies the causes and finally forms system function fusion safety model based on research and analysis of function capability model.

2.3 Safety Problem of IMA-based Task Organization Synthesis

In application space, the integration of new generation of avionics system is reflected in mission organization synthesis. Mission synthesis involves the significance of two layers: on the one hand, in line with total user demands, combine avionics system resource and function capability to disintegrate demands to the resources and functions with corresponding capability (the capability to meet demands), i.e. generation process of mission execution plan; on the other hand, based on mission plan and current resource and function capacities as well as system state, select suitable path to execute the mission, i.e. execution process. Essentially, mission synthesis is process integration. Mission synthesis achieves system mission organization optimization and improves mission system application efficiency. Although it can improve mission efficiency, it meanwhile causes such problems that it is hard to confirm failure state and diagnose failure constitution. Aiming at the difficulty of failure diagnosis caused by mission synthesis, this paper starts from the three aspects (environmental conditions, mission object and mission process) influencing mission completion and focuses on mission organization decision process and improving mission execution management ability to reveal the relationship among the elements influencing system safety at the organization layer of system mission, establishes system application safety evaluation system and provides theoretical support for analysis and evaluation of application safety of IMA system.

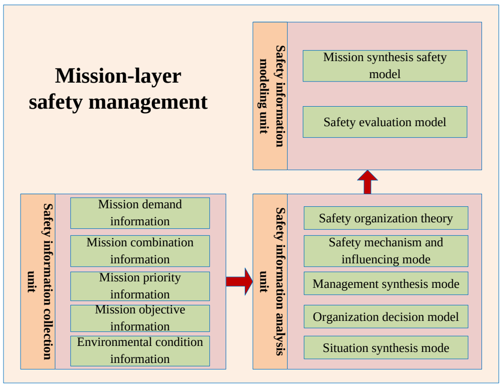

Figure 4. Structure Chart of Mission-layer Safety

To achieve the above functions, mission-layer safety management is also classified into three parts logically. The logic structure is shown in Fig.4. Safety information collection unit includes environmental condition information, mission objective information, mission priority information, mission combination information and mission operation information. Similar to function capability information and resource capacity information, mission demand information is the demand set of mission operation behaviors transforming to system space; mission objective information is the specific description information of mission purpose; mission priority information refers to the priority relation of a series of missions; mission combination information is the organization relationship among missions, i.e. which missions can be executed simultaneously, which missions are of mutual exclusion during execution; mutual exclusion ability condition information covers multi-source sensor information, navigation information and corresponding 3D terrain data.

Information analysis unit is in charge of analyzing situation synthesis mode, organization decision model and management synthesis mode, analyzing the influencing modes on safety and exploring safety mechanism on this basis and finally forming safety organization theory of mission synthesis. Situation synthesis aims at environmental condition elements of the mission to provide perceiving ability for the mission. It is the base to boost mission synthesis validity. Situation synthesis can be deemed as longitudinal capability synthesis of system application mission. The synthesis result corresponds to a snapshot in time series of application mission execution. Organization decision aims at mission environment ability to provide decision reference for mission synthesis and finally provides safety design and verification of avionics system. Mission organization synthesis is lateral organization synthesis of system application mission. In essence, the process is the generation process of a mission plan. Management synthesis refers to element management in the process of mission execution. Under the conditions where the resource has fault, the function goes wrong or the mission breaks down, the effects on system safety can be eliminated or reduced through effective state organization and management.

Information modeling unit contains the modeling for safety evaluation and mission synthesis safety. Mission synthesis safety model studies the elements influencing mission objective, environmental conditions and mission process and their relations from application perspective and guides mission execution plan and generation process of mission execution. Safety evaluation model aiming at integrated system application safety requirements researches the technique to assess the influences of mission synthesis on system safety and achieves evaluation of mission synthesis safety. Evaluation of mission-layer safety can be finished through function usage allocation relation of one or multiple missions and combining function safety model and resource safety model.

3. Mining Maximal Constant Row Biclusters

3.1 Problem Definition

Function-resource matrix is defined as a two-dimensional real matrix $D = R \times F$. Row set R represents the set of resources and column set F refers to the set of functions. Element D_{ij} of matrix D is a real number which represents the ability validity or usage rate of resource i supporting function j. |R| is the number of resources in data set D and |F| is the number of functions in data set D. For the convenience of mining, the domain of definition of the original effective value in resource effectiveness matrix is [0,1], where '0' means that this resource is not required during the implementation of some function; '1' means that this resource must be used during the implementation of some function, as shown in table 1.

