Temporal Bayes net information & knowledge entropy

KENNETH J. HINTZ
STEVEN DARCY

Various information measures have been defined on Bayes Nets (BN) with the assumption that the Bayes Net is stationary. Our interest is in the utilization of a BN as a component of a real-time, information-based sensor management system wherein the dynamics of the situation cause changes both in the structure and underlying probabilities of the nodes in the BN. If a BN is used to represent the situation assessment (SA) of an environment as a result of our observations of that environment, we can say that the BN represents our knowledge about the situation in the form of a temporal Bayes net (TBN). If one were to not observe the processes in an environment with additional sensor observations, then the underlying probabilities of at least some of the BN nodes diffuse at a rate dependent on the dynamics of the process whose uncertainty is represented by that node, hence the use of the modifier *temporal*. This loss of knowledge in the form of increasing uncertainty results in information flow from the TBN, or, as we refer to it here, temporal information loss. In order to compensate for this temporal information loss and maintain or improve our knowledge of an environment, the environment needs to be observed by obtaining data. We focus in this paper on choosing a global TBN information measure In doing so, we differentiate between aleatory nodes with stationary uncertainties and epistemic nodes with temporal uncertainties, as well as formulate a dynamic representation of these temporal uncertainties. We provide several examples of temporal information loss under different dynamic assumptions.

Manuscript received February 19, 2018; revised July 27, 2018, February 2, 2019; released for publication February 19, 2019.

Refereeing of this contribution was handled by Paulo Costa.

Authors' addresses: K. J. Hintz, Department of Electrical and Computer Engineering, George Mason University, Fairfax, VA 22030 USA (E-mail: khintz@gmu.edu). S. Darcy, Systems Engineering and Operations Research Department, George Mason University, Fairfax,VA 22030 USA (E-mail: sdarcy1@gmu.edu).

This paper results from research supported by the Naval Postgraduate School Contract No. N62271-17-P-1154 awarded by the Naval Postgraduate School, Director Contracting and Logistics. The views expressed in written materials or publications, and/or made by speakers, moderators, and presenters, do not necessarily reflect the official policies of the Naval Postgraduate School nor does mention of trade names, commercial practices, or organizations imply endorsement by the U.S. Government.

1557-6418/18/$17.00 © 2018 JAIF

1. INTRODUCTION

This document introduces the concept of information loss over time from Bayes nets (BN) due to the dynamics associated with epistemic nodes in a BN. Epistemic and aleatory uncertainties will be defined in the sequel and adapted to BN. We call this phenomenon *temporal Bayes information (TBI)* loss to distinguish it from the term *dynamic* used in dynamic Bayes nets (DBN) by Kjaerulff [1] and Chang & Sun [2]. DBN can also have a temporal component by incorporating changes in the network structure itself over time rather than just changes in the uncertainty. Furthermore, it will be shown that the decision as to which epistemic node to update in order to maximize the information rate can change with time. This is due to the fact that the uncertainties associated with the processes represented by different epistemic nodes do not change at the same rate. We will relate the concept of TBI to two different interpretations of Bayes nets, one of which is purely aleatory in that it provides a graphical representation of the joint probability distribution of random variables, and one used in target detection and tracking problems which is composed of both epistemic and aleatory nodes. There are two examples presented later in this paper which differentiate between aleatory (yellow) and epistemic (blue) nodes as shown in Figure 2 and Figure 4. The latter BN is representative of a causal Bayes net as introduced by Pearl in his fundamental book [3].

Our interest in TBI is intimately tied with our use of a BN as an underlying component of our method of information based sensor management (IBSM). IBSM will not be discussed further here as it is has been presented in previous papers by Hintz & McVey [4], and Hintz & Kadar [5]. Briefly, the IBSM situation information expected value net (SIEV-net) takes an information measure defined on a situation assessment Bayes net and combines it with mission values and the probability of obtaining information to compute the expected situation information value rate. We use the resulting expected situation information value rate ($EIVR_{sit}$) to choose from among the several situation information needs that information request which will yield the highest value of $EIVR_{sit}$. Our interest in TBI stems from the fact that the predictable loss of information from a BN will yield different values of the maximum $EIVR_{sit}$ depending on the delay in fulfilling that information request. The different values of $EIVR_{sit}$ at different times results in different choices of which information to request.

As a brief preliminary example of how TBI can affect the amount of information which could be obtained from mutually exclusive sensing actions which could be taken at two different times in the future, let's assume that we are tracking 2 targets, the state of each one being represented as individual nodes in a BN. Let's further hypothesize that one has highly dynamic kinematics, e.g., a fighter aircraft, with a large process noise, and

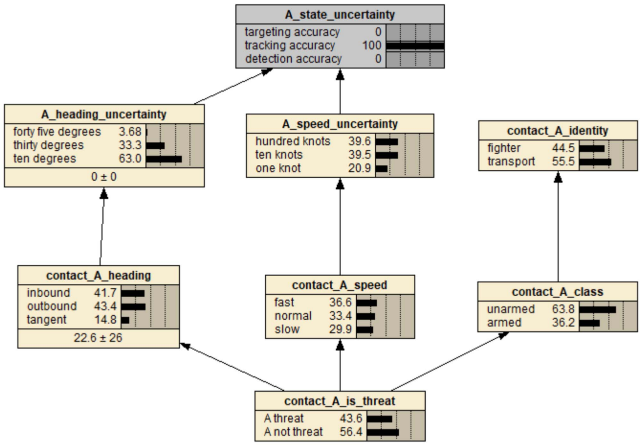

Fig. 1. BN showing the use of the target kinematic state produced by an external K-filter fusion process to populate nodes of a BN in order to estimate whether a contact is a threat.

a second target with slow dynamics, e.g., a helicopter, and smaller process noise. If we extrapolate the Kalman filter (K-filter) error covariance matrix of each of these to some proximate time in the future, it may be that we will obtain more situation information if we choose to observe the helicopter rather than the fighter, as in the case where we may have just detected the helicopter and started tracking it. However, if we were to wait to make an observation to some later time, then the high kinematic dynamics of the fighter may, through the extrapolation of both the helicopter and fighter error covariance matrices to this later time, result in the fact that more situation information will be gained by observing the fighter.

This loss of information over time has been recognized, but not explicitly evaluated by Ciftcioglu et al. [6] in dealing with maximizing information from multiple sensors. They say that "[t]he main property of QoI [Quality of Information] is that it is a composite metric which deteriorates with age and increases with time due to additional information gathered. The amount of information that sensors collect varies randomly throughout time, which leads to uncertainty in the QoI utility evolution."

1.1. Aleatory vs epistemic definitions

Winkler [7] states that "[a]t a fundamental level, uncertainty is uncertainty, yet the distinctions [aleatory and epistemic, reducible and irreducible, stochastic and subjective] are related to very important practical aspects of modelling [sic] and obtaining information." Costa, et al. [8], state that "Uncertainty Type is a concept that focuses on underlying characteristics of the information that make it uncertain. Its subclasses are Ambiguity, Incompleteness, Vagueness, Randomness, and Inconsistency…" Shafer [9], in discussing the distinction between belief and chance, provides a simple example by writing that "[c]hances arise only when one describes an aleatory (or random) experiment, like the throw of a die or the toss of a coin." We focus on two particular uncertainties, aleatory and epistemic, as they apply to BN in order to differentiate between those nodes that participate in an information measure and those that don't.

