AFIT/GCS/ENG/96D-30

On Unifying Time and Uncertainty: The Probabilistic Temporal Network

THESIS

Joel D. Young
Captain, USAF

AFIT/GCS/ENG/96D-30

AFIT/GCS/ENG/96D-30

On Unifying Time and Uncertainty: The Probabilistic Temporal Network

THESIS

Presented to the Faculty of the School of Engineering

of the Air Force Institute of Technology

Air University

In Partial Fulfillment of the

Requirements for the Degree of

Master of Science

Joel D. Young, B.S.

Captain, USAF

17 December, 1996

Acknowledgements

First and foremost, I would like to thank my wife, Wendy Snow, for tolerating my long nights spent at the lab. Her constant support is deeply appreciated. I would also like to thank my daughter, Alix Elizabeth, for her unceasing joy and love. I would like to thank my research committee, Dr. Santos, Dr. Lamont, and Maj. Banks for giving me the freedom to pursue my research as I chose with special thanks to Dr. Santos for the guidance that made it possible to bring my work to fruition. Finally, I would like to thank Donald Knuth and Leslie Lamport for developing, and giving to the world, TEX and LATEX—the finest typesetting system around.

Joel D. Young

All Nature is but Art, unknown to thee;
All Chance, Direction, which thou canst not see;
All Discord, Harmony not understood;
All partial Evil, universal Good:
And, spite of Pride, in erring Reason's spite,
One truth is clear, WHATEVER IS, IS RIGHT.

Alexander Pope, in *An Essay On Man*

Table of Contents

Page

List of Figures

Figure Page

List of Tables

AFIT/GCS/ENG/96D-30

Abstract

Complex real-world systems consist of collections of interacting processes/events. These processes change over time in response to both internal and external stimuli as well as to the passage of time. Many domains such as real-time systems diagnosis, (mechanized) story understanding, planning and scheduling, and financial forecasting require the capability to model complex systems under a unified framework to deal with both time and uncertainty. Existing uncertainty representations and existing temporal models already provide rich languages for capturing uncertainty and temporal information, respectively. Unfortunately, these partial solutions have made it extremely difficult to unify time and uncertainty in a way that cleanly and adequately models the problem domains at hand. This difficulty is compounded by the practical necessity for effective and efficient knowledge engineering under such a unified framework. Existing approaches for integrating time and uncertainty exhibit serious compromises in their representations of either time, uncertainty, or both. This thesis investigates a new model, the Probabilistic Temporal Network, that represents temporal information while fully embracing probabilistic semantics. The model allows representation of time constrained causality, of when and if events occur, and of the periodic and recurrent nature of processes.

On Unifying Time and Uncertainty:
The Probabilistic Temporal Network

I. Introduction

The field of Artificial Intelligence is at a nexus in its progress in the modeling of human cognition and in the performance of useful tasks[1]. The critical capability for passing through the nexus is a single coherent structure unifying both time and uncertainty. This thesis investigation advances the field of Artificial Intelligence by providing the requisite unifying structure.

1.1 Overview

In the evolution of expert systems, many techniques have been developed to represent human knowledge. One of the earliest (and still used) techniques is to represent knowledge as a logical system of if-then style rules (rule-based systems [5,11]). A more recent approach is to represent knowledge (including uncertainty) of a situation, or "domain," as a network of states and probabilities (Bayesian Networks [22]).

Many domains, whether they are rule-based, probabilistic, or other, require a representation of time and of the temporal relationships between events. Most systems rely on a mechanism in which a date is associated with each piece of knowledge. Relationships are then determined simply by the date ordering. In more complicated domains, such as emergency room diagnosis, the date mechanism is not sufficient; one must be able to represent situations with relative knowledge like "precedes" or "during."

Real-world domains requiring a unified model of time and uncertainty include dealing with real-time system diagnosis, (mechanized) story understanding, planning and scheduling, logistics, resource management, as well as financial forecasting. For example, consider the following scenario found in computer security analysis:

[1]At the recent Twelfth Conference on Uncertainty in Artificial Intelligence [14] (August 1996) held in Portland, Oregon, a panel of experts on "UAI by 2005: Reflections on critical problems, directions, and likely achievements for the next decade" identified the need for unifying time and uncertainty as among the top priorities necessary for advancing the entire field of AI.

> The computer operations center has a secure vault with a time-coded lock. This time-lock allows the vault to be opened from 0900 hours to 0905 hours and from 2100 to 2105. The center has critical operations from 0855 to 1805. Access to the vault is needed during the day and during critical operation making the vault likely to be open at those times. However, if the vault is closed, it cannot be reopened until the time-lock allows.

This scenario provides a detailed description of the causal and temporal relationships necessary to properly model the secure vault. As part of the computer security analysis, one must be able to translate this description and capture the knowledge in a form that can be correctly processed and reasoned with.

Once the knowledge representation is captured, inferences can be made. Inferences can be of several types including prediction and explanation. Prediction is concerned with extending forward from the known past and present to the unknown future (statistical syllogism [15]). Explanation involves the determination of causality by extending from known data back to hypotheses (abduction) [15].

1.2 The Problem

Complex systems consist of collections of interacting processes. These processes change over time in response to both internal and external stimuli as well as to the passage of time itself. There is great variety in the behavior of processes. Some processes are simple events such as opening a door or flipping a switch. Others are complex. For example, consider a communication channel where errors may occur due to lightning strikes and faults are more likely to occur given previous errors. Processes can also be recurrent or periodic, such as the passing of day into night or shifts in a work schedule.

What is needed is a model capable of representing complex systems changing over time. Given evidence about the past and present state of a system, one must be able to predict the system's future state. Also, given a future state, one must be able to determine the most probable causes for that state. As knowledge about such systems is bound to be incomplete and as the systems themselves may not be deterministic, the model must be able to represent uncertainty. This uncertainty permeates all areas: the duration of events, the strength of causal influence, the precise temporal relationship between events, and so forth.

1.3 Prior Work

Bayesian networks [22] provide a robust, probabilistic method of reasoning with uncertainty. Bayesian networks, however, do not provide a direct mechanism for representing temporal dependencies. For example, it is difficult to represent a situation such as the variability of the time an employee arrives at work and the causal relationships between the time of arrival and later events.

Prior temporal modeling techniques have made trade-offs in expressiveness between semantics for time and semantics for uncertainty. Significant research has been done exploring time nets (also called time-slice Bayesian networks) [12, 16, 17]. These approaches build on the strong probabilistic semantics of Bayesian networks for expressing uncertainty. The discrete time net approach developed by Kanazawa models time as a series of points [16]. Events are considered to occur at an instant of time while facts are considered to occur over a series of time points. Both events and facts are represented by random variables. If dependencies only connect between random variables at the same or consecutive time points, then the net is said to be a Markov time net. In other words, the Markov property holds for a model when the future is conditionally independent of the past, given the present [17].

Hanks et al, [12] is especially interesting for this work due to the emphasis on both endogenous and exogenous change [12]. Endogenous change is triggered by internal action, such as the progression of disease, and exogenous change is triggered by external change such as the administration of drugs. In the temporal model presented in this thesis, individual processes within a system must be able to respond to both endogenous (internal) and exogenous (external) stimuli.

The time-sliced approaches mentioned above are based on point models of time and, as such, require that events occur instantaneously. Often it is more natural to consider events as taking place over intervals of time. Also, the relationships between events that occur over intervals can be quite difficult to represent with only the three point relations (precedes, follows, equals).

Santos' Temporal Abduction Problem (TAP) [26] uses an interval representation of time. In the TAP, each event has an associated interval during which the event occurs. Relationships between events are expressed as directed edges from cause to effect within a weighted and/or directed acyclic graph structure. Edges are qualified with the possible interval relations. This allows great flexibility in expressing the relationship between events. For example, if event A must occur either before or after event B then the relation is written $\{<, >\}$. The TAP is an extension of cost based abduction [7] using a numeric cost to indicate the uncertainty of an event's occurrence. These costs are generally determined in an ad hoc manner by the domain expert. The TAP trades strong semantics of uncertainty for a powerful and flexible temporal representation.

1.4 Thesis Contribution

This thesis investigation presents a new model, the Probabilistic Temporal Network (PTN), for representing temporal and atemporal information while remaining fully probabilistic. The model allows representation of time constrained causality, of when and if events occur, and of the periodic and recurrent nature of processes. Bayesian networks lie at the foundation of the system and provide the probabilistic basis. Allen's interval system [2] and his thirteen relations provide the temporal basis.

PTNs focus on directly modeling processes and the interaction between them. The state of a process is represented by a value at a given time interval. A process can be defined over any number of such intervals. Random variables from traditional probability theory are used to model a process' value over each time interval.

The next chapter (Chapter II) discusses temporal reasoning and Bayesian networks. From this foundation, the theoretical structure and probabilistic nature are developed and proven in Chapter III. A linear constraint system for performing belief revision is developed in Chapter IV as well as a polynomial solvable subclass. Chapter V develops the process of extending an existing knowledge base into the temporal domain as well as recommendations for knowledge engineering with the probabilistic temporal network. The investigation concludes, in Chapter VI, with recommendations for further research. Along the way, several examples are developed including the secure vault scenario introduced previously.

II. Background

To represent complex, dynamic systems, the probabilistic temporal network draws from both temporal and Bayesian reasoning. This chapter introduces the foundations from which the PTN is built. Section 2.1 briefly develops temporal reasoning, emphasizing the aspects relevant to this thesis. Section 2.2 introduces Bayesian networks from which the PTN draws probabilistic semantics for representing uncertainty.

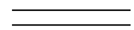
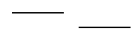
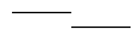
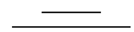
2.1 Temporal Reasoning

Temporal reasoning has been defined as the ability to reason about the relationships in time between events [11]. It is necessary to reason about time in many domains including planning, simulation, natural language understanding, and diagnosis. Temporal reasoning has been considered in philosophy and logic since Thales and Zeno [19]; however, it is only in the last two decades that temporal reasoning has been explicitly considered in artificial intelligence. McDermott and Allen, with their work in the early nineteen-eighties [2–4, 20], brought temporal reasoning into the AI mainstream. Other models for temporal reasoning include point algebras [32], semi-intervals [10], temporal constraint networks [9], and weak representations of interval algebras [18].

McDermott provides one of the earliest temporal representations [20]. In his approach, time is divided into a series of states with each state having an associated date, i.e., point in time. Facts are expressed as being true during particular states.

Allen introduced interval temporal reasoning to the AI community [2, 4]. Allen's interval algebra is governed by 13 relations on the intervals. Each event has an associated interval, denoted $[a, b]$, where a is the starting time point and b is the termination point. Temporal relationships between events are expressed as relations between their intervals. The relations between intervals, denoted $\mathcal{A}$, are $\{=, <, >, m, mi, d, di, s, si, f, fi, o, oi\}$ [2] (see Table 2.1). For example, event $A = [a, b]$ preceding event $B = [c, d]$ is denoted $A < B$ indicating that $a < b < c < d$. These relations are mutually exclusive and exhaustive. Note that, while there are thirteen relations between intervals, only three relations exist between points: precedes, equals, and follows.

Table 2.1 The thirteen possible interval-interval relations.

Symbol:		Name:	Relation:
=	=	equals	
<	>	precedes	
m	mi	meets	
d	di	during	
s	si	starts	
f	fi	finishes	
o	oi	overlaps	

Of special importance is Allen's use of disjunctive sets to express uncertainty in the exact relationship between intervals. For example, "interval A precedes or meets interval B" is written as $A\{<, m\}B$. Some commonly used disjunctions are *disjoint*, written $\{<, >, m, mi\}$, and *contains*, written $\{di, si, fi\}$ [2]. These relationships between events can be represented in a graphical form where nodes represent events and the arcs are labeled with a disjunction of relations. The goal is to determine whether there exists an interval assignment to all the events that satisfy the disjunctive relations. If such a solution exists, then the given knowledge base is consistent.

While there is debate, in both philosophy and artificial intelligence, as to which representation, points or intervals, is most appropriate; the expressive power of the two methods is generally considered equivalent [2, 16] as intervals can be represented with beginning and end points in a point based approach. Allen points out, however, some paradoxes that can occur when points are allowed as the fundamental unit of time [2]. The problems arise from the durationless nature of points. Durationless intervals are not allowed, i.e., for any interval $[t_1, t_2]$, $t_2 > t_1$. If $t_1 = t_2$ is allowed then the thirteen interval-interval relations are not mutually exclusive. For example $[t_1, t_2]$ *starts* $[t_2, t_3]$ is indistinguishable from $[t_1, t_2]$ *meets* $[t_2, t_3]$ when $t_1 = t_2$. Mathematically, point relations should be expressed as $t_1 \mathcal{R} [t_2, t_3]$ and as such, there is a different set of point-interval relations which would add unnecessary overhead if used in our model. Our model strictly adheres to the philosophy that intervals are primitive and have non-zero duration.

Definition 1. *A* temporal interval *is a closed interval* $[a, b]$ *on the reals*[1] *such that* $a < b$.

Axiom 1. *The temporal interval is the primitive temporal individual.*

Since all intervals must have non-zero duration, how can point intervals be expressed? The standard approach is to use $[t_0, t_0 + \epsilon]$ where ϵ is arbitrarily close, but not equal, to zero. Note that ϵ can either be added to the end or subtracted from the beginning or both. This approach is adopted in the PTN. To facilitate specifying the relationships between intervals, ϵ is deemed constant across an entire model. Thus $[t_0, t_0 + \epsilon]\{m\}[t_0, t_1]$ does not hold while $[t_0, t_0 + \epsilon]\{m\}[t_0 + \epsilon, t_1]$ does.

Aside from the temporal domain, neither Allen's nor McDermott's method can explicitly model uncertainty. Uncertainty arises from many sources including missing or unavailable data as well as over generalization of rules [11]. For example if we have the rule "Birds Fly" and "Ostriches are birds" we conclude that "Ostriches fly." To prevent such a conclusion, additional rules must be added such as "Some birds fly" or "Ostriches don't fly" to cover each exception. These additional rules can add significant complexity to a knowledge base.

2.2 Bayesian Networks (BNs)

Approaches to dealing with uncertainty include fuzzy logic [34], cost based techniques [7], certainty factors [29, 30], Dempster-Shafer theory [27], and probabilistic methods [22]. These approaches can be used both extensionally and intensionally. Extensional systems, such as rule-based systems, attach some sort of truth value to each rule or formula. The truth-value for formulae are calculated functionally from the truth-value of sub-formulae. Intensional systems, such as model-based systems, attach uncertainty to the possible states of the system itself [22]. Extensional systems are generally computationally efficient but their uncertainty measures are semantically weak. Intensional systems, on the other hand, are generally computationally expensive and semantically strong [22]. By

[1] *Temporal intervals can be defined over the rational numbers if countability is an issue, perhaps in proving some property of the model.*

carefully restricting which parts of an intensional system are relevant to the other parts, the computational limitations can, to some degree, be overcome.

In probabilistic reasoning, random variables (RVs) are used to represent events and/or objects in the world. By assigning various values to these RVs, we can model the current state of the world and weight the states according to the joint probabilities.