Table 1. An Example of Function-resource Matrix

	F_1	F_2	F_3	F_4	F_5
R_1	0.8	0.1	0.12	0.09	0.9
R_2	0.2	0.9	0.19	0.21	1
R_3	0.9	0.3	0.29	0.28	0.55
R_4	0.58	1	0.2	0.21	0.9

Table 2. An Example of Low Usage Rate

	F_1	F_2
R_1	0.8	0
R_2	0. 2	0.1
R_3	0.5	0
R_4	0	0.1

The significance of bicluster to be mined from function-resource matrix as shown in Table 1 is to mine a group of functions executed; under this group of functions, the usage rate of the resource is higher, i.e. which resources can reach the highest usage rate when used together. In other words, the resources have the highest effectives when all functions are executed. For example, for a group of functions F_1F_2 (F_1==>$R_1R_2R_3$, F_2==>R_2R_4), these three functions may be called simultaneously. For resource R_2, there are three situations for supporting F_1 and F_2. For both F_1 and F_2, the usage rate of R_2 is low, as shown in Table 2, the health degree is higher. Since the resource R_2 can serve F_1 and F_2 at the same time. This paper proposed algorithm *LowCluster* aims to mine all the maximal low usage rate constant row biclusters in real-valued function-resource matrix. We will give definitions of low usage rate of resources in real data below:

Definition 1. D is a function-resource usage rate matrix; α is a user-defined parameter used for measuring the degree of association of functions in resources; β is a parameter restricting low usage rate of resources; r is any resource in function-resource usage rate matrix D; F_1 and F_2 are any two functions in D; r should meet the following conditions for relevance in F_1 and F_2 [$\forall r \in R \mid (\max_{f\in\{F_1,F_2\}} D_{r,f} - \min_{f\in\{F_1,F_2\}} D_{r,f}) \le \alpha(\min_{f\in\{F_1,F_2\}} |D_{r,f}|)$ and $\max_{f\in\{F_1,F_2\}} D_{r,f} \le \beta$] . If all resources and functions meet the conditions above in a bicluster, this bicluster is one with low usage rate of resources.

It can be obtained from the description in definition 1 that α and β are used to restrict resources with a low usage rate producing each function, e.g. bicluster $F_2F_3F_4$ (R_1R_3) in table 1.

Therefore, each resource in bicluster mined by *LowCluster* algorithm satisfies the definition 1. To improve the mining efficiency of the algorithm, *LowCluster* algorithm mines all the maximal low usage rate constant row biclusters without candidate maintenance using sample-growth method in real-valued function-resource matrix. The mining process of this algorithm will be introduced in detail in the next section.

3.2 The LowCluster Algorithm

The mining process of *LowCluster* algorithm can be divided into two steps: firstly, scan original function-resource matrix; according to the definition of biclusters with low usage rate and constant row definition, the sample weighted graph are produced; then, all the maximal constant row biclusters with low usage rate are generated using sample-growth method.

(1) Construct the Sample Weighted Graph

LowCluster algorithm in this paper will adopt undirected sample relational weighted graph (hereinafter referred to as sample weighted graph) to mine maximal constant row biclusters with low usage rate.

Definition 2. Sample weighted graph can be denoted as the set $G = \{E, V, W\}$. Each node in the node set V in the weighted graph represents a function. If an edge exists between a pair of nodes, this means the resource with variant usage rate or low usage rate exists below two functions represented by this pair of nodes. The set of the edges is expressed as E. The weights of each edge are the resource set meeting the definition of variant usage rate or the definition of low usage rate under the two functions connected with this edge. The set of the weights is expressed as W.

According to the description in Definition 1, when the resources among functions satisfy the definition of low usage rate, the weight between two functions meets commutativity. For instance, the weight under F_3F_4 is $R_1R_2R_3R_4R_5$, while the weight under F_3F_4 is $R_1R_2R_3R_4R_5$. Figure 5 shows sample weighted graph corresponding to Table 1.