Aleatory and epistemic are terms used in seismic hazard analysis, reliability engineering, system safety, structural reliability, and risk analysis, but are not common in the information fusion literature. The general meaning of aleatory and epistemic can be taken from the Oxford English Dictionary as:

aleatory: Dependent on uncertain events or occurrences; haphazard, random [10]

epistemic: Of or relating to knowledge, or to its extent, linguistic expression, or degree of validation [11]

Unfortunately these are not very satisfying definitions for our intended use in information fusion and, in particular, situation assessment utilizing BN.

To facilitate the discussion we first make clear what we mean by aleatory and epistemic by quoting from Der Kiureghian and Ditlevsen [12] in the field of structural reliability or risk analysis:

> The word aleatory derives from the Latin *alea*, which means the rolling of dice. Thus, an aleatoric uncertainty is one that is presumed to be the intrinsic randomness of a phenomenon. Interestingly, the word is also used in the context of music, film and other arts, where a randomness or improvisation in the performance is implied. The word epistemic derives from the Greek $\epsilon\pi\iota\sigma\tau\epsilon\mu\epsilon$ (episteme), which means knowledge. Thus, an epistemic uncertainty is one that is presumed as being caused by lack of knowledge (or data).

In Abrahamson's paper related to seismic hazard analysis we find [13]

> Aleatory variability is the natural randomness in a process. For discrete variables, the randomness is parameterized by the probability of each possible value. For continuous variables, the randomness is parameterized by the probability density function.
>
> Epistemic uncertainty is the scientific uncertainty in the model of the process. It is due to limited data and knowledge. The epistemic uncertainty is characterized by alternative models. For discrete random variables, the epistemic uncertainty is modelled [sic] by alternative probability distributions. For continuous random variables, the epistemic uncertainty is modelled [sic] by alternative probability density functions. In addition, there is epistemic uncertainty in parameters that are not random by have [sic] only a single correct (but unknown) value.

While other authors have presented alternative views of these two terms, we believe that Abrahamson's meaning best serves our purposes in the field of information fusion and situation assessment.

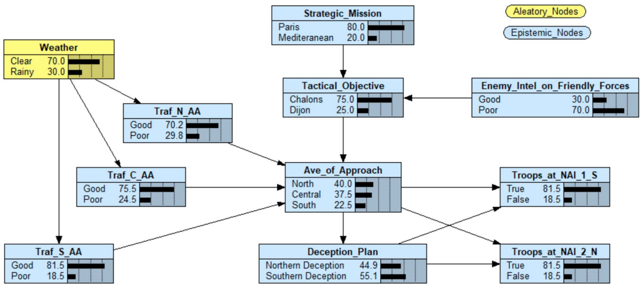

Fig. 2. Bayesian network representing enemy intent showing both aleatory (shown in yellow) and epistemic (shown in blue) nodes. After Buede, et al. [22].

According to DK&D [12], "[a]ny discussion on the nature and character of uncertainties should be stated within the confines of the model universe." They further suggest this determination should be a pragmatic choice based on how the modeler intends to use the uncertainty in the model. Since BN situation models in the information fusion world are quite diverse, it would seem that the allowance for both aleatory and epistemic chance nodes is appropriate. Furthermore, there is the possibility that the nodes change from aleatory to epistemic over time as the model is used. For example, DK&D [12] use the concept of the strength of concrete in a building having a known statistical uncertainty before the building is built (aleatory); however, after the building is built, measurements of the strength can be taken over the time the concrete is curing leading to an epistemic statistical uncertainty. Note that the process is the curing of the concrete with an associated uncertainty which can be reduced if measured, but remains the same or increases in uncertainty if not measured. In the case of situation assessment, an example is converting from the probability that a target is going to enter a volume of space (aleatory) to the probability that a target has been detected (epistemic) once a detection is made. The fact that there has been a detection does not mean that there is a target in that volume with absolute certainty since each detection has associated with it a probability of detection less than 1 ($P_d < 1$) and a probability of false alarm of greater than zero ($P_{fa} > 0$). The uncertainty about whether a target is actually in the volume is reduced with repeated measurements. Another example is shown in Figure 2 wherein the weather is an explicit aleatory node during mission planning, but becomes an epistemic node during its use when particular values of the weather can be acquired as evidence.

If we relate the above to Pearl's causal networks [14] [3] which require a directed relation between nodes, epistemic is a straightforward uncertainty which is added to either the linear or nonlinear functional relation between nodes as in the linear relationship below with the additive aleatory random variable u_i:

$$x_i = \sum_{k \neq i} \alpha_{ik} x_k + u_i \quad i = 1, \ldots, n \tag{1}$$

The nodal value of interest, x_i, is epistemic as its uncertainty can be refined with repeated measurements thereby reducing the uncertainty introduced by the random additive component, u_i.

1.2. Aleatory or epistemic: stationary or nonstationary?

Aleatory uncertainties may change, but cannot be improved with repeated measurements as they are associated with a naturally occurring randomness. A counter argument to this, which we will ignore without loss of generality, can be best exemplified by the probabilities associated with the roll of a die. Mathematically, a fair die has equal probability of the single event comprised of a face of a die. No amount of experiments on this mathematically fair die will change that. In reality, no physical die is perfect and hence, not fair. That is, repeated rolling of the die will show that some faces will occur more than others due to the imperfections in the physical die. This is not to be confused with the typical gambler's mistake most easily associated with the flip of a coin. If the coin toss results in an unusually long run of heads or tails, one wants to think that the next toss will be the opposite even though we know that there is equal probability of the two faces of the coin occurring as a result of the next toss.

Aleatory uncertainties may change over time, but not due to measurements. In the case of weather and whether or not it is going to rain, the aleatory uncertainty changes if there are observed clouds, but repeated measurements to determine if clouds are present do not change the uncertainty about whether it is going to rain or not.

Epistemic uncertainties can be non-stationary or changing over time due to observations of, or changes in, the process dynamics. An interesting example is a situation assessment node which represents the kinematic state of a target in track. If this nodal estimate is derived from a Kalman filter, then we can see both aleatory and epistemic statistics depending on how the modeler utilizes the K-filter. The equations from the discrete K-filter are [15]

system model,

$$\vec{x}_k = \mathbf{\Phi}_{k-1}\vec{x}_{k-1} + w_{k-1}, \quad w_k \sim N(0, \mathbf{Q}_k) \tag{2}$$

in which Φ_{k-1} is the state transition matrix and w_{k-1} is the process noise, both having subscripts indicating that they are non-stationary and may change over time,

measurement model,

$$\vec{z}_k = \mathbf{H}_k \vec{x}_k + v_k, \quad v_k \sim N(0, \mathbf{R}_k) \tag{3}$$

in which $\vec{z}_k$ is the observation vector, $\mathbf{H}_k$ is the observation matrix, and v_k is the additive, white, Gaussian measurement noise, all having subscripts indicating that they are non-stationary and may change over time, state estimate extrapolation,

$$\hat{\vec{x}}_k^- = \mathbf{\Phi}_{k-1}\hat{\vec{x}}_{k-1}^+ \tag{4}$$

error covariance matrix extrapolation,

$$\mathbf{P}_k^- = \mathbf{\Phi}_{k-1}\mathbf{P}_{k-1}^+\mathbf{\Phi}_{k-1}^T\mathbf{Q}_{k-1} \tag{5}$$

Kalman gain matrix,

$$\mathbf{K}_k = \mathbf{P}_k^-\mathbf{H}_k^T[\mathbf{H}_k\mathbf{P}_k^-\mathbf{H}_k^T + \mathbf{R}_k]^{-1} \tag{6}$$

state estimate update,

$$\hat{\vec{x}}_k^+ = \hat{\vec{x}}_k^- + \mathbf{K}_k[\vec{z}_k - \mathbf{H}_k\hat{\vec{x}}_k^-] \tag{7}$$

and, error covariance matrix update.

$$\mathbf{P}_k^+ = [I - \mathbf{K}_k\mathbf{H}_k]\mathbf{P}_k^- \tag{8}$$

Notice the similarity in form of the K-filter equations (2) and (3) to (1) of Pearl in that there is an additive random component to both the system model (2), and the measurement (3) which ripples through the other state estimator equations. The random variable in the system equation, (2), $\vec{w}(t)$, generally called the process noise, represents the unmodeled uncertainties (Pearl's latent variables) associated with the process dynamics including the random maneuvers of the target. The random variable in the measurement equation, $\vec{v}(t)$, represents the additive noise due to the fact that no observation is perfect and there are uncertainties associated with it.