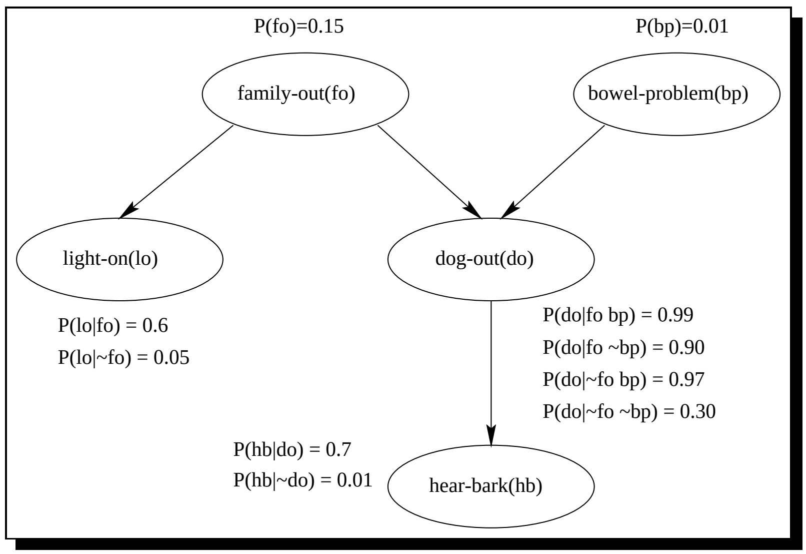


Figure 2.1 "Suppose when I go home at night, I want to know if my family is home before I try the doors. Now often when my wife leaves the house, she turns on an outdoor light. However, she sometimes turns on this light if she is expecting a guest. Also, we have a dog. When nobody is home, the dog is put in the back yard. The same is true if the dog has bowel troubles. Finally, if the dog is in the backyard, I will probably hear her barking." [6]

Bayesian networks are probabilistic intensional systems in which independence assumptions are used to restrict relevance. A Bayesian network is a directed acyclic graph (DAG) of random variable (RV) relationships. Directed arcs between RVs represent conditional dependencies. When all the parents of a given RV are instantiated, that RV is said to be conditionally independent of the remaining, non-descendent RVs given knowledge of its parents. For a more formal description of the independence semantics in Bayesian networks, see *d-separation* and *I-maps* in Charniak [6] and Pearl [22]. Figure 2.1 presents a simple example of a Bayesian network which demonstrates the nomenclature used in the following paragraphs.

In general, we are searching for the world state with highest likelihood. This is called *belief revision* [22]. Belief revision is best used for modeling explanatory/diagnostic tasks. Basically, some evidence or observation is given to us, and our task is to come up with a set of hypotheses that together constitute the most satisfactory explanation/interpretation of the evidence at hand. Belief revision is a form of *abductive reasoning* [7,13,23]. More formally, if W is the set of all RVs in our given Bayesian network and e is our given evidence[2], any complete instantiation to all the RVs in W that is consistent with e is called an *explanation* or *interpretation* of e. The problem, then, is to find an explanation w^* such that

$$P(w^*|e) = \max_w P(w|e). \tag{2.1}$$

w^* is known as the *most-probable explanation*. The joint probability of any explanation w,

$$w = (X_1 = x_1) \wedge (X_2 = x_2) \wedge \ldots \wedge (X_m = x_m) \tag{2.2}$$

(where $X_1 \ldots X_i \ldots X_m$ is an arbitrary ordering of random variables in W, and x_i is some assignment to random variable X_i) is found using the chain rule [22]:

$$P(w) = P(x_m|x_{m-1}, \ldots, x_1) \cdot P(x_{m-1}|x_{m-2}, \ldots, x_1) \cdots P(x_2|x_1) \cdot P(x_1) \tag{2.3}$$

Bayesian networks take the chain rule one step further by making the important observation that certain RV pairs may become uncorrelated once information concerning other RV(s) is known. More precisely, we may have the following independence condition:

$$P(A|X_1, \ldots, X_n, U) = P(A|X_1, \ldots, X_n) \tag{2.4}$$

for some collection of RVs U. Intuitively, we can interpret this as saying that given knowledge of $X_1, \ldots, X_n$ knowledge of U is irrelevant to the state of A.

Combined with the chain rule, these conditional independencies allow us to replace the terms in the chain rule with smaller conditionals. Thus, instead of explicitly keeping the joint probabilities, all we need are smaller conditional probability tables, from which the joint probabilities can then be calculated.

[2]That is, e represents a set of instantiations made on a subset of W.

For example, an application of the chain rule for computing the probability of an explanation for the Bayesian network in Figure 2.1 is

$$\begin{aligned} P(hb, do, lo, fo, bp) &= P(hb|do, lo, fo, bp) \cdot P(do|lo, fo, bp) \cdot \\ & P(lo|fo, bp) \cdot P(fo|bp) \cdot P(bp) \end{aligned} \tag{2.5}$$

Using the dependencies, we can simplify this to

$$P(hb, do, lo, fo, bp) = P(hb|do) \cdot P(do|fo, bp) \cdot P(lo|fo) \cdot P(fo) \cdot P(bp) \tag{2.6}$$

Since these conditional probabilities needed for the simplified chain rule are exactly those provided for each random variable in the Bayesian network, computation of joint probabilities is straightforward.

Bayesian networks [22] are an intuitive method for representing uncertainty. Bayesian networks, however, do not provide a direct mechanism for representing temporal dependencies. For example, it is difficult to represent a situation such as the variability of the time of an employee's arrival at work and the causal relationships between the time of arrival and later events.

III. Theoretical Structure

This chapter, in Section 3.1, defines the formal structure of the probabilistic temporal network. The PTN's ability to model periodic and recurrent processes is presented in Section 3.2. Section 3.2 also proves the probabilistic nature of the PTN. The chapter concludes with a discussion of a closely related temporal model which has appeared very recently in the literature.

3.1 Combining Time and Probability

As previously discussed in Chapter I, the time-sliced approaches provide strong probabilistic semantics for representing uncertainty; however, they are constrained in their temporal expressiveness. The temporal abduction problem, on the other hand, has strong interval-based temporal semantics, but lacks strong probabilistic semantics.

What is needed, then, is a combined approach integrating strong probabilistic and temporal semantics. While much research has been done on point-based probabilistic temporal network models, little or no research has been identified using interval methods, specifically Allen's interval relations, for intensional probabilistic reasoning [22]. As mentioned earlier, the interval representation of time is important for the expressive set of relations available. The closest research is the temporal abduction problem discussed above which does not have strict probabilistic semantics. Recent work by Young and Santos [33][1] does present a starting point, defining the network structure for a new model.

The nodes of the probabilistic temporal network are temporal aggregates and the edges are the causal/temporal relationships between aggregates. Each aggregate represents a process changing over time. A temporal aggregate contains every interval of interest for the process. Each interval has an associated random variable giving the state of the process over that interval. Figure 3.1 depicts an example temporal aggregate modeling when a "vault" is open. The 'Vault-Open' TA is dependent on itself (VO) and two other processes (TU and CO). This example is expanded into a full network next.

[1] In which Probabilistic Temporal Networks (PTNs) are termed Temporal Bayesian Networks (TBNs) and Temporal Aggregates (TAs) are termed Temporal Random Variables (TRVs)

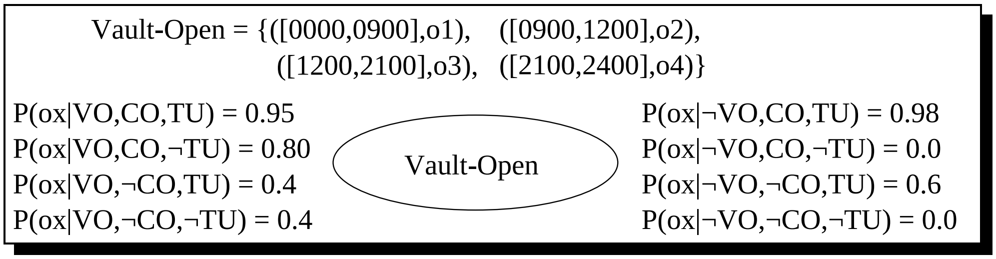


Figure 3.1 A simple temporal aggregate, 'Vault-Open,' defined over four intervals. The conditional probability tables show 'Vault-Open' to be dependent on itself through some temporal causal relationship.

As is the case in the real world, the apparent state of a process is dependent on the temporal perspective of observation. An observation made in the middle of the night as to whether or not someone is at work may return different results than if the observation is made during the day. A switch can be turned on only if, at some previous time, the switch was turned off; the light can be on only when the switch is on.

To model the effects of different perspectives on the apparent state of a process, edges in the network consist of a disjunctive set of interval relations and a schema to map the random variables of the intervals to a single value. This allows the precise selection of those intervals during which the state of one process affects another.

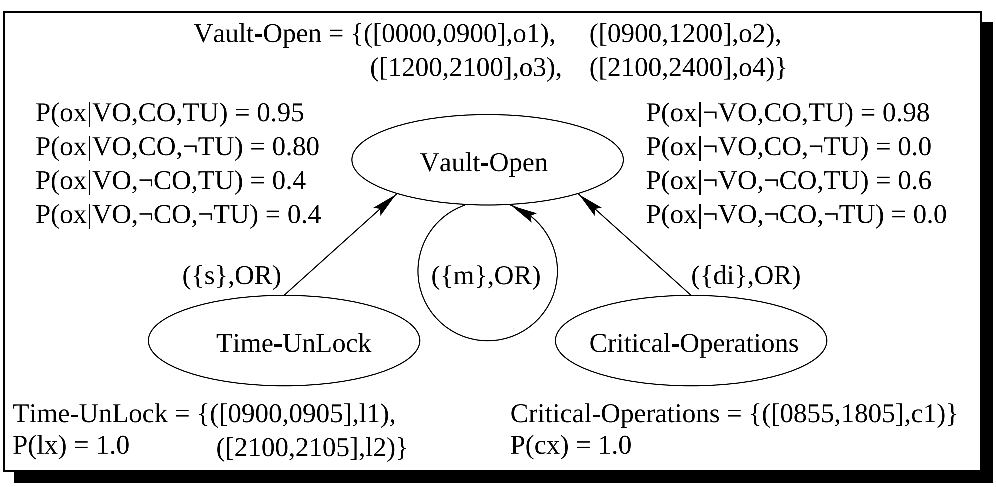


Figure 3.2 A probabilistic temporal network modeling a secure vault. This example extends the 'Vault-Open' temporal aggregate in Figure 3.2. Note that 'ox,' 'lx,' and 'cx' above are instantiated with $o_1 \ldots o_4$, $l_1 \ldots l_2$, and c_1 respectively.

Figure 3.2 shows a probabilistic temporal network modeling our secure vault scenario detailing the various components and their interactions. These network components are discussed and defined in the following paragraphs.

3.1.1 Temporal Aggregates. A process, such as 'Vault-Open' in Figure 3.2, is represented in the PTN by a temporal aggregate. Intuitively, a temporal aggregate consists of the set of states, e.g., $\{\mathsf{true}, \mathsf{false}\}$, $\{1, 2, 3\}$, or $\{\mathsf{false}\} \cup \{Red, Blue\}$, that the process can take on, and a set of temporal intervals each having an associated random variable. Each such RV has a conditional probability table defined over the states of the process.

Definition 2. *A* temporal aggregate *(TA) is an ordered pair* (T, Σ) *in which* Σ *is a set of states and* T *(pronounced* Tau*) is a set of ordered pairs* (i, r) *where* i *is a temporal interval and* r *is a random variable defined over* Σ*. For all pairs* (i_1, r_1) *and* (i_2, r_2) *in* T*,* $r_1 = r_2$ *iff* $i_1 = i_2$*. The dependencies for each random variable in the TA are defined only by temporal causal relationships between TAs.*

In the authors prior work [33], temporal aggregates (there termed temporal random variables) were allowed to have internal dependencies to model endogenous change. This was found to be a source of temporal inconsistency and better represented through self loops as demonstrated in Figure 3.2. Endogenous change is explicitly modeled in the PTN with cyclic temporal causal relationships. Endogenous change can be seen in the 'Vault-Open' process in Figure 3.2 in which the vault is more likely to stay open, given that it is open. Also note that this definition allows T to contain a potentially infinite number of interval-RV pairs. It is assumed that temporal aggregates are finite, both in T and in Σ.

'Vault-Open' is formally written, according to Definition 2, as $VO = \{\mathrm{T}, \Sigma\}$ where

$$\mathrm{T}_{VO} = \{([0000, 0900], o_1), ([0900, 1200], o_2), ([1200, 2100], o_3), ([2100, 2400], o_4)\} \quad (3.1)$$

and

$$\Sigma_{VO} = \{\mathsf{true}, \mathsf{false}\} \quad (3.2)$$

with the conditional probability table being

$$\begin{array}{lcllcl} P(o_x|VO,CO,TU) & = & 0.95 & P(o_x|\neg VO,CO,TU) & = & 0.80 \\ P(o_x|VO,CO,\neg TU) & = & 0.80 & P(o_x|\neg VO,CO,\neg TU) & = & 0.0 \\ P(o_x|VO,\neg CO,TU) & = & 0.4 & P(o_x|\neg VO,\neg CO,TU) & = & 0.6 \\ P(o_x|VO,\neg CO,\neg TU) & = & 0.4 & P(o_x|\neg VO,\neg CO,\neg TU) & = & 0.0 \end{array} \tag{3.3}$$

for all RVs o_x where $o_x \in \{o_1, o_2, o_3, o_4\}$. The $\neg$ symbol (as in $\neg TU$ above) indicates that the RV is assigned false, a non-negated RV (as in TU) indicates that the RV is assigned true.

Since $\Sigma = \{\mathsf{true}, \mathsf{false}\}$, $P(\neg o_x|VO,CO,TU) = 1 - P(o_x|VO,CO,TU)$. This condition holds for the other probabilities as well. In general, the probabilities are not explicitly shown when the probability of the true case is zero, e.g., $P(o_x|\neg VO,\neg CO,\neg TU) = 0.0$ would not be shown. Symbols used for temporal aggregates are uppercase letters from the end of the alphabet, e.g., X or Y, or uppercase abbreviations from the text name of the process being modeled, e.g., process 'Vault-Open' has a temporal aggregate denoted VO. Random variables within temporal aggregates are denoted with lowercase letters, e.g., a, b, and c or y_1 and y_2. Since the possible states of the aggregate are often evident from the conditional probability tables, Σ is often not explicitly shown. To differentiate between components of different temporal aggregates, the symbol of the component can contain the subscripted symbol of the associated TA, e.g., Σ_{VO} or $o_{1_{VO}}$.

An assignment to a temporal aggregate consists of an assignment to each interval-RV pair.

Definition 3. *A is an* aggregate assignment *(AA) iff A is a set of ordered pairs (τ, σ) where $\tau \in \mathrm{T}$ and $\sigma \in \Sigma$ such that $\forall \tau \in \mathrm{T}$, there exists an unique $\sigma \in \Sigma$ such that $(\tau, \sigma) \in A$. In other words, an* aggregate assignment *is a function from T into Σ.*

For example,

$$A_{VO} = \left\{ \begin{array}{c} ([0000, 0900], \mathsf{false}), ([0900, 1200], \mathsf{true}), \\ ([1200, 2100], \mathsf{true}), ([2100, 2400], \mathsf{false}) \end{array} \right\} \tag{3.4}$$

is an AA for the temporal aggregate VO from Figure 3.1. A_{VO} might be read "The vault was closed from 0000 hours to 0900 hours, open from 0900 hours to 2100 hours, and closed from 2100 hours to 2400 hours." The use of past tense here is arbitrary, *is closed* or *will be closed* would be equally appropriate. Aggregate assignments are denoted by uppercase letters from the beginning of the alphabet, e.g., A or B, subscripted if necessary by the symbol for the associated temporal aggregate.

Sometimes the entire state of a TA is not known. For example, we may only know that the vault was closed from 0000 to 0900. A partial aggregate assignment, which is simply a subset of an aggregate assignment, expresses this.

Definition 4. *P is a* partial aggregate assignment *(PAA) for some temporal aggregate, X, iff there exists an A such that $P \subseteq A$ where A is an aggregate assignment for X. In other words, a partial aggregate assignment is a partial function from* T *into* Σ.

Our example, where the vault is only known to be closed over one interval is thus written:

$$P_{VO} = \{([0000, 0900], \mathsf{false})\} \tag{3.5}$$

Note that P_{VO} is a subset of aggregate assignment A_{VO} above. PAAs are usually denoted by capital letters from the middle of the alphabet; however, since, by definition an aggregate assignment is also a PAA, some uppercase letters from the beginning of the alphabet may sometimes be used for PAAs.

