Available online at www.sciencedirect.com
ScienceDirect
Reliability Engineering and System Safety 92 (2007) 121–130
RELIABILITY ENGINEERING & SYSTEM SAFETY
www.elsevier.com/locate/ress

Information integration for complex systems

Alyson G. Wilson[a,*], Laura A. McNamara[b], Gregory D. Wilson[a]

[a]Los Alamos National Laboratory, P.O. Box 1663, Ms F600, Los Alamos, NM 87545-0600, USA
[b]Sandia National Laboratories, P.O. Box 5800, Albuquerque, NM 87185-0839, USA

Available online 22 August 2006

Abstract

This paper develops a framework to determine the performance or reliability of a complex system. We consider a case study in missile reliability that focuses on the assessment of a high fidelity launch vehicle intended to emulate a ballistic missile threat. In particular, we address the case of how to make a system assessment when there are limited full-system tests. We address the development of a system model and the integration of a variety of data using a Bayesian network.

Keywords: Bayesian network; Prior elicitation; Reliability; Ethnography

1. Introduction

At 8:40 p.m. on February 25, 1991, parts of an Iraqi Scud missile destroyed the barracks housing members of the US Army's 14th Quartermaster Detachment. This was the single, most devastating attack on US forces during the First Gulf War: 29 soldiers died and 99 were wounded. In the aftermath of this attack, there has been great focus on developing air defense systems capable of defending against ballistic missile attacks. The Critical Measurements and Counter Measures Program (CMCM) run by the US Army Space and Missile Defense Command conducts exercises to replicate projected ballistic missile threats. These exercises help the US military collect realistic data to evaluate potential defensive measures. The high-fidelity hardware and realistic scenarios created for the exercises provide extensive optical, radar, and telemetry data [1].

CMCM is organized into campaigns. Each campaign chooses a new ballistic missile threat and develops two to four high-fidelity launch vehicles that emulate the threat as closely as possible given intelligence information. That is, if country ‘‘A’’ has a ballistic missile that might be used against the US, a CMCM campaign may involve building a small number of replicas that can be launched to test and train US tracking and intercept capabilities. Assessing the reliability of these missile targets is difficult for a variety of reasons. While there is some reuse across campaigns, each set of launch vehicles is essentially a complex, one-of-a-kind, one-time-use system built for a specific data collection purpose. Typically, due to cost and schedule constraints, there are no ‘‘risk reduction’’ flights performed, so there is no full-system checkout before the actual flights. The systems are designed and built in a distributed fashion, with scientists and engineers from different companies designing, building, and integrating various parts of the vehicle. These campaigns are expensive (millions of dollars) and politically high profile.

The issue that we address in this paper is how to determine a preflight probability of mission success and how to assess areas of risk to the flight. Since there are no preflight full-system tests, this involves careful system modeling and the integration of as much component, historical, and engineering data as possible. The applied problem described here is large and complex. The system itself is well-understood in some dimensions by the groups working on the project, but not in terms of its overall reliability and performance. Knowledge of the total system is distributed across two primary research and development contractors and several subcontractors, all of which are located in different parts of the country. Each research group understands its area of responsibility at a local/granular level, and there is working knowledge of how to

*Corresponding author.
E-mail addresses: agw@lanl.gov (A.G. Wilson), lamcnam@sandia.gov (L.A. McNamara), gdwilson@lanl.gov (G.D. Wilson).

0951-8320/$ - see front matter
doi:10.1016/j.ress.2006.07.003

build a missile that will fly, but the project teams do not have methodology or tools to assess or predict full-system performance or reliability. The government agency that sets technical and scheduling requirements and oversees budgets is in yet another location, limiting its opportunity to assess the problems and progress of the project to weekly conference calls and periodic technical meetings where all of the contractors gather at a single location to brief the status of their efforts.

Modeling a system of this type presents challenges. There is heterogeneous data that explains different aspects of component and subcomponent performance, but very little sense of how that data is interrelated or how to sensibly combine the data and propagate reliability estimates and their uncertainties to understand overall system reliability. There are hundreds of components and subcomponents that all perform differently. Our approach to grappling with this problem was to first build a qualitative model of the problem space (its parts and relationships) and then to migrate that qualitative model to a graphical statistical model. The project involved collaboration between a social scientist who studies technical communities and a statistician who studies reliability and information integration for complex systems. We used ethnographic interviewing and observation techniques to elicit the problem structure, which was then represented in conceptual graphs. The framework that we use to quantitatively model the system and integrate the data is Bayesian networks (BN).

Ethnographic methods were originally developed by Western anthropologists to study foreign cultures. In the 20th century, these methods were deployed to study a variety of subcultures of Western society as well [2] (e.g., street gangs [3], long distance truck drivers [4], single parents [5], cigar smokers [6], endocrinologists [7], nuclear weapons designers [8]). In all instances, the goal of the anthropologist is largely the same: to better understand the internal logic of particular cultures, including beliefs, rituals, rules, problem solving strategies, and ways of producing and preserving knowledge. Modern ethnographic methods include interviewing, observation, and textual analysis with an eye towards understanding the culture in its own terms [9–11]. Considerable effort goes into not imposing outside preconceptions that would color interpretation of the information and data collected during fieldwork.

For several reasons, we employ ethnographic methods to create initial qualitative models of complex systems like the one discussed in this article. First, we want to capture how the technical community understands both the problem and their technical system and let that drive the statistical analysis that is ultimately performed. American industry is littered with effective statistical models that had a short useful life because they were unintelligible to the client or never gained cultural buy-in from the organization. Second, much technical knowledge is tacit and not explicit. To appropriately model the reliability of a complex system, we need to capture the sort of things that are left off the wiring schematics and engineering block diagrams. Third, we want to understand what the technical experts think is important to system performance, where they collect data, what that data means to them, and what their engineering judgment tells them about the system. When an engineer says "When it's cold this part doesn't work well, and when it's really cold it never works," we want to make sure we capture the logic of that knowledge in the system model.

