SPACE SHUTTLE RTOS BAYESIAN NETWORK

A. Terry Morris, NASA Langley Research Center, Hampton, Virginia

Peter A. Beling, Dept. of Systems Engineering, University of Virginia, Charlottesville, Virginia

Abstract

With shrinking budgets and the requirements to increase reliability and operational life of the existing orbiter fleet, NASA has proposed various upgrades for the Space Shuttle that are consistent with national space policy. The cockpit avionics upgrade (CAU), a high priority item, has been selected as the next major upgrade. The primary functions of cockpit avionics include flight control, guidance and navigation, communication, and orbiter landing support. Secondary functions include the provision of operational services for non-avionics systems such as data handling for the payloads and caution and warning alerts to the crew. Recently, a process to selection the optimal commercial-off-the-shelf (COTS) real-time operating system (RTOS) for the CAU was conducted by United Space Alliance (USA) Corporation, which is a joint venture between Boeing and Lockheed Martin, the prime contractor for space shuttle operations. In order to independently assess the RTOS selection, NASA has used the Bayesian network-based scoring methodology described in this paper. Our two-stage methodology addresses the issue of RTOS acceptability by incorporating functional, performance and non-functional software measures related to reliability, interoperability, certifiability, efficiency, correctness, business, legal, product history, cost and life cycle. The first stage of the methodology involves obtaining scores for the various measures using a Bayesian network. The Bayesian network incorporates the causal relationships between the various and often competing measures of interest while also assisting the inherently complex decision analysis process with its ability to reason under uncertainty. The structure and selection of prior probabilities for the network is extracted from experts in the field of real-time operating systems. Scores for the various measures are computed using Bayesian probability. In the second stage, multi-criteria trade-off analyses are performed between the scores. Using a prioritization of measures from the decision-maker, trade-offs between the scores are used to rank order the available set of RTOS candidates.

1 Introduction

The Space Shuttle (figure 1) is a unique vehicle with unrivaled capabilities and, with a 98.9 percent success rate, is the most capable, versatile and reliable space-faring vehicle in the world. As a major launch vehicle for all U.S. and many international components of the International Space Station (ISS), the Shuttle has kept the United States on the cutting edge of space exploration and scientific discovery for the last two decades. As space transportation needs continue to evolve, access to space must be assured. The continuing development and upgrade of the Space Shuttle -- which was designed to have a 100 mission life -- poses the most effective means for assuring access to space for missions requiring its unique operational capabilities. Upgrades to the Space Shuttle increase assurance by improving safety and reliability, reducing operating costs, enhancing performance, and incorporating new and better technologies.

The CAU is a critical high-priority Space Shuttle safety upgrade. The project is designed to implement new Orbiter cockpit avionics hardware and software to meet the man-machine interface requirements identified by the Space Shuttle

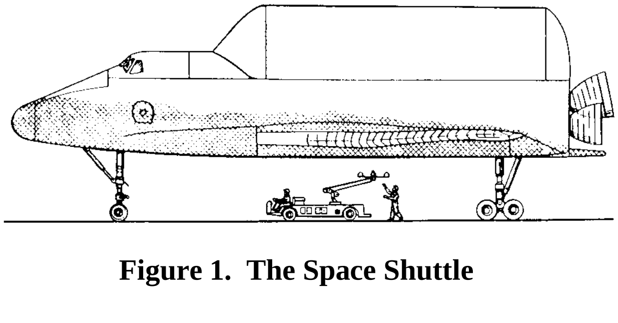


Figure 1. The Space Shuttle

Cockpit Council in an effort to enhance overall crew safety [1]. Orbiter cockpit displays and crew interface capabilities will be significantly improved by replacing the existing integrated display processors with higher performance command and display processors. These units will provide expanded processing performance to enable dramatic improvements in information access and display capability as well as the implementation of new caution and warning software functions. The CAU will increase crew situational awareness and decrease crew workload in the cockpit to enable more timely and accurate crew decisions. Improving the crew's ability to manage information during critical flight operations will significantly impact safety and reliability.

1.1 USA's CAU RTOS Evaluation

The prime Space Shuttle contractor, USA Corp., was responsible for CAU operation and development, including integration of hardware and software components into the vehicle. One of the highest priority CAU program objectives involved the evaluation and selection of the optimal COTS RTOS for the CAU. A trade study was performed by USA Corp. to determine the most applicable RTOS for the CAU from available products on the market [2]. USA utilized a linear three-filter process for the RTOS trade study, which started with over 100 RTOSs and ended with one. The first filter eliminated niche RTOSs and narrowed the available set to ten potential candidates: VxWorks, OSE, Integrity, Nucleus+, OS-9, Precise/MQX, Real-Time Craft, SMX, Supertask, and VRTX. The second filter applied technical and industry performance conditions that helped to narrow the ten candidates down to two: VxWorks and OSE. The final filter employed benchmark testing which narrowed the RTOS selection to VxWorks [3].

1.2 The Independent Assessment Process

There are three Independent Assessment (IA) teams participating in the CAU project. One of the teams is NASA's Independent Program Assessment Office (IPAO), a Langley Research Center based organization that reports to the NASA chief engineer and chief financial officer. The IA process performed by the IPAO is a peer validation of a proposed advanced aerospace systems concept design using the best available, independent systems analysis expertise and methodology in accordance with NASA policy [4].

In order to perform an IA for the CAU RTOS selection, the IPAO decided to use a Bayesian Network-based scoring methodology. This two-stage methodology was utilized in order to achieve an objective, non-advocate, in-depth study of the RTOS candidates. This methodology also served to verify RTOS performance, design integrity, life-cycle costs, inherent risks, and technology-related issues, thereby validating or invalidating USA's RTOS selection. The methodology used an approach that assessed key product features (measures) from functional, non-functional, and performance perspectives. Scores, computed using probabilistic inference, were then used to rank feature performance. Trade-off analyses between the feature rankings determined the acceptability of the RTOS.

