BAYESIAN BELIEF NETWORKS TO INTEGRATE MONITORING EVIDENCE OF WATER DISTRIBUTION SYSTEM CONTAMINATION

BY

WESLEY J. DAWSEY

THESIS

Submitted in partial fulfillment of the requirements
for the degree of Master of Science in Environmental Engineering in Civil Engineering
in the Graduate College of the
University of Illinois at Urbana-Champaign, 2011

Urbana, Illinois

Adviser:
Professor Barbara Minsker

ABSTRACT

A Bayesian belief network (BBN) methodology is proposed for combining evidence to better characterize contamination events and reduce false positive sensor detections in drinking water distribution systems. A BBN is developed that integrates sensor data with other validating evidence of contamination scenarios. This network is used to graphically express the causal relationships between events such as operational changes or a true contaminant release and consequent observable evidence in an example distribution system. In the BBN methodology proposed here, multiple computer simulations of contaminant transport are used to estimate the prior probabilities of a positive sensor detection. These simulations are run over multiple combinations of possible source locations and initial mass injections for a conservative solute. This approach provides insight into the effect of uncertainties in source mass and location on the detection probability of the sensors. In addition, the simulations identify the upstream nodes that are more likely to result in positive detections. The BBN incorporates the probabilities that result from these simulations, and the network is updated to reflect three demonstration scenarios – a false positive and two true positive sensor detections.

To my wife and family

ACKNOWLEDGEMENTS

I would like to thank the many people who assisted and supported me in this project including my advisor Professor Barbara Minsker, Vicki Van Blaricum and Tim Perkins of the Army Corps of Engineers Construction Engineering Research Laboratory, and the members of Professor Minsker's research group. These people were instrumental in helping me to develop the research that is presented in this thesis. I would also like to thank the faculty of the Civil and Environmental Engineering Department at the University of Illinois for their dedication to the many graduate students that pass through their offices and classrooms. Finally, I am deeply thankful for the support and patience of my wife.

TABLE OF CONTENTS

INTRODUCTION

There is a great deal of uncertainty in real time characterization of water distribution system contamination events. Much of this uncertainty is due to the lack of targeted sensors, which makes it necessary to use surrogate water quality parameters to indirectly measure the presence of a contaminant. A positive sensor detection can often be validated by other pieces of evidence observed in a distribution system. This paper proposes a Bayesian belief network (BBN) methodology for integrating evidence associated with a potential contamination event.

Drinking water utilities face many challenges in recognizing, characterizing, and responding to contamination events. Chief among these is coping with the daunting number of dangerous chemical, biological, and radiological substances that may be accidentally or purposefully introduced into a drinking water system. Furthermore, real-time analytical tools for characterizing many of these contaminants in-situ do not currently exist, or are prohibitively expensive (ASCE, 2004). As a result, current online contaminant monitoring system design makes use of surrogate water quality measures such as total organic carbon (TOC), turbidity, pH, chlorine concentration and others which may be correlated to "fingerprint" contamination events (ASCE, 2004).

False positive sensor detections pose an important threat to online contaminant monitoring even when sensors with relatively 'good' false positive rates are installed. For example, assume that the probability of a contamination event was known a priori to be 0.01%. If a sensor with a 0.3% false positive rate were to take 10,000 measurements, it would return 30 false positives and only 1 true positive. This problem will occur for any case where the true probability of an occurrence is much less than the false positive rate of the sensor. In reality, the true probability of contamination is likely to be much less and the true false positive rates much greater than this example (ASCE, 2004). When considered within the context of an entire water collection, treatment, and distribution system, it is apparent that there are numerous other system events that could obscure or validate such a contaminant detection. Related pieces of information such as physical security alarms, distribution system model topology, and contaminant measurements by sensors need to be tied in to expert knowledge of potential contamination scenarios. A BBN is a useful framework for representing the causal relationships between events and observations that comprise a contamination scenario.

LITERATURE REVIEW

Bayesian methodologies, such as BBNs used in this project, have been used to combine diverse data inputs for numerous applications including battlefield strategy, optical recognition, fault detection, advanced driver assistance systems, and sensor network data fusion (Sanzotta and Sherrill, 1997; Fox et al., 2003; Liu and Zhang, 2002; Coue et al., 2003; Karlsson et al., 2002). Bayesian networks have been used to fuse data from multiple sensor networks in many applications (Bonci et al., 2002; Beckerman, 1992; Brown et al., 1995). In the energy domain, Bayesian modeling methods were used to analyze the distribution of the failure rate at nuclear power plants (Chu, 1995). Schlumberger et al. (2002) used coupled dynamic models with a BBN to assess the voltage stability limits for part of France's subsystem and to identify more efficient rules of operation.

In the battlefield strategy application area, significant research has been conducted to apply Bayesian networks and decision trees to support battlefield decision-making. The CoRaven system described by Jones et al. (1998) and Hayes et al. (2000) used a BBN structure to make inferences from data observations to a commander's information requirements. Franzen (1999) modeled the decision structure of battle damage assessment (BDA) within a BBN. Das et al. (2002) presented an approach to battlefield situation assessment based on the real-time combination of small Bayesian network components to form a BBN for a specific high-level scenario. Therrien (2002) used a BBN to model human and environmental parameters influencing risk assessment and stress in combat scenarios. Information sources for the BBN included observations, training, orders, and reports.

Substantial research in the Environmental Engineering domain has utilized Bayesian approaches for a number of applications, but few studies have used BBNs specifically and none have involved applications to water supply protection as illustrated here. In groundwater remediation, BBNs coupling an expert knowledge base with process models have been used to evaluate the potential of naturally occurring reductive dechlorination at sites contaminated with TCE (Stiber et al., 1999, 2004a, 2004b). Marcot (2001) combined expert knowledge with ecological data in a BBN to model the causal relationships between planning decisions and impacts on at-risk wildlife species habitats. Stow et al. (2003) compared a BBN approach with two deterministic models for predicting the effect of nitrogen loading on estuarine *chlorophyll a*

concentrations. Murray et al. (2005) predicted health risks associated with a contamination event using a dynamic disease model linked with a distribution system flow and transport model.