Finally, many of the systems we work with are untestable or very expensive to test. There is not much data that can be used to indicate overall system performance. In the case of the CMCM campaigns, no integrated system test data is available before the missile is actually flown. What might be available for a system like this is subcomponent data, computer models, historical data on related parts/systems, and expert judgment. Working from these data sets necessitates a model that looks at reliability at the part and component level, and then rolls up that granular information into a reliability number for the overall system.

Because ethnographic methods explicitly require the researcher to avoid imposing external conceptual frames on a subject of study, ethnographic methods are well-suited to studying parts of a larger whole to document the rules and logic by which the parts relate to each other. Recognizing the value of ethnographic field methods in technology development, an increasing number of high-tech companies are using ethnographers in their design process [12–14].

What is unique about this project is the use of ethnographic methods to understand an engineering design process as part of a larger effort to design a statistical model. In this case, ethnographic methods were particularly useful because the CMCM missile was being developed by several stakeholder communities, each of which had very specific areas of expertise. Ethnographic field methods ensure that perceptions and views of one group of experts are not privileged over those of another group; in other words, no single community is allowed to define the technology in its local terms.

Instead, an integrated view of the relationships among parts, functions, and outcomes emerges iteratively as the researcher engages with, questions, and documents the responses of individual communities. The end result is a graphical system model that not only makes sense to all members of the project, but that is owned by all members of the expert community developing the technology. This graphical system model provides a big picture view of the system's functionality that can form the basis for a statistical analysis.

When the researcher understands the logical relationships of the components and the data that is available for each component, there are classes of graphical statistical models that can be used to understand overall system performance. In our case, the information was well-modeled using a BN. There is limited literature on the use of BNs in failure modes and effects analysis [15] and reliability [16,17], although there is quite a broad literature on using BNs for probabilistic modeling (e.g., [18–21]).

This paper will describe our approach to the assessment of Campaign 4 of the CMCM program (CMP-4). Section 2 details the development of the system representation using both qualitative and quantitative methods. Section 3 discusses the statistical model for the system and the information and data available to populate the model. Section 4 shows how the information was combined to make estimates. Section 5 contains conclusions and discussion.

2. Representing the system

The week before a previous campaign mission was supposed to fly, the project manager looked around the table at the subcontractors who had built various parts of the missile. She asked the question, "What's the probability this thing's going to fly?" The people at the table could only answer "My part is going to work!" There was no perspective, however, on how the integrated collection of these parts was going to perform. The authors of this paper had been working at Los Alamos National Laboratory (LANL) to develop methods for understanding the reliability of the US nuclear stockpile—a set of complex systems for which integrated system testing is no longer performed and where a multitude of traditional and nontraditional component data exists. The CMCM managers were eager for a collaboration to see if the methods under development at Los Alamos could help them understand the likely performance of their campaign missiles. In an initial kick-off meeting, the first priority was to identify and define goals.

2.1. Defining mission success

The questions of interest for CMP-4 included assessing the probability of mission success and identifying areas of technical risk. For large collaborative technical efforts to develop specialized technology, it is often difficult for those involved to have an integrated view of the project space, its goals, and the metrics for success. Likewise, at the beginning of the CMP-4 project, there was no clear definition of overall mission success. Given the occasional high-profile failure of a government test, managers often half-jokingly state "A mission is successful if I can write a press release that saves my job." The CMP-4 mission was a multimillion dollar effort, and it was important for the managers to understand what was likely to go right and what could possibly go wrong. Armed with this knowledge they could make better decisions and brief their supervisors about likely outcomes. There were clearly some things that would make a mission unsuccessful—for example, if the vehicle explodes on the launch pad. However, there was no clear enumeration of intermediate negative events that would render a mission more or less successful, nor a clear understanding of how certain events in combination would lead to undesirable outcomes.

To develop a definition of mission success, we worked with experts at CMCM to understand what events occur to make up the mission. The CMP-4 team used a diagram that mapped high-level mission events (i.e., launch, boost, booster separation, etc.) and the event timeline to the nominal missile trajectory. Through further discussions, we discovered that each of those events was much more complicated than the diagram suggested. Separation of booster stages involved performing numerous functions (i.e., the sending of electrical signals, the firing of pyrotechnic squibs, etc.) and many parts had to operate in concert to accomplish those functions.

Following this kick-off meeting with CMP-4 managers, the Los Alamos ethnographer working on the project made multiple trips to the three main cities where project activities were being conducted—one city in the Southeastern US, one in the Southwestern US, and one in New England. She attended technical briefings and conducted interviews with project managers and engineers working on specific components to develop a decomposition model of the system. The decomposition mapped out the system in terms of high-level mission events, the functions that had to transpire for those events to occur, and the parts that performed the requisite functions. This events-functions-parts approach to modeling the technical system allowed us to model both the pieces of the system and how they came together as a functional system. We could characterize both the overall probability of mission success and localize the areas of risk across the system. Unacceptably high risk or uncertainty could be traced to the event, function, part, or interaction of parts that was the root cause.

For CMP-4, the events that make up the mission fall into three categories: the threat-representative flight, the data collection, and the auxiliary experiments. Failure in any of these categories causes a mission to be unsuccessful. Fig. 1 summarizes the events that make up the mission. The threat-representative events are on the left side of the diagram. Notice that there are nine different data collection streams; some start immediately after ignition, and others start after later events.