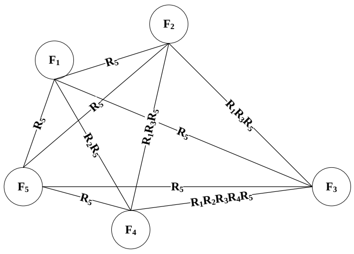

Figure 5. The Sample Weighted Graph Constructed from Table 1

(2) Mining Maximal Constant Row Biclusters with Low Usage Rate

After the sample relational weighted graph is constructed, this section will introduce how *LowCluster* algorithm mines all the maximal constant row biclusters with low usage rate in the sample weighted graph without candidate maintenance in detail. According to the description in definition 1 and theorem 1, biclusters with low usage rate extended meet anti-monotonicity, i.e. if the bicluster obtained by extension of $F_1F_2...F_n$ does not meet constraint conditions, neither does any superset $F_1F_2...F_nF_m$. Therefore, biclusers with a greater scale can be obtained by extension of the weight on each edge in the weight graph in terms of intersection. According to descriptions in definitions 1, when a new function is introduced in bicluster, it is necessary to calculate the intersection of all edges of the function newly introduced and the resource collection of bicluster extended, thus ensuring that the resource collection under the function newly introduced and that under existing functions meet constraint conditions in definition 1.

Theorem 1. The sample range support measure is anti-monotonic.

Theorem 1 states that our low usage rate satisfies the anti-monotonic property, which guarantees that we can use *Apriori*-like efficient pattern mining framework to discover biclusters. Therefore, our *LowCluster* algorithm adopts *Aprioir*-like procedure to produce constant row bicluster with low usage rate using sample-growth in sample weighted graph. The following lemma can guarantee that *LowCluster* algorithm mines all the maximal low usage rate constant row biclusters without candidate maintenance.

Lemma 1. Given P be the current extending constant row bicluster with low usage rate, M is the candidate sample set of P and N is the priori candidate sample set of P. Supposed the current candidate sample is $M_i, M_i \in M$, and N_j is a priori candidate sample where $N_j \in N$. If it is satisfied the following criteria: (1) PN_jM_i.*Resources* is the same as PM_i.*Resources*; (2) For each other candidate sample M_p in M, PN_jM_i.*Resources* is the subset of PN_jM_p.*Resources*, all the constant row bicluster with low usage rate generated by PM_i can also be produced by extending PN_j.

Lemma 1 states how to escape of producing non-maximal low usage rate constant row biclusters without candidate maintenance. According to above lemma, *LowCluster* algorithm exploits the following pruning technique to achieve mining maximal low usage rate constant row biclusters without candidate maintenance in real-valued function-resource matrix.

Pruning 1. Given P be the current extending constant row bicluster with low usage rate, M is the candidate sample set of P and N is the priori candidate sample set of P.

Supposed the current candidate sample is M_i, $M_i \in M$, and N_j is a priori candidate sample where $N_j \in N$. If it is satisfied the following criteria: (1) *PN_jM_i.Resources* is the same as *PM_i.Resources*; (2) For each other candidate sample M_p in M, *PN_jM_i.Resources* is the subset of *PN_jM_p.Resources*, M_i should be pruned.

We will explain the algorithm mining process through an example. The data in the example are function-resource use relationship matrix shown in Table 1. Firstly, construct the sample weighted graph among functions, as shown in Fig.5. The mining process for *LowCluster* mining table 1 expressed matrix is shown in Fig.6. The specific description of *LowCluster* algorithm is as follows:

Algorithm 1: *LowCluster* algorithm

Input: number threshold: n, coherent threshold：α, low usage rate threshold: β, function-resource matrix: D

Output: all the maximal constant row biclusters with low usage rate meeting the threshold

Initial value: sample weight graph: G =Null, current bicluster to be extended Q =Null, S_i=Null and S_j=Null.

Algorithm description: LowCluster(n，α，β，D，Q，S_i，S_j)

(1) If G is null, scan data set D and make its weight graph. S_i is the first sample in the weight graph;
(2) For each sample S_j connected with sample S_i
(3) If all resource linked lists in S_j meet pruning technique, then
(4) Continue;
(5) Else
(6) For resource linked lists not meeting pruning conditions, *Q.Sample*= *Q.Sample*$\cup S_j$; *Q. Resource*= *Q.Resource*$\cap S_iS_j$*.Resource*;
(7) LowCluster(n，α，β，D，Q，S_i，S_j);
(8) Endif
(9) Endfor
(10) If Q meets output conditions, then
(11) Output Q
(12) Endif;
(13) S_i = S_i->next;
(14) Return