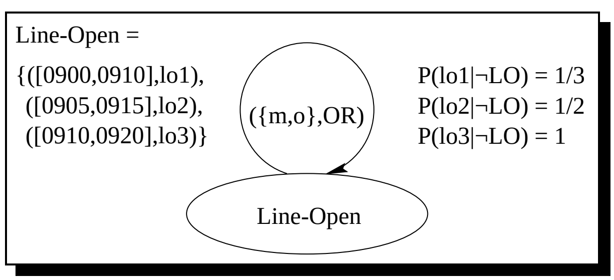


Figure 3.3 A simple, one-process probabilistic temporal network enforcing a mutual exclusion relationship. A communication line can only be opened given that it has not previously been opened.

3.1.2 Temporal Causal Relationships. How are the aggregates interconnected? The example network in Figure 3.3 shows a directed edge from 'Line-Open' to itself labeled $(\{m, o\}, \mathbf{OR})$. The edge combined with the conditional probability tables enforce a mutual exclusion constraint on 'Line-Open,' i.e., the communication line can only be opened over one of the three possible intervals. Mutual exclusion in time is an important characteristic of many processes which can occur over only one of several different intervals. Edges in the probabilistic temporal network are temporal causal relationships or TCRs.

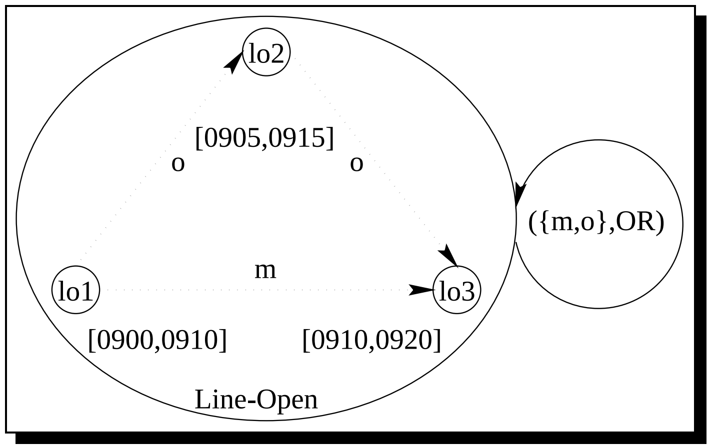


Figure 3.4 The probabilistic temporal network from Figure 3.3 decomposed to explicitly show the intervals (small circles) and the temporal relationships between intervals (dotted lines).

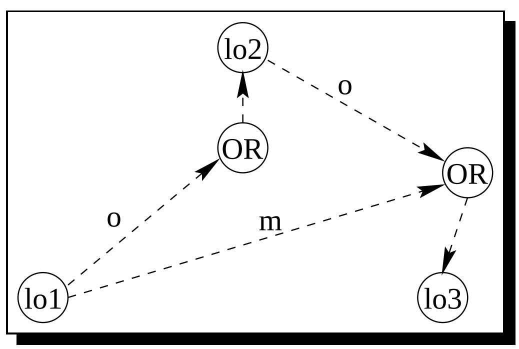


Figure 3.5 The network in Figure 3.4 with the temporal causal relationship replaced with the TCR induced random variables. The induced random variables are labeled with the name of the corresponding RV-schema, in this case, **OR**.

While portrayed graphically as a labeled edge between temporal aggregates, the TCR is actually shorthand for a set of induced random variables that enforce the temporal constraints. These random variables combine the intervals selected by a disjunctive set of interval relations (see Table 2.1), e.g., $\{m, o\}$, using the probability distribution specified by a schema, e.g., **OR**, **XOR**, **PASSTHROUGH**. Figure 3.4 shows the example network from Figure 3.3 with the intervals and temporal relations explicitly shown. For example, the dotted line from interval lo_1 to interval lo_2 shows that lo_1 overlaps lo_2.

Figure 3.5 shows the network with the TCR replaced by the appropriate induced RVs. The figure shows that the probability of the line being open over $[0910, 0920]$ is dependent on the probability of the line being open over $[0900, 0910]$ and $[0905, 0915]$, mediated by the induced **OR** random variable. Likewise, the probability of the line being open over interval $[0905, 0915]$ is dependent on the line being open over $[0900, 0910]$ again mediated by **OR**. The probability of the line being open over $[0900, 0910]$ is independent of the other probabilities.

What are the semantics behind the temporal causal relationship? The probability of some TA Y taking on some particular state over each interval is dependent on TA X taking on some state over interval(s) fitting the temporal relation, e.g., "no interval in Y can have state `true` unless that interval follows some interval in X having state `true`." This is written $X(\{<\}, \mathbf{OR})Y$ with every $(i, r) \in \mathrm{T}(Y)$ having conditional probabilities of the form $P(r|\ldots, \neg X) = 0.0$. Schemas in general and the **OR** schema in particular are further discussed below.

Definition 5. *A* temporal causal relationship *(TCR) describes a relationship between two temporal aggregates $X = (\mathrm{T}_X, \Sigma_X)$ and $Y = (\mathrm{T}_Y, \Sigma_Y)$ where X is considered the "cause" and Y the "effect." Textually, the TCR is written $X(\mathcal{R}, \mathcal{M})Y$ where $\mathcal{R}$ is a nonempty set of interval relations and $\mathcal{M}$ is a schema for describing random variables. Graphically, the TCR is presented as a directed edge from the node for X to the node for Y, labeled with $(\mathcal{R}, \mathcal{M})$. Formally, the relationship is written as the four-tuple $(\mathcal{R}, \mathcal{M}, X, Y)$.*

The TCR induces, for each interval-RV pair, (i_Y, r_Y) in T_Y, a random variable $\mathcal{M}_r$, defined over Σ_X, such that

1. *r_Y is directly dependent on $\mathcal{M}_r$.*
2. *for each $(i_X, r_X) \in \mathrm{T}_X$ where $i_X \mathcal{R} i_Y$, $\mathcal{M}_r$ is directly dependent on r_X.*
3. *for each random variable x such that $\mathcal{M}_r$ is directly dependent x, there exists an i_X such that $(i_X, x) \in \mathrm{T}_X$.*
4. *the conditional probability table for $\mathcal{M}_r$ is defined by the schema $\mathcal{M}$.*

Temporal causal relationships are rarely given explicit names. Notationally, the random variables in the interval-RV pairs in the effect TA are usually written, in the conditional probability tables, as being dependent simply on the cause TA. This can be seen in the tables for the 'Vault-Open' temporal aggregate in Figure 3.2. In cases where there is more than one TCR between two TAs, some appropriate name or symbol can be associated with the TCR and the dependencies in the effect TA can be written as the name of the cause TA subscripted with the name of the TCR.

The random variable schema algorithmically defines the conditional probability tables for the random variables induced by the temporal causal relationship.

Definition 6. *A* random variable schema $\mathcal{M}$ *takes as parameters a set of states* Σ*, a set of interval-RV pairs* T *with RVs defined over* Σ*, a single interval-RV pair* (i, r)*, and an algorithm* Δ *which together define the conditional probability table for a random variable* $\mathcal{M}_r$ *with states* Σ *such that for each* $(i_\mathrm{T}, r_\mathrm{T}) \in \mathrm{T}$*,* $\mathcal{M}_r$ *is directly dependent on* r_T*.* $\mathcal{M}_r$ *is directly dependent on nothing else. The conditional probability table for* $\mathcal{M}_r$ *is constructed with an algorithm,* Δ*.* Δ *can be either declarative or procedural.*

For many models, these schemas are extremely simple, e.g.,

$$\mathbf{OR} : \left(\begin{array}{c} \mathrm{T}, \\ \Sigma = \{\mathsf{true}, \mathsf{false}\}, \\ (i, r), \\ \Delta_{\mathbf{OR}} \end{array} \right) \rightarrow \mathbf{OR}_r \tag{3.6}$$

where $\Delta_{\mathbf{OR}}$ is defined as

Algorithm 1: ($\Delta_{\mathbf{OR}}$)

1. *Let* $(i_{\mathrm{T}_1}, r_{\mathrm{T}_1}) \ldots (i_{\mathrm{T}_n}, r_{\mathrm{T}_n})$ *be an arbitrary ordering of the elements of* T
2. *Create random variable* $\mathbf{OR}_r$ *such that for each assignment* A *to* $\{r_{\mathrm{T}_1}, \ldots, r_{\mathrm{T}_n}\}$
   (a) *If there exists an* $r \in A$ *such that* $r = \mathsf{true}$

   Let $$\begin{array}{rcl} P(\mathbf{OR}_r = \mathsf{true}|A) & = & 1 \\ P(\mathbf{OR}_r = \mathsf{false}|A) & = & 0 \end{array}$$

   (b) *else*

   Let $$\begin{array}{rcl} P(\mathbf{OR}_r = \mathsf{true}|A) & = & 0 \\ P(\mathbf{OR}_r = \mathsf{false}|A) & = & 1 \end{array}$$

Exclusive-or, **XOR**, can be defined by changing "there exists an $r \in A$" in step 2a above to "there exists a unique $r \in A$." The other logical operations are also easily defined.

The schema **PASSTHROUGH**, defined:

$$\mathbf{PASSTHROUGH} : \begin{pmatrix} \mathrm{T} = (i_{\mathrm{T}}, r_{\mathrm{T}}), \\ \Sigma, \\ (i, r), \\ \Delta_{\mathbf{PASSTHROUGH}} \end{pmatrix} \rightarrow \mathbf{PASSTHROUGH}_r \tag{3.7}$$

with $\Delta_{\mathbf{PASSTHROUGH}}$ defined as

Algorithm 2: ($\Delta_{\mathbf{PASSTHROUGH}}$)

1. *Create random variable* $\mathbf{PASSTHROUGH}_r$ *such that for each* $\sigma \in \Sigma$

$$P(\mathbf{PASSTHROUGH}_r = \sigma | r_{\mathrm{T}} = \sigma) \quad = \quad 1$$

$$P(\mathbf{PASSTHROUGH}_r \neq \sigma | r_{\mathrm{T}} = \sigma) \quad = \quad 0$$

produces a random variable for a causal relationship from a singleton TA (only one interval-RV pair in T). The temporal causal relationship

$$X(\mathcal{A}, \mathbf{PASSTHROUGH})Y, \tag{3.8}$$

read "X exerts direct causal influence on Y under all temporal relationships" is analogous to the causal relation in Bayesian networks. This type of relationship is useful when 'temporalizing' existing Bayesian networks.

3.1.3 Probabilistic Temporal Networks. A probabilistic temporal network is a directed graph in which the nodes are TAs and the edges are temporal causal relationships.

Definition 7. *A* probabilistic temporal network *(PTN) is an ordered pair (R, E) where R is a set of temporal aggregates and E is set of temporal causal relationships such that, for each TCR in E from some temporal aggregate, X, to some temporal aggregate, Y, both X and Y are in R.*

If each temporal aggregate in a probabilistic temporal network is assigned, then that PTN is said to be completely assigned. The set of all of the assignments and associated temporal aggregates forms a complete assignment.

Definition 8. *The set* $\mathscr{C}$ *containing (temporal aggregate, aggregate assignment) pairs is a* complete assignment *(CA) of some PTN* (R, E) *iff*

1. $\forall (X, A) \in \mathscr{C}$, $X \in R$ *and* A *is an aggregate assignment of* X.
2. $\forall (X, A), (Y, B) \in \mathscr{C}$, $X = Y \Rightarrow A = B$.
3. $\forall X \in R \exists (Y, A) \in \mathscr{C}$ *such that* $X = Y$.

Complete assignments are denoted by uppercase script letters from the beginning of the alphabet, e.g., $\mathscr{A}$,$\mathscr{B}$, or $\mathscr{C}$.

When inferencing over a probabilistic temporal network, incomplete evidence as to the state of the network may be held. Such evidence is represented with a partial assignment. In the simplest form, any subset of a complete assignment is a partial assignment. A more complicated case arises when only a partial aggregate assignment is known for some temporal aggregate. Since a PAA is a subset (possibly improper) of an aggregate assignment, a partial assignment to a PTN consists of a subset of the variables of the PTN and associated partial aggregate assignments for the TAs. More formally:

Definition 9. *The set* $\mathscr{P}$ *containing (temporal aggregate, aggregate assignment) pairs is a* partial assignment *(PA) of some PTN* (R, E) *iff*

1. $\forall (X, P) \in \mathscr{P}$, $X \in R$ *and* P *is a partial aggregate assignment of* X.
2. $\forall (X, P), (Y, Q) \in \mathscr{P}$, $X = Y \Rightarrow P = Q$.

PAs are usually denoted with uppercase script letters from the middle of the alphabet, e.g., $\mathscr{P}$ or $\mathscr{Q}$. As a complete assignment is a subset of itself, by definition any complete assignment is also a partial assignment.

Notation. *A partial assignment,* $\mathscr{P}$, *is said to be a subset of another partial assignment,* $\mathscr{Q}$, *(denoted* $\mathscr{P} \sqsubseteq \mathscr{Q}$*) if every* (X, P) *in* $\mathscr{P}$ *(except those having* $P = \emptyset$*) has a corresponding* (Y, Q) *in* $\mathscr{Q}$ *such that* $X = Y$ *and* $P \subseteq Q$. *A complete assignment, say* $\mathscr{C}$, *is said to be* compatible *with a partial assignment,* $\mathscr{P}$, *if* $\mathscr{P} \sqsubseteq \mathscr{C}$, *otherwise* $\mathscr{C}$ *is said to be* incompatible *with* $\mathscr{P}$. *If* $\mathscr{C}$ *is incompatible with* $\mathscr{P}$, *then at least one temporal aggregate in* $\mathscr{C}$ *has a different assignment than that in* $\mathscr{P}$.

The goal of belief revision is to find the most probable state of the world given some evidence. This is the *most probable explanation.*

Definition 10. *Let B be a PTN, let $\mathscr{P}$ be partial assignment (evidence) of B, and let $\mathscr{C}$ be some complete assignment (explanation) of B. $\mathscr{C}$ is a* most probable explanation *(MPE) given $\mathscr{P}$ iff for all $\mathscr{A}$ where each $\mathscr{A}$ is a complete assignment of B compatible with $\mathscr{P}$, $P(\mathscr{C}|\mathscr{P}) \geq P(\mathscr{A}|\mathscr{P})$.*

Since $P(\mathscr{A}|\mathscr{P}) = P(\mathscr{A}, \mathscr{P})/P(\mathscr{P})$ and an incompatible complete assignment can not be a MPE (unless the evidence $\mathscr{P}$ is itself contradictory in which case all CAs are MPEs), we only need to consider as candidates those complete assignments for which $\mathscr{P} \sqsubseteq \mathcal{A}$. Thus since $\mathscr{P} \sqsubseteq \mathscr{A}$, we derive $P(\mathscr{A}|\mathscr{P}) = P(\mathscr{A})/P(\mathscr{P})$. Furthermore, since $1/P(\mathscr{P})$ is a factor in the conditional probability of each explanation $\mathscr{A}$, to find the MPE, we need only compute the probability of each complete assignment, i.e., $P(\mathscr{A})$. $P(\mathscr{A})$ is calculated with the chain rule.

3.2 Cycles and Temporal Ordering

Now that the basic definitions and properties have been introduced, this section briefly explores the probabilistic temporal network in Figure 3.3 and considers a potential alternate representation. Figure 3.3 shows a network using a cyclic dependency to represent the internal dependencies in process 'Line-Open,' i.e., a cyclic TCR has been used to explicitly model the endogenous temporal relationships. For 'Line-Open' to be `true` over some interval, 'Line-Open' must not be `true` over any earlier intervals.