The exact set of mission events is fluid almost up until launch. Engineers are trying to figure out what data collection is feasible and whether the weight of the collection equipment will fit within the weight limit imposed upon the mission. The same decisions must be made for the auxiliary experiments. Engineers and managers debate the data value versus weight versus technical viability tradeoffs late into the mission planning. At the same time, the contractors building the booster and payload sections are trying to design and build a missile that can fly a threat-representative trajectory while carrying the additional instrumentation and experiments. The requests for specific data collection or auxiliary experiments came from many communities, and the campaign

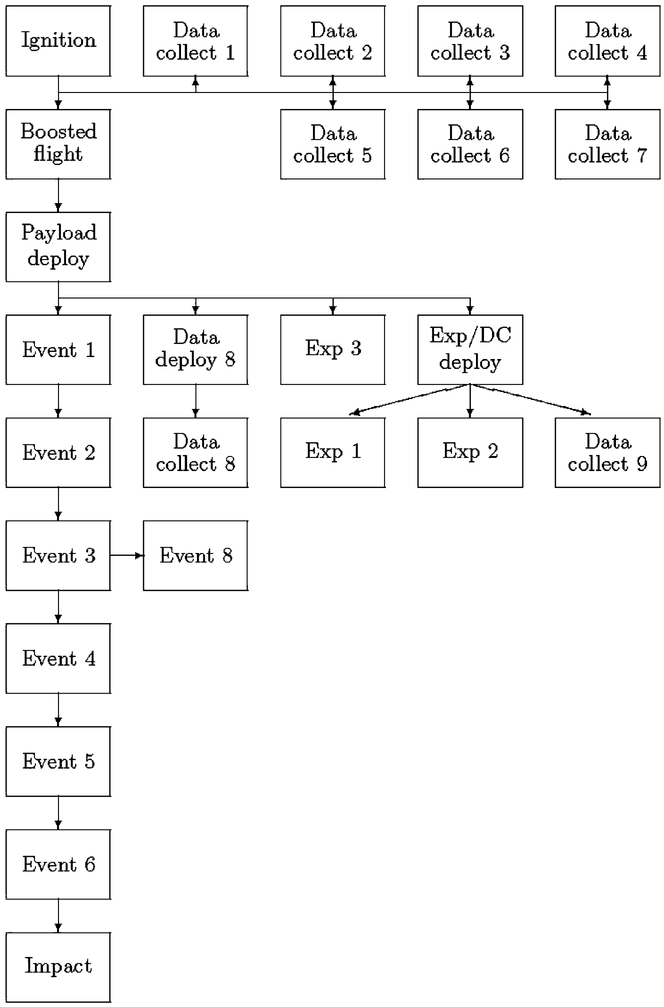


Fig. 1. Bayesian network showing conditional dependencies between threat-representative events, data collection, and auxiliary experiments.

managers wanted to satisfy as many of the "stakeholders" of the flight as possible.

For modeling purposes, it was possible to pin down most of the event sequence and determine the likely variations. Once the mission was defined in terms of the events that made it up, then the question of mission success could be revisited. One issue that had to be addressed was whether mission success was a discrete or continuous quantity. It was decided that the degree of mission success could be defined in terms of three categories: catastrophic failure (RED), degraded (YELLOW), or nominal (GREEN). Our ethnographic interviews indicated that this red, yellow, green (also known as a "stop-light chart") representation was natural for the CMCM staff and contractors working on the program, as it is commonly used in the Department of Defense to describe categories of outcomes in technical and military missions. Each of the events in Fig. 1 was also defined to have RED, YELLOW, or GREEN states. Table 1 summarizes what event states can cause catastrophic mission failure (RED); Table 2 summarizes the event states that cause a degraded mission (YELLOW).

Table 1
States where mission success is RED

Event	State
Ignition	RED
Boosted flight	RED
Payload deploy	RED
Event 3	RED

Table 2
States where mission success is YELLOW

Event	State
Boosted flight	YELLOW
Data collection 1	RED
Data collection 4	RED
Data collection 5	RED
Data collection 7	RED
Data collection 8	RED
Data collection 8	YELLOW
Data collection 9	RED
Experiment 1	RED
Event 1	RED
Event 2	RED
Event 4	RED
Event 5	RED

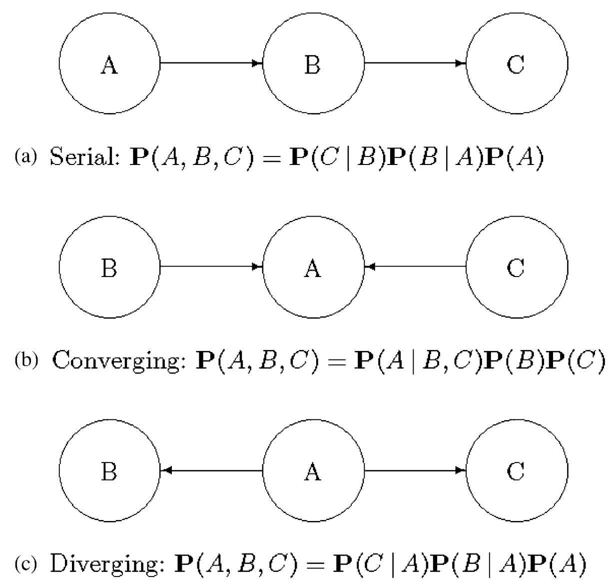
For the events in Tables 1 and 2, each is contingent upon the previous events (from Fig. 1) not failing catastrophically. For example, ignition fails catastrophically when the vehicle blows up on the launch pad. If the vehicle blows up before launch, all other events do not occur. The conditional specification of relationships in the mission suggests that a BN is an appropriate representation for the probability of mission success.

2.2. Constructing Bayesian networks

Informally, a BN is useful when the structure of the model and information is "local," meaning that each variable depends on a "few" other variables and affects a "few" more. The information that we elicited about mission success for CMP-4 (Fig. 1) has this structure. BNs have two parts: a qualitative part specifying the conditional probabilistic relationships among variables, and a quantitative part specifying particular conditional probabilities. Fig. 1 is an example of the qualitative part of a BN; we discuss more quantitative details in Section 3.

Formally, a BN is a pair $N = \{(V, E), P\}$, where (V, E) are the nodes and edges of a directed acyclic graph and P is a probability distribution on V. Each node contains a random variable, and the directed edges between them define conditional dependences/independences among the random variables. Fig. 2 summarizes the three probabilistic relationships that can be specified in a BN.