In the field of drinking water supply, several researchers have focused on characterizing intentional contamination of distribution systems, but not with Bayesian methodologies. Nilsson et al. (2005) simulated the exposure of a population to a deliberate contamination attack using a Monte Carlo approach over a wide range of uncertain demands and operation scenarios. Allmann and Carlson (2005) analyzed the intentional introduction, spread and detection of several known contaminants in a distribution system. Others including Lee and Deininger (1992), Kessler et al. (1998), and Ostfeld (2004), have sought to optimize the placement of water monitoring stations in distribution systems.

METHODOLOGY

Our research seeks to use BBNs to integrate sensors and other relevant data to better characterize a system for real-time response. This task is similar to military and studies in other fields that utilize Bayesian approaches to combine diverse sources of data and intelligence in a battlefield setting or other response scenario. This paper presents a BBN approach that integrates information from multiple sensors with records of operational changes. The approach is implemented on a hypothetical case study. Simulations over numerous possible combinations of upstream locations and source mass values are used to estimate the probability of a detection given that a contaminant release has occurred. This probability along with others estimated by expert judgment define the variables that comprise the BBN. The following sections present an introduction to Bayesian belief networks, a description of the hypothetical case study, a methodology for implementing the BBN, and a demonstration using the BBN to infer knowledge of the system state for three positive sensor detection scenarios.

Bayesian Belief Networks

Bayes theorem states that:

$$P(h \mid D) = \frac{P(D \mid h)P(h)}{P(D)} \qquad (1)$$

where:

P(h) = prior probability of h with no knowledge of observation D,

P(D) = prior probability of D with no knowledge of h,

P(h|D) = posterior probability of h after observation D, and

P(D|h) = probability of observation D given that h is true.

The Bayesian prior probability P(h) is updated to a posterior probability P(h|D) that reflects an observation D. Conceptually, this updating process mimics the reasoning of people presented with new information about uncertain phenomena. For a classic illustration of this idea, presented in The Economist (September, 2000), consider a newborn infant in her first day. After observing her first sunset, she wishes to determine the probability that the sun will rise again using Bayes theorem. The infant puts a black marble and a white marble into a bag to

represent her initial estimate that there is an equal probability that the sun will rise again or not. Each day that the sun rises, she puts another white marble into the bag. After one morning, the probability of sunrise increases from 0.5 (1 white marble/2 total marbles) to 0.75 (2 white marbles/3 total marbles). The probability is 0.8 after the second day, 0.833 after the third, and so on until the child has near certainty that the sun will continue to rise every day. In much the same way, the prior probability of a hypothesis, P(*h*) is revised to a posterior probability after an observation has been made P(*h*|*D*).

Bayesian learning in a complex, interconnected system can be represented with a BBN, which represents the conditional independence assumptions among a set of variables, thus specifying the joint probability distribution. A BBN is typically presented as a directed acyclic graph with nodes representing variables and arcs representing assertions of conditional independence. A node is conditionally independent of its non-descendants. In Figure 1, variable d can be said to be conditionally independent of variable c, given a and b.

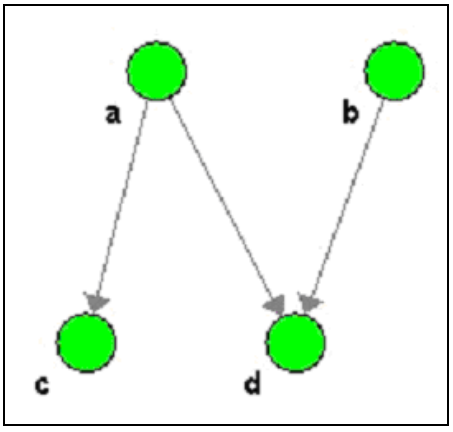

Figure 1. Bayesian belief network structure

Bayes theorem defines the relationships among variables. The joint probability for the assignment of any set of values ($x_1,\ldots,x_n$) to the set of variables ($X_1,\ldots,X_n$) in a BBN can be determined by:

$$P(x_1,...,x_n) = \prod_{i=1}^{n} P(x_i \mid Parents(X_i)) \qquad (2)$$

where, Parents(X_i) is the set of values for preceding nodes in the network. The joint probability of any set of variables can be inferred from observed values or distributions for any subset of

remaining variables. There are a number of exact and approximate algorithms that have been developed to infer posterior probabilities for BBNs (Jenson, 1996). The work presented in this study utilizes a generalized variable elimination approach in which posteriors are derived from the marginal probabilities for a set of variables in the BBN. Further details on this algorithm are provided by Cozman (2000) and Dechter (1999).

A powerful attribute of BBNs is their ability to integrate expert knowledge with models and data to define both the conditional independence structure between variables and the associated probabilities. The BBN can then be recursively updated as new evidence becomes available, in much the same way that the newborn infant adds additional white marbles to her bag as she observes each sunrise. The BBN presented in this study utilizes expert knowledge to determine the network structure. Prior probabilities were determined using modeling for some nodes and expert judgment for others.

Application of BBNs to a Distribution System Case Study

In this section, a BBN methodology is proposed to represent a distribution system and possible pieces of evidence that might be available to inform a system operator in the event of a contamination event. Evidence includes sensors and operational records. The contaminant release is assumed to occur from a single location at a service connection, hydrant, storage tank, or other water infrastructure component. The approach is general and should be readily adaptable to other types of scenarios and distribution systems.

The distribution system used in this study is a hypothetical campus-type facility that was created for research, development, and demonstration purposes only. The relevant water infrastructure components are extracted from GIS data for the model facility. The distribution system is modeled with 235 nodes and 261 pipes ranging from 5.08 cm (2 in) to 30.48 cm (12 in) in diameter. Water demands are assigned to nodes based on typical per capita consumption values in the assortment of buildings that are present in the model facility. Figure 2 shows the demand patterns used in the model. Pipes, pumps, and storage facilities are sized and configured so that water pressures are between 276 kPa (40 psi) and 552 kPa (80 psi) and water velocities are less than 1.53 m/sec (5 ft/sec). EPANET (Rossman, 2000) is used to solve the distribution system flow and transport equations.