Examining the intervals, "earlier" turns out to be either *meets* or *overlaps*. This is represented with a disjunctive set containing *meets* and *overlaps*: $\{m, o\}$. The conditional dependencies are represented using the **OR** schema. The TCR, $LO(\{m, o\}, \mathbf{OR})LO$, describes the random variable $\mathbf{OR}_{lo_3}$ such that

$$P(\mathbf{OR}_{lo_3}|\neg lo_1, \neg lo_2) = 0 \quad \text{and} \quad P(\neg\mathbf{OR}_{lo_3}|\neg lo_1, \neg lo_2) = 1. \tag{3.9}$$

$\mathbf{OR}_{lo_3}$ replaces LO in $P(lo_3|\neg LO) = 1$ to yield $P(lo_3|\neg\mathbf{OR}_{lo_3}) = 1$. By using cyclic TCRs to explicitly represent the temporal relationships within a process, the knowledge engineer can more clearly "see" the nature of the system being modeled.

Figure 3.6 shows an attempt to simplify the conditional dependencies in process 'Line-Open.' The conditional probability tables for each random variable in process LO

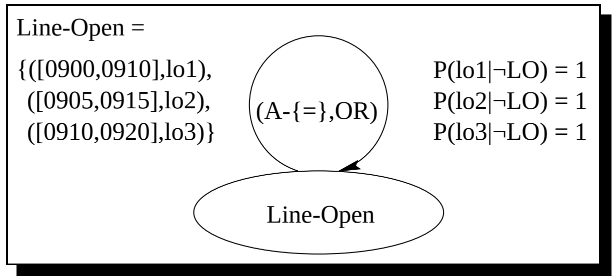


Figure 3.6 The network in Figure 3.3 rewritten using a cyclic dependency such that the conditional probability table for each RV can be written with the same probability 1 instead of the dependent probabilities 1/3, 1/2, and 1 (not well-formed).

are identical. This simplification is accomplished using the TCR $LO(\mathcal{A} - \{=\}, \mathbf{OR})LO$[2], which states that the random variable in each interval-RV pair is dependent on the random variables in all the other interval-RV pairs. While visually similar to the network in Figure 3.3, this network has a serious problem.

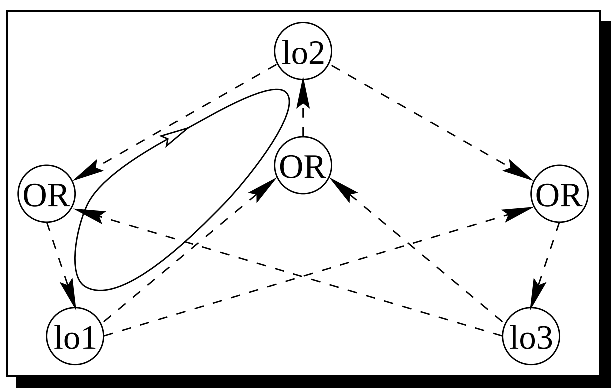


Figure 3.7 Process 'Line-Open' from Figure 3.6 drawn with the temporal causal relationship expanded. The loop shows a cycle in the dependencies.

The problem is exposed in Figure 3.7 which shows process 'Line-Open' with the TCR expanded into the induced random variables. Notice that this expanded structure reveals violations of the conditional independence assumptions discussed in the presentation of Bayesian networks. Random variable lo_2 is dependent on $\mathbf{OR}_{lo_2}$ which is dependent on lo_1 which is dependent on $\mathbf{OR}_{lo_1}$ which is dependent on lo_2 which is .... lo_2 is separated

[2]The set, $\mathcal{A} - \{=\}$, consists of all thirteen interval relations sans *equals*

from itself by random variables $\mathbf{OR}_{lo_2}$, lo_1, and $\mathbf{OR}_{lo_1}$ indicating that given knowledge of each of these variables that lo_2 is independent of itself which is clearly contradictory.

Figure 3.3 demonstrates an example in which a cycle in the PTN provided a useful representation of the internal dependencies within a process. Figure 3.6, on the other hand, shows a case in which the cycle, while intuitively satisfying, violates the requirements of conditional independence. This raises the question: "Under what circumstances are cycles appropriate in probabilistic temporal networks?"

Definition 11. *An* expanded probabilistic temporal network *(EPTN) is the directed graph created by expanding all temporal causal relationships in some PTN.*

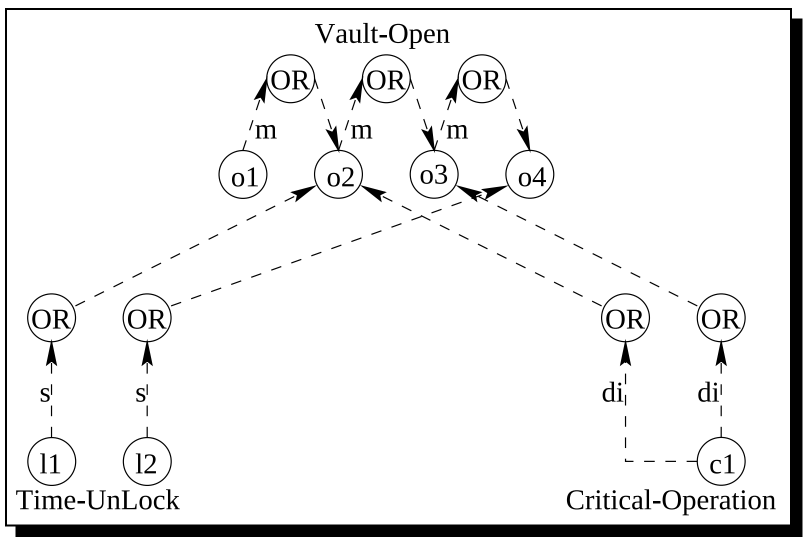


Figure 3.8 Expanded Probabilistic Temporal Network for PTN in Figure 3.2. Labels on arcs indicate temporal relation; during inverse, starts, meets.

Figure 3.8 shows the expanded probabilistic temporal network for the PTN from Figure 3.2. The **OR** node for o_1 is not shown as it has no parents and does not affect the probability distribution, i.e., $P(\mathbf{OR}_{o_1} = \mathsf{false}) = 1.0$. Note that a given EPTN is not necessarily a Bayesian network. Cycles can exist or extraneous arcs can be present, i.e., not a minimal I-map. Redundant induced RVs may also be present. Figure 3.9 presents an optimized network with an equivalent joint distribution as that of Figure 3.8. This optimization process is an avenue of further research.

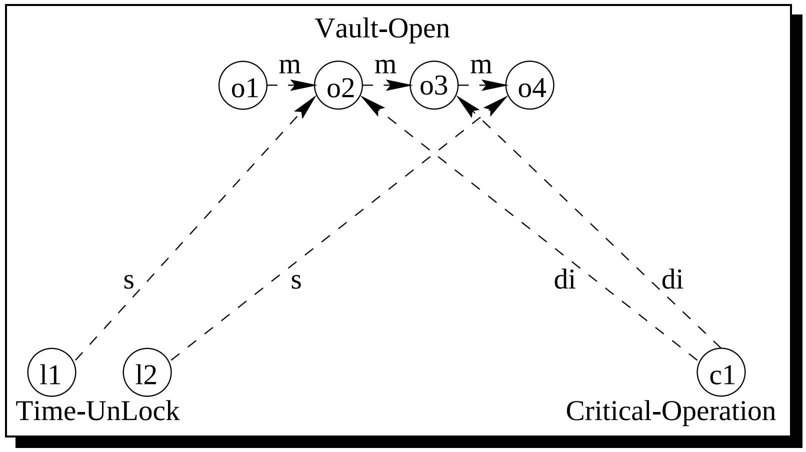


Figure 3.9 Optimized network for EPTN in Figure 3.8.

Definition 12. *A probabilistic temporal network is said to be* well-formed *iff the corresponding expanded probabilistic temporal network contains no directed cycles, i.e., the EPTN of a well-formed PTN is a directed acyclic graph.*

Figure 3.7, shown previously, gives an example EPTN with cycles. As discussed, cycles in the expanded structure are problematic. A well-formed probabilistic temporal network does not contain any such directed cycles.

Lemma 1 ([22]). *For any DAG D there exists a probability distribution P such that D is a perfect map of P relative to d-separation, i.e., P embodies all the independencies portrayed in D, and no others.*

This lemma, combined with Definition 12, leads directly to

Theorem 1. *For each well-formed, finite PTN (R, E) there exists a probability distribution P such that P embodies all the independencies in (R, E), and no others.*

Theorem 1 indicates that if we have a well-formed, finite PTN, then we have an associated probability distribution. How can we guarantee that a given PTN is well-formed and finite? If there are a finite number of temporal aggregates in the PTN and each aggregate contains only a finite number of interval-RV pairs, then the PTN is finite. As mentioned earlier, finite PTNs are assumed. Clearly if the PTN structure itself contains no cycles then there can be no cycles in the EPTN and our PTN is well-formed. The problem with this restriction is that we lose significant expressive power. Networks such as that in Figure 3.2 would not be allowed.

Cycles in the EPTN occur when an interval-RV pair becomes self-dependent. If only temporal relations which are strictly one directional are used, an interval-RV pair can not possibly be self-dependent. For example, if only $\{<\}$ is used in a PTN, no cycles are possible. Santos, in the development of the temporal abduction problem, defined the concept of *monotonicity* [25] as applied to temporal relations.

Definition 13. *A set $\mathcal{R}$ of temporal relations is said to be* monotonic *if and only if for all R in $\mathcal{R}$, $\overline{R}^c \cap (\overline{R}^c)^{-1} = \emptyset$ where $\overline{R} = \cup_{R \in \mathcal{R}} R$ and $\overline{R}^c$ is the transitive closure of R and $\overline{R}^{c-1}$ is the inverse of the transitive closure of R.*

In the same work [25], Santos introduced the following monotonic set:

Proposition 1. *The subset of relations $\mathcal{C} = \{<, o, s, fi, di, m\}$ from the original thirteen is a monotonic set.*

Intuitively, a monotonic set, such as $\mathcal{C}$ above, can be said to temporally 'point in only one direction.' This is compatible with Suppes' probabilistic theory of causality [31] and Shoham's criteria for causation [28] (both point based approaches) in which causation can only extend forward in time. For this reason, $\mathcal{C}$ is said to be the *causal* set of temporal relations. The network in Figure 3.3 holds to $\mathcal{C}$.

Theorem 2. *If, for probabilistic temporal network (R, E), there exists a* monotonic *set, $\mathcal{Q}$, of temporal relations such that for each $(\mathcal{R}, \mathcal{M}, X, Y) \in R$, $\mathcal{R} \subseteq \mathcal{Q}$; then the PTN (R, E) is* well-formed*.*

Proof. Since the only temporal relations used in the PTN are drawn from $\mathcal{Q}$ and $\mathcal{Q}$ is *monotonic*, no interval-RV pair can ever relate to itself temporally (otherwise $\overline{Q}^c \cap (\overline{Q}^c)^{-1} \neq \emptyset$) and as there can be no cycles within the TAs themselves, there can be no cycles in the EPTN; thus the PTN is well-formed. □

Combining Theorem 2 and the *causal* set $\mathcal{C}$ from Proposition 1 leads us to the following definition:

Definition 14. *A causal probabilistic temporal network (CPTN) is a PTN for which Theorem 2 holds with $\mathcal{R} = \mathcal{C}$.*

The causal PTN model enforces the constraint that causality flows forward in time. Each link in the network advances in time. When following a cycle from a temporal aggregate back to itself, one always returns to a different interval-RV pair. The CPTN model enforces, through local constraints, a consistent theory of time.

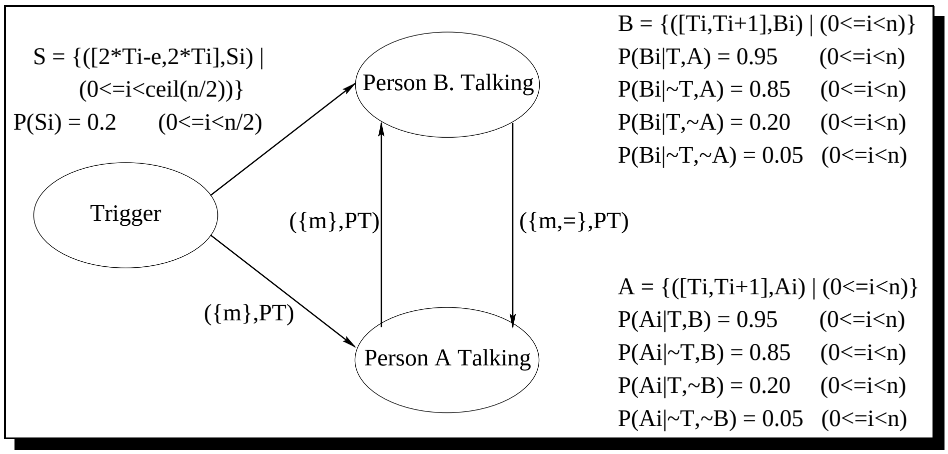


Figure 3.10 PTN modeling two people chatting with an an occasional conversational trigger. Note the use of set-builder notation.

The equals relation, '=,' is not a member of $\mathcal{C}$, and cannot be a member of any *monotonic* set of relations as '=' is its own inverse. Equals is, however, useful for expressing simultaneity. Figure 3.10 shows an example in which two people are chatting. Talker A tends to 'talk over' Talker B. To model this example, the TCR from B to A includes equals as well as meets. Figure 3.11 shows the EPTN for Figure 3.10.

To insure that a CPTN extended to use equals is well-formed, each directed cycle must have at least one TCR in which equals is not used. This guarantees 'temporal progression' in each cycle. A probabilistic temporal network limited to $\mathcal{C} \cup \{=\}$ with this broken cycle property is said to be S-Causal (SCPTN) ('S' for simultaneity).

3.3 A Related Model

In addition to the other temporal representations mentioned in Chapter II of this thesis, Aliferis and Cooper [1] have developed, in parallel with the work presented in this thesis, a preliminary temporally extended Bayesian network formulation termed the *Modifiable Temporal Bayesian Network-Single Granularity* (MTBN-SG). Their research,

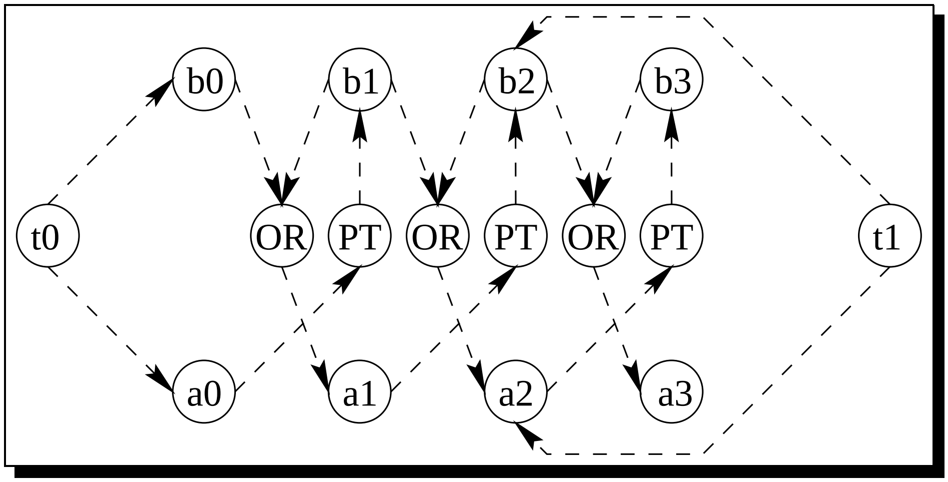


Figure 3.11 EPTN for PTN in Figure 3.10 with $n = 4$.

published shortly after the author's initial results [33], is discussed in some detail here as they also introduce the idea of representing state over several points in time as a single node in a network structure with arcs between nodes representing temporally qualified causation.