The conditional dependence/independence structures in Fig. 2 may be useful in developing the network, for

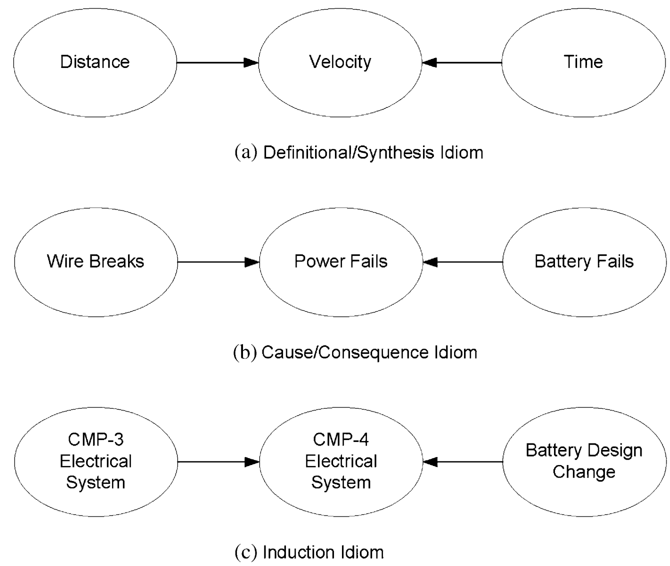


(a) Serial: $\mathbf{P}(A,B,C)=\mathbf{P}(C\mid B)\mathbf{P}(B\mid A)\mathbf{P}(A)$

(b) Converging: $\mathbf{P}(A,B,C)=\mathbf{P}(A\mid B,C)\mathbf{P}(B)\mathbf{P}(C)$

(c) Diverging: $\mathbf{P}(A,B,C)=\mathbf{P}(C\mid A)\mathbf{P}(B\mid A)\mathbf{P}(A)$

Fig. 2. Serial, converging, and diverging structures in a Bayesian network.

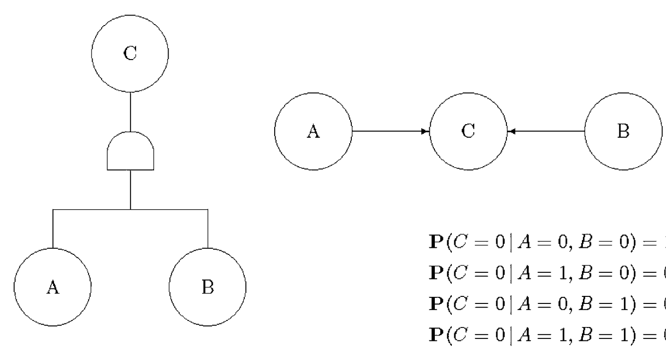


(a) Definitional/Synthesis Idiom

(b) Cause/Consequence Idiom

(c) Induction Idiom

Fig. 3. Bayesian network idioms.



$\mathbf{P}(C=0\mid A=0,B=0)=1$
$\mathbf{P}(C=0\mid A=1,B=0)=0$
$\mathbf{P}(C=0\mid A=0,B=1)=0$
$\mathbf{P}(C=0\mid A=1,B=1)=0$

Fig. 4. Fault tree with AND gate (left) conversion to Bayesian network (right).

example, when the network represents a hierarchical Bayesian model. There are other heuristics that can be used to construct a BN. Neil et al. [19] identifies five *idioms* or patterns that appear frequently in BNs.

The first is the *definitional/synthesis idiom*. This idiom is used when a child node is defined by its parents. For example, in Fig. 3a, velocity = distance/time. Since velocity is not random given distance and time, it technically does not have to be represented in the BN, but including it can clarify the model.

The second useful idiom is the *cause/consequence idiom*. This idiom is used when the parent nodes are inputs to a process, and the child node is the output. For example, in Fig. 3b, wire breaking or battery failing causes the power to fail. The cause/consequence idiom is ordered chronologically, with the parent nodes (inputs) occurring before the child node (outputs).

The third useful idiom is the *induction idiom*. This idiom is used when the parents contain historical, similar, or exchangeable data, and inductive reasoning is used to make inference (with uncertainty) about the child. In Fig. 3c, a historical attribute (perhaps a population parameter) and a measure of similarity to the historical data are used to make a forecast.

The fourth idiom is the *measurement idiom*, which is used to capture the uncertainty about the accuracy of a measurement instrument. The fifth idiom is the *reconciliation idiom*, which is used to capture the reconciliation of results from different measurement systems.

The development of the system model for CMP-4 primarily used the cause/consequence idiom, although we also used the definitional/synthesis idiom, measurement idiom, and induction idiom. In addition to the Neil et al. [19] idioms, another probability structure that can be captured by a BN is a fault tree [22]. The fault tree translation to a BN is like the definitional/synthesis idiom, with two basic events that contribute to an intermediate event represented as two parents and a child. Fig. 4 shows the correspondence between a fault tree AND gate and a BN AND node.

3. System model

The joint distribution of V, the set of nodes in a BN, is given by

$$\prod_{v\in V} P(v|\text{parents}[v]), \tag{1}$$

where the parents of a node are the set of nodes with an edge pointing to the node. For example, in the serial structure in Fig. 2a, the parent of node C is node B, and node A has no parents.

Eq. (1) shows that the joint distribution of the nodes in the BN is determined by a set of conditional distributions. For example, in Fig. 1, one of the probabilities that needs to be assessed to determine the joint distribution of all of the events is **P(**Data Collect 1 = RED|Ignition = GREEN). Notice that the conditional dependence/independence structure of the BN greatly decreases the total number of probabilities that have to be specified. If the random variables are discrete and there is no conditional structure, then every possible combination of values of the random variables must be assessed.