As an example of using the state estimate produced by a K-filter to populate or update the parameters of a node in a BN, we present the simple BN of Figure 1. This network shows how the various uncertainties in the components of the kinematic state vector can affect the uncertainty in a situation assessment node which is not directly determined by the kinematic state.

In these most general K-filter equations (2) through (8) Φ_{k-1}, w_{k-1}, $\vec{z}_k$, $\mathbf{H}_k$, and v_k all have subscripts indicating that they are non-stationary and may change over time, indicating that the process and the resulting state estimates are not stationary. The system model propagates based on the previous state with a time-dependent random component added to it and a reduction in uncertainty based on noisy observations. It is important for our purpose here to note that if an observation of the system is not taken, then the uncertainty of the extrapolated state variable, $\hat{\vec{x}}_k^-$, as represented by the extrapolated error covariance matrix, $\mathbf{P}_k^-$, grows. The uncertainty in the extrapolated state estimate is represented by some norm of the error covariance matrix. The trace will not do as a norm as it is dimensionally nonconformal; the determinant, while dimensionally conformal, and monotonically related to information, does not have meaningful dimensions but may still be a useful norm. Alternatively, the error covariance matrix can be normalized to meaningful spatial units by pre- and post-multiplying by a dimension conforming matrix.

It can be seen from the error covariance extrapolation (5) that the state estimate (7) depends on the propagation of the previous state estimate (4) plus the Kalman gain (6), $\mathbf{K}(t)$, multiplying the difference between the previous estimate and the observation. If there is no observation (3), $\vec{z}(t)$, then the uncertainty continues to grow. This is our first hint that without continual observations of the state of a process corrupted by random process (latent variable) noise (2), our uncertainty about its state (5) grows, and hence its entropy.

Whether the elements of the K-filter are treated as stationary or non-stationary depends on the modeler's understanding of the process and how the model is to be used. One might consider the state propagation matrix in (2), $\mathbf{\Phi}_{k-1}$, representing the physics of the target's trajectory, to be stationary and unchanging over time yielding a constant $\mathbf{\Phi}$. If the same sensor is used to obtain a measurement of the target, then the observation matrix of (3) also becomes a constant, $\mathbf{H}_k = \mathbf{H}$. Even with this simplifying assumption, we still need to deal with the additive random components in (2) and (3), the process noise, w_k, and the measurement noise, v_k which are characterized in (2) and (3) by their covariance matrices, $\mathbf{Q}_k$ and $\mathbf{R}_k$, respectively. Typically target trackers include multiple model (e.g., IMM, Interacting Multiple Model [16]) methods with different process noise covariances $\mathbf{Q}_k$ at different stages in the target tracking process. Since $\mathbf{Q}_k$ is directly involved in the computation of the extrapolated error covariance matrix of (5), $\mathbf{P}_k^-$, which is used to compute the error covariance matrix update of (8), $\mathbf{P}_k^+$, the amount of information change associated with target observations under different model assumptions will change independently of the observation noise.

The measurement noise covariance matrix in (3), $\mathbf{R}_k$, also directly affects the amount of information associated with the computation of the state estimate update as well as the Kalman gain of (6), $\mathbf{K}_k$, which is used to compute the error covariance matrix update of (8), $\mathbf{P}_k^+$. Of course, the measurement covariance may change from observation to observation, but let's assume it is constant for the sake of discussion. The point here is to show that while the process and measurement noises can be considered stationary and hence aleatory, the resulting uncertainties in the target state updates as measured by the updated error covariance matrix are epistemic.

1.3. Aleatory and epistemic BN nodes

Concerning the modeler's view of the K-filter state estimate of a target in track as part of a BN representing the situation assessment, the random components can be viewed as either aleatory or epistemic. Continuing with our K-filter example which uses observation data of (3), $\vec{z}_k$, to reduce our uncertainty about the kinematic state of a target in track, we can look at the sources of randomness and see that they are, in the general model, non-stationary as they are all functions of time.

We do not discuss here the choices made by the modeler, but will, in our example later, show how evolving process dynamics of a target model affect the amount of information that one can extract from a measurement. We do recognize that in a situation assessment there will be a combination of both aleatory and epistemic nodes. Our concern with respect to a global entropy of a BN is limited to the epistemic nodes since there is no change in uncertainty in the aleatory nodes.

As we will see in the sequel, entropy changes reflect a gain or loss of information. We will extend this epistemic entropy change to a BN and see that there is a global gain or loss of information over time but only due to epistemic nodes.

Section 1 is the introduction which provides some necessary background information. Section 2 differentiates and makes clear the distinction among data, information, and knowledge. Section 3 investigates properties of hard (i.e., physics based) and soft (i.e., human-derived) data to conclude that there exist probabilistic measures which can be used to compute entropy of both hard and soft data. In section 4 we present sample computations which may help to clarify some of the newly introduced concepts.

2. DATA, INFORMATION, AND KNOWLEDGE

Waltz [17] distinguishes among three levels of abstraction of knowledge: data, information, and knowledge.

- Data are "individual observations, measurements, and primitive messages [which] form the lowest level. Human communication, text messages, electronic queries, or scientific instruments that sense phenomena are the major sources of data."
- Information is "organized sets of data…The organization process may include sorting, classifying, or indexing and linking data to place data elements in relational context for subsequent searching and analysis."
- Knowledge or foreknowledge (predictions or forecasts) is "information once analyzed, understood, and explained…"

For our purposes we take a slightly different approach by considering *information* to be a change in our uncertainty about processes in the environment which result from temporal changes, an observation, or the acquisition of relevant data (evidence). *Knowledge* in our context is expressed in the form of a Bayes net because the BN contains both the causal processes in the environment as well as the uncertainties associated with them. Furthermore, the fact that BN uncertainty increases over time is already known as Singhal & Brown [18] note in their discussion of dynamic Bayes nets, "[a] decay function is associated with the PDFs that increases the variance of the beliefs when they have not been updated for a period of time." They also recognize that observations do not have to be regular or synchronous. "To relax synchronization issues and constraints, we employ an asynchronous update policy that uses dynamic Bayesian networks to create new probability density functions (PDF)." [18]

If we consider the BN as the repository of knowledge about the situation, then the changes in uncertainty associated with the BN can be considered as information gain or loss. It is common to think about information gain as a result of obtaining data, but it is less common to think about the information loss associated with increases in our uncertainty about a process as the process evolves over time. Generally the BN is considered to be stationary until more data, i.e., evidence, are obtained to decrease the remaining uncertainty, but that is not the case when we are dealing with processes. As we saw in the K-filter target tracking example previously presented in Figure 1, our uncertainty about the kinematic state estimate grows as time advances if we do not observe the process due to the additive process noise. In the case of K-filters, the loss of information can be computed as a change in the entropy of our state estimate over time [4]. If we make an observation and obtain data, then the difference in uncertainties represented by the entropies of a norm of the extrapolated error covariance matrix (5), $\mathbf{P}_k^-$, and the norm of the error covariance update (8), $\mathbf{P}_k^+$, is a measure of the amount of information gain. This information gain represents the increase in knowledge as a result of obtaining data and decreases the uncertainty in the BN.