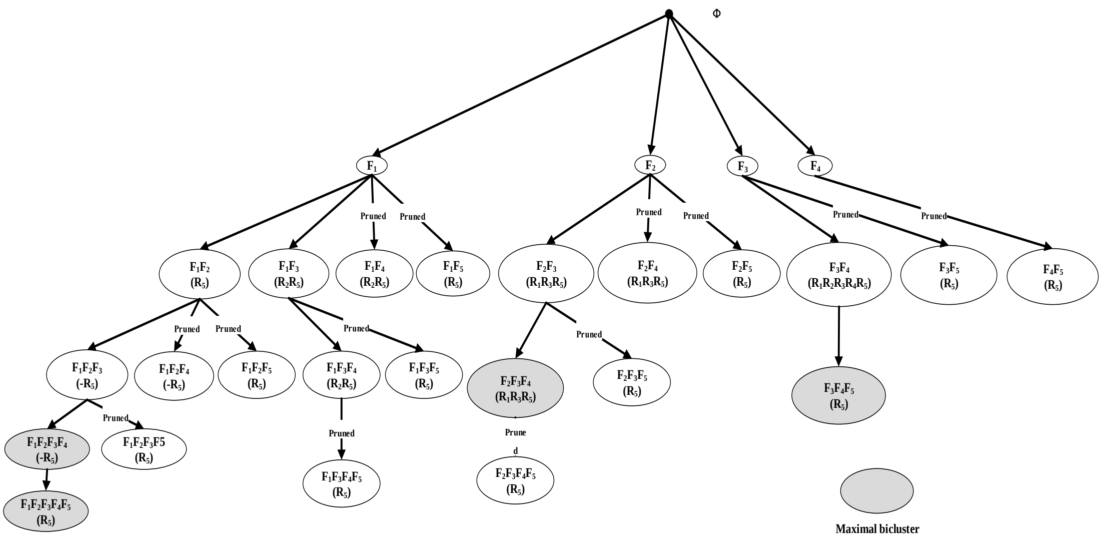

Figure 6. The Example Mining Procedure of *LowCluster* Algorithm

4. Experimental Results

4.1 Performance Analysis

In this section, we will make an experimental comparison on the mining efficiency and result of the algorithm above and existing algorithms. The hardware environment of the experiment is Intel(R) Core(TM)2 Duo 2.53GHz CPU and 4G internal memory; the software environment is Microsoft Windows 7 SP1 operating system; the algorithm programming and operating environment is Microsoft Visual C++ 6.0 SP6. Experimental data used in this paper are simulation data. To fully test the performance of the algorithm, we produce three data sets randomly, each of which contains 20 sampling sites and 1000 resources. Table 5 describes proportions of 0, 0.1, 0.2 and 0.8 in each row in each data set.

Table 3. The Proportion of Each Value in Three Data Set

	0	0.1	0.2	0.8
D_1	0.2	0.2	0.2	0.4
D_2	0.2	0.3	0.3	0.2
D_3	0.4	0.2	0.2	0.2

In this section, we will compare the mining efficiency between *LowCluster* algorithm and *RAP* algorithm [11]. In order to fully compare the extendibility of algorithms, we produce multiple groups of data sets with different numbers of resources and functions in allusion to three data sets in table 2. The selection of resources and functions are based on the order of resources and functions in data set. The parameter of low usage rate is 0.5. Figs 7(a)-7(b) provide the comparison of performance period when the number of functions of two algorithms above is 10, 20 and 30 respectively and the number of resources is 200, 400, 600, 800 and 1000 respectively and the parameter of relevancy under data set D_1 is 1. It can be seen from these figures that the mining time of both algorithms increases progressively with the increase of number of resources in data set. Meanwhile, the mining efficiency of *LowCluster* algorithm is higher than that of *RAP* algorithm under each data size. Especially when the number of resources in data set is high, the mining efficiency of *LowCluster* algorithm is nearly 20 times higher than that of *RAP* algorithm. Meanwhile, according to the increase of the total of functions, *LowCluster* algorithm is nearly 100 times higher than that of *RAP* algorithm. The reason is that *RAP* algorithm mines bicluster with the method of resource extension. With the increase of number of resources in data set, this algorithm needs more iterations to mine all biclusters meeting threshold conditions. However, *LowCluster* algorithm uses high-efficiency pruning strategies for mining and will produce more maximal biclusters especially when the number of resources in data set is high and data are dense. Therefore, *LowCluster* algorithm has a higher pruning efficiency. Figures 8(a)-8(b) provide the comparison of performance period under data sets with different resources of functions and resources when the parameter of relevancy of three algorithms above is 2 in data set D_1. Similar to the description in figs 7(a)-7(c), the mining efficiency of *LowCluster* algorithm is higher than that of RAP algorithm under each data size.