If the BN structure is derived from a fault tree, then some of the conditional probabilities have known 0/1 values. For an AND gate, some of the known probabilities are given in Fig. 4; we also know, for example, that $P(C = 1|A = 0, B = 0) = 0$. However, $P(A = 0)$ and $P(B = 0)$ are not specified by the fault tree structure. In a BN not specifically capturing fault-tree structure, the conditional probabilities are not constrained by the qualitative structure.

Consider again Fig. 1 and Tables 1 and 2. These summarize the events that make up the mission and the states of these events that define mission success. To make a quantitative assessment of the probability of mission success, all of the conditional probabilities in Fig. 1 need to be assessed. For CMP-4, these probabilities could not be elicited directly, nor were test data collected that addressed the probabilities directly. Consequently, once mission success was defined, the definition process began again for each of its component probabilities. This is where the events-functions-parts approach to modeling the system is relevant.

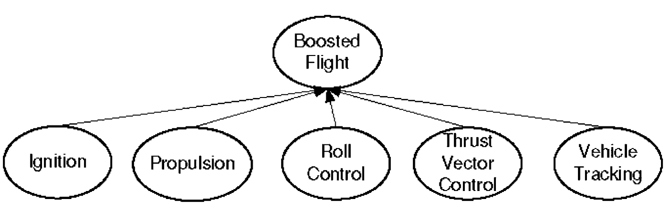


Fig. 5. Bayesian network decomposing boosted flight.

Consider, for example, the event boosted flight. Boosted flight can be decomposed into the BN given in Fig. 5. In particular, boosted flight is made up of a series of functions. Functions correspond to the idea of network *fragments*, which are small groups of related variables that help structure the BN [19,20].

Again, these functions are not at the right granularity, as there is no data or information about the conditional probabilities. Fig. 6 is the BN for roll control, which is a further decomposition of part of Fig. 5. The conditional probabilities for some of the parts can be estimated from existing data for those parts, many of which were used in past missions. The conditional probabilities can also be elicited using standard expert judgment elicitation techniques [23], which is what we did for newer parts. Because parts make up functions, and functions comprise events, the conditional probabilities for the corresponding functions and events can then be calculated. This process was completed for the entire set of events in Fig. 1 and resulted in a BN with approximately 600 nodes.

Eliciting the structure for a 600 node BN is not trivial. The LANL ethnographer spent many hours in the offices of subject matter experts and many days traveling between Los Alamos, NM, and the three cities where the contractors and managers worked. The ethnographic approach was successful in capturing how the component and part engineers understood the performance of their pieces of the system; the mapping of the connections between components sometimes revealed engineering frames of reference that a statistician working from design documents would not have otherwise recognized. For example, ethnographic interviews with the engineers who had developed parts indicated that focusing on the components, in and of themselves, would not be sufficient for understanding total system performance or reliability.

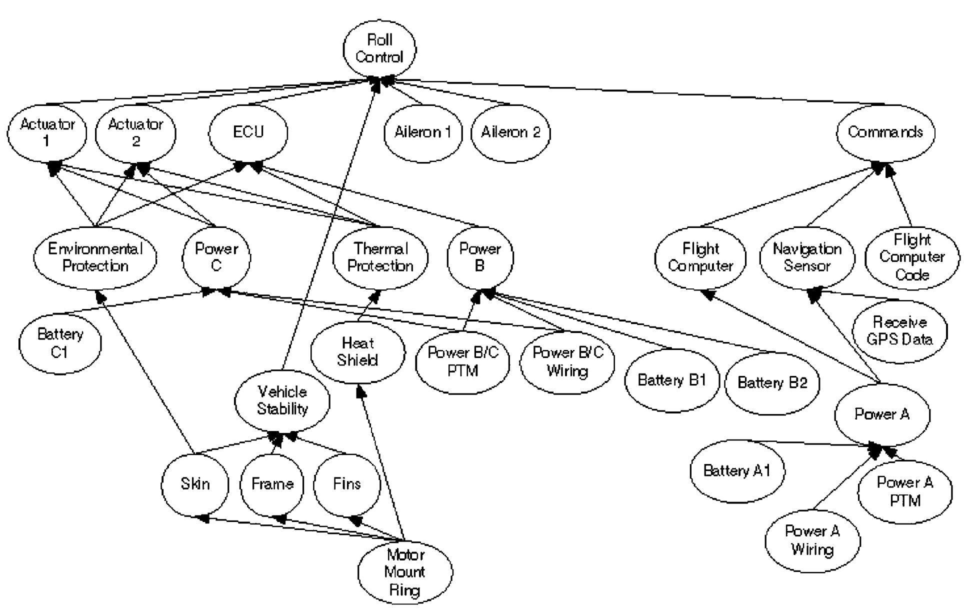


Fig. 6. Bayesian network decomposing roll control.

Instead, the engineers indicated that they were concerned about the specific function that the component, or a set of components, would perform. These functions resided in the collective knowledge of the engineers and were not represented on any of their standard engineering diagrams. For the statistical model, the importance of these functions was in connecting specific data sources about a specific part, to the overall performance of a system during flight.

Eliciting the functional relationships among components allowed the ethnographer to develop a large conceptual graph [24,25]. This resulted in a set of drawings that revealed relationships between parts, sets of parts, their functional relationships, and the key events these systems would produce during the flight. These diagrams were then formalized into conceptual graphs that the ethnographer reviewed iteratively with the system's engineers. The missile system had hundreds of parts, so the system modeling required many trips to interview contractor experts. With each trip, the ethnographer could conduct some introductory interviews with a few experts, iterate diagrams with a few other experts, and discuss mission planning and system modeling progress with project managers.

The analysis of the CMP-4 was conducted along with the design, construction, and testing of the system itself. Consequently, at various points during the analysis, there was no "data" available, at least in terms of repeated observations of particular parts, functions, or events. Often, the experts were unable to provide precise conditional probabilities, and frequently they did not have access to historical data that they could use to formally estimate the probabilities. Consequently, values were elicited within the ranges given in Table 3. The experts were asked to identify failure modes for each part, and were then asked to estimate the chance of part failure.