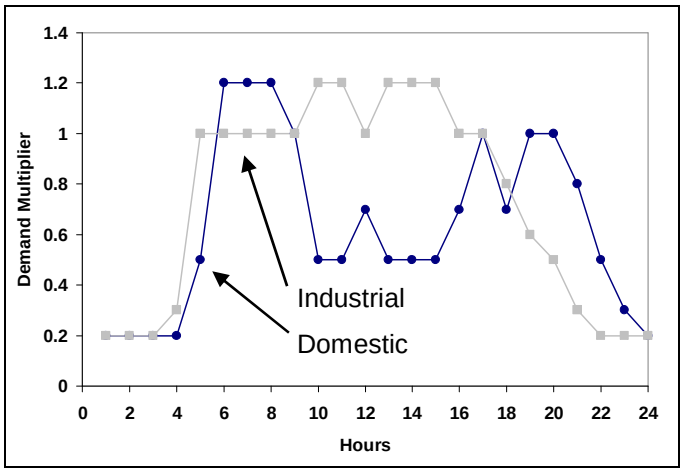

Figure 2. Demand patterns used in distribution system model

Three sensor locations were identified based on a qualitative inspection of the model flow patterns in extended time period hydraulic simulations. Figure 3 shows a schematic of the distribution system model and the sensor locations. These locations were selected from within a limited region of the distribution system to reduce complexity of the analysis. An attempt was made to maximize the sensor's upstream coverage, and distribute sensors evenly across the target area, however, no rigorous mathematical optimization was performed. Others have conducted research into algorithms to determine the optimal placement of sensors (e.g., Lee and Deininger, 1992; Kessler et al., 1998; Ostfeld and Salomons, 2004), which is a topic beyond the scope of this study.

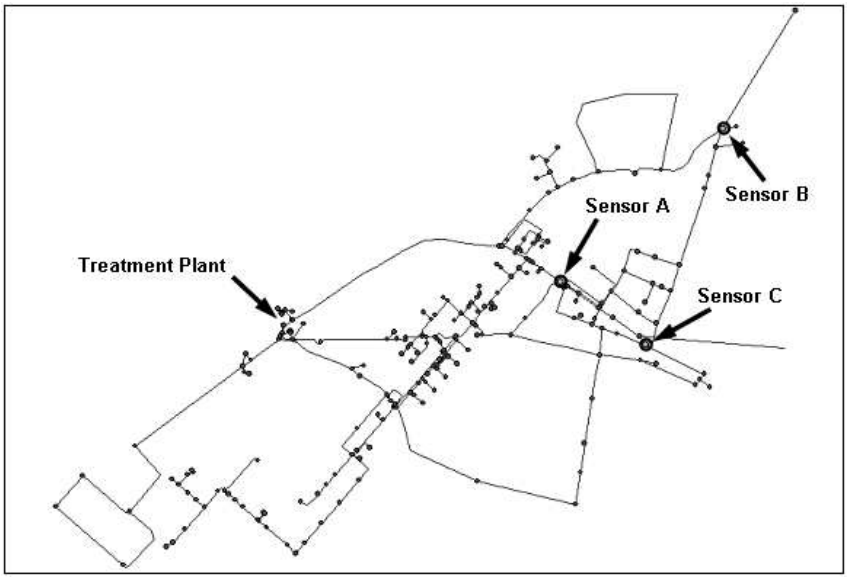

Figure 3. Distribution system model showing location of treatment plant and sensors

Each sensor was assumed to be capable of identifying contaminants from the region of the distribution system upstream of the sensor's location. The upstream region of each sensor was determined by simulating the release of a unit contaminant from each node in the system and observing which source nodes resulted in non-zero detections at the sensors. Four mutually exclusive upstream areas were defined by the overlapping coverages of the downstream sensors. Area 1 is upstream of all three sensors, Area 2 is upstream of Sensor A and Sensor C, Area 3 is upstream of Sensor B and Sensor C, and Area 4 is only upstream of Sensor C (Figure 4). Sensors A and B are upstream of Sensor C and completely contained in its coverage area. Therefore there are no areas defined to be exclusively upstream of A or B but not C. Sensors were assumed to provide a simple yes/no indication of the presence of a contaminant above a threshold concentration with known false positive and false negative rates. These sensors could be devices that measure the contaminant directly, or they could measure a surrogate parameter that is then related to the contaminant through postprocessing using a statistical relationship, learning algorithm (e.g. artificial neural network), clustering, or other method (ASCE, 2004). Surrogate water quality parameters are not included in this analysis, however, they could easily be integrated into a BBN as additional evidence nodes.

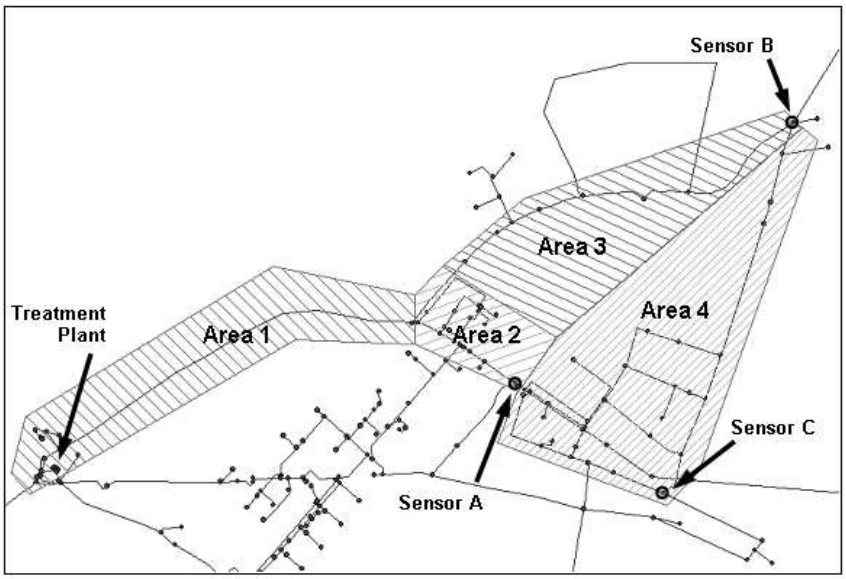

Figure 4. Distribution system showing locations of sensors and upstream areas

A BBN representing the joint probabilities of several contamination scenarios is shown in