Figures 9(a)-9(b) and Figures 10(a)-10(b) respectively provide the comparison of performance period of both algorithms above under data sets with different numbers of sampling sites and resources when their parameters of relevancy under data set D_2 are respectively 1 and 2. It can be seen that, as proportions of 0.1 and 0.2 in data set D_2 increase compared to those in data set D_1, according to descriptions of the definition of variant usage rate and low usage rate, mining data set D_2will produce more biclusters than mining data set D_1 under the same parameter. Therefore, when the number of functions is 20, *RAP* algorithm cannot mine data sets with the number of resources higher than 400 in limited memory space, but *LowCluster* algorithm can complete all mining processes within 10 seconds. Figures 11(a)-11(b) and Figures 12(a)-12(b) respectively provide the comparison of performance period of both algorithms above under data sets with different numbers of sampling sites and resources when their parameters of relevancy under data set D_3 are respectively 1 and 2.

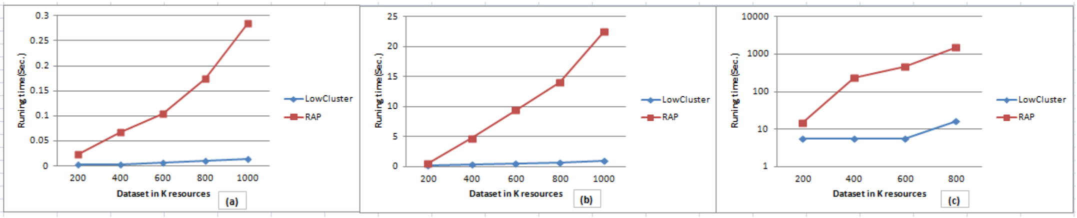

Figure 7. The Running Time Comparison between Two Algorithms under Different Number of Resources and Functions in D_1 when α=1: (a) 10 Functions; (b) 20 Functions

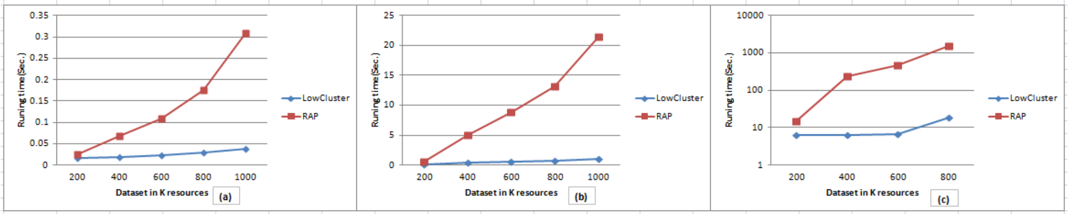

Figure 8. The Running Time Comparison between Two Algorithms under Different Number of Resources and Functions in D_1 when α=2: (a) 10 Functions; (b) 20 Functions

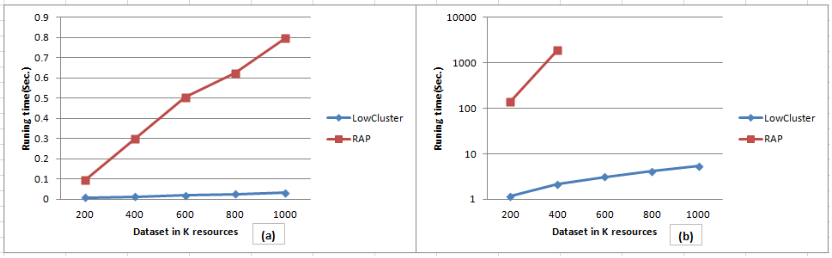

Figure 9. The Running Time Comparison between Two Algorithms under Different Number of Resources and Functions in D_2 when α=1: (a) 10 Functions; (b) 20 Functions

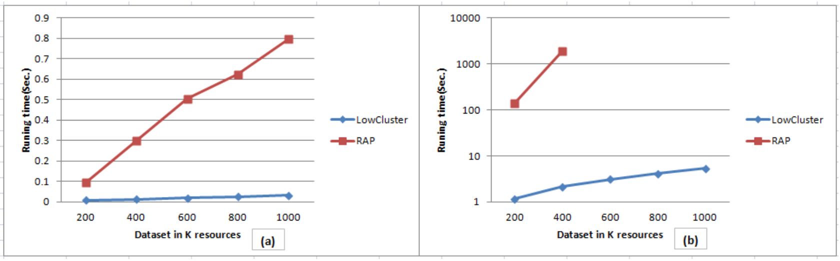

Figure 10. The Running Time Comparison between Two Algorithms under Different Number of Resources and Functions in D_2 when α=2: (a) 10 Functions; (b) 20 Functions

4.2 Safety Analysis

In this section, we will use a simple experiment to show how to use our proposed algorithm to improve system safety analysis. Figure 13 shows a simple composition of electronic aircraft fuel provision system, which is composed of a pump, a backup pump, a pump monitor, a fuel valve, and a fuel tank. The working concept of electronic aircraft fuel provision system is, the pump monitor monitors the conditions of the pump and the backup pump; when the fuel is not enough, the fuel valve would shut the pump and open the backup pump.