We reviewed the various biases associated with estimating probabilities with the respondents and asked them to remain aware of bias issues throughout the discussions. With this type of expert judgment elicitation, there are a few key biases to mitigate: availability, anchoring and adjustment, and the management bias [26,27]. The availability bias is a subject's tendency to over-estimate the future probability of events they have seen in the past. That is, if subjects have seen a part fail in a particular way in the past, they will mistakenly assign a higher chance of seeing that failure again than to seeing another failure mode that has not been observed yet. Anchoring and adjustment biases refer to a subject's tendency to anchor on the last probability that was discussed and adjust the next probability elicited based on that previous information. For example, when this bias is operating, the elicitation process may result in getting strings of identical estimates for failure modes that may, in actuality, have very different chances of occurrence. The management bias is when the subject is answering based on a goal rather than an actual probabilistic belief. That is, if the subjects know that the whole system performs poorly if the chance of part failure is above 10%, they may say that the chance of part failure is 10% or lower, when in fact they believe the chance of part failure to be higher.

Table 3
Elicited probability ranges

Elicited value (Φ)	Range
0	0
1	(0,0.01)
2	[0.01,0.1)
3	[0.1,0.25)
4	[0.25,0.5)
5	[0.5,1)
6	1

When conducting expert judgment elicitations, it is a good practice to make the subjects aware of such biases in an attempt to inoculate the elicitation against their interference. There are other measures to take as well. Making subjects explicitly talk about the full range of failure modes helps avoid the availability bias. Breaking up the questioning and changing the phrasing of the questions will also help address the anchoring and adjustment biases. Many of the interviews for this project were conducted in groups, which also helped to address bias. Group interviews prompted discussion among colleagues and consensus estimates that weeded out individual biases that might have come into play. Group interviews are less productive, however, where status or personality differences make it difficult for the experts to freely offer and discuss opinions. To mitigate groupthink bias, we also took the probabilities the experts provided, propagated them through the BN, and fed the initial answers back to the experts to check the coherence of the set of judgments and the structure of the BN itself. It is also always important to make sure that the experts being interviewed are the right people to be questioning within the organization—i.e., the ones who can give you the best information. For this project, the experts who had given us the structural information for a particular part were the ones (we discovered) who had spent their lives designing batteries, or actuators, or squibs. These individuals were the experts we went back to when we needed to elicit quantitative performance estimates. The model structures, the elicited probabilities, and the localized model results were briefed to and reviewed by the component and system experts as well as the system and program managers in an effort to validate the information that was being used for the model.

As examples, some of the elicited information for Fig. 6 had the following form:

- The flight computer code is RED with $\Phi = 2$.
- If the navigation sensor fails RED, commands are YELLOW with $\Phi = 3$.
- If the flight code is RED, then commands are RED.
- If Power B is RED, the electronics control unit is RED.
- If environmental protection is RED, the electronics control unit is RED with $\Phi = 4$.

Another potential approach that we did not pursue is fuzzy logic with approximate reasoning [28].

4. Statistical inference

Suppose that we have developed a system model and identified data sources using the methodology described in Section 3. We now need to calculate the marginal probabilities for various nodes of interest in the BN. We illustrate the methodology using two simple examples.

Suppose that we have the BN in Fig. 7. Each node has three possible states RED (R), Yellow (Y), and GREEN (G). We have the following data and information:

- A is RED with $\Phi = 1$, A is YELLOW with $\Phi = 2$.
- B is RED with $\Phi = 1$, B is YELLOW with $\Phi = 3$.
- C is RED if A and B are both RED.
- C is YELLOW if either (but not both) A or B is RED, or if either is YELLOW.
- C is GREEN if A and B are both GREEN.

We model the information about A as a Dirichlet(0.5,10,200). Let $p_A^{(R)}$ denote the probability that A is RED and $p_A^{(Y)}$ the probability that A is YELLOW. This implies that the marginal distribution for $p_A^{(R)}$ is Beta(0.5,210), and the marginal distribution for $p_A^{(Y)}$ is Beta(10,200.5). Choosing a specific probability distribution adds information to that elicited in Table 3, so it is important to do sensitivity analysis and show the consequences of this choice back to the subject matter experts. For example, we could also have chosen to model $p_A^{(R)}$ and $p_A^{(Y)}$ as independent Uniform(0,0.01) and Uniform(0.01,0.1). Suggestions for choosing specific beta and gamma distributions to describe information like that in Table 3 are found in Mosleh and Apostolakis [29].

We model the information about B as a Dirichlet(0.1,10,50). We have

$$p_C^{(R)} = p_A^{(R)} p_B^{(R)},$$

$$p_C^{(Y)} = p_A^{(R)} p_B^{(Y)} + p_A^{(R)} p_B^{(G)} + p_A^{(Y)} p_B^{(R)} + p_A^{(G)} p_B^{(R)} + p_A^{(Y)} p_B^{(Y)} + p_A^{(Y)} p_B^{(G)} + p_A^{(G)} p_B^{(Y)},$$

$$p_C^{(G)} = p_A^{(G)} p_B^{(G)}.$$

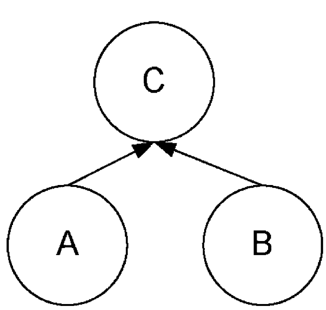


Fig. 7. Simple example: converging Bayesian network.

We calculate the joint distribution of $p_C^{(R)}$ and $p_C^{(Y)}$ using Monte Carlo simulation. The marginal distribution for $p_C^{(Y)}$ is given as the solid line in Fig. 8.

Now suppose that we have somewhat different data and information:

- A is RED with $\Phi = 1$, A is YELLOW with $\Phi = 2$.
- B is RED with $\Phi = 1$, B is YELLOW with $\Phi = 3$.
- C is RED if A and B are both RED.
- C is RED with $\Phi = 2$ if either (but not both) A or B is RED; otherwise it is YELLOW.
- C is GREEN if A and B are both GREEN.