2.1. Bayes net information

We can extend this concept from the kinematic state of a single target in track to the global knowledge of a situation as represented in a BN [18]. Situation information and sensor information are differentiated by the authors [19], and we only focus here on situation information as represented by the global change in uncertainties in a BN, namely entropy changes among the situation assessment nodes. The Shannon entropy [20], H, of a discrete random variable, X, with possible values $\{x_1, x_2, \ldots x_n\}$ and probability mass function $P(X)$ is computed in the usual manner as

$$H(X) = \mathcal{E}[I(X)] \tag{9}$$

$$H(X) = \mathcal{E}[-\ln(P(X))] \tag{10}$$

with $\mathcal{E}$ being the expectation operator and I being the information content of the RV. Alternatively, and letting the RV, X, be a BN node, N_j, the node entropy can be computed as follows

$$H(X) = \sum_{i=1}^{n} P(x_i) I(x_i) \tag{11}$$

$$H(N_j) = -\sum_{i=1}^{n} P(x_i) \log_b P(x_i) \tag{12}$$

here n is the number possible values that can be taken on by a single node of a BN, N_j, and b is the radix of the logarithm used. Utilizing the radix 2 yields the entropy measured in bits.

After an observation which changes the probability distribution associated with node N_j, the single node BN information, I^{+j}, [19] is, therefore,

$$I^{+}(N_j) = H^{-}(N_j) - H^{+}(N_j) \quad (13)$$

where j is the index number of a single node of the BN, the superscript "+" indicates the values associated with the jth node after the effects of the observation have changed the probability within the jth node (the *a posteriori* value), and the superscript "−" reflects the probability of the jth node before the observation (the *a priori* value).

The global BN information gain or loss is the sum of the information gain/loss of all the nodes. Since we assume a mixture of aleatory and epistemic nodes in the BN, and furthermore that the aleatory nodes are stationary, there is no information gain/loss associated with them. That is, we only have to sum the information gain/loss of epistemic nodes. Without observations, there will be a net increase in our uncertainty of each process node with an associated loss of information over time. Assuming mutually exclusive sensor observation opportunities, there may be either a net global gain or net global loss of information in the situation assessment BN with the observation of a single node. For the observed node, there may be either a gain or a loss of information based on whether the observation decreases the uncertainty more than it had increased since the last observation. For the non-observed process nodes there may be a loss of information since they are not being observed and their uncertainty may have grown since their last observation.

The global temporal Bayes net information at the kth observation is

$$I_k^{\text{TBN}} = \sum_{\substack{\text{all epistemic} \\ \text{nodes}}} [H_k - H_{k-1}] \quad (14)$$

which can be reduced to

$$I_k^{\text{TBN}} = \sum_{j-1}^{m} I^{+}(N_j) \quad (15)$$

where m is the number of epistemic nodes in the BN and $I^{+}(N_j)$ is defined in (13).

2.2. Hard/soft knowledge and information

The implication until now in this paper is that the situation assessment in the form of a BN represents only kinematic uncertainty of processes in the domain of concern. We take a more egalitarian view of situation assessment in that nodes can represent kinematic uncertainties as well as *intentional* (not purposeful, but rather motivational as in the node *contact_A_is_threat* of Figure

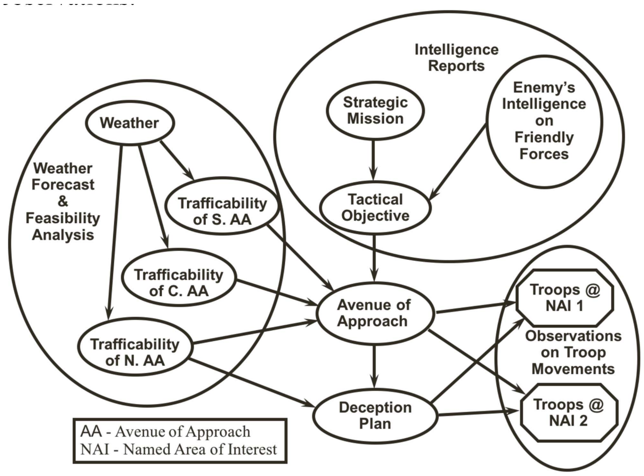

Fig. 3. Relational diagram representing enemy intent. After Buede, et al. [22].

1) uncertainties about hypotheses such as whether the enemy is going to attack or not. Hypothesis nodes like this can be partially resolved if there is overt physical action or observable preparatory action on the part of the enemy; however, it is more likely that actionable intelligence is derived from the interception and analysis of communications intelligence (COMINT) or other automatically processed natural language communications, i.e., soft data. According to Dragos, [21] "[s]oft data are a mix of both facts and opinions" the difference between the two being the source and the probability associated with each.

Yet it doesn't matter whether the source of data is hard or soft, but rather whether the acquisition of data changes our uncertainty about a particular aspect of the situation being assessed as reflected by a changed probability in one or more nodes of the BN.

2.3. Changing BN structure, information gain/loss?

Since we consider a BN as a knowledge representation structure with uncertainties associated with each node, the addition of another node, be it aleatory or epistemic, does not inherently add any *information* unless the addition of the node connects to other nodes which are conditioned on it. The addition of a node may affect the amount of temporal information that is gained or lost as time progresses or observations are made since the BN information is the sum of the information lost over time and likely regained with observations.

But the question arises as to how much *knowledge* is contained in a BN and whether adding or deleting a node changes that amount of knowledge. If we take the entropic view of uncertainty and the information theoretic view of information being a change in entropy, then we can consider the maximum uncertainty, the total entropy, in a BN to be the sum of the entropies of all the nodes *as if* each of these entropies were at its maximum uncertainty. If we measure the entropy of a node, we can view this as its *potential information (PI)* since

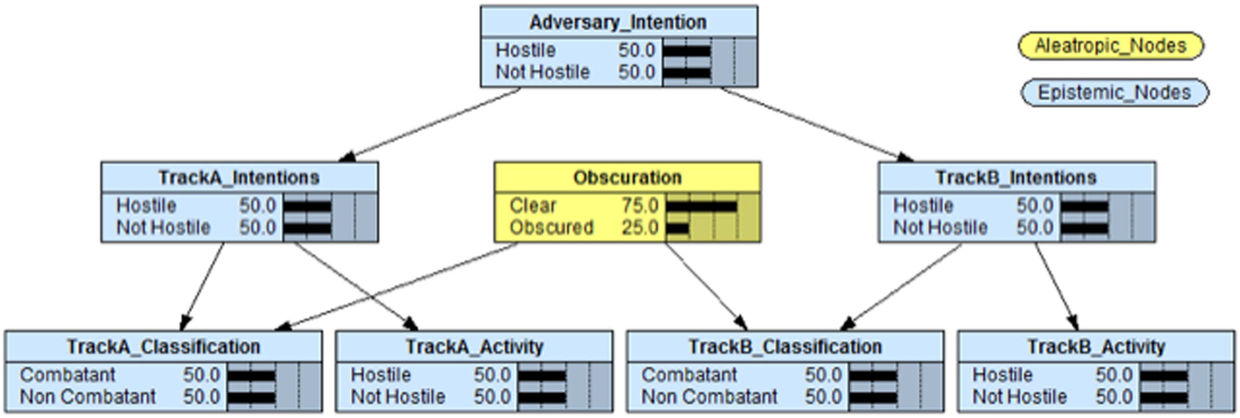

Fig. 4. Example BN containing uniform probabilities (no evidence) and both aleatory (yellow) and epistemic (blue) nodes, $KEn = 7$.

it represents the amount of uncertainty that can be resolved through measurements if it is an epistemic node. The maximum PI of a BN node occurs when all values are equally probable and represents the maximum information that can be obtained from a BN node resulting in perfect knowledge of that node since the entropy of a node with no uncertainty is zero. If we extend this to the BN itself, then we can talk about the maximum potential information of a BN as the sum of the maximum PIs of the individual nodes. Note, however, that the probabilities of a node may not be at their maximum uncertainty since we may have some *a priori* knowledge which skews the probabilities. In this case, the PI is the entropy of that skewed distribution associated with the node.