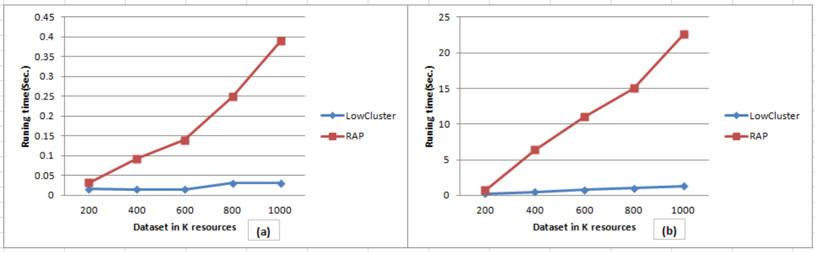

Figure 11. The Running Time comparison between Two Algorithms under Different Number of Resources and Functions in D_3 when α=1: (a) 10 Functions; (b) 20 Functions

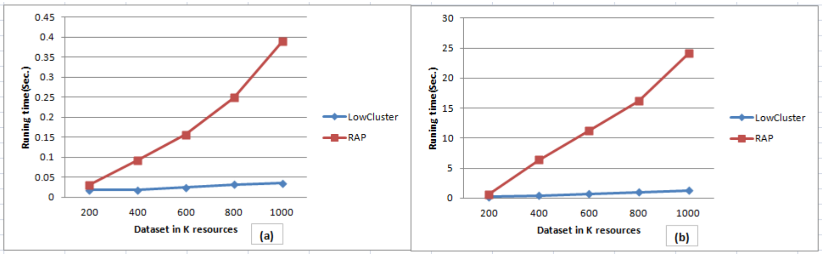

Figure 12. The Running Time Comparison between Two Algorithms under Different Number of resources and Functions in D_3 when α=2: (a) 10 functions; (b) 20 functions

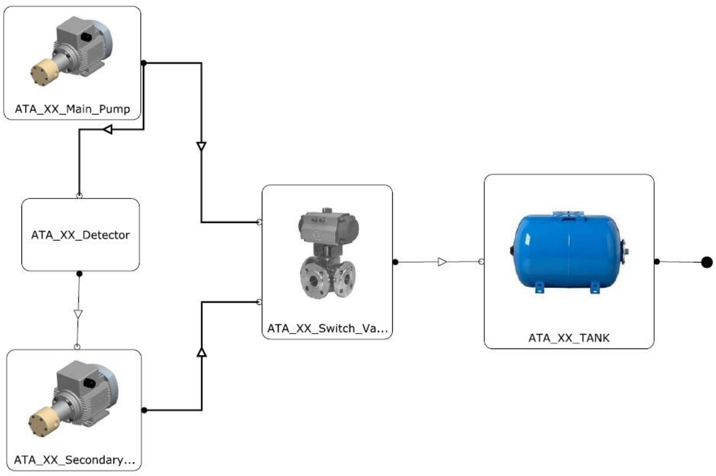

Figure 13. The Electronic Aircraft Fuel Provision System

However, the electronic aircraft fuel provision system may not have enough fuel at uncertain time, which is shown in Figure 14. It cannot be found by using fault-tree analysis. According to the process of electronic aircraft fuel provision system, we generate a function-resource usage matrix, which is shown in Table 4, where R1 to R5 represent the pump, the backup pump, the pump monitor, the fuel valve, and the fuel tank; F1* to F5* represent shutting pump, opening backup pump, the value after comma represents the sampling value under one sampling time.

Using our proposed *LowCluster* algorithm can mine the abnormal resource errors of the pump and the backup pump when executing opening backup pump function and shutting pump function. Through the history dataset analysis, using our algorithm can mine the errors which are the pump monitor and the fuel valve cannot meet the requirements of shutting pump function and opening backup pump function, when opening backup and shutting pump. After the analysis, we found the reason is that the pump was not shut, at the same time, the backup pump opened. It causes the faults. Therefore, using our proposed *LowCluster* algorithm can find the error reason when executing more functions, which will help to improve system safety analysis.