Figure 5. The bottom level nodes of the BBN represent observable evidence such as that provided by sensors and operational logs. Top-level nodes represent causal events that are not directly observable. Some of the top nodes are 'true positive' events such as a contaminant release in an upstream area, while some would be 'false positives' such as a change in operation. The origin of the contaminant release is not specified and could be intentional or unintentional. These are simplifications and abstractions of the relevant components that might exist in an actual distribution system. It is expected that the 'operation change' node would be subdivided into many individual nodes that each represents the state of a system component such as pumps or storage tanks. Additional observable evidence such as specific threat intelligence could easily be incorporated into the BBN through additional nodes. The structure of the BBN is defined by the topology of the distribution system – events upstream cause observable evidence downstream. Additional structural elements such as physical security evidence, power grid events, hospital diagnosis patterns, and others could be integrated using expert knowledge of causal relationships between events and evidence.

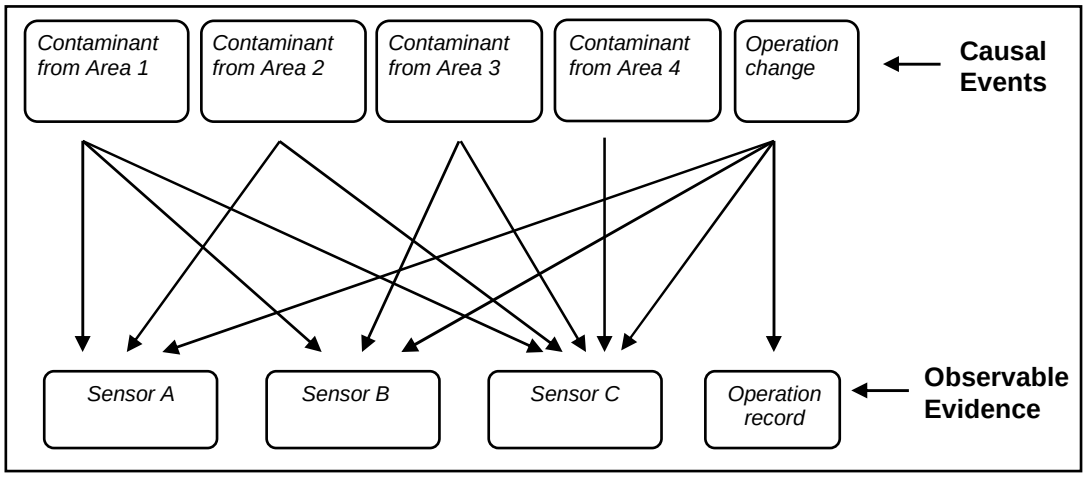

Figure 5. Bayesian belief network for distribution system contamination

Each node of the BBN utilizes a table containing conditional probabilities of discrete Boolean states given the state of that node's parents, or $P(x_i \mid Parents(X_i))$. For top level nodes, this table is simply the prior probability of that event, $P(x_i)$ and the prior probability of *not* that event, $P(\neg x_i)$. Table 1 shows the matrix of conditional probabilities estimated for S*ensor B* given the state of its parent nodes: *Contaminant from Area 1, Contaminant from Area 3,* and *Operation*

change. *Contaminant from Area 1* refers to the introduction of a contaminant into the region that is upstream of all three sensors,and *Contaminant from Area 3* refers to the introduction of a contaminant into the region that is covered by *Sensor B* and *Sensor C* (Figure 4). *Operation change* refers to actions such as system flushing, booster pump activation, valve maintenance, seasonal changes, or others that might cause false sensor detections. This node could be subdivided and refined during implementation to reflect further knowledge of system operational characteristics. The relevant operational activities would be dependent on a specific system's hydraulic characteristics and the sensor technology used. The number of combinations of parent node states is 2^i, where i = number of parent nodes. For the case of *Sensor B* there are 8 possible combinations of parent states. It is not possible for both *Contaminant from Area 1* and *Contaminant from Area 3* to be 'true' since we assume a single injection scenario.

The prior probabilities for each variable in the network may be determined in a number of ways. False positive and false negative rates for the sensor due to internal causes may be provided by the manufacturer or determined during project implementation. For this study, the false positive and false negative rates for the sensors are assumed to be 0.05 and 0.03, respectively. The prior probability of a contaminant release may be somewhat more difficult to rationally estimate. In contrast to rare natural phenomena such as severe weather events, intentional terrorist attacks are not random events with a historical record on which to base a frequency analysis. Priors for unintentional contamination events, however, may be estimated based on the historical record, expectation of infrastructure failures, or other inputs. For this study, the prior probability of a contaminant release is arbitrarily assumed to be 0.0001. The prior probability of *Operation change* was also assumed to be 0.0001. For a real application, it may be possible to determine the prior probabilities for operational changes represented in the BBN from analysis of historical data.

The conditional prior probabilities for observable nodes given the state of parent nodes can be determined by computer simulation. In this study, the prior probabilities of detection by the sensors given an upstream release are estimated using repeated simulations with EPANET. The contaminant is modeled as the single injection of a pure conservative solute for six hours using the mass booster option in EPANET. This release is simulated from all possible source locations in each of the upstream areas covered by the sensor, and the initial mass is varied from 0 to 300 g/min in 0.25 g/min increments. The upper limit reflects practical considerations for a

likely intentional terrorist attack. The amount of contaminant that could be concealed at a residential location and the capacity of a 1.905 cm (3/4 in) service connection would likely preclude larger initial mass values. An injection of 300 g/min over 6 hr would require a mass of approximately 108 kg of pure solute. We assume a single release scenario, so multiple injections were not simulated.