Since we have made a distinction in this paper between aleatory and epistemic nodes, we need to define which nodes need to be included in our definition of potential information. We assert that only epistemic nodes should be included as, by definition, the uncertainty in aleatory nodes cannot be reduced by measurements and those in epistemic nodes can. The decision between which ones are epistemic nodes and which ones are aleatory nodes is a modeling decision and can change with point of view and over time and there is no general rule that can be applied other than whether observations of any other node in the network changes a node's probabilities.

We also can view the observing of a node to reduce its uncertainty as gaining information about the node. We call this *kinetic information (KI)* because it results from a physical or cyber action and a change in the BN as well as a reduction in the potential information yet available to be gained. We can actively observe the process associated with a node to obtain kinetic information. Alternatively, by not observing a random variable related to a dynamic process represented by a node, the BN can leak KI over time which increases its PI.

Currently, there is no unit for the uncertainty of knowledge in a BN. We propose to use units of *Knowledge Entropy (KEn)* to represent the uncertainty in a BN. Zero KEn results when there is no uncertainty in any of the epistemic nodes. There is precedence for this new use of an old (if not archaic) word if one examines definitions found in the Oxford English Dictionary (Oxford English Dictionary, 2017) which defines *ken* as:

- *ken*, v.1, 11. a. To know (a thing); to have knowledge of or about (a thing, place, person, etc.), to be acquainted with; † to understand. Now chiefly *Sc*.
- *ken*, v.1, 12. a. *intr.* or *absol*. To have knowledge (*of* or *about* something). † Also with inf.: To know how to, to be able to (*obs.*).

So we can refer to the KEn of a BN at any time as measured in bits of uncertainty in our situation knowledge. The KEn can change over time due to the leakage of KI or the acquisition of KI through observations, and can be computed as the sum of the entropies of all epistemic nodes in the BN. Formally, the knowledge entropy of a BN, KEn, is

$$KEn(t) = \sum_{\substack{\text{all epistemic} \\ \text{nodes}}} H(t) \tag{16}$$

and the amount of temporal Bayes information, TBI, which results from a change in nodal probabilities or network structure from time t_0 to t_1, is

$$TBI(t_1) = KEn(t_0) - KEn(t_1) \tag{17}$$

As previously mentioned, TBI may be zero, positive, or negative. Zero TBI means that no network information was gained or lost over the time period although there could be individual nodal information changes whose net sum is zero. Positive or negative TBI indicates a gain or loss of information respectively with a concomitant change in our situation knowledge as represented by the temporal BN.

One would like to hypothesize that there is a conservation of knowledge law associated with BN, i.e., the *KNowledge Entropy (KEn)* is conservative, and that there is a one-for-one exchange between KI and PI, but this does not appear to be the case due to the conditional probabilities relating nodes. Increasing or decreasing the

TABLE 1.
Entropy of uniformly distributed probabilities based on the number of bins.

# of bins in X	Uniform probability	$\log_2(p_i)$	$PI = H(X)$ (bits)
2	0.500	−1.000	1.000
4	0.250	−2.000	2.000
5	0.200	−2.322	2.322
8	0.125	−3.000	3.000

TABLE 2
Conditional Probability Table (CPT) of the TrackA_Classification node.

TrackA_Int	Obscured	Combatant	NonCombatant
Hostile	Clear	75	25
Hostile	Obscured	50	50
Not Hostile	Clear	25	75
Not Hostile	Obscured	50	50

uncertainty in one node may increase or decrease the entropy in other nodes, but our preliminary investigations lead us to conjecture that the gain in KI is not offset by an equal loss in PI. This is a topic that bears further investigation but is not the main point of this paper so we leave it for now.

$$KEn_{\text{total}} \neq \sum PI + \sum KI \tag{18}$$

We can now answer the question about what to do about a node which is added to, or deleted from, a dynamic BN and that is to simply consider it to be adding or deleting potential information to or from the knowledge represented in the BN. Adding a node will increase the KEn, but by less that the PI of the node if its connections affect the other probabilities. Furthermore, the loss of information, KI, results in an increase in the PI although not on an equivalent basis. The acquisition of information, KI, by observing a relevant RV decreases the PI. The total knowledge, the KEn, in a structurally stationary BN is knowable and computable. As with the question of conservation of information in a BN, we defer this topic to a later paper.

3. BN HARD/SOFT INFORMATION

We have already described the uncertainty associated with kinematic state estimates utilizing the K-filter formulation. This epistemic uncertainty is fully described by the increase in a norm of the error covariance matrix as it propagates over time or decreases with an observation. Other similar physical state estimates have continuous or discretized uncertainties that are straightforward to work with. Soft knowledge in BN, on the other hand, requires additional explanation since it includes other forms of uncertainty as noted by Dragos [21] namely

- Intrinsic uncertainties such as ambiguity, vagueness, and precision
- Source related uncertainties which are a mixture of facts and opinions
- Relational uncertainties which are concerned with inaccuracies, overlappings, and contradictions in analysis

Dragos [21] continues with methods for estimating all of these uncertainties which will not be repeated here. For our purposes, we will assume that soft uncertainties can be estimated allowing us to compute entropies of soft data.

The suitability of uncertainty in any form…hard or soft, social or physical, quantitative or fuzzy…has been shown by Kjaerulff [23] to be applicable to formulation as a BN. "Note, that the method in this paper can be applied to other evidential frameworks where independent pieces of evidence are combined into a joint evidence e.g., Bayesian combination. For more high-dimensional problems, i.e., when it could be more suitable to utilize a graph structure for modeling dependencies e.g. Bayesian Networks,…" That is, relations among aleatory and epistemic processes such as the example of Buede et al., [24] as shown in Figure 2, can be represented in a causal BN, also from Buede, et al., as shown in Figure 3.

4. EXAMPLE GLOBAL TEMPORAL BN COMPUTATIONS

The following simplified examples will demonstrate some of the concepts introduced in this paper. First, looking at an individual node and computing the entropy of a uniform distribution of discrete values in accordance with (11), we see as exemplified in Table 1 that the potential information of the jth node is simply the $\log_2$(*number of bins in* jth *node*) and the potential information of the BN consisting of m nodes each having k_j values associated with each nodes.

$$PI(N_j) = \sum_{j=1}^{m} -\log_2 k_j \text{ bits} \tag{19}$$

If each node were a true/false hypothesis node, then there would be one bit of PI/node resulting in an m-node BN containing an upper bound of m-bits of PI since the connectivity of the BN reduces the actual amount of information that is available. Clearly as one changes the number of uniformly distributed bins/node, the summation of PI is easily calculated as well as the amount of PI if a node is added or deleted.