A MTBN-SG is primarily an extended time-sliced Bayesian network defined over a range of time points. Each *ordinary* node in a MTBN-SG is indexed over this entire range. Edges between nodes are represented by *mechanism* variables. A mechanism variable is a Boolean true/false random variable indicating whether the link is active, i.e., whether a dependency exists between the connected variables. Each such mechanism has an associated *lag* random variable (Delta TAs in the PTN) defined over the range of time points indicating the delay between the "cause" and the "effect." Atemporal or *abstract* random variable nodes are supported and are not instantiated for each time point. The resultant graph can have cycles to allow expressions of recurrence and feedback. As long as all cycles in the underlying joint distribution have zero probability, the graph is said to be *well-defined*.

Since the edges, both *mechanism* and *lag* components, are represented by random variables, the edges can be both dependent on and causal too other random variables in the network. This representation allows the knowledge engineer to express conditions where a relationship exists between variables only under certain circumstances. The problem with this approach is that joint distributions can be described which are not compatible with the Bayesian model. Maintaining consistency in the local probability tables across random variables then becomes a concern.

As indicated in the name, the MTBN-SG model only supports a single granularity for the size of the time step in any given network. Extending the model to support multiple granularities appears problematic, especially in the case when the granularities are not multiples, e.g., g_1 is every 10 minutes and g_2 is every 15 minutes. A perhaps more difficult problem arises in the model if the start time for one granularity is not the same as that for another as the granularities may be forever out of phase. This problem is not an issue for our model. Individual processes or temporal aggregates can be modeled with arbitrary sets of intervals. There is no requirement that the intervals in one TA match those in other TAs as the temporal causal relationship describes the desired relationships.

Intervals can be modeled in the MTBN with *abstract* variables, INT_START and INT_END, representing the start and end points of the interval respectively. INT_END is dependent on INT_START such that the end time will never be before the start time. The duration of an interval can be acquired from a third variable, INT_DUR, dependent on both INT_START and INT_END. One problem with this representation arises from the need to use *abstract* instead of *time indexed* variables. If one needs to reason with both a blend of time-sliced and interval data, then dependencies will exist between the *abstract* variables and the *time-indexed* ones.

The semantics of such arcs and the deployment transformations (conversion to BN form) are not clear. Presumably, if, in the MTBN, an *abstract* variable was dependent on a *time indexed* variable, then, in the deployed graph, the *abstract* variable would be dependent on each copy of the *time indexed* variable for each time index. If the *time indexed* variable is dependent on the *abstract* variable, then the condition is similar in that each copy of the *time indexed* variable is dependent on the *abstract* variable. These dependencies result in high degrees of fan-in and fan-out in the deployed graph leading to excessive number of needed probabilities and high complexity.

The MTBN-SG formulation, introduced by Aliferis and Cooper, is interesting in its high-level similarity to the probabilistic temporal network. Their point based approach, semantic difficulties arising from the *abstract* variables, and the single granularity restriction are problems which the PTN does not have.

IV. Reasoning

This chapter develops an approach for reasoning over probabilistic temporal networks. The first section introduces the method of calculating the probability of some state of, or explanation for, the system being modeled. Section 4.2 extends these calculations to find the most probable state of the system. Unfortunately, finding the most probable state is $\mathcal{NP}$-hard. Section 4.3 presents a subclass of the probabilistic temporal network with polynomial time solvability.

4.1 Constructing a Partial Order and Using the Chain Rule

In Section 3.1, we discussed finding the *most probable explanation.* The most probable explanation is the complete assignment with the greatest joint probability. As mentioned, this joint probability is calculated using the chain rule. To efficiently use the chain rule, a partial ordering (from effect to cause) of the random variables must exist. The ordering is drawn from the expanded PTN and can only be found when the PTN is well-formed and finite. The following algorithm finds a partial ordering for a well-formed and finite PTN:

Algorithm 3: (Partial Ordering)

1. *First, find the EPTN of a well-formed and finite PTN.*
2. *From the EPTN, select all RVs with no children. Place these first in the ordering in arbitrary order.*
3. *Find all RVs among all those not yet ordered such that all children thereof are ordered. Place these next in the ordering, again in arbitrary order.*
4. *Repeat Step 3 until no unordered RVs remain.*

For example, the PTN in Figure 3.3 expands to the EPTN in Figure 3.5. A partial ordering of the RVs is found in the following steps:

1. Order: () RVs: $\{lo_1, lo_2, lo_3, \mathbf{OR}_{lo_2}, \mathbf{OR}_{lo_3}\}$
2. Order: (lo_3) RVs: $\{lo_1, lo_2, \mathbf{OR}_{lo_2}, \mathbf{OR}_{lo_3}\}$
3. Order: $(lo_3, \mathbf{OR}_{lo_3})$ RVs: $\{lo_1, lo_2, \mathbf{OR}_{lo_2}\}$
4. Order: $(lo_3, \mathbf{OR}_{lo_3}, lo_2)$ RVs: $\{lo_1, \mathbf{OR}_{lo_2}\}$
5. Order: $(lo_3, \mathbf{OR}_{lo_3}, lo_2, \mathbf{OR}_{lo_2})$ RVs: $\{lo_1\}$

6. Order: $(lo_3, \mathbf{OR}_{lo_3}, lo_2, \mathbf{OR}_{lo_2}, lo_1)$ RVs: {}

yielding $(lo_3, \mathbf{OR}_{lo_3}, lo_2, \mathbf{OR}_{lo_2}, lo_1)$ as a partial ordering.

Since a partial ordering exists for the network, the chain rule can be used to find the joint probability of each assignment. Table 4.1 shows the probability distribution defined by the example in Figure 3.3. Only non-zero probability assignments are shown (but one).

Table 4.1 The possible complete assignments to the network in Figure 3.3 with associated probabilities. One 'impossible' assignment is also shown.

Joint Probability Table for Figure 3.3					
Line-Open					**Assignment Probability:**
[0910,0920]	**OR**$_{lo_3}$	**[0905,0915]**	**OR**$_{lo_2}$	**[0900,0910]**	
true 1	false 1	false 1/2	false 1	false 2/3	1/3 (1)
false 1	true 1	true 1/2	false 1	false 2/3	1/3 (2)
false 1	true 1	false 1	true 1	true 1/3	1/3 (3)
true 0	true 1	false 1	true 1	true 1/3	0 (4)
				Total:	1

Each joint probability in Table 4.1 is calculated using the chain rule [22]. For example, the probability of the complete assignment

$$\left\{ \left(LO, \left\{ \begin{array}{c} (([0900, 0910], lo_1), \mathsf{true}), \\ (([0905, 0915], lo_2), \mathsf{false}), \\ (([0910, 0920], lo_3), \mathsf{false}) \end{array} \right\} \right) \right\} \tag{4.1}$$

is calculated from

$$\left(\begin{array}{c} P(lo_3 = \mathsf{false} | \mathbf{OR}_{lo_3} = \mathsf{true}) \\ \cdot \\ P(\mathbf{OR}_{lo_3} = \mathsf{true} | lo_1 = \mathsf{true}, lo_2 = \mathsf{false}) \\ \cdot \\ P(lo_2 = \mathsf{false} | \mathbf{OR}_{lo_2} = \mathsf{true}) \\ \cdot \\ P(\mathbf{OR}_{lo_2} = \mathsf{true} | lo_1 = \mathsf{true}) \\ \cdot \\ P(lo_1 = \mathsf{true}) \end{array} \right) = \left(\begin{array}{c} 1 \\ \cdot \\ 1 \\ \cdot \\ 1 \\ \cdot \\ 1 \\ \cdot \\ \frac{1}{3} \end{array} \right) = \frac{1}{3} \tag{4.2}$$

4.2 Constraint Satisfaction

The previous section showed how to calculate the probability of a complete assignment to a probabilistic temporal network. This section presents a method for finding the most probable complete assignment, i.e., performing belief revision on probabilistic temporal networks. A constraint satisfaction approach is used with mixed Boolean linear programming. Constraint satisfaction has three main advantages; first, constraints can be formed to take advantage of the inherent structure of the PTN; second, very efficient algorithms developed by the operations research community are available; and finally, alternate explanations, e.g., second or third best, can be found using techniques presented in [24].

Definition 15. *A* constraint system *is a 3-tuple* (Γ, I, ψ) *where* Γ *is a finite set of variables,* I *is a finite set of linear inequalities based on* Γ*, and* ψ *is a cost function from* $\Gamma \times \{\mathsf{true}, \mathsf{false}\}$ *to* $\Re$*.*

Our probabilistic temporal network model can be considered to have a layered structure. The layers consist of temporal aggregates and temporal causal relationships. For this reason, the system of constraints is presented in two parts, those for TCRs and those for TAs. For some well-formed PTN $P = (R, E)$, the following steps produce the constraints, variables, and costs for the temporal causal relationships in E and those for the temporal aggregates in R, i.e., the following steps produce $L(P) = (\Gamma, I, \psi)$.

1. For each TCR $(\mathcal{R}, \mathcal{M}, (\mathrm{T}_X, \Sigma_X), (\mathrm{T}_Y, \Sigma_Y))$ in E,
   (a) For each $(i_Y, r_Y) \in \mathrm{T}_Y$ construct variables $\mathcal{M}^{r_Y}_{\sigma_{X_1}} \ldots \mathcal{M}^{r_Y}_{\sigma_{X_n}}$ in Γ where $\sigma_{X_1} \ldots \sigma_{X_n}$ are states in Σ_X. Set costs for each variable as

$$\psi(\mathcal{M}^{r_Y}_{\sigma_{X_i}}, \mathsf{false}) = \psi(\mathcal{M}^{r_Y}_{\sigma_{X_i}}, \mathsf{true}) = 0. \tag{4.3}$$

   where $1 \le i \le n$ and add the following constraint to I:

$$\sum_{i=1}^{n} \mathcal{M}^{r_Y}_{\sigma_{X_i}} = 1. \tag{4.4}$$

   (b) For each $(i_Y, r_Y) \in \mathrm{T}_Y$ and each $\sigma_X \in \Sigma_X$ let $(i_{X_1}, r_{X_1}) \ldots (i_{X_j}, r_{X_j}) \in \mathrm{T}_X$ be those pairs for which $i_{X_h} \mathcal{R} i_Y$ with $1 \le h \le j$, then

i. for each conditional probability of the form

$$P(\mathcal{M}_{r_Y} = \sigma_X | r_{X_1} = \sigma_{X_1} \ldots r_{X_j} = \sigma_{X_j}) \tag{4.5}$$

as induced by schema $\mathcal{M}$, construct a variable

$$q[\mathcal{M}_{r_Y} = \sigma_X | r_{X_1} = \sigma_{X_1} \ldots r_{X_j} = \sigma_{X_j}] \tag{4.6}$$

(denoted q in following steps) in Γ such that

A.

$$\psi(q, \mathsf{false}) = 0, \tag{4.7}$$

$$\psi(q, \mathsf{true}) = -\log\left(P\left(\mathcal{M}_{r_Y} = \sigma_X \;\middle|\; \begin{array}{c} r_{X_1} = \sigma_{X_1} \\ \ldots \\ r_{X_j} = \sigma_{X_j} \end{array}\right)\right) \tag{4.8}$$

B. with the following constraint in I:

$$q \geq \sum_{h=1}^{j} \mathcal{T}_{\sigma_{X_h}}^{r_{X_h}} + \mathcal{M}_{\sigma_X}^{r_Y} - j \tag{4.9}$$

(c) Let $\Upsilon_{\mathcal{M}_{\sigma_X}^{r_Y}}$ be the set of all q constructed in step (1b) for variable $\mathcal{M}_{\sigma_X}^{r_Y}$. For each such variable, add the following constraint to I:

$$\mathcal{M}_{\sigma_X}^{r_Y} = \sum_{q \in \Upsilon_{\mathcal{M}_{\sigma_X}^{r_Y}}} q. \tag{4.10}$$

2. For each TA $X = (\mathrm{T}_X, \Sigma_X)$ in R

(a) For each $(i_X, r_X) \in \mathrm{T}_X$ construct variables $\mathcal{T}_{\sigma_{X_1}}^{r_X} \ldots \mathcal{T}_{\sigma_{X_n}}^{r_X}$ in Γ where $\sigma_{X_1} \ldots \sigma_{X_n}$ are states in Σ_X. Set costs for each variable as

$$\psi(\mathcal{T}_{\sigma_{X_i}}^{r_X}, \mathsf{false}) = \psi(\mathcal{T}_{\sigma_{X_i}}^{r_X}, \mathsf{true}) = 0. \tag{4.11}$$

where $1 \leq i \leq n$ and add the following constraint to I:

$$\sum_{i=1}^{n} \mathcal{T}_{\sigma_{X_i}}^{r_X} = 1. \tag{4.12}$$

(b) For each $(i_X, r_X) \in \mathrm{T}_X$ and each $\sigma_X \in \Sigma_X$ let $\mathcal{M}_1 \ldots \mathcal{M}_j$ be those random variables induced by TCRs $(\mathcal{R}_h, \mathcal{M}_h, Y_h, Z_h)$ for which $1 \leq h \leq j$ and $Z_h = X$. Then

i. for each conditional probability of the form

$$P(r_X = \sigma_X | \mathcal{M}_1 = \sigma_{Y_1} \ldots \mathcal{M}_j = \sigma_{Y_j}), \tag{4.13}$$

construct a variable

$$q[r_X = \sigma_X | \mathcal{M}_1 = \sigma_{Y_1} \ldots \mathcal{M}_j = \sigma_{Y_j}] \tag{4.14}$$

(denoted q in following steps) in Γ such that

A.

$$\psi(q, \mathsf{false}) = 0, \tag{4.15}$$

$$\psi(q, \mathsf{true}) = -\log \left(P \left(r_X = \sigma_X \,\middle|\, \begin{array}{c} \mathcal{M}_1 = \sigma_{Y_1} \\ \ldots \\ \mathcal{M}_j = \sigma_{Y_j} \end{array} \right) \right) \tag{4.16}$$

B. with the following constraint in I:

$$q \geq \sum_{h=1}^{j} \mathcal{M}^{r_X}_{\sigma_{Y_h}} + \mathcal{T}^{r_X}_{\sigma_X} - j \tag{4.17}$$

(c) Let $\Upsilon_{\mathcal{T}^{r_X}_{\sigma_X}}$ be the set of all q constructed in step (1b) for variable $\mathcal{T}^{r_X}_{\sigma_X}$. For each such variable, add the following constraint to I:

$$\mathcal{T}^{r_X}_{\sigma_X} = \sum_{q \in \Upsilon_{\mathcal{T}^{r_X}_{\sigma_X}}} q. \tag{4.18}$$

In this construction, constraints (4.4) and (4.12) ensures that each random variable, either induced or in a TA, can take on one and only one value. Constraints (4.9) and (4.10) guarantee that each of the probabilities for TCR induced variables is computed in concordance with the appropriate temporal relations and schema. Constraints (4.17) and (4.18) guarantee that the probability of a temporal assignment to a TA is computed with the appropriate set of conditional probabilities. Variables of the form $q[r_X = \sigma_X | \mathcal{M}_1 = \sigma_{Y_1} \ldots \mathcal{M}_j = \sigma_{Y_j}]$ are called *conditional variables* in that they explicitly represent the dependencies between RVs and are the mechanism for computing the probability of any complete assignment.