The Bayesian Network scoring methodology was selected for its ability to incorporate causal relationships between the often-competing measures of interest, and its proclivity for fusing *a priori* knowledge from RTOS experts and product specific evaluations. Using Bayes' Theorem, this approach performs statistical inference by taking into account the probabilistic nature of the causal influences of the measures of interest, thereby encoding, to some degree, the uncertainty in the decision analysis problem. The approach is coherent, admissible, and consistent with an acceptance methodology for COTS software [5].

1.3 Problem Description

NASA's primary rationale for using a COTS RTOS is to lower development costs and time, to reduce maintenance effort, and to take advantage of advances in technology. The integration of a COTS RTOS in a safety-critical avionics system poses inherent problems, such as hazardous threats involving unreliability, software maintenance issues, and obsolescence within the CAU life cycle. The primary issue confronting the IA is the conflict between cost and mission assurance.

The IPAO's problem of evaluating COTS RTOSs for the CAU can be expressed mathematically in the following way. Given a set,

X, of alternative software products, let $X(\alpha)$ represent the set of objective attributes for each alternative x, where the set of m attributes is expressed as the vector $\alpha = [a,b,c,...,m]$. The function $f : X(\alpha) \rightarrow \Re^p$ is an aggregation scoring function. Let Y represents the space of judgmental inputs with respect to context or product fitness. Then, the fitness function rule

$$F : f\{X(\alpha)\} \times Y \rightarrow \Re^q$$

describes how well the alternatives fit the decision maker's values and environmental constraints. The COTS RTOS comparison and selection ranking problem is expressed as

$$\begin{aligned} &max\text{-}dominate\ F[f\{x(\alpha)\},Y] \\ &such\ that\ \ x \in X \end{aligned}.$$

This is a standard multi-objective decision problem characterized by an r-dimensional vector of objective functions, i.e.,

$$X(\alpha) = [X_1(\alpha), X_2(\alpha),...,X_r(\alpha)]$$

and a feasible region defined by the scoring function f. The solution seeks to find a set of non-dominated solutions, which will be a subset of the feasible region. For the RTOS problem, the α's represent the RTOS data and r represents the measures of interest.

2 Background

To grasp the complexities involved with selecting a COTS RTOS for the Space Shuttle, background information in four primary areas is required. The first area involves the current state of COTS software and the dilemma that it introduces into project management decisions. The second area involves specific safety-critical concerns that need to be addressed when integrating commercially available software products into high-priority safety-related systems. The third area of concern involves understanding specifics of USA's RTOS acceptance process in order for the IPAO to ensure process independence for the IA. The final source of background information involves understanding the appropriateness and applicability of Bayesian networks. Each of these areas will be discussed in succession.

2.1 The COTS Software Dilemma

The proliferation and increased use of COTS and modified-off-the-shelf (MOTS) software stems from the realization that pre-existing software products can be a means of lowering development costs, shortening development time, and keeping pace with the changing software market. The Federal government has found that, particularly with regards to safety-critical systems, COTS software is currently not plug-and-play, has significant tradeoffs (e.g., performance, safety, costs, etc.), and usually contains a "cradle-to-grave" dependence on the software manufacturer. Various risks inherent with using COTS software include incompatibility of the product with other COTS products, lack of control over the product's current and future functionality, and immaturity of product or vendor [6]. In addition, most COTS software components have no warranty and are not usually subject to rigid development or verification processes [7]. In an effort to reduce the risks of COTS software, various software standards, maturity models, and software development frameworks have been developed, but they provide no specific guarantee of eliminating software faults. Additionally, vendors prefer not to be responsible for guaranteeing their software [8]. This is because many of the faults found in COTS software are due to business decisions made by the software manufacturer. For instance, time-to-market is commonly seen as the primary factor for a product's success [9]. In order to be first-to-market, some COTS vendors will release software with known defects [10]. The primary driver in many of the trade-off decisions involves minimizing costs.

The government's increased dependence on COTS software has introduced substantial risks, particularly those involved with safety and mission-critical systems. Trade-offs, therefore, are necessary in order to address the current software dilemma, that is, the apparent conflict between minimizing costs and improving mission assurance.

2.2 RTOS Safety-Critical Concerns

COTS software may offer important advantages in the creation of new systems or the upgrading of existing systems. Unfortunately, projects often find that unforeseen costs and technical issues associated with COTS software products offset the benefits they hoped to achieve from the use of these products. With respect to the Space Shuttle, COTS usage is governed by policies outlined in the Shuttle Master Verification Plan [11]. The policy states that COTS software usage in safety-critical systems should be evaluated and reported as part of the normal risk assessment process to determine any risk-related exposures. A commercially available RTOS for use in the CAU is considered "high criticality" software. Guidelines state that a COTS certification process is expected to make maximum use of prior vendor testing results as well as analysis based on actual prior "field usage" of software. The data can be procured by test reports provided by the vendor or third party sources. For the RTOS software, proof of extensive analysis and/or testing must be required as part of the certification process to meet the criticality level of the system. At a minimum, COTS RTOS data must include information related to vendor support, reliability, product deficiencies, compatibility, maintenance, visibility of code, and life cycle concerns.

2.3 USA's RTOS Acceptance Process

In order to meet Shuttle requirements for COTS integration, USA Corp. employed a three-filter process to select the best RTOS [3]. In addition to eliminating niche RTOS products (cell phone, game, automotive, etc.), the first filter selected products that supported target platforms. The second filter applied technical and industry criteria to narrow the number of RTOS candidates. These criteria involved acquiring data in several specific categories. The category areas included RTOS process control, scheduler algorithms, process coordination, multiple CPUs, memory management, I/O support, network support, error handling, company information, interrupt support, support tools and certification. VxWorks and OSE were the only two remaining candidates at this point. The third filter involved bringing in both vendors to test each RTOS on a target host used to mimic the characteristics of the proposed system. This was done to analyze the responsiveness of the vendors to problems, real-time product stability, and ease of use of the support tools. The three-filter approach yielded VxWorks as the clear favorite.