The solute concentration is measured at the sensors over the 36-hour simulation period for each combination of source mass and location. This duration was chosen to ensure that the simulation would capture the peak concentration at each sensor. A positive detection is assumed to have occurred if the sensor concentration exceeds an arbitrarily selected threshold of 100 mg/L. The prior conditional probability of a sensor detection is determined by,

$$P(positive \mid release) = \frac{n_{et} - FN(n_{et}) + FP(n_{tot} - n_{et})}{n_{tot}} \qquad (3)$$

where, n_{et} is the number of simulations that result in a concentration exceeding the threshold for a single sensor, *FN* is the false negative rate, *FP* is the false positive rate, and n_{tot} is the total number of releases from the upstream area over all locations and initial mass values. The conditional prior probabilities for operation change variables, such as P(*Operation record* | *Operation change*) or P(*Sensor B* | *Operation change*), are assumed in this study. However, these conditional priors could be reasonably estimated by analysis of historical records or additional simulations. The value of P(*Sensor A* | *Operation change*) is estimated to be 0.7, P(*Sensor C* | *Operation change*) is 0.05, and P(*Sensor B* | *Operation change*) is 0.05. These conditional prior probabilities were set to different values to reflect a closer relationship between *Operation change* and *Sensor A* than the other sensors. This introduces an additional complexity to the BBN similar to that likely to occur in a real world application. The conditional prior probabilities for combinations of Operation change with other parent nodes are determined by calculating the union of the individual probabilities, since the individual probabilities are mutually exclusive. For example, P(*Sensor B* | *Contaminant from Area 1, Operation change*) is simply P(*Sensor B* | *Contaminant from Area 1*) OR P(*Sensor B* | *Operation change*). Table 1 shows the conditional prior probabilities for *Sensor B.*

Table 1: Matrix of prior conditional probabilities for *Sensor B* and its parent nodes

Parent node states			Probability of detection	
Contaminant from Area 1	*Contaminant from Area 3*	*Operation Change*	*Sensor B* Positive	*Sensor B* Negative
True	False	True	0.522	0.478
True	False	False	0.497	0.503
False	True	True	0.973	0.0266
False	True	False	0.972	0.0280
False	False	True	0.0965	0.9035
False	False	False	0.0500	0.950

RESULTS

The BBN is used to explore hypothetical combinations of sensor detections and other evidence that might occur in a contamination event. The posterior probability that a contaminant has been introduced is inferred from changes to observable nodes in the BBN. When a node is 'observed', its value becomes fixed, and the probability of that node's parent(s) is inferred to reflect the new observation. The posterior probabilities of the causal event nodes can provide useful guidance for interpreting a positive sensor detection. Because prior probabilities are set to an arbitrarily low value (0.0001) for the causal event nodes, the *change* in probability is used to indicate causality. Three possible scenarios are explored for illustration purposes below: a false positive and two true positive detections.

In the first example, *Operation record* is observed to be 'true', *Sensor A* is observed to be 'true', *Sensor B* is observed to be 'false', and *Sensor C* is observed to be 'false'. The probability of the positive sensor detection being caused by *Contaminant from Area 2* is updated from a prior of 0.0001 to a posterior of 0.00047 (Figure 6).

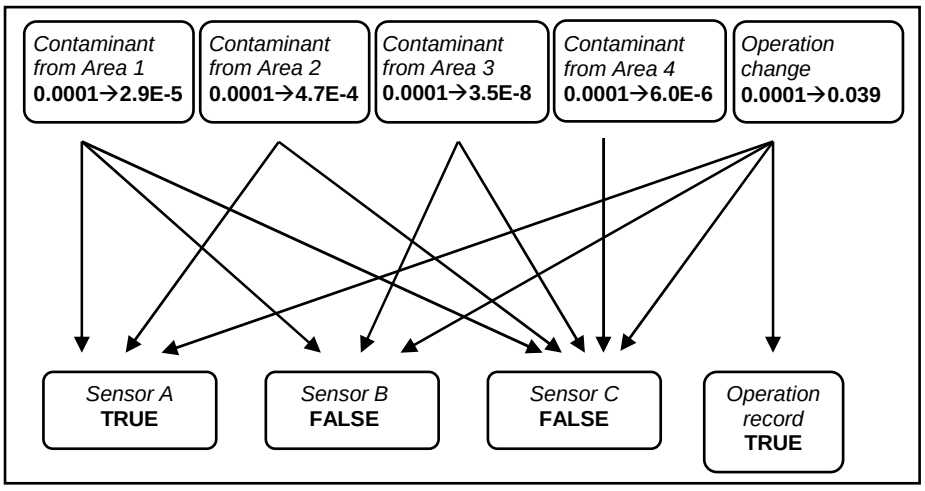

Figure 6. Example of BBN updating for a false positive sensor detection

The probabilities of releases in the two other upstream areas also increase slightly. However, the probability that the positive sensor detection was caused by *Operation change* is updated from a prior of 0.0001 to a posterior of 0.0394. This posterior value is quite small due to the effect of the very low prior probability that was assigned to this node. However, the probability for

Operation change increased by a factor of 394 from its prior value, which is a much greater increase than other causal nodes. Taking the change in probabilities into account, this result suggests that the sensor detection is a false positive.

In the second example, *Operation record* is observed to be 'false', *Sensor A* is observed to be 'true', *Sensor B* is observed to be 'false', and *Sensor C* is observed to be 'true'. The probability of *Contaminant from Area 2* is updated from a prior of 0.0001 to a posterior of 0.0201, an increase by a factor of 201 (Figure 7). The probability of *Operation record* decreases for this scenario from 0.0001 to 7.76×10^{-5}. This result suggests that the sensor detection is a true positive.