We have done 15 tests of C, and 3 were RED, 5 were YELLOW, and 7 were GREEN.

We choose the same Dirichlet distributions to model our information about $p_A^{(R)}$, $p_A^{(Y)}$, $p_B^{(R)}$, and $p_B^{(Y)}$. Let ρ denote the probability that C is RED if either A or B is RED. We model $\rho \sim$Uniform(0.01,0.1). We have

$$p_C^{(R)} = p_A^{(R)} p_B^{(R)} + \rho (p_A^{(R)} p_B^{(Y)} + p_A^{(R)} p_B^{(G)} + p_A^{(Y)} p_B^{(R)} + p_A^{(G)} p_B^{(R)}),$$

$$p_C^{(Y)} = (1-\rho)(p_A^{(R)} p_B^{(Y)} + p_A^{(R)} p_B^{(G)} + p_A^{(Y)} p_B^{(R)} + p_A^{(G)} p_B^{(R)}) + p_A^{(Y)} p_B^{(Y)} + p_A^{(Y)} p_B^{(G)} + p_A^{(G)} p_B^{(Y)},$$

$$p_C^{(G)} = p_A^{(G)} p_B^{(G)}.$$

In addition, we have multinomial data, $X_i \sim$Multinomial ($p_C^{(R)}$, $p_C^{(Y)}$, $p_C^{(G)}$). We can no longer use Monte Carlo simulation to calculate the distribution of $p_C^{(R)}$ and $p_C^{(Y)}$. Instead, if we use a Bayesian inferential approach, we can use the Dirichlet distributions on $p_A^{(R)}$, $p_A^{(Y)}$, $p_B^{(R)}$, and $p_B^{(Y)}$ and the Uniform distribution on ρ as prior distributions, the multinomial data to specify a likelihood, and Markov Chain Monte Carlo [30] to calculate a posterior distribution for $p_C^{(R)}$ and $p_C^{(Y)}$ (and, of course, $p_A^{(R)}$, $p_A^{(Y)}$, $p_B^{(R)}$, $p_B^{(Y)}$, and ρ). Details of the Bayesian approach for fault trees

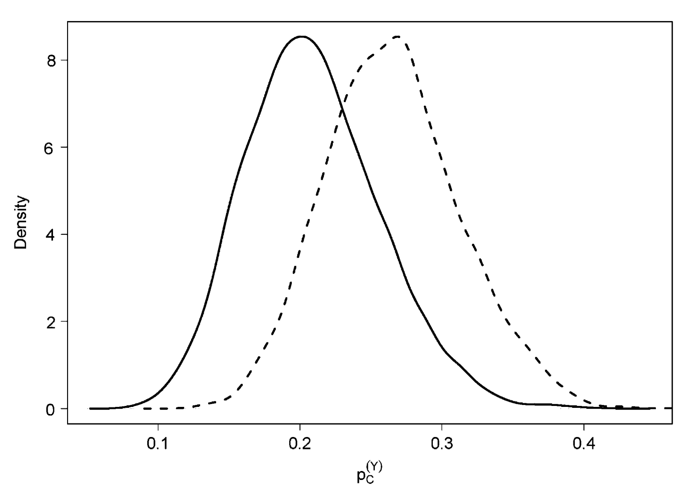


Fig. 8. Marginal distribution of probability C is YELLOW given first set of information (solid) and second set of information (dashed).

with multi-level data can be found in [31]. The marginal distribution for $p_C^{(Y)}$ using the additional information is given as the dashed line in Fig. 8. Note that in this example, since the data suggests more failures than one might have expected given the prior information, the choice of prior is important.

5. Conclusions

The final BN for CMP-4 contained approximately 600 nodes. Neil et al. [19] summarizes many of the issues that surround working with a model of this size:

> Large knowledge-based systems, including BNs, are subject to the same forces as any other substantial engineering undertaking. The customer might not know what they want; the knowledge engineer may have difficulty understanding the domain; the tools and methods applied may be imperfect; dealing with multiple ever-changing design abstraction is difficult, etc. In the end these issues, along with people, economic and organizational factors, will impact on the budget, schedule, and quality of the end product. (Neil et al. [19], p. 265)

There are inherent challenges in working on a project like this. For the people in charge of such a system, even articulating an overarching definition for success can be difficult—getting a workable statistical model out of it, even more so. The authors of this paper represent a collaboration between a statistician who studies reliability and information integration for complex systems and an ethnographer with expertise in eliciting model structures and expert judgment. They worked together to develop first a qualitative model of the events, functions, and parts of the CMP-4 system, and second a graphical statistical model that added elicited and available data to provide quantitative answers for CMP-4 managers.

There are no end-to-end computer tools that help with system representation, statistical model formulation, and inference. The original drawings were done in Microsoft Visio, which is a fairly agile graphics package. However, the model structure elicitation and iteration involved drawing and redrawing hundreds of diagrams; even the final set of component and part diagrams numbered nearly a hundred. Hence, it was often difficult to keep track of parts across the diagrams because each one was drawn as a separate file; e.g., it was sometimes difficult to discern if the "Battery A" on diagram 34 was the same as "battery-a" on diagram 64, especially since the same battery could be powering two or three different components. Moreover, Visio does not support statistical inference, so all diagrams had to be redrawn into Analytica, the statistical software package that we used to integrate the diagrams into a single model with logical relationships between the components/parts. The inference was then coded in C. In retrospect, the inference likely could have been done within Analyitica, but it was easier to think through some of the statistical analysis issues in C.

At LANL, we are working on more integrated tools for system representation [32] and inference [33]. We hope that eventually these tools will be part of an end-to-end suite of tools.