For example, consider again the simple probabilistic temporal network in Figure 3.3. Section 4.1 showed how to calculate the probability of an assignment to this network using the chain rule (see Table 4.1 and Equation 4.2). Now, if we take the complete assignment

$$\left\{\left(LO, \left\{\begin{array}{c}(([0900, 0910], lo_1), \mathsf{true}),\\(([0905, 0915], lo_2), \mathsf{false}),\\(([0910, 0920], lo_3), \mathsf{false})\end{array}\right\}\right)\right\} \tag{4.19}$$

we expect our variable assignments to be

$$\left.\begin{array}{rcc} LO^{lo_1}_{\mathsf{true}} & = & q[lo_1 = \mathsf{true}|\mathbf{OR}_{lo_1} = \mathsf{false}] \\ LO^{lo_2}_{\mathsf{false}} & = & q[lo_2 = \mathsf{false}|\mathbf{OR}_{lo_2} = \mathsf{true}] \\ LO^{lo_3}_{\mathsf{false}} & = & q[lo_3 = \mathsf{false}|\mathbf{OR}_{lo_3} = \mathsf{true}] \\ \mathbf{OR}^{lo_1}_{\mathsf{false}} & = & q[\mathbf{OR}_{lo_1} = \mathsf{false}] \\ \mathbf{OR}^{lo_2}_{\mathsf{true}} & = & q[\mathbf{OR}_{lo_2} = \mathsf{true}|lo_1 = \mathsf{true}] \\ \mathbf{OR}^{lo_3}_{\mathsf{true}} & = & q[\mathbf{OR}_{lo_1} = \mathsf{true}|lo_1 = \mathsf{true}, lo_2 = \mathsf{true}] \end{array}\right\} = 1 \tag{4.20}$$

with all other variables being zero. Since the only variables which encrue costs are the $q[\ldots]$ variables, the cost of the assignment is $-\log(1/3) - \log(1) - \log(1) - \log(1) - \log(1) - \log(1) = -\log(1/3)$ and thus the probability of the assignment is 1/3 as expected. As informally demonstrated in this example, the cost of a variable assignment is found by summing the product of each variable in Γ and its corresponding cost in ψ.

Definition 16. *A* variable assignment *for a constraint system* $L = (\Gamma, I, \psi)$ *is a function* s *from* Γ *to* $\Re$*. Furthermore,*

1. *If the range of* s *is* $\{0, 1\}$*, then* s *is a* 0-1 assignment*.*
2. *If* s *satisfies all of the constraints in* I*, then* s *is a* solution *for* L*.*
3. *If* s *is a solution for* L *and is also a 0-1 assignment, then* s *is a* 0-1 solution *for* L*.*

Definition 17. *Given a constraint system* $L = (\Gamma, I, \psi)$*, we construct a function* Θ_L *from variable assignments to* $\Re$ *as follows:*

$$\Theta_L(s) = \sum_{\gamma \in \Gamma} s(\gamma)\psi(\gamma, \mathsf{true}) + (1 - s(\gamma))\psi(\gamma, \mathsf{false}) \tag{4.21}$$

Θ_L *is called the* objective function *of* L*.*

Definition 18. *An* optimal 0-1 solution *for a constraint system* $L = (\Gamma, I, \psi)$ *is a 0-1 solution which minimizes* Θ_L.

By finding an optimal 0-1 solution for a constraint system, we find the most probable explanation for the corresponding PTN. Santos [24] presents a customized algorithm using the *cutting plane method* [21] for finding the optimal 0-1 solution. Since any Bayesian network can be represented as a PTN[1], we know that, in general, belief revision over PTNs is $\mathcal{NP}$-hard [8, 22].

4.3 Polynomial Time Belief Revision—The Generalized Temporal Polytree

The previous section presented a method for performing belief revision on probabilistic temporal networks. In general, this problem is $\mathcal{NP}$-hard. However, for singly-connected PTNs (polytrees), belief revision can be done in polynomial time. A polytree is a directed acyclic graph in which no more than one path exists between any two nodes. The lack of undirected cycles in the graph structure allows for efficient local decisions. This section presents the generalized temporal polytree (GTP); a PTN model with a restricted graph and temporal structure. The EPTN for a GTP is guaranteed to be a polytree.

First, a pair of additional restrictions on the probabilistic temporal network are introduced. These two restrictions force the expanded PTN to be a causal tree, i.e., all nodes (except root nodes) have one and only one incoming edge (cause)[2]. A causal tree structure allows for very easy belief updating and revision. The first requirement is that the only interval-interval relation allowed is *meets*. *Meets* enforces a strictly monotonic progress in time and, unlike *precedes*, does not allow "temporally remote causation [31]." The second requirement is that all intervals across the network have different end-points. Together, these two requirements impose a causal tree structure on the expanded network. A probabilistic temporal network holding to these two requirements is termed a Generalized Causal Temporal Tree.

[1] Treat each RV in the BN as a TA with a single interval-RV pair, using the $(\{=\}, \textbf{PASSTHROUGH})$ TCR, and make all intervals in the TAs equivalent.

[2] Note that by this definition, the model actually allows a collection of such unconnected trees.

Definition 19. *A* generalized causal temporal tree *(GCTT) is a probabilistic temporal network in which*

1. $\mathcal{R} = \{m\}$ *for each* $(\mathcal{R}, \mathcal{M}, X, Y) \in E$*, i.e., meets is the only temporal relation allowed.*
2. *All intervals in all temporal aggregates must have unique end-points.*

Theorem 3. *The expanded probabilistic temporal network of any generalized causal temporal tree is a causal tree.*

Proof. By Contradiction. Let $P = (R, E)$ be some generalized causal temporal tree. Let N be the EPTN of P. Assuming that N is not a tree, we know by the definition that there exists a node, a, such that at least two different directed edges enter a from two different causal nodes (ignoring intervening induced RVs), say b and c. Each of these nodes (a, b, c) have associated intervals, say, $([a_s, a_e], [b_s, b_e], [c_s, c_e])$ respectively. Since, by the definition of generalized causal temporal tree, $[b_s, b_e] meets [a_s, a_e]$ and $[c_s, c_e] meets [a_s, a_e]$; $b_e = a_s$ and $c_e = a_s$ and thus $b_e = c_e$. However, again from the definition of generalized causal temporal tree, all end-points are unique and thus $b_e \neq c_e$. □

Corollary 1. *The EPTN of a GCTT in which constraint 2 in Definition 19 is changed to start-points instead of end-points, has an inverted tree structure.*

By connecting together regions with varying end-points (*out-regions*) with regions of varying start-points (*in-regions*) a PTN with polytree structure is formed. A region, then, is a collection of TAs in which all interval end or start points are different. Regions join together at a set of TAs referred to as a *join-region.* All TAs in a join-region are members of both regions being joined. For example, if an in-region and an out-region are joined, then all end-points in the join-region must be different from all end-points in the out-region and all start-points in the join-region must differ from all start-points in the in-region.

Definition 20. *A set of temporal aggregates, R, forms an* out-region *if for each*

$$([s_1, e_1], r_1) \in \bigcup_{(\mathrm{T}, \Sigma) \in R} \mathrm{T} \tag{4.22}$$

there does not exist another

$$([s_2, e_2], r_2) \in \bigcup_{(\mathrm{T}, \Sigma) \in R} \mathrm{T} \tag{4.23}$$

such that $r_1 \neq r_2$ and $e_1 = e_2$, i.e., all intervals in all temporal aggregates have unique end-points.

Definition 21. *A set of temporal aggregates, R, forms an* in-region *if for each*

$$([s_1, e_1], r_1) \in \bigcup_{(\mathrm{T},\Sigma)\in R} \mathrm{T} \tag{4.24}$$

there does not exist another

$$([s_2, e_2], r_2) \in \bigcup_{(\mathrm{T},\Sigma)\in R} \mathrm{T} \tag{4.25}$$

such that $r_1 \neq r_2$ and $s_1 = s_2$, i.e., all intervals in all temporal aggregates have unique start-points.

Definition 22. *A set of temporal aggregates, R, forms a* join-region *for two in- or out-regions, R_1 and R_2 if $R = R_1 \cap R_2$*

To prevent undirected cycles (directed cycles are prevented by the *meets* restriction), out-regions are not permitted to join to out-regions. In-regions can join with both in-regions and out-regions. No temporal causal relationships can extend, however, from a join-region back into an in-region. This prevents undirected cycles by enforcing the constraint that all inverted trees in an in-region must end in the join-region (or not enter the join-region).

Definition 23. *A* generalized temporal polytree *(GTP) is a probabilistic temporal network $P = (R, E)$ for which there exist sets I (in-regions), O (out-regions), and J (join-regions) such that*

1. *$\mathcal{R} = \{m\}$ for each $(\mathcal{R}, \mathcal{M}, X, Y) \in E$, i.e., meets is the only temporal relation allowed.*
2. *Each TA in the PTN is in some in- or out-region and vice versa.*
3. *Each join-region in J connects two in-regions or connects an in-region with an out-region. Out-regions can not join with other out-regions.*
4. *For each TCR, $(\mathcal{R}, \mathcal{M}, X, Y) \in E$, exactly one of the following must hold:*
   (a) *there exists one and only one $r \in I \cup O$ such that $X, Y \in r$, or*
   (b) *there exists a $j \in J$ such that $X, Y \in j$, or*
   (c) *there exists a $j \in J$ and an $o \in O$ such that $X \in j$ and $Y \in o$.*

In no case can X be in a join-region and Y be in an in-region outside of the join.

Theorem 4. *The expanded probabilistic temporal network of any generalized temporal polytree is a polytree.*

Proof. By Contradiction. Let $P = (R, E)$ be some generalized temporal polytree. Let N be the EPTN of P. Assuming that N is not a polytree, we know by definition that there exists at least two nodes such that two unique undirected paths exist between them. These two paths form an undirected cycle. Based on Theorem 3 and Corollary 1, there can not exist more than one unique path between any two nodes within any give in- or out-region. Also, different regions can only connect together in join-regions. Thus at least two nodes on the undirected cycle must be in the join-region. Let these two nodes be a and b. Since all nodes in the join-region belong to both in- or out- regions and no cycles can exist within any single in- or out- regions, at least one node on the cycle, say c, must exist outside of the join-region. This leads to two cases: either c is in an in-region or c is in an out-region. Either way if c is in one region and a and b in the join-region, there must be a fourth node, d, in the other region from c, otherwise the cycle would lie entirely within one in- or out-region.

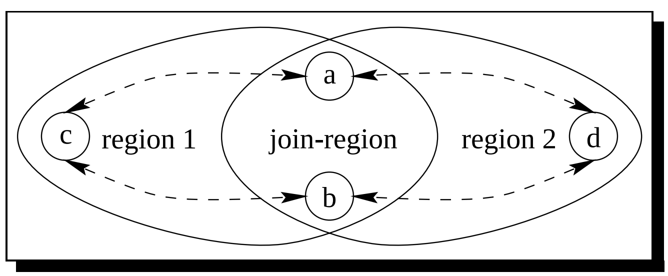


Figure 4.1 (Im)possible shape of an undirected cycle in a generalized temporal polytree.

This gives us four nodes on our cycle, a, b, c, and d. We know that a and b are both in the join-region and we know that both c and d are outside of the join-region and each in different regions. This gives us a structure as in Figure 4.1. Since out-regions can not join to out-regions, either node d or node c must lie in an in-region. Let us assume that this is node d. Since a TCR can not extend from the join-region out into an in-region, a TCR must extend from the TA containing d into the join-region. This TCR must be such that the interval associated with d *meets* two nodes in the join-region, however since all nodes in the join-region are also in the same in-region as d, no two nodes in the join-region can have the same start point and thus d can not meet these two nodes and thus an undirected cycle can not exist. □

Although not stated in the formal definition of the generalized temporal polytree, interval start-points in evidence TAs and end-points in leaf TAs do not need to be different from other start- or end-points as evidence nodes are not dependent on anything and nothing is dependent on leaf nodes.

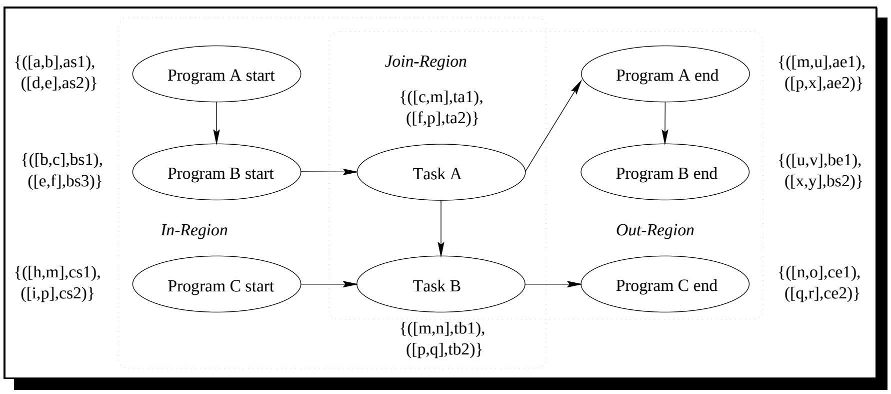


Figure 4.2 A Generalized Temporal Polytree depicting a program execution scenario.

Figure 4.2 shows a GTP modeling a program execution scenario. Program-A executes Program-B to complete Task-A. Program-C must complete Task-B. Task-B, however, requires that Task-A complete immediately prior. The start and task TAs form an in-region and the task and end TAs form an out-region. Task-A and Task-B together form a join-region. Figure 4.3 shows the expanded probabilistic temporal network for this GTP.

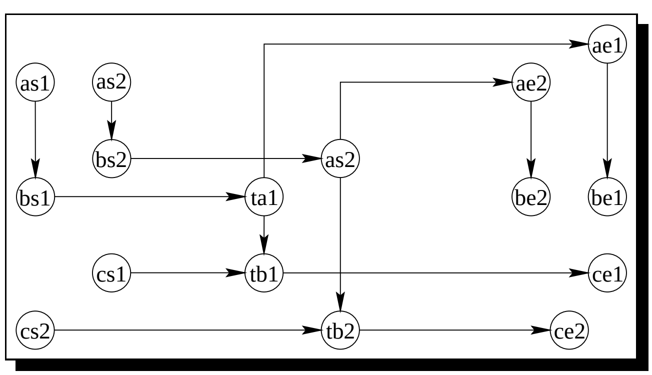


Figure 4.3 The EPTN for the GTP in Figure 4.2.

V. Knowledge Engineering

The probabilistic temporal network provides the knowledge engineer with a powerful tool. This chapter discusses techniques for applying the PTN to particular problems. The first section is focused on extending an existing knowledge base into the temporal domain. In the second section, further techniques appropriate for completely temporal models are discussed.

5.1 Extending a Bayesian Network with Time

Probabilistic temporal networks provide an easy migration from a *timeless* Bayesian representation to a fully temporal representation. For example, consider the Bayesian network in Figure 5.1 representing the following scenario:

> Tech support is only available if the phones are working and the support technician has arrived at work. The probability that the phones are working is 0.95 and that the support technician has arrived is 0.875.

This scenario is easily and adequately modeled with a Bayesian network. Suppose that we also have the following additional requirement:

> The support tech has a fifty percent chance of starting work between 7:15am and 7:45am, 25 percent chance between 7:45am and 8:15am, and a 12.5 percent chance between 8:15am and 8:45am. If the tech is not in by 8:45am, she is not coming in at all.

To reflect this change, the Bayesian network in Figure 5.1 would have to be modified to explicitly contain each of the above three intervals with support-available being dependent on all three. The probabilistic temporal network approach provides a cleaner alternative.

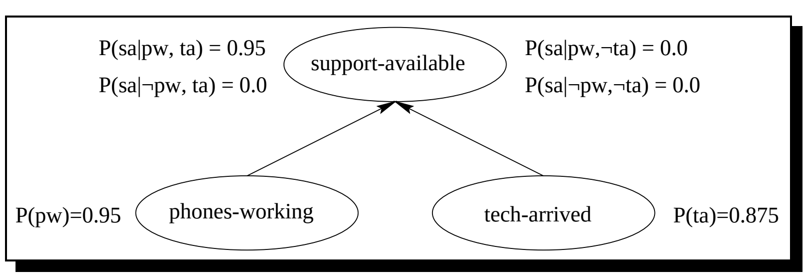


Figure 5.1 A Bayesian network for a simple tech support scenario.