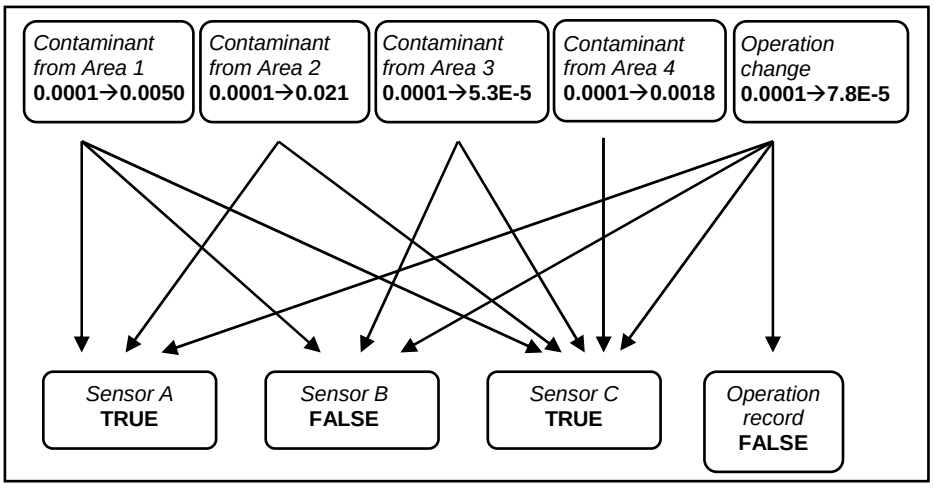

Figure 7. Example of BBN updating for a true positive sensor detection

The change in probability is used to suggest causality where priors are set to an arbitrarily low value. An analysis was performed to determine the sensitivity of this change to different prior probabilities. When the prior for *Contaminant from Area 2* is changed from 0.0001 to 0.001, the resulting posterior for the true positive scenario is 0.175, an increase by a factor of 175. Further trials with different prior probabilities for this variable result in a relatively consistent increase in the posterior when the prior is 0.001 or less. Figure 8 shows the factor by which the posterior increases in this sensor detection scenario for different prior probabilities of *Contaminant from Area 2.* This result suggests that the change in probability is insensitive to the initial value for rare events.

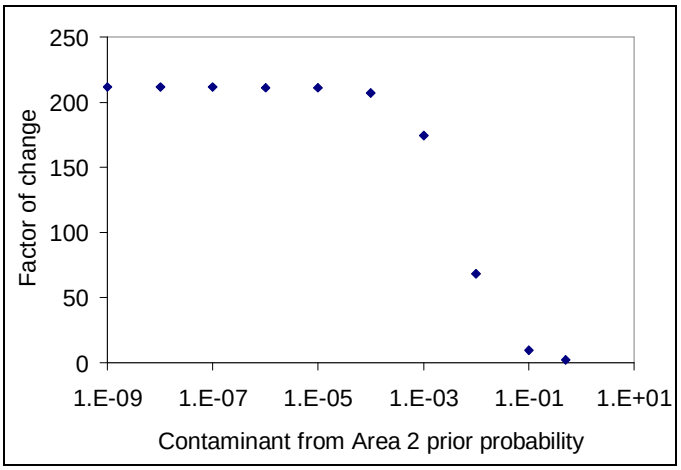

Figure 8. Factor by which the probability of *Contaminant from Area 2* increases in the true positive scenario for different prior probabilities

The final example shows a less intuitive combination of evidence in which two sensors are positive and an operation change has occurred. *Sensor A* is observed to be 'false', *Sensor B* is 'true', *Sensor C* is 'true', and *Operation record* is 'true'. The posterior probability for *Operation change* in this case is 0.00367, a change by a factor of 36.7. The posterior for *Contaminant from Area 2* is 0.0343, an increase by a factor of 343. While somewhat less conclusive than the other examples, this combination of evidence clearly implicates *Contaminant from Area 2* as the most likely source of a true positive detection. This result is reasonable since *Operation change* is more closely related to *Sensor A* with a higher conditional probability than the other two sensors.

The methodology proposed in the previous sections also provides insight into the probability that a contaminant released at a given node would be detected by a downstream sensor. The probability for each upstream node was determined using summing positive detections by a sensor over all initial mass values while keeping location constant. This information would be useful in identifying the location of a contaminant release in response to a positive sensor detection. Figure 9 shows the estimated probability of contaminant detection at *Sensor C* for releases at each upstream location. In this case the probabilities for each upstream location are all relatively high. It is likely that a model of a contaminant in a more complex full-scale system would result in a larger range of probability values.

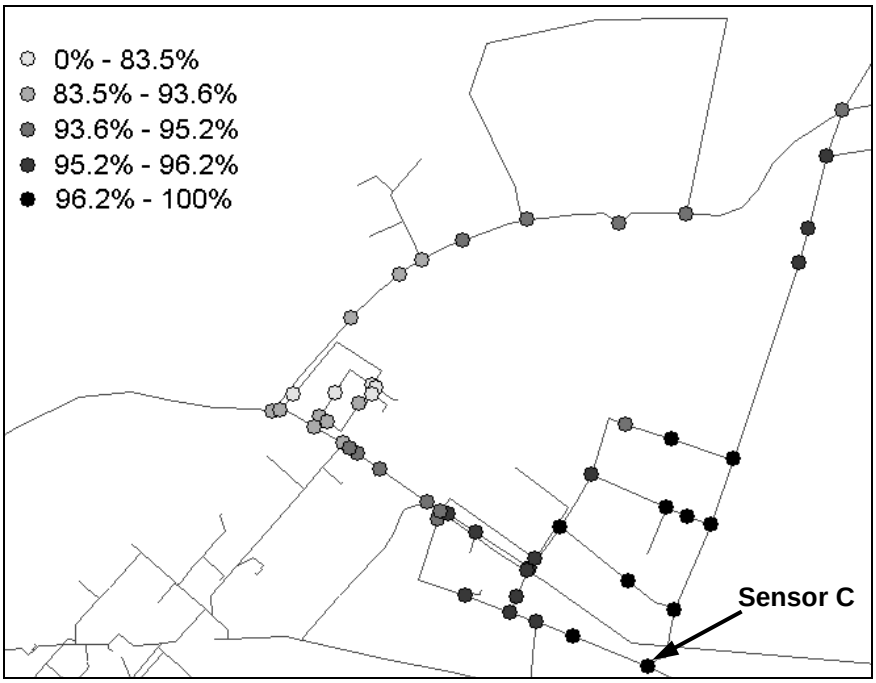

Figure 9. Estimated probability of contaminant detection at *Sensor C* for releases at each location