4.1. Simple BN PI and KI example for epistemic nodes

In order to instantiate some of the concepts introduced here, we perform PI, KI, and KEn com-

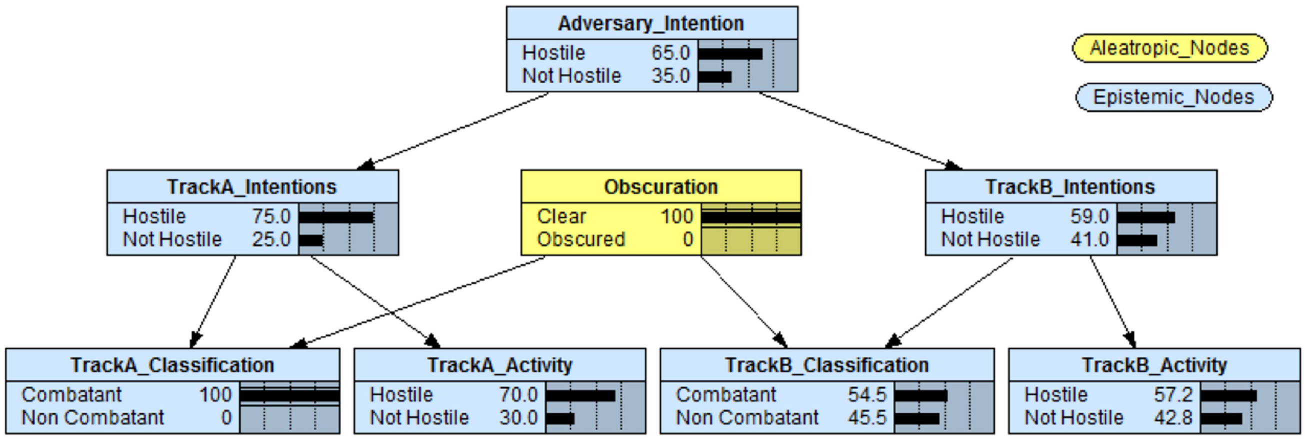

Fig. 5. Classification evidence at time T_0.

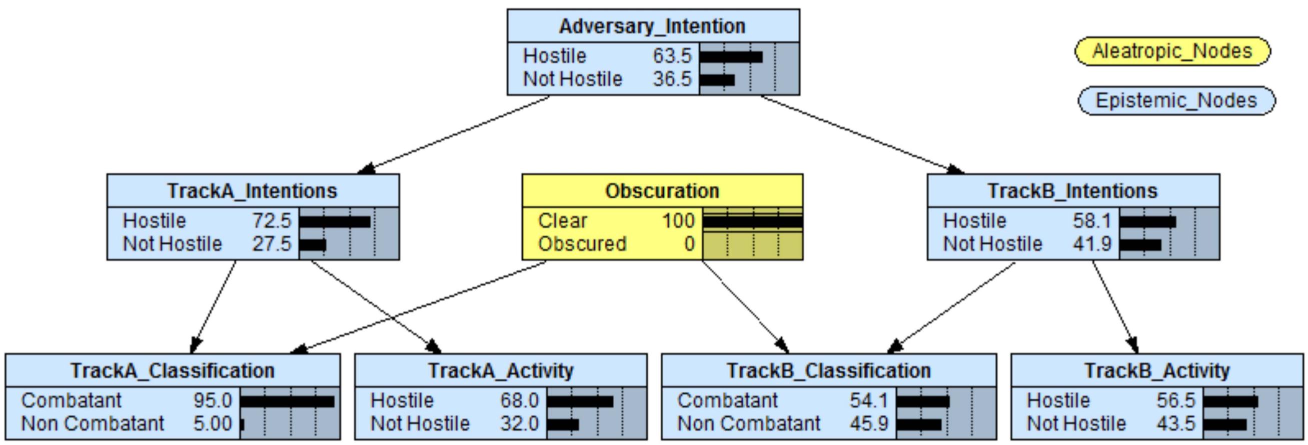

Fig. 6. Classification evidence at time T_{30} showing a change in the BN's initial knowledge with a KEn = 5.58 and the result of a loss of information in the TrackA_Classification node resulting in an increase in KEn to 5.95.

putations on a simple 8-node BN as shown in Figure 4, with all the nodes being epistemic and uniform probability nodes, except one (Obscuration) which is aleatory. Associated with the TrackA_Classification and TrackB_Classification nodes is a Conditional Probability Table (CPT) as shown in Table 3.

For our numerical example, we use the sensitivity as computed in Norsys Netica [25] BN program. The mathematical formulations utilized by Netica are documented in their on-line documentation [26]) and are the same as (19) above. Furthermore, we have done sample calculations outside of Netica utilizing the net of Figure 4 without the aleatory "obscurations" node and the results of our calculations match those produced by the Netica sensitivity analysis.

Referring to Figure 4, the initial entropy of the Adversary_Intention node at time t_0 is the expected 1.0 bits with uniform distributions in the other two nodes. If the TrackA_Classification node of Figure 4 is set to 100% as shown in Figure 5 and Figure 6, the Adversary_Intention changes to 65% Hostile/35% Non-Hostile and the BN KEn changes from 7 to 5.58 indicating a global network decrease in uncertainty (or increase in KI) of 1.42 bits as a result of the sensing action which provided the Combatant classification with 100% certainty.

We demonstrate the temporal increase in uncertainty in the TBN by changing the probabilities of the TrackA_Classification node. At some later time, t_{30}, we assume the uncertainty has decreased from 100% Combatant to 95% Combatant/5% Non-Combatant. This temporal loss of information results in the Adversary_Intention changing to 63.5% Hostile/36.5% Non-Hostile and the BN KEn increasing from its t_0 of 5.58 to its t_{30} value of 5.95 of a KI loss of 0.07 bits.

As an example of an alternative type of information loss related to a different sensing node, TrackA_Activity, Figure 7 and Figure 8 shows the BN of Figure 4 with initial, non-uniform uncertainties in the TrackA_Activity node of 100% Hostile/0% non-hostile. With this initial condition, the Adversary_Intention becomes Hostile 74%/Non-Hostile 26% for an initial KEn at t_0 of 5.06 bits.

We model the temporal change in our certainty of the TrackA_Activity node by decreasing the uncertainty at some later time, t_{30}, from 100% Hostile to 80%/20% as shown in Figure 7 and Figure 8. This temporal loss of information results in the Adversary_Intention changing

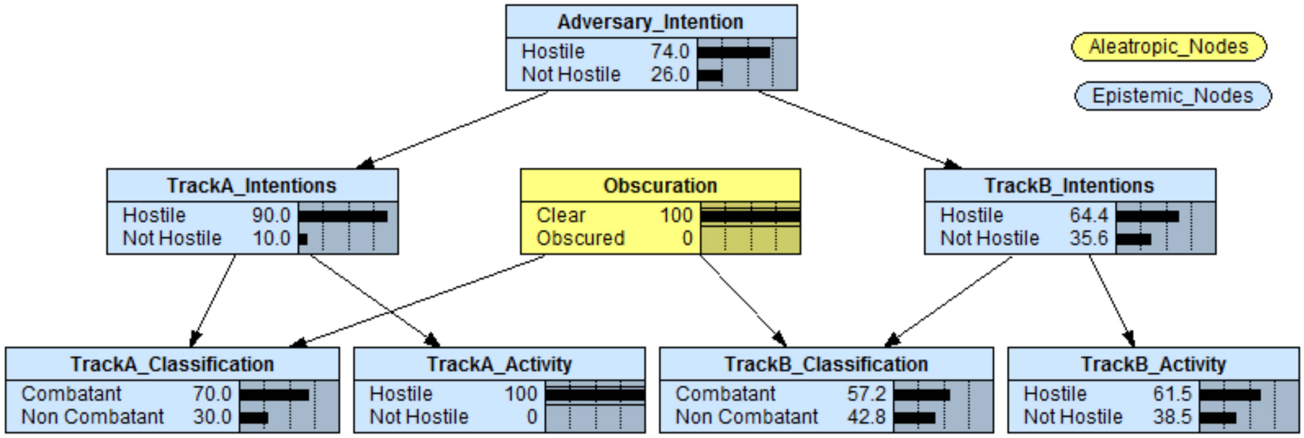

Fig. 7. Activity evidence at time T_0.