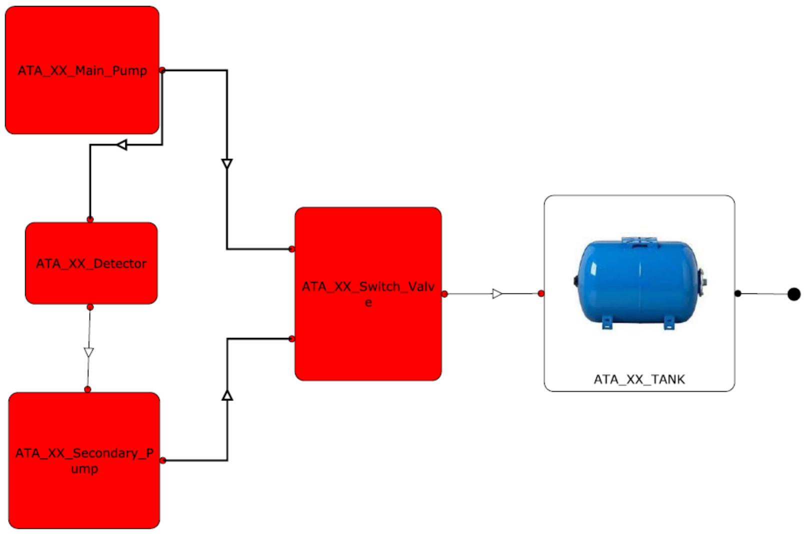

Figure 14. The Wrong Electronic Aircraft Fuel Provision System

Table 4. The Function-resource Usage Matrix of Electronic Aircraft Fuel Provision System

	$F_{1,1}$	$F_{2,1}$	$F_{1,2}$	$F_{2,2}$	…	…	$F_{1,n}$	$F_{2,n}$
R_1	0.93	0.12	0.93	0.12	…	…	0.93	0.72
R_2	0.82	0.11	0.82	0.11	…	…	0.82	0.81
R_3	0.7	0.1	0.7	0.1	…	…	0.7	0.1
R_4	0.72	0.1	0.72	0.1	…	…	0.72	0.1
R_5	0.8	0.12	0.8	0.12	…	…	0.8	0.12

5. Conclusion

In this paper, we use data mining technology to construct resource-layer safety model, function-layer safety model and task-layer safety model, firstly; secondly, we proposed an efficient bicluster mining algorithm: *LowCluster*, to effectively mine all the maximal constant row biclusters with low usage rate in real-valued function-resource matrix for IMA safety analysis. In *LowCluster* algorithm, a sample weighted graph is constructed firstly, it includes all resource collections between both samples which meet the definition of low usage rate; then, all the maximal constant row biclusters with low usage rate are mined using sample-growth and depth-first method in the sample weighted graph. In order to improve the mining efficiency, *LowCluster* algorithm uses pruning strategy to ensure the mining of maximal bicluster without candidate maintenance.

Acknowledgements

This paper is supported by National Key Basic Research Program of China (No. 2014CB744900).