This problem and these methods are of particular interest to the Department of Energy national laboratories. Since the US ended its full-scale nuclear test program in 1992, LANL has been developing statistical methodology to certify the reliability and performance of the US nuclear stockpile. The assessment issues are quite similar to those faced by the CMCM program: no full-system testing and the need to develop complete system models and integrate all available information and data. We believe that these methods are applicable to a variety of complex systems and that they can provide a traceable and defensible estimate of system metrics, which can facilitate other planning and problem-solving efforts.

References

[1] United States Army Space and Missile Defense Command Technical Center. Critical Measurements Countermeasures Program Fact Sheet, 2004 ⟨http://www.smdc.army.mil/FactSheets/CMP.pdf⟩.
[2] Deegan MJ. The Chicago school of ethnography. In: Paul Atkinson K, et al., editors, Handbook of ethnography. Thousand Oaks, CA: Sage Publications; 2002. p. 11–25.
[3] Barron SW, Tindall DB. Network structure and delinquent attitudes within a juvenile gang. Soc Networks 1993;15(3):255–73.
[4] Gardner AM. Long haul from deregulation: truck drivers and social capital in the Louisiana oilpatch. Hum Organ 2002;61(4):390–8.
[5] Coles RL. Black single fathers: choosing to parent full-time. J Contemp Ethnogr 2002;31:411–39.
[6] Desantis AD. A couple of white guys sitting around talking: the collective rationalization of cigar smokers. J Contemp Ethnogr 2003;32:432–66.
[7] Latour B, Woolgar S. Laboratory life. Princeton, NJ: Princeton University Press; 1986.
[8] Gusterson H. Nuclear Rites: a Weapons Laboratory at the end of the Cold War. Berkeley, CA: University of California Press; 1996.
[9] Agar MH. The professional stranger: an informal introduction to ethnography. New York: Academic Press; 1996.
[10] Weiss RS. Learning from strangers: the art and method of qualitative interview studies. New York: The Free Press; 1994.
[11] Sunstein BS, Chiseri-Strater E. Field working: reading and writing research. Boston: Bedford/St. Martins; 2002.
[12] Suchman L. Plans and situated actions: the problem of human–machine communication. New York: Cambridge University Press; 1987.
[13] Suchman L. Building bridges: practice-based ethnographies of contemporary technology. In: Schiffer M, editor. Anthropological perspectives on technology. Albuquerque: University of New Mexico Press; 2001. p. 163–77.
[14] Orr J. Talking about machines: an ethnography of a modern job. Ithaca, NY: Cornell University Press; 1996.
[15] Lee B. Using Bayesian belief networks in industrial FMEA modeling and analysis. In: Proceedings of the annual reliability and maintainability symposium, International symposium on product quality and integrity, Philadelphia, 2001. p. 7–15.
[16] Sigurdsson J, Walls L, Quigley J. Bayesian belief nets for managing expert judgement and modelling reliability. Qual Reliab Eng Int 2001;17:181–90.
[17] Portinale L, Bobbio A, Montani S. From artificial intelligence to dependability: Modeling and analysis with Bayesian networks.

In: Wilson A, Limnios N, Keller-McNulty S, Armijo Y, editors. Modern statistical and mathematical methods in reliability. Singapore: World Scientific; 2005. p. 365–82.

[18] Spiegelhalter D. Bayesian graphical modeling: a case-study in monitoring health outcomes. Appl Stat 1998;47(1):115–33.

[19] Neil M, Fenton N, Nielsen L. Building large-scale Bayesian networks. Knowledge Eng Rev 2000;15(3):257–84.

[20] Laskey K, Mahoney S. Network engineering for agile belief network models. IEEE Trans Knowledge Data Eng 2000;12(4):487–98.

[21] Jensen F. Bayesian networks and decision graphs. New York: Springer; 2001.

[22] Bobbio A, Portinale L, Minichino M, Ciancamerla E. Improving the analysis of dependable systems by mapping fault trees into Bayesian networks. Reliab Eng Syst Saf 2001;71(3):249–60.

[23] Meyer M, Booker J. Eliciting and analyzing expert judgment: a practical guide. Philadelphia: ASA-SIAM; 2001.

[24] Sowa J. Knowledge representation: logical, philosophical, and computational foundations. Pacific Grove, CA: Brooks Cole Publishing, 2000.

[25] McNamara L, Leishman D. Interlopers, translators, scribes, and seers: anthropology, knowledge representation and Bayesian statistics for predictive modeling in multidisciplinary science and engineering projects. In: Malcolm G., editors. Multidisciplinary studies of visual representations and interpretations. Amsterdam: Elsevier; 2004. p. 211–28.

[26] Wilson A. Cognitive factors affecting subjective probability assessment. Duke University ISDS Discussion Paper 94-02, 1994.

[27] Booker J, McNamara L. Expert knowledge in reliability characterization: A rigorous approach to eliciting, documenting, and analyzing expert knowledge. In: Nikolaidis E, Ghiocel DM, Singhal S, Nikolaidis N, editors. Engineering design reliability handbook. CRC Press: New York; 2005.

[28] Klir G. Uncertainty and information. Hoboken, NJ: Wiley; 2006.

[29] Mosleh A, Apostolakis G. Some properties of distributions useful in the study of rare events. IEEE Trans Reliab 1982;R-31(1):87–94.

[30] Gelman A, Carlin JB, Stern HS, Rubin DB. Bayesian data analysis. London: Chapman & Hall; 1995.

[31] Hamada M, Martz H, Reese CS, Graves T, Johnson V, Wilson A. A fully Bayesian approach for combining multilevel failure information in fault tree quantification and optimal follow-on resource allocation. Reliab Eng Syst Saf 2004;86(3):297–305.

[32] Koehler A, Klamann R. GROMIT: a tool for systems ethnography. Los Alamos National Laboratory Technical Report LA-UR-04-4370, 2004.

[33] Graves TL. A framework for expressing and estimating arbitrary statistical models using Markov chain Monte Carlo. Los Alamos National Laboratory Technical Report LA-UR-03-5934, 2003.