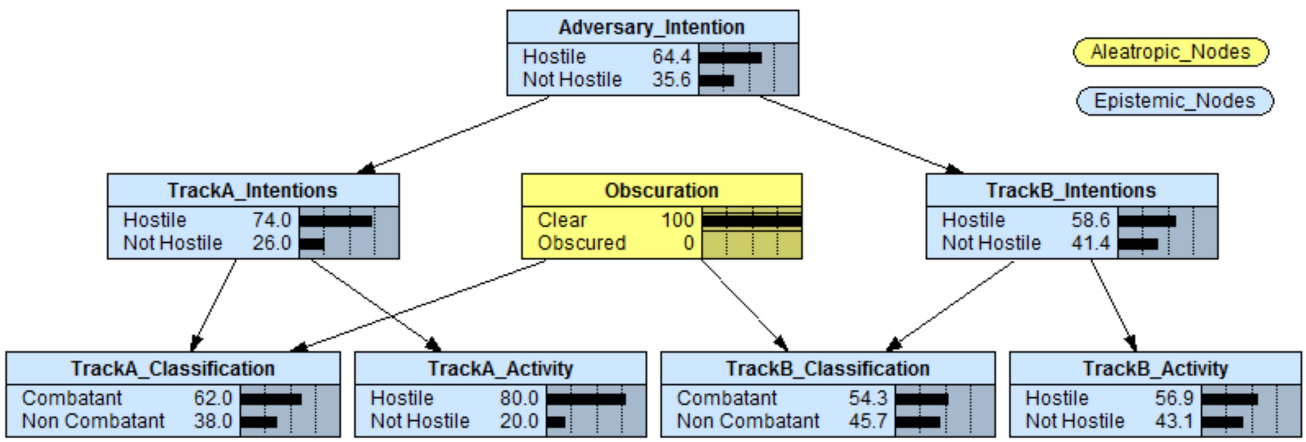

Fig. 8. Activity evidence at time T_{30} showing a change in the BN's initial knowledge with a KEn = 5.06 and the result of a loss of information in the TrackA_Activity node resulting in an increase in KEn to 6.41.

TABLE 3
Table summarizing the results of information loss over time due to decreased uncertainty in classification and, alternatively, identity. No obscuration in the aleatory node.

Scenario with no obscuration	Total Knowledge Entropy (epistemic only)
No evidence	7.00
Classify TrackA, t_0	5.58
Classify TrackA, t_{30}	5.95
Identify Activity TrackA, t_0	5.06
Identify Activity TrackA, t_{30}	6.41

to 64.4% Hostile/35.6% Non-Hostile and the BN KEn increasing from its t_0 of 5.06 to its t_{30} value of 6.41 of a KI loss of 1.35 bits. The loss in KEn with time is summarized in the table of Table 3.

This example shows that if we were to use a BN with no evidence and a KEn of 7.0 with the expected information gains at t_0 of 1.42 bits if we choose to classify as opposed to 1.94 bits if we choose to identify, we would choose to classify since it yields the maximum information. If, on the other hand, if we chose to wait until t_{30} the expected information gain from the initial KEn of 7.0 would yield 1.05 bits for classify and 0.59 bits for identify showing that accounting for the temporal loss of information from t_0 to t_{30} results in a different choice of which sensor function to use.

Other findings have been computed which result in higher losses of information while most result in a positive flow of KI into the BN.

4.2. Information in the presence of aleatory node

Remembering that the computation of KEn only includes epistemic nodes, the question arises as to the effect of an aleatory node on the amount of information gain and choice of sensor function if one makes different assumptions about the probabilities in an aleatory node. If the aleatory Obscuration node is changed from its 100% clear as used for the previous example to 75% clear/25% obscured, the following results. The results are shown in Table 4.

Referring to Figure 4, the initial entropy of the Adversary_Intention node at time t_0 is still the expected 1.0 bits with uniform distributions in the other two nodes. If the TrackA_Classification node is set to 100%, the Adversary_Intention changes to 61.3% Hostile/38.8% Non-Hostile and the BN KEn changes from 7 to 5.77

TABLE 4
Table summarizing the results of information loss over time due to decreased uncertainty in classification and, alternatively, identity. Twenty-five percent obscuration in the aleatory node.

Scenario with 25% obscuration	Total Knowledge Entropy (epistemic only)
No evidence	7.00
Classify TrackA, t_0	5.77
Classify TrackA, t_{30}	6.10
Identify Activity TrackA, t_0	5.12
Identify Activity TrackA, t_{30}	6.43

indicating a global network decrease in uncertainty (or increase in KI) of 1.23 bits as a result of the sensing action which provided the Combatant classification with 100% certainty.

As before, we model the temporal change in our certainty of the TrackA_Classification node by decreasing the uncertainty at some later time, t_{30}, from 100% Combatant to 95%/5%. This temporal loss of information results in the Adversary_Intention changing to 60.1% Hostile/39.9% Non-Hostile and the BN KEn increasing from its t_0 of 5.77 to its t_{30} value of 6.10 of a KI loss of 0.33 bits.

As an example of an alternative type of information loss under aleatory uncertainty related to a different sensing node, the TrackA_Activity, node is changed to 100% Hostile/0% non-hostile. With this initial condition, the Adversary_Intention becomes Hostile 74%/Non-Hostile 26% for an initial KEn at t_0 of 5.12 bits.

Again, modeling the temporal change in our certainty of the TrackA_Activity node by decreasing the uncertainty at some later time, t_{30}, from 100% Hostile to 80%/20%. This temporal loss of information results in the Adversary_Intention changing to 64.4% Hostile/35.6% Non-Hostile and the BN KEn increasing from its t_0 of 5.12 to its t_{30} value of 6.42 of a KI loss of 1.30 bits. The loss in KEn with time is summarized in Table 4.

This aleatory example, Table 4, shows that if we were to use a BN with no evidence and a KEn of 7.0, choosing to classify would yield an expected information gains at t_0 of 1.23 ($7.0 - 5.77$). If, instead, choosing to identify would yield 1.88 bits ($7.0 - 5.12$). Because of this expected differential information gain, would choose to identify since it yields the maximum information. If, on the other hand, we choose to wait until t_{30} the expected information gain from the initial KEn of 7.0 would yield 0.9 bits ($7.0 - 6.1$) for classify and 0.57 bits ($7.0 - 6.43$) for identify, leading us to choose to classify as the maximum information choice. That is, accounting for the temporal loss of information from t_0 to t_{30} results in a different choice of which sensor function to use in order to maximize the information gain for a single observation.

For this example of changes in our *a priori* assumption about the probabilities associated with an aleatory node, there is a change in the expected information gain even though the entropy of the aleatory node is not included in the information measure. This shows that our model assumptions about the unmeasurable causal probabilities can affect our choice of sensing actions since they may affect our expected situation information expected value rate ($EIVR_{sit}$).