References

[1] N. Leveson, "Engineering a Safer World: Systems Thinking Applied to Safety", MIT Press **(2011).**
[2] L. Meshkat, J. B. Dugan and J. D. Andrews, "Dependability Analysis of Systems With On-Demand and Active Failure Modes Using Dynamic Fault Trees", IEEE Trans. Reliability, vol. 51, no. 2, **(2002),** pp. 240–251.
[3] SAE Standard J-1739, Potential Failure Mode and Effects Analysis in Design (Design FMEA) and Potential Failure Mode and Effects Analysis in Manufacturing and Assembly Processes (Process FMEA) and Effects Analysis for Machinery (Machinery FMEA) **(2002).**
[4] Y. Cheng and G.M. Church, "Biclustering of Expression Data," Proc. 8th Int'l Conf. Intelligent Systems for Molecular Biology (ISMB00), **(2000)** ACM Press, pp. 93–103.
[5] C. Becquet, S. Blachon, B. Jeudy, J. F. Boulicaut and O. Gandrillon, "Strong-association-rule mining for large-scale gene-expression data analysis: a case study o human SAGE data. Genome Biol, vol. 12, **(2003),** pp. 1-16.
[6] S. C. Madeira and A. L. Oliveira, "Biclustering algorithms for biological data analysis: a survey", IEEE/ACM TCBB, vol. 1, no. 1, **(2004),** pp. 24–45.
[7] G. Pandey, G. Atluri, M. Steinbach, C. L. Myers, and V. Kumar, "An association analysis approach to biclusting", In Proc. ACM Conf. on Kownlege Discovery and Data Mining, **(2009),** pp. 677-686.
[8] Ben, "Discovering local structure in gene expression data: the order-preserving submatrix problem", J. Comput. Biol, vol. 10, **(2003),** pp. 373-384.
[9] Cheng, "Bivisu: software tool for bicluster detection and visualization", Bioinformatics, vol. 23, **(2007),** pp. 2342-2344.
[10] L. Zhao and M. J. Zaki, "MicroCluster: An Efficient Deterministic Biclustering Algorithm for Microarray Data", in IEEE Intelligent Systems, special issue on Data Mining for Bioinformatics, vol. 20, no. 6, **(2005),** pp. 40-49.
[11] G. Pandey, G. Atluri, M. Steinbach, C. L. Myers and V. Kumar, "An association analysis approach to biclusting", In Proc. ACM Conf. on Kownlege Discovery and Data Mining, **(2009),** pp. 677-686.
[12] R. H. Torgeir, L. Astrid and K. Jan, "Learning rule-based models of biological process from gene expression time profiles using Gene Ontology", Bioinformatics, vol. 19, **(2003**), pp. 1116-1123.
[13] I. Olovnikov, A. Le Thomas and A A. Aravin, "A Framework for piRNA Cluster Manipulation. PIWI-Interacting RNAs", Humana Press, **(2014),** pp. 47-58.
[14] H. Fotso, S. Yang and H. Hafermann, "Extended Correlation in Strongly Correlated Systems", Beyond Dynamical Cluster Approximation. Bulletin of the American Physical Society, vol. 57, **(2012).**
[15] U. Maulik, A. Mukhopadhyay, M. Bhattacharyya, L. Kaderali, B. Brors, S. Bandyopadhyay, and R. Eils, "Mining Quasi-Bicliques from HIV-1-Human Protein Interaction Network: A Multiobjective Biclustering Approach", IEEE-ACM Transactions on Computational Biology and Bioinformatics, vol.10, **(2013),** pp. 423-435.
[16] A. Király, J. Abonyi and A. Laiho, "Biclustering of High-throughput Gene Expression Data with Bicluster Miner", Data Mining Workshops (ICDMW), IEEE 12th International Conference on. IEEE, **(2012),** pp. 131-138.
[17] B. Desai, P. Andhale and M. Rege, "Biclustering and feature selection techniques in bioinformatics. Data Engineering and Management", Springer Berlin Heidelberg, **(2012),** pp. 280-287.
[18] M. Wang, X. Shang, S. Zhang and Z. Li, "FDCluster：Mining frequent closed discriminative bicluster without candidate maintenance in multiple microarray datasets", ICDM 2010 workshop on Biological Data Mining and its Applications in Healthcare, **(2010),** pp. 779-786.
[19] M. Wang, X. Shang, M. Miao, Z. Li and W. Liu, "FTCluster: Efficient Mining Fault-Tolerant Biclusters in Microarray Dataset", Proceedings of ICDM 2011 workshop on Biological Data Mining and its Applications in Healthcare, **(2011),** pp. 1075-1082.

Authors

Miao Wang is an engineer at science and technology on avionics integration laboratory and China aeronautical radio electronics research institute. He completed his doctor and master degree from northwestern polytechnical university in 2013 and 2018, respectively. He is a member of China computer federation. His research interests mainly include data mining, PHM, avionics and safety.

Zhiyong Xiong is research fellow and the director of science and technology on avionics integration laboratory office. His research interests is IMA.

Liang Xu is an engineer at China aeronautical radio electronics research institute.

Lihua Zhang is an engineer at science and technology on avionics integration laboratory. She completed his doctor and master degree from northwestern polytechnical university in 2014 and 2018, respectively. Her current research interests are PHM, avionics, data mining and safety.

Qingfan Gu is a senior engineer at China aeronautical radio electronics research institute. He had been an association professor at Nanjing University before 2013. His current research interests are IMA, system safety, and software.

Guoqing Wang is a professor and a supervisor of Ph.D. student in Northwestern Polytechnical University. He was born in 1956 and received his M.S. and Ph.D. degrees in computer science and technology from the Northwestern Polytechnical University in 1984 and 1991 respectively. He was the institute director of China aeronautical radio electronics research institute, and the director of science and technology on avionics integration laboratory. He has long been engaged in the related technical research of avionic system integration, distributed parallel processing, high reliable fault-tolerant system, network and bus system etc. He serves as the vice director of national serve environment computer academy.