Though limitations were present in USA's trade study, the process appeared to be simple and straightforward. Some of the limitations involved the lack of several key items. These items included a clear and complete justification process, a clear delineation of Shuttle CAU software specifications, a priority ranking of the selection criteria, an explicit list of necessary and sufficient conditions the selection criteria must satisfy, and a determination of necessary constraints. Regardless of these limitations, the IPAO's task was still to independently validate or invalidate USA's RTOS selection. In order to produce an objective, non-advocate, in-depth study of the RTOS candidates, a Bayesian network scoring methodology was adopted.

2.4 Bayesian Networks

Bayesian networks are directed acyclic graphs (DAGs) in which the nodes represent variables, the arcs signify the existence of direct causal influences between the variables, and the strengths of these influences are expressed by forward conditional probabilities [12]. An advantage of a Bayesian network is its natural perception of causal influences thus making it an unambiguous representation of dependency. This is useful for the RTOS problem in that it allows for the explicit identification of influences between the attributes of each product. Another advantage of a Bayesian network is the requirement of strict positivity, which allows it to serve as an inference instrument for logical and functional dependencies. Moreover, its ability to quantify the influences with local, conceptually meaningful parameters allows it to serve as a globally consistent knowledge base.

The independence criteria for Bayesian networks, *d*-separation [12], allows the joint probability distribution (JPD) to be efficiently computed using small prototypical clusters of variables forming local probability distributions. In this way, Bayesian networks are a natural tool for dealing with uncertainty and complexity. Characterizations of Bayesian networks generally involve determining whether the structure of the model is known and whether the data is complete. Bayesian networks with unknown structure and incomplete data are known to be NP-hard [13]. The

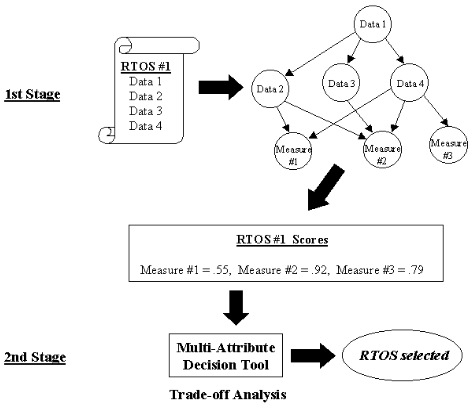


Figure 2. Bayesian Network Scoring Methodology

main use of Bayesian networks is in situations that require statistical inference. In a typical inference application, a user has some observed evidence and wishes to infer the probabilities of other events, which have not as yet been observed. Additionally, Bayesian networks are a way of dealing with complex probabilistic reasoning with their ability to accommodate both subjective probabilities from domain experts as well as probabilities based on objective data.

3 Bayesian Network Scoring

In this section, we provide a brief overview of the Bayesian network acceptance methodology. Detailed analyses of each of the two major stages of the methodology are provided. Results of the CAU RTOS IA are also discussed.

3.1 Bayesian Network Acceptance Methodology

The Bayesian network scoring methodology is a two-stage process used for RTOS software acceptance. The first stage involves a scoring process and the second stage involves trade-off analyses. The first stage inputs data from each RTOS and computes probabilistic scores for measures of interest using causal influences and conditional probability distributions (CPD) extracted from the database of RTOS products. The probabilistic scores represent how favorable the RTOS performed in a particular measure. High scores represent performance that is both favorable and certain. Low probabilistic scores represent performance that is unfavorable or uncertain. In this way, the probabilistic scores provide a performance ranking for each RTOS candidate. In the second stage of the process, trade-offs between the scores determine the feasibility of acceptance. Figure 2 depicts the components and stages of the Bayesian network scoring acceptance methodology. Discussions of both stages follow.

3.2 Bayesian Net Scoring Process (1st Stage)

Elements of the Bayesian network scoring process include a database of RTOS information, the establishment of measures, model development, model validation, and the computation of probabilistic scores. Figure 3 provides a pictorial representation of the interrelations among these elements. As seen in this diagram, a Bayesian network is developed from existing data in a database of RTOS products. The data in the database are characterized as either evidential data or measure data. Evidential data (nodes) are objective or computed values that represent product attributes. Measure data (nodes) are objective or estimated values that represent measures of interest to the decision-maker. Initially, a macro model is developed. The macro model represents only the causal influences between the measures of interest. Then a micro model is developed by adding evidential nodes and influences to the macro model structure. Generally, expert domain knowledge is involved in the construction of the RTOS Bayesian network at both the macro and micro model levels. Expert domain knowledge helps to establish how causal influences contribute to the measures of interest when there is not enough objective product data to justify the connection.

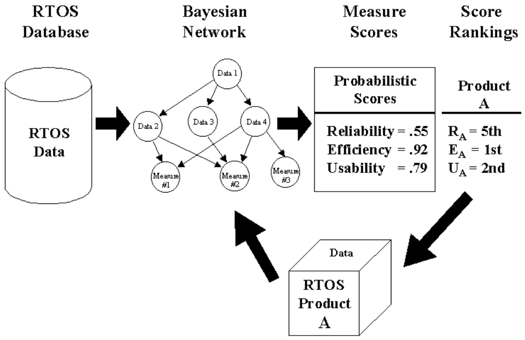
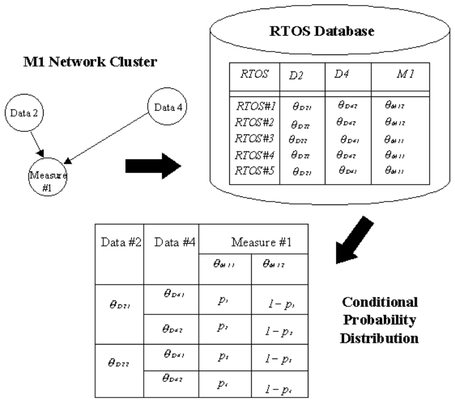


Figure 3. Bayesian Net Scoring Process (1st Stage)

Given the network structure, the CPDs of the nodes are extracted from the data in the database.