5. CONCLUSION

The differentiation between aleatory and epistemic nodes in Bayes nets has been defined and illustrated. It is also shown that BN are not limited to hard data as the analogy to Kalman-filter shows, but soft data nodes can be included since there exist soft data entropy measures. The fact that both hard and soft data uncertainty measures can be expressed as entropies allows one to put the two types of knowledge in the same BN and apply information measures based on entropy changes. Potential information and kinetic information are defined and it is conjectured that a conservation of knowledge law exists, but the details of this will be the subject of further research. Finally, a simple example of a temporal BN was presented showing how the leakage of information over time could lead to increases in entropy over time which could affect the choice of expected situation information gain when utilizing an information based sensor measurement (IBSM) approach to sensor management.

REFERENCES

[1] U. Kjaerulff
"A computational scheme for reasoning in dynamic probabilistic networks,"
in *Eighth Conference on Uncertainty in Artificial Intelligence*, 1992.

[2] K. C. Chang and W. Sun
"Performance modeling for dynamic Bayesian networks,"
in *Signal Processing, Sensor Fusion, and Target Recognition XIII*, Orlando, FL, 2004.

[3] J. Pearl
Probabilistic Reasoning in Intelligent Systems: Networks of Plausible Inference,
San Francisco, CA: Morgan Kaufmann, 1988.

[4] K. J. Hintz and E. S. McVey
"Multi-process constrained estimation,"
IEEE Transactions on Systems, Man, and Cybernetics, vol. 21, no. 1, pp. 237–244, 1991.

[5] K. Hintz and I. Kadar
"Improving ORM utilizing implicit collaboration & context sensitive fusion,"
in *Proceedings International Conference on Information Fusion, Fusion2015*, Washington, DC, 2015.

[6] E. N. Ciftcioglu, A. Yener and M. J. Neely
"Maximizing quality of information from multiple sensor devices: the exploration vs exploitation tradeoff,"
IEEE Journal of Selected Topics in Signal Processing, vol. 7, no. 5, pp. 883–894, 2013.

[7] R. L. Winkler
"Uncertainty in probabilistic risk assessment,"
Reliability Engineering and System Safety, vol. 54, pp. 127–132, 1996.

[8] P. C. G. Costa, K. B. Laskey, E. Blasch and A.-L. Jousselme "Towards Unbiased Evaluation of Uncertainty Reasoning: The URREF Ontology," in *15th International Conference on Information Fusion, Fusion2012*, Singapore, Singapore, 2012.

[9] G. Shafer A Mathematical Theory of Evidence, Princeton: Princeton University Press, 1976.

[10] Oxford University Press "OED Online," March 2018. [Online]. Available: http://www.oed.com/view/Entry/4795?redirectedFrom=aleatory. [Accessed 04 June 2018].

[11] Oxford University Press "OED Online," March 2018. [Online]. Available: http://www.oed.com/view/Entry/63541?rskey=rMvb1N&result=1&isAdvanced=false. [Accessed 04 June 2018].

[12] A. Der Kiureghian and O. Ditlevsen "Aleatory or epistemic? Does it matter?," *Structural Safety*, vol. 31, pp. 105–112, 2009.

[13] N. A. Abrahamson "5: Aleatory Variability and Epistemic Uncertainty," [Online]. Available: http://www.ce.memphis.edu/7137/PDFs/Abrahamson/C05.pdf. [Accessed 20 January 2018].

[14] J. Pearl Causality, Models, Reasoning, and Inference, 2nd Edition, New York: Cambridge University Press, 2009.

[15] A. Gelb Applied Optimal Estimation, Cambridge, Massachusetts: MIT Press, 1974.

[16] E. Mazor, A. Averbuch, Y. Bar-Shalom and J. Dayan "Interacting multiple model methods in target tracking: a survey," *IEEE Transactions on Aerospace and Electronic Systems*, vol. 34, no. 1, pp. 103–123, 1998.

[17] E. Waltz Knowledge Management in the Intelligence Enterprise, Boston: Artech House, 2003.

[18] A. Singhal and C. Brown "Dynamic Bayes Net Approach to Multimodal Sensor Fusion," in *Sensor Fusion and Decentralized Control in Autonomous Robotic Systems*, Pittsburgh, PA, 1997.

[19] K. J. Hintz and S. Darcy "Valued situation information in IBSM," in *20th International Conference on Information Fusion, Fusion2017*, Xi'an, China.

[20] C. E. Shannon "A Mathematical Theory of Communication," *Bell System Technical Journal*, vol. 27, pp. 623–657, 1948.

[21] V. Dragos "An ontological analysis of uncertainty in soft data," in *16th International Conference on Information Fusion, Fusion2013*, Istanbul, Turkey, 2013.

[22] Oxford English Dictionary "OED," Oxford University Press, 2017. [Online]. Available: http://www.oed.com. [Accessed 03 February 2018].

[23] H. J. Steinhauer, A. Karlsson and S. F. Andler "Traceable uncertainty," in *16th International Conference on Information Fusion, Fusion 2013*, Istanbul, Turkey, 2013.

[24] D. M. Buede, J. A. Tatman and T. A. Bresnick *Introduction to Bayesian Networks, Monterey*, CA, 1998.

[25] Norsys Software Corp. "Netica Application," 2018. [Online]. Available: https://www.norsys.com/netica.html. [Accessed 06 June 2018].

[26] Norsys "Sensitivity Equations, Entropy Reduction," [Online]. Available: https://www.norsys.com/WebHelp/NETICA/X_Sensitivity_Equations.htm#entropy. [Accessed 25 June 2018].

Kenneth J. Hintz (BSEE '67–MSEE '79–PhD '81) Dr. Hintz has been an Associate Professor in the Department of Electrical and Computer Engineering at George Mason University since 1987. He designed and established the Bachelor and Masters in Computer Engineering Degree Programs at GMU and teaches courses in sensor engineering, image processing, and computer engineering.

Dr. Hintz' current research interest is in Orchestrated Resource Management (ORM) and Information Based Sensor Management (IBSM) supported by the Naval Postgraduate School. He also developed a new method for pre-shot detection of barreled weapons based on his discovery of cavity induced modulation (CIM). Additional signal processing developments include syntactic pattern recognition applied to ground penetrating radar images for landmine detection and classification. His other research interests are in information-based real-time sensor management, X-ray, thermal, and visual, image processing, as well as algorithms for real time signal processing.

Before joining GMU, Dr. Hintz was with the Naval Surface Warfare Center, Dahlgren, VA, working in electronic warfare and radar signal processing where he designed and built the original AN/ULQ-16 pulse analyzer. Prior to working at NSWC, Dr. Hintz was with the U. S. Navy as a designated Naval Aviator stationed for 3 years in Rota, Spain flying Electronic Warfare Reconnaissance with Fleet Air Reconnaissance Squadron Two (VQ-2). During that time he became designated Electronic Warfare Aircraft Commander (EWAC) in both the EC-121 and EP-3E aircraft. Dr. Hintz holds 24 patents, 6 patents pending, is a Fellow of SPIE, a Senior Life Member of IEEE, and lead author on a book on Microcontrollers. He received his B.S. degree in Electrical Engineering from Purdue University, West Lafayette, Indiana in 1967 and his M.S. and Ph.D. degrees in Electrical Engineering from the University of Virginia in 1979 and 1981 respectively.

Steven Darcy is an analyst with Innovative Decisions Inc. He has extensive experience as an OR analyst and software developer. He has helped develop over a dozen significant computer models. He is a Ph.D. candidate in the Systems Engineering and Operations Research Department of George Mason University. He has a BA in Physics from Rutgers University and a MS in Operations Research from George Mason University.