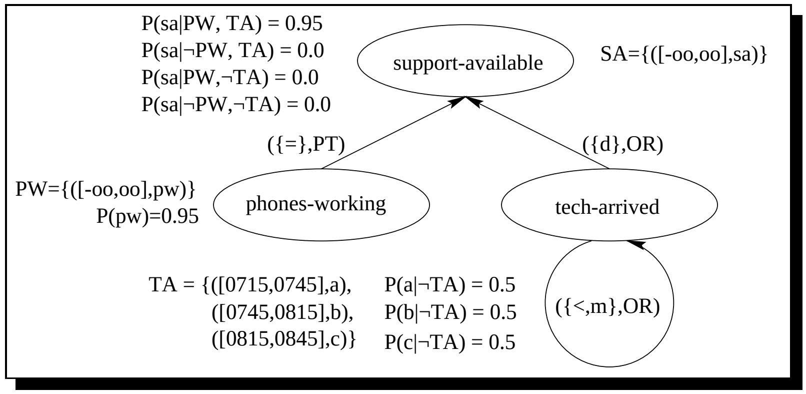


Figure 5.2 Tech-support Probabilistic Temporal Network. The probabilities used in the figure are the dependent probabilities rather then the break-out used in the text description, e.g., $(0.5, 0.25, 0.125)$ becomes $(0.5, 0.5, 0.5)$. **PT** is shorthand for **PASSTHROUGH**.

5.1.1 Temporalizing a Bayesian Network. First, however, a technique must be found to provide a temporal binding for the nodes in the Bayesian network. Since BNs usually do not contain explicit temporal information, we represent the nodes in Bayesian networks as temporal aggregates defined over a single interval from negative infinity to positive infinity[1]. Thus each RV X with states Σ from a BN becomes the TA $X = (\{([-\infty, \infty], x)\}, \Sigma)$. Then, under the assumption that $[-\infty, \infty] = [-\infty, \infty]$, the temporal relationship between TAs X and Y, where an edge exists from X to Y in the BN, is simply $X(\{=\}, \textbf{PASSTHROUGH})Y$.

Then, if random variable x in temporal aggregate X holds state σ, we can interpret this to mean that X holds state σ for *all time*[2] (or at least for the time of discourse). If TA X represents a boolean proposition, then we could instead interpret $x = \mathsf{true}$ to mean that at *some time* our boolean proposition holds and if $x = \mathsf{false}$ then at *no time* does the proposition hold. This is the interpretation used in our examples here.

Notation. *For convenience, a temporal aggregate so adapted from a Bayesian network is termed an* adapted temporal aggregate (ATA).

[1]While an open interval may be more proper, the closed interval $[-\infty, \infty]$ is used for consistency of notation. Practically, $[-\infty, \infty]$ could be replaced by any interval containing the time of discourse.

[2]Keep in mind that the temporal interval is the *primitive temporal individual* and thus when we talk about a 'time' we are talking about an interval and not a point. If we say 11 AM, an interval such as $[1100, 1101]$ is implied.

Figure 5.2 shows the PTN modeling the above scenario. Since explicit temporal information is provided, 'tech-arrived' is represented by a temporal aggregate defined over three intervals. The edge between 'tech-arrived (TA)' and 'support-available (SA)' is replaced with $(\{d\}, OR)$ indicating that "Support is available if the tech arrives." TA is also dependent on itself with the TCR $TA(\{<, m\}, \mathbf{OR})TA$ constraining TA, when combined with the conditional probability tables, to be `true` over only one interval.

Using this technique, any Bayesian network can be represented with a probabilistic temporal network! As additional temporal information is gathered, temporal aggregates can be modified to contain the actual times rather than $[-\infty, \infty]$. Semantically, the transformation can be awkward since the direction of causality within BNs can raise implicit temporal constraints. It remains the task of the knowledge engineer to complete the 'temporalization' of the model.

5.1.2 The time of reference. Something is still missing. The network in Figure 5.2 can tell us *if* tech support is available but we can't tell *when*. In other words, Figure 5.2 can answer the question "Is tech-support ever available?" but not the question "It is 12pm. Is tech-support available now?" A time of reference is needed.

As mentioned previously, each TA X, adapted from the original Bayesian network, can be interpreted as indicating if the proposition associated with X holds at *some time*. We need a mechanism to determine what that time is. Consider 'support-available' and 'phones-working' in Figure 5.2. We could change the interval from $[-\infty, \infty]$ to something like $[t, t + \epsilon]$ but then our reasoning algorithms and the structure of the network would have to be changed to constrain t. Instead, we take a different approach.

Consider again the original Bayesian network in Figure 5.1. If one asks "At what time is support available?" Intuitively, we answer "When the phones are working and the tech has arrived." If one then asks "It is between 0715 and 0745 and the phones are working. Is support now available?", 'support-available' is only dependent on 'phones-working' and 'tech-arrived' during the interval specified even though this interval is not expressed anywhere in the network.

Now consider the PTN in Figure 5.2. If one again asks "It is between 0715 and 0745 and the phones are working. Is support now available?", we can't answer the questions as 'support-available' is not dependent on only 'tech-arrived' during $[0715, 0745]$ but also for $[0745, 0815]$ and $[0815, 0845]$. Tech support is available *some time.* If, however, 'tech-arrived' is forced to be false after $[0745, 0815]$ (zero probability), then 'support-available' is effectively no longer dependent on 'tech-arrived' after $[0745, 0815]$. The following calculations show this:

Let $\mathscr{P}$ be the partial assignment for our query.

$$\mathscr{P} = \{(PW, \{([-\infty, \infty], \mathsf{true})\}), (TA, \{([0745, 0815], \mathsf{false}), ([0815, 0845], \mathsf{false})\})\}. \quad (5.1)$$

Let $\mathscr{C}$ be some complete assignment compatible with $\mathscr{P}$. Using the chain rule, we derive

$$P(\mathscr{C}|\mathscr{P}) = P(sa|pw, TA) \cdot P(pw) \cdot P(TA) \quad (5.2)$$

$$P(TA) = P(\mathbf{OR}_{sa}|c, b, a) \cdot P(c|b, a) \cdot P(b|a) \cdot P(a) \quad (5.3)$$

Since we know $b = \mathsf{false}$, $c = \mathsf{false}$, and $pw = \mathsf{true}$ we can simplify the calculation to

$$P(\mathscr{C}) = P(sa|pw, TA) \cdot P(TA) \quad (5.4)$$

$$P(TA) = P(\mathbf{OR}_{sa}|c, b, a) \cdot P(a) \quad (5.5)$$

and since $P(\mathbf{OR}_{sa}|c, b, a) = 1$ if $a = \mathsf{true}$ and $P(\mathbf{OR}_{sa}|c, b, a) = 0$ if $a = \mathsf{false}$, $P(TA) = P(a)$. We can now write

$$P(\mathscr{C}) = P(sa|pw, a) \cdot P(a) \quad (5.6)$$

Thus 'support-available' is only dependent on 'tech-arrived' during $[0715, 0745]$.

The idea of forcing falseness for future propositions is compatible with our intuitions about causality. If we are interested in the state of the world at present, it can not be dependent on what hasn't yet occurred.

This research does not, however, take the approach of simply clamping the future states to false as that approach only allows forward reasoning. One also wants to reason backwards, e.g., to find the most probable time for support to be available. So instead, a different approach is taken—introducing a new temporal aggregate, *Now*, containing an

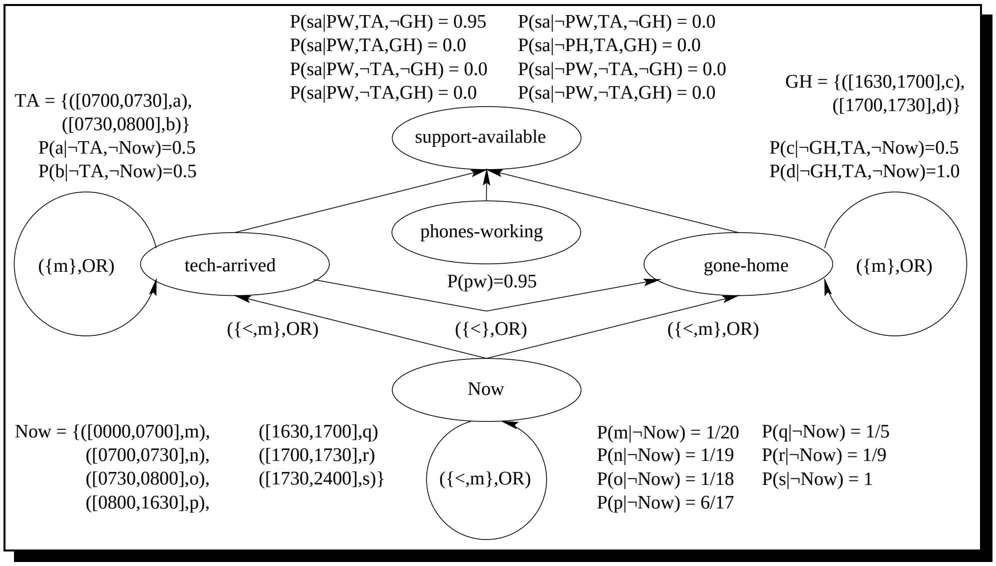


Figure 5.3 Probabilistic temporal network demonstrating time of reference. Empirical evidence of the density of support calls is used to assign probabilities associated with intervals in *Now*.

interval for each time of interest. By making other (non adapted) temporal aggregates dependent on *Now* in such a way that a given TA can not be true over an interval unless *Now* is false for all earlier intervals, we can use *Now* to block future events. If the time of reference is known, *Now* can be clamped to true for that interval, preventing TAs from being true at any time after the current time. Also, if some state of the world is clamped, then belief updating can be used to determine what the most probable time is.

Figure 5.3 models the following scenario using the *Now* construct.

> Tech support is only available if the phones are working and the support technician has arrived at work and is not at lunch. The phones almost always work. The support tech has a fifty percent chance of arriving between 07:00 and 07:30 and a 25 percent chance between 07:30 and 08:00. If the tech is not in by 8:00am, she is not coming in at all. The tech has a fifty percent chance of going home between 16:30 and 17:00 and a fifty percent chance between 17:00 and 17:30, given that she comes in at all.

The *Now* temporal aggregate allows what-if queries where *Now* is unclamped to predict at which time events are most likely to happen as well as queries where the time is clamped to determine the most probable state of the system. Different strategies can be used for the conditional probabilities in *Now*. The PTN in Figure 5.3 uses probabilities

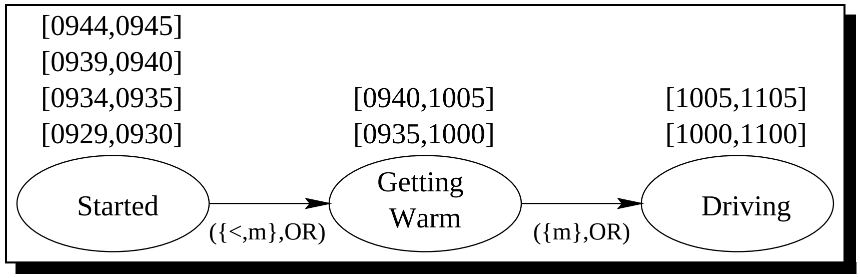


Figure 5.4 A simple network using a delta TA. Event A can occur over many different intervals, however B is only dependent on A if A occurs over at least 30 minutes.

for Now representing the density of support calls. Since Now is weighted by call density, we can make queries such as "What is the most likely time for a client to call, but support not be available?" It does not,however, allow meaningful predictive queries of the form "Given that it is before 0700, when is the most likely time for the tech to go home?" since if the time of reference is given, the future is clamped to false.

5.1.3 Temporally quantified causation. It is straightforward to model that cause must precede effect. It is not, however, straightforward to model by how much the cause must proceed the effect. With the thirteen basic interval-interval relations there is no direct way to quantify the temporal distance between cause and effect. Should our representation support temporally remote causation at all? Patrick Suppes, in his *A probabilistic theory of causality* strongly rejects the concept: "There is almost a feeling of ludicrousness in the idea of one body acting on another at a slow and leisurely pace from remote time and space. [31]" In principle the author concurs with this philosophy, however, in practice holds that the infinity of causes and effects lying between a remote cause and the resulting effect, can not be represented efficiently in a computational model. These myriad underlying mechanisms are merely another source of uncertainty.

In a probabilistic temporal network, to model that one process relates to another, an edge is created from the causal TA to the effect TA with a temporal causal relationship containing a set of temporal relations and a random variable schema. One could create new random variable schemas such as **DELTA_OR** such that the induced random variables are independent of those intervals in the causal process that do not satisfy the additional quan-

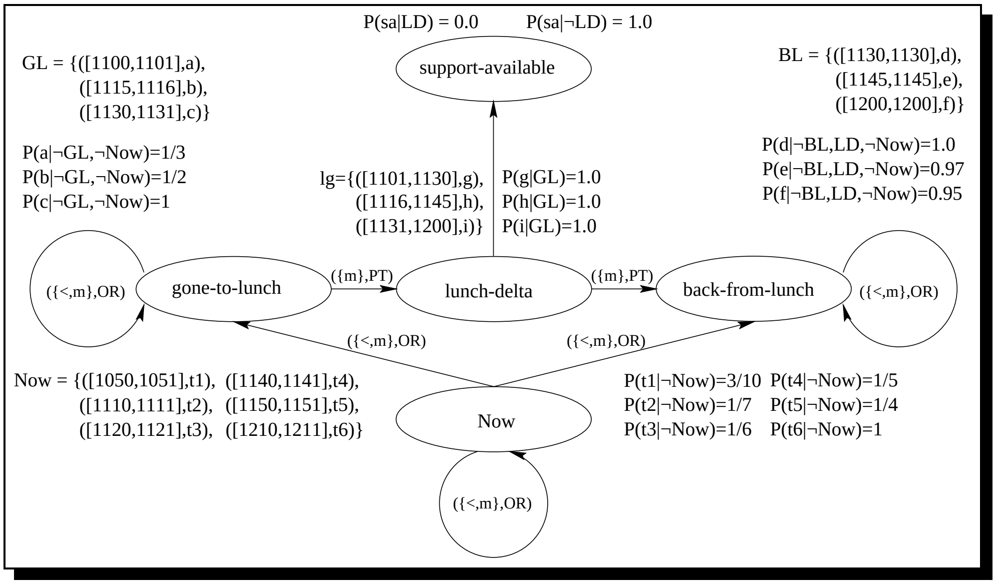


Figure 5.5 Probabilistic temporal network modeling a tech-support representative going to lunch and maybe not returning to work.

tified constraints but do satisfy the temporal relations. The problem with this approach is that many different RV schemas would be needed (for duration, overlap, precedes, etc.). Instead of creating new schemas, simply introduce a new temporal aggregate between the cause and effect to enforce the quantification.

The new temporal aggregate, termed a *delta TA*, lies between the the cause and effect TAs. The delta TA contains a set of intervals enforcing the quantification. A TCR from the cause TA to the delta TA selects the appropriate intervals in the cause TA. A TCR from the delta TA to the effect TA passes on the causal information. Figure 5.4 shows a simple example modeling an old car starting on a winter morning. The car must have been started at least twenty-five minutes before it can start moving. In this simple example the delta TA is named 'Getting Warm'. Figure 5.5, modeling the following scenario, shows an application of the delta TA to our tech support realm in which lunch is always 30 minutes. The delta TA is a useful tool for designing probabilistic temporal networks in general, not just for extending Bayesian networks.

> Tech support is only available if the support technician is not at lunch. The tech always goes to lunch with equal likelihood at 11:00am, 11:15am, or 11:30am. Lunch lasts exactly one half-hour. Going to lunch later in the day slightly increases the chances that the tech will not return to work. In particular we are

interested in times 10:50am (or earlier), 11:10am, 11:20am, 11:40am, 11:50am, and 12:10am (or later).

5.2 Some Guidelines for Building a Temporal Knowledge Base

Using PTNs to extend an existing knowledge base with explicit temporal information is effective, however, the disadvantages of carrying along the atemporal semantics of BNs are significant. These disadvantages include the necessity of constructs such as Now to provide temporal reference. Ultimately, the knowledge base should be built ground up with explicit temporal information. This section briefly presents a few guidelines for developing probabilistic temporal networks.