RTOS	*D2*	*D4*	*M1*
RTOS#1	θ_{D21}	θ_{D42}	θ_{M12}
RTOS#2	θ_{D22}	θ_{D42}	θ_{M12}
RTOS#3	θ_{D22}	θ_{D41}	θ_{M11}
RTOS#4	θ_{D22}	θ_{D42}	θ_{M11}
RTOS#5	θ_{D21}	θ_{D41}	θ_{M11}

Data #2	Data #4	Measure #1	
		θ_{M11}	θ_{M12}
θ_{D21}	θ_{D41}	p_1	$1-p_1$
	θ_{D42}	p_2	$1-p_2$
θ_{D22}	θ_{D41}	p_3	$1-p_3$
	θ_{D42}	p_4	$1-p_4$

Figure 4. Conditional Probability Distribution

Probabilistic scores for a particular RTOS are then computed by incorporating the product's data into the evidential nodes of the network and acquiring the propagated probabilities from the measure nodes. After computing the probabilistic scores for all the RTOS candidates, score rankings can be determined by comparing each product's scores to those of the other products (see figure 3).

A brief example of the Bayesian network scoring process can be shown using the generic model shown in figures 2 and 3. In this generic structure, evidential nodes are labeled with *D's* (D1, D2, D3, D4) and measure nodes are labeled with *M's* (M1, M2, M3). In the macro version of the network, there are only three measure nodes that are all independent since there are no direct causal influences between them. Therefore, the generic network in figure 3 is the micro model. Given this micro model (M), the JPD can be computed using

$$p(X = x_k \mid M) = \prod_{i=1}^{n} p(X_i = x_{ik} \mid \Pi_i = \pi_{ij}, M),$$

where π_i = the parents of node x_i .The JPD for the Bayesian network in figure 3 is

$$\begin{aligned} p(X = x_k \mid M) = \; & p(m_1 \mid d_2, d_4)p(m_2 \mid d_2, d_3, d_4) \\ & p(m_3 \mid d_4)p(d_2 \mid d_1)p(d_3 \mid d_1) \\ & p(d_4 \mid d_1)p(d_1) \end{aligned}$$

The JPD can then be grouped into clusters of conditional probabilities using

$$p(X \mid M) = \prod g(x_i, \pi_i) = g(x_1, \pi_1) \bullet\bullet\bullet\, g(x_n, \pi_n),$$

where $g(x_1, \pi_1) = p(m_1 \mid d_2, d_4)$, and so on. The CPD for the local cluster $g(x_1, \pi_1) = p(m_1 \mid d_2, d_4)$ can be computed from the data in the database. As seen in figure 4, values for D2, D4 and M1 are shown in the RTOS database for five products. The options for D2 (e.g., multi-cpu capability) could be θ_{D21} = "high" and θ_{D22} = "low". Also, two options for data D4 (e.g., latency) could be θ_{D41} = "fast" and θ_{D42} = "slow". Values for the measure M1 (e.g., efficiency) may either be known directly from the vendor or extracted from RTOS domain experts. The CPD table for this cluster is shown in figure 4 where

$$p_1 = p(m_1 = \theta_{M11} \mid d_2 = \theta_{D21}, d_4 = \theta_{D41})$$

can be computed from the data in the database using Bayes Theorem. Details of each step of the scoring process for the CAU RTOS selection problem will follow.

3.2.1 Database Population

The Bayesian network scoring process begins with the acquisition of RTOS data from USA Corp. [2, 3]. Optimally, the RTOS database should be accurate, complete, and mature. Unfortunately, as in most practical applications, this was not the case. Though USA Corp.'s RTOS trade study investigated 10 primary operating systems following the first filter of their process, careful inspection of their data revealed that product information for two of the ten candidates was extremely sparse. For this reason, data from only 8 RTOSs was used to populate the database. In order to develop a more robust data set, the RTOS data was updated and augmented to include data from additional sources. These sources included the Securities and Exchange Commission (SEC), interviews with previous RTOS principle investigators who used them on space-related flights, user groups, the ISO compliance database, and various benchmarking RTOS analysis organizations. A data fusion approach was utilized to incorporate data from various sources.

3.2.2 Measure Development

Measures are continuous or discrete random variables used to assess a particular feature of a software product. In this methodology, measures are grouped into functional, non-functional, and performance classifications. The measures are further segregated into product and environmental categorizations depending on whether the measure

Table 1. Measure Characterization

Assessment	Component	Measure
Functional	*Product*	Functionality
	Environment	Interoperability
		Legal
Performance	*Product*	Usability
		Efficiency
		Portability
		Correctness
		Reliability
		Certifiability
	Environment	Software Design Process
		Business
		Product History
Non-functional	*Product*	Life Cycle
	Environment	Cost

was deemed to be intrinsic to the product or primarily influenced by external factors. The measure characterization is fairly consistent with measures listed in the international standard ISO 9126 [14]. Fourteen (14) RTOS measures of interest are listed in Table 1.

3.2.3 Model Development Process

Bayesian networks are constructed by determining the nodes to represent as random variables, the causal influences between the linked variables (model structure), and the strengths of the influences (CPDs). Steps for determining the model structure and CPDs for the RTOS network were developed using the iterative procedure shown in figure 5. In this process, model development starts with an initial macro model and then proceeds to compute several micro models, which represent clusters of random variables separated using the *d*-separation criterion. After extracting the CPDs for the micro model clusters (as expressed in an earlier example), the process iterates to improve the overall macro model until no improvement can be made in the posterior probability. Explanations of this process follow.