DISCUSSION

This work suggests that BBNs have the potential to be a useful tool for interpreting the sometimes confusing information from online contaminant monitoring and other sources that characterize a water system. However, there are additional challenges that require further research. The spatial and temporal characteristics of the sensor data and distribution system model are not directly expressed in the BBN presented here. This approach does not, for example, provide information regarding the exact location within an upstream region that a contaminant was released. Others such as Shang et al. (2002) and Zierolf et al. (1998) have presented more rigorous mathmatical approaches for particle backtracking that utilize timing information to link output to source input. It is possible that the location probability map shown in Figure 9 could be further refined in a real-time response scenario by accounting for the time difference between sensor detections. The timing of these observations could be used with real time flow data to pinpoint the location of an attack. Figure 9 is also beneficial in that it identifies nodes that are not well covered by the sensor network. This information could be useful for network planning purposes.

Timing would also be relevant to the probability that observations are evidence of the same causal event. In addition, only contamination from a single ideal source is represented in the BBN demonstrated here. Further research is needed to extend this work beyond these simplifying assumptions to additional contamination scenarios, additional reactive solutes, larger and more complex distribution systems, and additional case studies. Future work will also explore the use of simulated data streams to update the Bayesian network in real time.

Very little computational time was required to solve the Bayesian belief network once the prior probabilities were defined. The contaminant simulation to calculate prior probabilities required approximately 80,000 runs, which took 2.5 hours to complete using a Windows desktop PC. Future implementations of this method could easily make use of a parallel computing cluster to reduce the computational time.

The BBN presented here is used as a framework for expressing the complex causal relationships and conditional probabilities that comprise contamination scenarios. A challenge in implementing this approach would be to imagine contamination scenarios that cover the wide range of possible vulnerabilities. These scenarios would ideally be the product of a diverse

group of experts, engineers, operators, responders, and others that have detailed knowledge of the system and current research in vulnerability analysis. However, the possibility would always exist that a terrorist could attack a water system in an unpredictable way that would not be characterized accurately by the BBN. During implementation, observed evidence would likely be augmented by data from field testing kits that could be deployed to a region of the distribution system. The evidence from these tests could be integrated into the BBN simply by adding additional observation nodes.

An approach for estimating some of the prior probabilities is shown in this study. However, there are additional variables for which probabilities must either be determined through expert judgment, regression, or additional modeling. The BBN presented here is intended to be a framework for expressing the conditional relationships between system variables and making inferences in response to observations. It is expected that any implemented BBN would be refined and augmented by real time data to define both probabilities and network structure. Mining these relationships and patterns in real time data is an area of active research that will likely complement the work presented here.

CONCLUSIONS

This study proposes a BBN methodology for expressing complex causal relationships among the events and observations that comprise contamination scenarios in water distribution systems. These scenarios can be better understood when explicitly visualized in a graphical probabilistic model such as a BBN. The methodology uses distribution system simulations to estimate conditional prior probabilities for contaminant introductions. Application of the approach to a hypothetical system illustrated how data from sensors and other sources can be interpreted with a BBN to better characterize a water system and distinguish between a routine false positive sensor detection and a true system contamination event. This approach has the potential to be incorporated into both security planning and real time response actions.

REFERENCES

Allmann, T. P. and Carlson K. H. (2005). "Modeling intentional distribution system contamination and detection." *J. Am. Water Works Assoc.*, 97(1), 58-61.

American Society of Civil Engineers (2004). *Interim Voluntary Guidelines for Designing an Online Contaminant Monitoring System*, Reston, VA.

Beckerman, M. (1992). "A Bayes-maximum entropy method for multi-sensor data fusion." *Proc. - IEEE International Conference on Robotics and Automation*, IEEE, Piscataway, NJ, 2, 1668-1674.

Bonci, A., Di Francesco, G., and S. Longhi (2002). "A Bayesian approach to the hough transform for video and ultrasonic data fusion in mobile robot navigation." *Proc. of the IEEE International Conference on Systems, Man and Cybernetics*, IEEE, Piscataway, NJ, 3, 350-355.

Brown, C., Marengoni, M., Kardaras, G. (1995). "Bayes nets for selective perception and data fusion." *Proc. of SPIE - The International Society for Optical Engineering*, SPIE, Bellingham, WA, 2368, 117-127.

Chu, T. L. (1995). "Estimation of initiating event distribution at nuclear power plants by Bayesian procedure." *ISSAT international conference on reliability and quality in design (2nd)*, ISSAT, Orlando, Fl.

Coue, C., Fraichard, T., Bessikre, P., and Mazer, E. (2003). "Using Bayesian programming for multi-sensor mu1ti-targe t tracking in automotive applications." *Proc. of the 2003 IEEE International Conference on Robotics & Automation*, IEEE, Piscataway, NJ, 2104 – 2109.

Cozman, F. G. (2000). "Generalizing variable elimination in Bayesian networks." *Workshop on Probabilistic Reasoning in Artificial Intelligence*, Atibaia, Br, <http://www-2.cs.cmu.edu/~javabayes/Home/.> (Nov. 10, 2005).

Das, S., Grey, R., and Gonsalves, P. (2002). "Situation assessment via Bayesian belief networks." *Proc. of the Fifth International Conference Information Fusion*, IEEE, Piscataway, NJ, 1, 664-671.