After constructing the macro model, several micro models are developed using local clusters of variables computed using the conditional independence criterion. The algorithm that finds the best micro model starts by decomposing the JPD of the macro model into local contributions, i.e.,

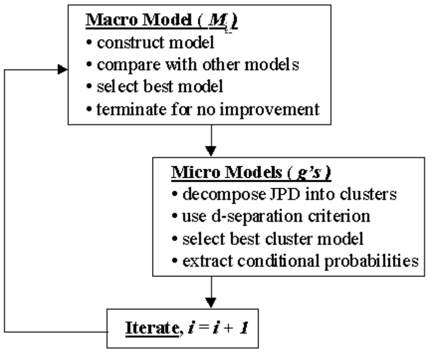

Macro Model (M_i)
• construct model
• compare with other models
• select best model
• terminate for no improvement
Micro Models (*g's*)
• decompose JPD into clusters
• use d-separation criterion
• select best cluster model
• extract conditional probabilities
Iterate, $i = i + 1$

Figure 5. Model Development Process

$$p(X \mid M) = \prod g(x_i, \pi_i) = g(x_1, \pi_1) \bullet\bullet\bullet\, g(x_n, \pi_n).$$

For each local contribution g, compute

$$g(x_i, \pi_i) = \frac{\Gamma(\alpha)}{\Gamma(\alpha + n)} \prod \frac{\Gamma(\alpha + n)}{\Gamma(\alpha)}$$

using conjugate analysis of exponential families for Dirichlet distributions. Also, for each "g", π_i should be expanded to include the parent nodes that give the largest contribution to $g(x_i, \pi_i)$. The algorithm stops when there is no increase due to local contributions.

Next, parametric learning commences with a given model M and decomposes the JPD with the assumption that the parameters θ are conditional probabilities that represent the observer's belief before observing the data, i.e., $\theta = \{\theta_1, \ldots, \theta_n\}$. Given the data in the database, the prior density $p(\theta)$ is updated in the posterior density using Bayes' Theorem. A Bayesian estimate of θ is computed, i.e., $E(\theta \mid D)$, where the θ's are assumed to be mutually independent with Dirichlet distributions.

Finally, hyperparameters [16] are used to update the posterior distribution using the frequency of data in the database. A significant problem encountered using this procedure involves missing data. There are, however, several stochastic procedures that can be used to estimate data for incomplete databases. Unfortunately, stochastic procedures are generally computationally expensive. Ramoni and Sebastiani have developed a deterministic technique called Bound and Collapse [16], which can learn the parameters of a Bayesian network from possibly incomplete databases thereby improving computational complexity.

By constructing a new macro model and comparing it to the initial model, the model development process can be iterated. The algorithm to select the best macro model is analogous to the one for the micro models in that it would chose the model with the highest posterior probability and then quit when there is no improvement due to changes in the model structure.

3.2.4 Macro Model Development

As stated earlier, the macro model describes the major causal influences between the measures of interest (table 1) as it relates to real- time operating systems. A macro model was determined with the prior belief of a particular dependence structure based on expert experience [15]. The macro level RTOS Bayesian network is shown in figure 6 where the measure nodes will serve as probabilistic output nodes for the network.

Given a model structure M, the joint probability of various random variables *X's* can be computed as

$$p(X = x_k \mid M) = \prod_{i=1}^{n} p(X_i = x_{ik} \mid \Pi_i = \pi_{ij}, M)$$

where π_i = the parents of node x_i .Given a database $D = \{x_1, ..., x_n\}$ from which to select a model M of conditional dependencies among the variables in the database, let $p(M)$ = our belief about a particular model M. The posterior probability of M given the data is

$$p(M \mid D) = \frac{p(M, D)}{p(D)}.$$

Given several rival models, the one with the highest posterior probability should be chosen using the Bayes' factor (BF), i.e.,

$$BF = \frac{p(M_1, D)}{p(M_2, D)},$$

where M_1 should be chosen if $BF > 1$, M_2 should be chosen if $BF < 1$, and either should be selected if $BF = 1$. A solution exists if the data are independent given the parameters associated with the model, the prior distribution of parameters is conjugate to the sampling model, and the parameters are marginally independent.

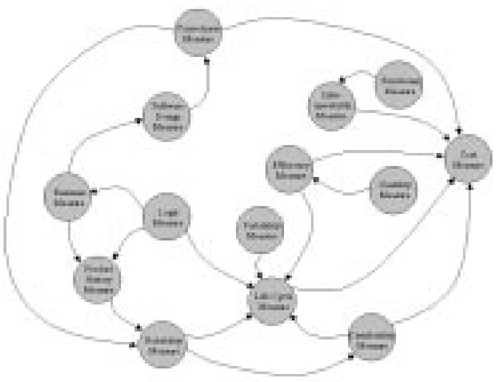


Figure 6. Macro Level RTOS Bayesian Network

3.2.5 Micro Model Development

Micro models are clusters of Bayesian networks that tie product attributes from the database to measures in the macro model. The process involves determining which RTOS attributes influence each measure and continues by calculating the strength of those influences from frequency counts in the database. Given a number of salient RTOS features and various choices for each feature, the RTOS problem can be constructed in a form using multinomial sampling. In multinomial sampling, let $X = \{X_1, ..., X_n\}$ be n discrete random variables, each having r possible states. The likelihood function is given by

$$p(X_i = x_i^r \mid \theta) = \theta_k, \quad k = 1, ..., r,$$

where $\theta = \{\theta_1, ..., \theta_r\}$ are parameters associated with physical probabilities. Also, let $D = \{x_1, ..., x_n\}$ be a database of observations and $\overline{M} = \{M_1, ..., M_z\}$ a set of models each containing dependency relationships among the random variables *X*.

For a class of distributions known as exponential families, each member in this class has sufficient statistics that are of fixed dimension for any random sample and a simple conjugate prior. Conjugate distributions are distributions connected with Bayesian networks such that the natural parametric family of posterior distributions belongs to the family of prior distributions [17]. In multinomial sampling, the simple conjugate prior is the Dirichlet distribution, i.e.,

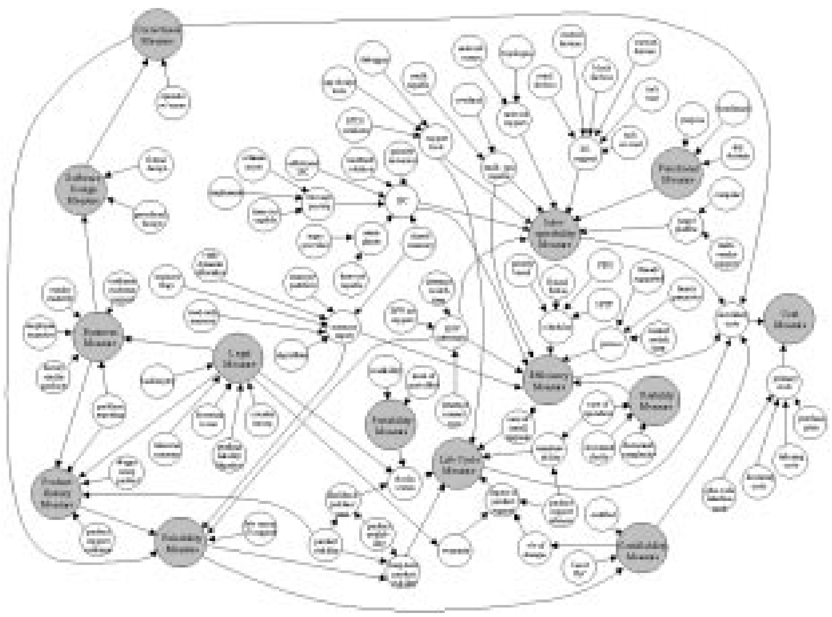


Figure 7. Micro Level RTOS Bayesian Network.

$$p(\theta|M)=Dir(\theta|\alpha_1,...,\alpha_r)\equiv\frac{\Gamma(\alpha)}{\prod_{k=1}^{r}\Gamma(\alpha_k)}\prod_{k=1}^{r}\theta_k^{\alpha_k-1}$$

where $\alpha=\sum_{i=1}^{r}\alpha_k$, $\alpha_k>0$, $k=1,...,r$, M = the model, and $\alpha' s$ are the hyperparameters. The posterior distribution is

$$p(\theta | D,M)=Dir(\theta|\alpha_1+N_1,...,\alpha_r+N_r).$$

The marginal likelihood or evidence, $p(D|M)$, is

$$p(D|M)=\frac{\Gamma(\alpha)}{\Gamma(\alpha+N)}\prod_{k=1}^{r}\frac{\Gamma(\alpha_k+N_k)}{\Gamma(\alpha_k)}.$$

Given a particular model M , the JPD is computed as before. Given a model M and θ ,

$$p(X|\theta,M)=\prod_{i=1}^{n}p(x_i|\pi_i,M).$$

Computation of the posterior distribution is conditioned on the assumptions that the data D is complete and the parameter vectors θ_{ij} are mutually independent, i.e.,

$$p(\theta|M)=\prod_{i=1}^{n}\prod_{j=1}^{9}p(\theta_{ij}|M).$$

Using these assumptions, the parameters remain independent given a random sample, i.e.,

$$p(\theta|D,M)=\prod_{i=1}^{n}\prod_{j=1}^{9}p(\theta_{ij}|D,M).$$

Using Bayes' Theorem and the assumptions (distributions are in an exponential family, parameters are mutually independent, conjugate priors have been selected, and there is complete data), the marginal likelihood is computed as

$$p(D|M)=\frac{p(\theta|M)p(D|\theta,M)}{p(\theta|D,M)}.$$

In accordance with conjugate analysis, various cluster networks were constructed for the measures with CPDs extracted from the RTOS database. Figure 7 displays the micro level Bayesian network. Calculations for the CPDs (not shown) were computed using commercially available software.

3.2.6 Model Validation

After incorporating both expert domain knowledge and objective data with respect to the macro model structure, various micro Bayesian networks were constructed and validated using the log-likelihood criterion. It should be noted that validation of the macro model was not performed and is an area of future research.

3.2.7 RTOS Score Rankings

By applying data from the augmented RTOS data set to the micro Bayesian network, scores for each of the eight RTOSs were computed for 14 measures of interest (table 1). The CPDs of the augmented data set were used as priors for the network. Score rankings for the 8 RTOSs are shown in Table 2. Longer RTOS names were abbreviated to fit the space provided, i.e., VxW = VxWorks, Int = Integrity, N+ = Nucleus +, and ST = SuperTask (see Table 2). As seen in the table, several candidates were ranked equally depending on their probabilistic scores for a particular measure.

Table 2. Score Rankings

Measure	Rankings							
	1	**2**	**3**	**4**	**5**	**6**	**7**	**8**
Product History	VxW Int	SMX	OS9	OSE	N+	ST	MQX	
Correctness	VxW Int	OSE	ST	N+	MQX	SMX	OS9	
Efficiency	VxW	Int	MQX	OS9	N+	OSE	SMX	ST
Reliability	Int	VxW	OSE	N+	SMX	OS9	MQX	ST
Functionality	VxW Int OSE OS9 ST	N+ MQX SMX						
Life Cycle	VxW Int	OS9	MQX	SMX	N+	OSE	ST	
Inter-Operability	VxW	OSE	Int	OS9	N+	MQX	ST	SMX
Usability	VxW Int N+ SMX	OSE	OS9	ST	MQX			
Business	VxW Int	SMX	OS9	OSE	ST	N+	MQX	
Software Design	VxW Int	MQX	OSE	SMX	OS9	ST	N+	
Certifiability	Int OSE	MQX	VxW	SMX	OS9	N+	ST	
Portability	VxW MQX N+	OSE OS9	Int SMX ST					
Cost	N+	Int	SMX ST	MQX	VxW OSE	OS9		
Legal	SMX	VxW Int OS9	ST	N+	OSE	MQX		