Dechter, R. (1999). "Bucket elimination: A unifying framework for probabilistic inference." *Learning in Graphical Models*, M. I. Jordan, editor, MIT Press, Dordecht, The Netherlands, 75–104.

The Economist (2000). "In Praise of Bayes," The Economist, Sept. 30, 2000, 356(8190), 83.

Fox, D., Hightower, J., Liao, L., Schulz, D., and Borriello, G. (2003). "Bayesian Filtering for Location Estimation." *Pervasive Computing*, IEEE, 2(3), 24-33.

Franzen, D. W. (1999). "Bayesian Decision Model for Battle Damage Assessment." Master's

thesis, Air Force Institute of Technology, Air University, Air Education and Training Command, WPAFB, OH.

Hayes, C., Penner, R., Ergan, H., Lu, L., Tu, N., Jones,P., Asaro, P., Bargar, R., Chernyshenko, O., Choi, I., Danner, N., Mengshoel, O., Sniezek, J., Wilkins, D. (2000). "CoRaven: model-based design of a cognitive tool for real-time intelligence monitoring and analysis." *Systems, Man, and Cybernetics 2000 IEEE International Conference,* IEEE, Piscataway, NJ, 2, 1117-1122.

Jenson, F. V. (1996). *An Introduction to Bayesian Networks*, Springer Verlag, New York.

Jones, P. M., Hayes, C. C., Wilkins, D. C., Bargar, R., Sniezek, J., Asaro, P., Mengshoel, O., Kessler, D., Lucenti, M., Choi, I., Tu, N., and Schlabach, J. (1998). "CoRAVEN: modeling and design of a multimedia intelligent infrastructure for collaborative intelligence analysis." *Systems, Man, and Cybernetics 1998 IEEE International Conference,* IEEE, Piscataway, NJ, 1, 914-919.

Karlsson, B., Jan-Ove, J., and Wide, P. (2002) "A fusion toolbox for sensor data fusion in industrial recycling" *IEEE Transactions on Instrumentation and Measurement*, IEEE, Piscataway, NJ, 51(1), 144-149.

Kessler, A., Ostfeld, A., and Sinai, G. (1998). "Detecting accidental contaminations in municipal water networks." *J. of Water Resour. Plan. Manage.*, 124(4), 192-198.

Lee, B. H. and Deininger, R. A. (1992). "Optimal Locations of Monitoring Stations in a Water Distribution System." *J. of Environ, Eng.*, 118(4).

Liu, E., and Zhang, D. (2002). "Diagnosis of component failures in the space shuttle main engines using Bayesian Belief Networks: a feasibility study." *Proc. of the 14th IEEE International Conference on Tools with Artificial Intelligence (ICTAI'02)*, IEEE, Piscataway, NJ, 181-188.

Marcot, B. G. (2001). "Using Bayesian belief networks to evaluate fish and wildlife population viability under land management alternatives from an environmental impact statement." *Forest Ecology and Management*, 153 (1-3), 29-42.

Murray, R., Uber, J., and Janke, R. (2005). "A Model for Estimating the Acute Health Impacts Resulting from Consumption of Contaminated Drinking Water." *ASCE World Water and Environmental Resources Congress*, ASCE, Reston, VA, 173, 31.

Nilsson, K. A., Buchberger, S. G., and Clark, R. M. (2005). "Simulating exposures to deliberate intrusions into water distribution systems." *J. Water Resour. Plan. Manage.*, 131(3), 228-236.

Ostfeld, A., and Salomons, E. (2004). "Optimal layout of early warning detection stations for water distribution systems security." *J. Water Resour. Plan. Manage.*, 130(5), 337-385.

Rossman, L. A. (2000). *EPANET2 Users Manual*, United States Environmental Protection Agency, Washington, DC. <http://www.epa.gov/ORD/NRMRL/wswrd/epanet.html> (Nov. 10, 2005).

Sanzotta, M. A. and Sherrill, E. T. (1997). *Approximation Probability of Detection in the Janus Model*, United States Military Academy, West Point, New York.

Schlumberger, Y., Pompee, J., and De Pasquale, M. (2002). "Updating operating rules against voltage collapse using new probabilistic techniques." *Transmission and Distribution Conference and Exhibition 2002: Asia Pacific*, IEEE, Piscataway, NJ, 2, 1139-1144.

Shang, F., Uber, J. G., and Polycarpou, M. M. (2002). "Particle Backtracking Algorithm for Water Distribution System Analysis." *J. Environ. Eng.*,128(5), 441-450.

Stiber, N. A., Pantazidou, M., and Small, M. J. (1999). "Expert system methodology for evaluating reductive dechlorination at TCE sites." *Environ. Sci. Tech.*, 33(17), 3012-3020.

Stiber, N. A., Pantazidou, M., and Small, M. J. (2004). "Embedding expert knowledge in a decision model: Evaluating natural attenuation at TCE sites." *J. Haz. Mat.*, 110(1-3), 151-160.

Stiber, N. A., Small, M. J., and Pantazidou, M. (2004). "Site-specific updating and aggregation of Bayesian Belief Network models for multiple experts." *Risk Analysis*, 24(6), 1529-1538.

Stow, C. A., Roessler, C., Borsuk, M. E., Bowen, J. D., and Reckhow, K. H. (2003). "Comparison of Estuarine water quality models for total maximum daily load development in Neuse River Estuary." *J. Water Resour. Plan. Manage.*, 129 (4), 307-314.

Therrien, S. S. (2002). *Bayesian Model to Incorporate Human Factors in Commanders' Decision Making*, Master's thesis, USNAVY Postgraduate School, Monterey, Ca.

Zierolf, M. L., Polycarpou, M. M., and Uber, J. G. (1998). ''Development and auto-calibration of an input-output model of chlorine transport in drinking water distribution systems.'' *IEEE Trans. Control Syst. Technol.*, 6(4), 543–553.