3.3 Trade-off Analysis (2^{nd} Stage)

Trade-offs provide a mechanism to mitigate the risks of RTOS integration by selecting a product that minimizes the incompatibilities between the product and the decision-maker's preferences and constraints. The process assists in finding a product that maximizes benefits while minimizing product-environmental misfit. The following sections describe how the RTOS scores for the measures of interest were used in conjunction with a preference ordering to rank the 8 RTOSs and select the optimal RTOS from the set. Multi-attribute decision analysis is the primary technique employed to select the best RTOS. Finally, sensitivity analysis results show the robustness of the final selection due to changes in the preference ordering.

3.3.1 Preference Ordering

A preference order for the RTOS measures was acquired from a USA Corp. decision-maker. The preference ordering, obtained from USA, was as follows (from highest to lowest preference): Product History, Correctness, Efficiency, Reliability, Functionality, Life Cycle, Interoperability, Usability, Business, Software Design, Certifiability, Portability, Cost, and Legal. This preference ordering described the weighted values of relative importance that the USA decision-maker placed on the 14 measures of interest.

3.3.2 Multi-Criteria Decision Analysis

After obtaining scores for each RTOS candidate as well as the preference ordering of measures (from USA), PRIME Decisions, a multi-criteria decision analysis tool was used to determine the optimal RTOS among the available candidates. PRIME Decisions is a decision analytic tool which implements a PRIME (Preference Ratios In Multi-attribute Evaluation) technique developed by Ahti A. Salo and Raimo P. Hamalainen at the Helsinki University of Technology [18]. PRIME seeks to strike a balance between the theoretical soundness of the trade-off method and the functionality of decomposed ratio judgments [18]. Towards this end, ratio elicitation is based on the comparison of preference differences in pairs of measures. Such comparisons may be specified either as exact point estimates or, more typically, as interval judgments which impose linear constraints on the single-attribute scores of the alternatives. Through these constraints, the preference model becomes increasingly specific so that more conclusive dominance results can be inferred. The use of imprecise preference statements, modeled through intervals, may be particularly appropriate for group decision support as the decision-maker's conflicting views can be combined into an aggregate preference model.

After incorporating scores for each of the 8 RTOS candidates, PRIME Decisions calculated the value intervals (figure 8) for each candidate based on the normalized weights for each measure. The value intervals represent the spread (mean and variance) for each candidate and the weights represent the prioritized order of importance for each measure.

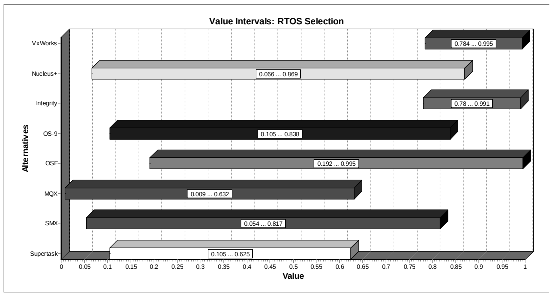
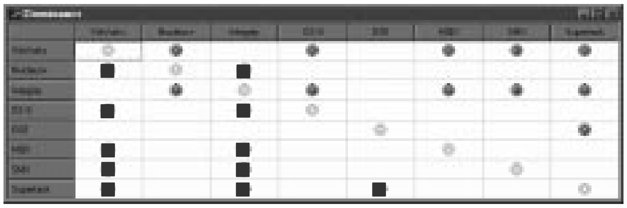


Figure 8. Value Intervals for each RTOS.



	VxWorks	Nucleus+	Integrity	OS-9	OSE	MQX	SMX	Supertask
VxWorks	○	●		●		●	●	●
Nucleus+	■	○	■					
Integrity		●	○	●		●	●	●
OS-9	■		■	○				
OSE					○			●
MQX	■		■			○		
SMX	■		■				○	
Supertask	■		■		■			○

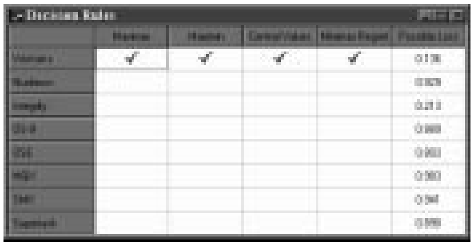
Figure 9. RTOS dominance matrix.

Next, a dominance matrix was determined (figure 9). Dominance is a situation in which one RTOS candidate is preferred to another for all the permissible combinations of measure preferences. Dominance is generally considered when the value intervals of two candidates overlap. PRIME uses two types of dominance – absolute and pairwise. Absolute dominance occurs when one RTOS candidate is preferred to another without any doubt. That is, alternative x is preferred to x' in the sense of absolute dominance if and only if the smallest value of x exceeds the largest value of x'. The set of RTOS dominated alternatives is determined by the other criterion of dominance - pairwise dominance. According to this dominance criterion, alternative x is preferred to x' if and only if the value of x exceeds that of x' for all feasible scores.

In figure 9, a black square indicates that the candidate on that row is dominated by the candidate in that column. Circles indicate just the opposite. As seen in figure 9, VxWorks and Integrity both completely dominate Nucleus+, OS-9, MQX, SMX, and Supertask while OSE completely dominates only Supertask. Empty elements indicate overlapping value regions. For example, neither OSE nor OS-9 dominate each other.

3.3.3 RTOS Rankings and Selection

Decision rules were used by PRIME Decisions to help the decision-maker in the determination of the optimal CAU RTOS candidate. Four decision rules (maximax, maximin, central values, and minimax regret) provided in this software are described as follows:



	Maximax	Maximin	Central Values	Minimax Regret	Possible Loss
VxWorks	✓	✓	✓	✓	[illegible]
Nucleus+					[illegible]
Integrity					0.213
OS-9					[illegible]
OSE					[illegible]
MQX					[illegible]
SMX					[illegible]
Supertask					[illegible]

Figure 10. RTOS decision analysis.

Utopian Decision Rule - Maximax (an optimistic decision rule) supposes that the most probable value lies at or near the greater bound of the candidates' value intervals, hence it selects the candidate with the greatest upper bound,

RISK Based Decision Rule - Maximin (a pessimistic decision rule) supposes that the worst case for the chosen candidate will happen and it selects the candidate with the greatest lower bound of the value interval,

Probabilistic Decision Rule (based on the Central Limit Theorem) - Central Values selects the candidate with the greatest midpoint, and

Pareto Optimal Decision Rule - Minimax Regret calculates the possible loss of value for each candidate by using dominance data and selects the candidate with the smallest possible loss.

As seen in figure 10, all decision rules unanimously computed VxWorks as the optimal CAU RTOS given the data provided. Based on the various decision rules provided by the multi-criteria decision analysis software, VxWorks should be selected as the optimal RTOS for the CAU. The entire ranking of the 8 RTOS candidates is shown in Table 3. IT should be noted that this selection is only valid for the CAU given the available data and the preference ordering from the decision-maker. Any changes in these items may produce a different optimal RTOS selection.

3.3.4 Sensitivity Analysis

Various preference orders were used to evaluate the robustness of the final RTOS candidate rankings. In all cases, VxWorks, Integrity, and OSE were consistently placed in the top three positions. Though most preference orderings produced practically identical rankings to those shown in table 3, two particular preference orders were found that changed the order of VxWorks and

Table 3. RTOS Rankings

Ranking	RTOS Candidate
1	VxWorks
2	Integrity
3	OSE
4	OS-9
5	Supertask
6	Nucleus+
7	SMX
8	MQX

Integrity. The first case involved switching the third and fourth measures in USA's preference ordering, that is, ranking Reliability higher than Efficiency. In this case, Integrity was selected as the best RTOS using risk-based and probabilistic decision rules. The second case involved selecting the following preference order (from highest to lowest preference): Certifiability, Reliability, Cost, Product History, Correctness, Functionality, Life Cycle, Business, Software Design, Usability, Efficiency, Legal, Portability, and Interoperability. Given this preference order, Integrity was selected as the best RTOS using all four decision rules.

4 Conclusions

A Bayesian network scoring methodology has been used to select the best RTOS for the Space Shuttle CAU. The process obtained scores for various RTOS measures using a Bayesian network and then performed trade-off analyses between the scores using multi-criteria decision analysis software to rank order the available set of RTOS candidates. Using this methodology, the IPAO validated USA's RTOS selection for the Shuttle CAU. The method allowed for an objective, non-advocate, in-depth study of the RTOS candidates and served to verify RTOS performance and integrity in order to provide a clear justification process to Space Shuttle decision-makers.

References

[1] Cockpit Avionics Upgrade Project Management Plan, NSTS 37346, NASA JSC, TX. Dec. 2000.

[2] Biekert, R., Peterson, B., and Ferguson, R., *CAU Real Time Operating System Evaluation*, USA Inc., October 20, 2000.

[3] *Avionics Upgrade RTOS Study*, USA Inc., December 6, 2000.

[4] *NASA Program and Project Management Processes and Requirements*, NPG 7820.5A, NASA Software Policies, Washington, DC.

[5] Morris, A.T., *COTS Score: An Acceptance Methodology for COTS Software*, 19th DASC, Philadelphia, PA., October 2000.

[6] Abts, Christopher, 1999, *COTS Software Life Cycle Cost Modeling*, Ph.D. Thesis Proposal, USC.

[7] Sha, Lui, John Goodenough, and Bill Pollak, April 1998, Simplex Architecture: Meeting the Challenges of COTS in High-Reliability Systems, *CrossTalk – The Journal of Defense Softw. Eng.*

[8] Voas, Jeffrey, April 1998, Software Certification Laboratories: To Be or Not To Be Liable?, *CrossTalk – The Jnl of Defense Softw. Eng*.

[9] Kaner, Cem, David Pels, 1998, *Bad Software – What To Do When Software Fails*, John-Wiley & Sons, Inc. New York, NY.

[10] Bach, J.S., 1996, The Challenge of 'Good Enough' Software, Software Test Labs, URL: www.stlabs.com/bach/good.htm.

[11] Shuttle Master Verification Plan, NSTS 07700-10-MVP, Vol. IX, Part 1, Rev. B, NASA JSC, TX.

[12] Pearl, J., *Probabilistic Reasoning in Intelligent Systems: Networks of Plausible Inference*, Morgan Kaufmann Publishers, San Mateo, CA., 1988.

[13] Cooper, G.F., The Computational Complexity of Probabilistic Inference using Bayesian Belief Networks, *Artificial Intell.*, 42, pp. 393-405, 1990.

[14] *ISO/IEC. Information Technology – Software Product Evaluation*, ISO/IEC 9126, 1991.

[15] Heckerman, D., A Tutorial on Learning With Bayesian Networks. Technical Report MSR-TR-95-06, Microsoft Research, Redmond, WA., 1995.

[16] Ramoni M, and Sebastiani, P., Learning Bayesian networks from incomplete databases, *Proceedings of the 13th Conference on Uncertainty in Artificial Intelligence*, San Mateo, CA., 1997.

[17] Bickel, P. and Doksum, K., *Mathematical Statistics: Basic Ideas and Selected Topics*, Prentice Hall, Englewood Cliffs, NJ., 1977.

[18] Salo, A., and Hamalainen, R., *PRIME – Preference Ratios in Multi-attribute Evaluation*, Systems Analysis Laboratory, Helsinki University of Technology, Finland, February 24, 1999.