5.2.1 General Guidelines. Unlike Bayesian networks, PTNs allow cycles. Cycles are very important for representing recurrence, periodicity, and endogenous change, however, they can be a two-edged sword as they introduce the need to avoid cycles in the underlying probability structure. By only using a monotonic set of temporal relations, the need to check for cycles can be avoided. Furthermore, by using using the *causal* set, $\mathcal{C} = \{<, o, s, fi, di, m\}$, one gets the added benefit of temporal consistency, i.e., causality only extends forwards in time. Networks restricted to just $\mathcal{C}$ are termed *Causal PTNs.*

While philosophically debatable, it is often necessary to represent simultaneity in practical systems. In point based temporal models, simultaneity is represented with the equals relation. This is also true for interval models. CPTNs, though, do not allow '=.' Since cyclic dependencies arise when a TA becomes dependent on itself over some interval, we can allow '=' as long as for every cycle in the CPTN, at least one TCR on the cycle, does not use '=.' Such a network is termed a *S-Causal PTN*. SCPTNs require that cycles in the PTN must be checked; however, this check is much simpler then that required for PTNs in general.

Inference over probabilistic temporal networks is $\mathcal{NP}$-hard [8, 22]. This constrains the size of networks that can be reasoned with. The *generalized temporal polytree* defines a class of PTNs for which inference is polynomial. These types of networks are useful for modeling systems in which can be grouped temporally by starting, working, and finishing.

Figure 4.2 shows an example in which three concurrent programs are being executed to complete two tasks.

5.2.2 Processes and Events. Intervals can model both processes[3] and events [3]. Processes are generally described by 'ing' words such as *walking* and *talking*. If $([a,b],w)$ is an interval-RV pair for a process such as 'tech arriving at work,' then w is true implies that 'arriving at work' holds for all intervals contained in $[a,b]$ also. An event, on the other hand, does not hold for all sub-intervals. Consider 'tech arrived at work,' again represented by $([a,b],w)$. Just because w is true, we can not assert that the tech arrived at work during some subinterval of $[a,b]$. Our model does not explicitly differentiate between events and processes. The knowledge engineer can represent either.

Often, the exact interval that an event occurs in is not known. In this case, the interval-RV pairs in a temporal aggregate represent intervals *during* which the event may take place. In other words the interval encapsulates small scale uncertainty that is not important to the situation being modeled.

5.2.3 Mutual Exclusion. Many situations contain events that are not recurrent, i.e., they either do not happen, or they happen exactly once. For example, consider a light-bulb that may burn out sometime in the scenario. The bulb can burn out only once (no replacement), and may not burn out at all. Such events are referred to as *one-shots* and can easily be represented with a temporal aggregate with a single, self dependent, temporal causal relationship.

To construct the temporal aggregate, we need to decide on Σ, the set of states, and on T, the set of intervals we are interested in. Since we are modeling something that can either happen or not, we have only two states. Let true indicate that the event happens *during* the interval and false indicate that the event does not happen. How many intervals are needed? This depends on the resolution needed to model the situation; we will use three consecutive interval-RV pairs, $\{([a,b],r_1),([b,c],r_2),([c,d],r_3)\}$, representing the intervals

[3] Processes is not used here in the sense of what a temporal aggregate represents.

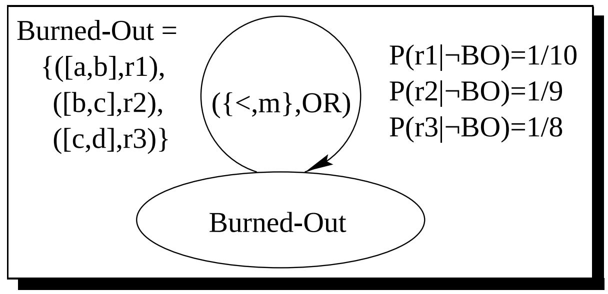


Figure 5.6 Probabilistic temporal network modeling a light-bulb. Demonstrates mutual exclusion relationship between intervals.

during which the light might burn-out. There is no requirement that these intervals be consecutive, they can overlap (see Figure 3.3) or be disjoint.

How probable is it the bulb will burn out? If it is certain that the bulb will burn out (during the scenario) then the sum of the independent probabilities must be one, in any case $P(r_1) + P(r_2) + P(r_3) \leq 1$. However, since there is a mutual exclusion relationship between the intervals, r_1, r_2, and r_3 are not independent random variables. Instead we must order the interval-RV pairs and make each pair dependent on all prior pairs. As far as the probabilities are concerned, there is no preference for the ordering, however, semantically, the ordering should be from 'earliest' to 'latest' where 'earliest' might be defined by the causal set, $\mathcal{C}$, of temporal relations (see Proposition 1 on Page 3-15). For our problem, r_3 is dependent on r_2 and r_1, r_2 is dependent on r_1, and r_1 is dependent on nothing.

The next step is to convert our independent probabilities to conditional probabilities. Let us assume that there is a 10% chance that the bulb will burn out during each of the intervals, e.g., $P(r_x) = 0.10$. Clearly, since r_1 is dependent on nothing, $P(r_1) = 1/10$. This leaves 90% left. r_2 will take the next 10% or 10/90, so $P(r_2|r_1) = 1/9$. This leaves 80% for r_3 so $P(r_3|r_1, r_2) = 10/80 = 1/8$. If the bulb burning-out was certain, that is $P(r_1) + P(r_2) + P(r_3) = 1$, then the conditional probability for the 'latest' RV would be 1.

We have now defined our temporal aggregate item $BO = (\mathrm{T}, \Sigma)$ such that

$$\mathrm{T} = \{([a,b], r_1), ([b,c], r_2), ([c,d], r_3)\} \quad \text{and} \quad \Sigma = \{\mathsf{true}, \mathsf{false}\} \tag{5.7}$$

and computed our conditional probabilities. The next step is to define the temporal causal relationship to capture the conditional dependencies. Since our three intervals are consecutive, only two temporal relations are needed, *meets* and *precedes*. Since the bulb can not burn-out in an interval if it burned-out in any prior interval, we can use the **OR** schema. This gives our TCR as

$$BO(\{<, m\}, \mathbf{OR})BO \tag{5.8}$$

and our final conditional probability tables as

$$\begin{aligned} P(r_1|\neg BO) &= 1/10 \\ P(r_2|\neg BO) &= 1/9 \\ P(r_3|\neg BO) &= 1/8. \end{aligned} \tag{5.9}$$

Figure 5.6 shows the corresponding network.

5.2.4 Sure Events. A sure event is a fact that we want to explicitly model in the PTN. An example of a sure event can be seen in Figure 3.2 in which 'Critical-Operations' is shown defined over only one interval, $[0855, 1805]$, with probability of one. We could, instead, have changed our conditional probability tables for 'Vault-Open' instead of even having the 'Critical-Operations' aggregate.

The reason 'Critical-Operations' is explicitly modeled is twofold. First, critical operations has causal influence on 'Vault-Open' and, as such, should be explicitly modeled. Secondly, by explicitly modeling 'Critical-Operations' we get the added benefit of having only one conditional probability distribution which applies to all interval-RV pairs in 'Vault-Open' rather than having a different one for each pair.

Why is only one interval needed in 'Critical-Operations?' Since when critical operations are not occurring, we would want the temporal aggregate to appear false, one would expect the additional intervals $[0000, 0855]$ and $[1805 - 2400]$ each having probability of true set to zero. However, because the **OR** schema is designed such that if no interval exists satisfying the temporal relation, the constructed random variable has a zero probability of being true. This property gives us the advantage of minimizing the number of intervals needed to express processes with true and false states.

VI. Recommendations and Conclusions

This chapter presents recommendations for several avenues of future research. None of these recommendations are show stoppers—the probabilistic temporal network, as it stands, is an excellent representation for complex, dynamic systems. However, since the PTN is unproven, the most important step for further effort is to demonstrate this excellence by implementing a real-world, large-scale model. Good domains for this large-scale model include security analysis and medical diagnosis.

6.1 Recommendations for Future Research

The probabilistic temporal network can represent very complicated and traditionally difficult domains. This research has focused on exploring recurrence and periodicity, temporal spacing between cause and effect, and modeling the time-of-reference. These are traditional problems for temporal models. Current and future efforts are focused on exploring these and other knowledge engineering issues.

This thesis introduces a constraint satisfaction formulation for performing belief revision (Section 4.2). This formulation needs to be extended to perform belief updating (finding the most likely state of a given interval-RV pair or temporal aggregate). The constraint set needs to be enhanced to take better advantage of the structure imposed by our network structure.

Performing belief revision is in general $\mathcal{NP}$-hard. To address this, the generalized temporal polytree was introduced, which, because of the polytree structure of its dependencies, allows polynomial time belief revision. We are currently investigating practical domains for which the GTP is tenable. The question also remains as to what exactly the maximal tractable class of PTNs is.

Overlapping intervals in a temporal aggregate are troublesome. The theory, as it stands, allows overlapping intervals so that events happening over intervals can be expressed. For example, if a switch could be on from 1000 to 1030 or 1015 to 1045, this condition could be represented as $\{([1000, 1030], S_0), ([1015, 1045], S_1)\}$ where S_0 and S_1 are random variables for the switches position. S_1 would be conditioned on S_0 to prevent the switch from being on over both intervals. The problem arises in that now the switch could be considered both on and off in the interval $[1015, 1030]$. Originally, this wasn't considered a problem as the temporal causal relation (TCR) resolved any ambiguity from the perspective of the caused process. One possibility is to make the interval itself random. For example $\{([1000, 1030], S_0), ([1015, 1045], S_1)\}$ might become $\{(I, On)\}$ where $P(I = [1000, 1030]) = P(S_O|\ldots)$ and $P(I = [1015, 1045]) = P(S_1|\ldots)$. This solution gets us to only one interval; however, there are now two sorts of probabilities to deal with when doing computation.

Most work to date has been within the discrete realm. Future research will focus on modeling continuous domains. Using continuous, rather than discrete, sets of states (Σ) in temporal aggregates is straightforward. For example, we might have a TA, $Temp = (\mathrm{T}_T, \Sigma_T)$ where $\mathrm{T}_T = \{([0000, 0100], t_1), \ldots, ([2300, 2400], t_{24})\}$ and $\Sigma_T = \Re$. $Temp$ models changes in the peak temperature over the course of a day. We could have a second TA, $NitDay = (\mathrm{T}_N, \Sigma_N)$ where $\mathrm{T}_N = \{([0000, 0700], n_1), ([0700 - 1900], n_2), ([1900, 2400], n_3)$ and $\Sigma_N = \{\texttt{night}, \texttt{day}\}$. With these two TAs, we would like to model peak temperature changing over the course of the day. Temperature during a given hour is dependent on whether or not it is day or night, on the temperature during the previous hour, and on the rate of change between the previous two hours. Constructing the network structure is trivial (see Figure 6.1).

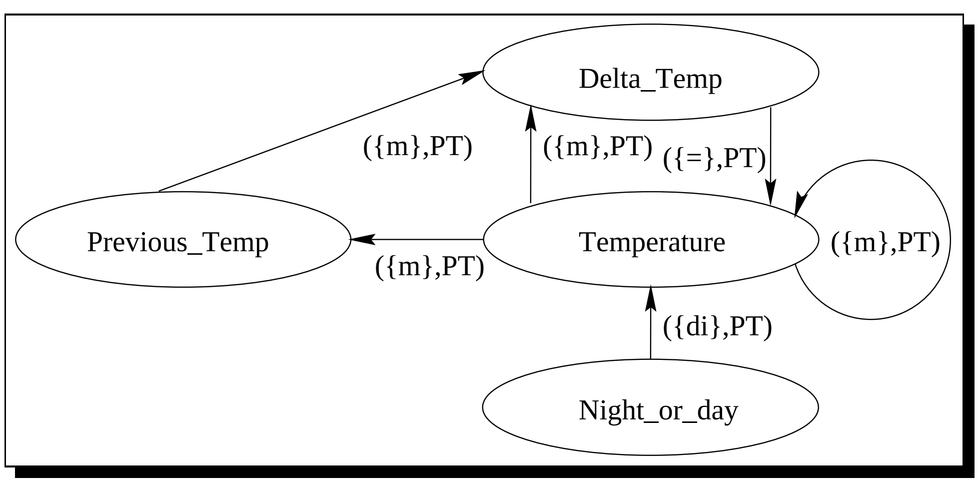


Figure 6.1 A probabilistic temporal network modeling peak temperature changing over the course of a day.

The difficulty arises in developing appropriate continuous distribution functions for the domain and representing the causal connection between processes (developing appropriate random variable schema) as well as the conditional dependencies in the caused process. Also, even with continuity in states, without continuity in time, continuous change can not truly be represented. A potential approach is to use a structure similar to the one discussed for dealing with overlapping intervals in which a continuous density function is used to give the probability distribution over the temporal interval space.

Among the avenues for further research discussed here, modeling continuous change is perhaps the most interesting. In a sense, being able to represent continuity would "complete" the probabilistic temporal network model, allowing the model to fully represent natural systems.

6.2 Conclusion

The research, presented here, develops a new knowledge representation unique in its ability to represent both time and uncertainty. The technique, the probabilistic temporal network, draws from the independence semantics of Bayesian networks and from the temporal representation in the interval algebra. The proven probabilistic nature allows knowledge engineers to drawn on previously developed statistical data as well as the entire field of probability theory. This property is crucial for developing well defined, non ad hoc models.

By directly representing processes as temporal aggregates and modeling the causal relationships between the processes with temporal causal relationships, complex systems of interacting processes can be modeled. Being able to model such systems is crucial to successfully automating domains such as medical diagnosis, story understanding, planning and scheduling, and financial forecasting. Mastery of these and related domains, such as security analysis and combat modeling, is crucial for the continued success of the United States Air Force. These domains all share in common the need to reason with both time and uncertainty—the domain of the probabilistic temporal network.

Bibliography

1. Aliferis, Constantin F. and Gregory F. Cooper. "A Structurally and Temporally Extendend Bayesian Belief Network Model: Definitions, Properties, and Modeling Techniques." *Proceedings of the Twelfth Annual Conference on Uncertainty in Artificial Intelligence (UAI–96)*. 28–39. 1996.
2. Allen, James F. "Maintaining Knowledge About Temporal Intervals," *Communications of the ACM*, *26*(11):832–843 (1983).
3. Allen, James F. "Towards a General Theory of Action and Time," *Artificial Intelligence*, *23*(2):123–154 (1984).
4. Allen, James F. and Johannes A. Koomen. "Planning Using a Temporal World Model." *Proceedings of the Eight International Joint Conference on Artificial Intelligence*. 741–747. August 1983.
5. Buchanan, Bruce G. and Edward H. Shortliffe. *Rule-Based Expert Systems*. Addison Wesley, 1984.
6. Charniak, Eugene. "Bayesian Networks Without Tears," *AI Magazine*, *12*(4):50–63 (1991).
7. Charniak, Eugene and Solomon E. Shimony. "Probabilistic Semantics for Cost Based Abduction." *Proceedings of the AAAI Conference*. 106–111. 1990.
8. Cooper, Gregory F. *Probabilistic Inference Using Belief Networks is NP-hard*. Technical Report KSL-87-27, Medical Computer Science Group, Stanford University, 1987.
9. Dechter, Rina, et al. "Temporal Constraint Networks," *Artificial Intelligence*, *49*:61–95 (1991).
10. Freska, Christian. "Temporal Reasoning Based on Semi-intervals," *Artificial Intelligence*, *54*:199–227 (1992).
11. Gonzalez, Avelino J. and Douglas D. Dankel. *The Engineering of Knowledge-Based Systems: Theory and Practice*. Prentice-Hall, Inc., 1993.
12. Hanks, Steve, et al. "Probabilistic Temporal Reasoning with Endogenous Change." *Proceedings of the Eleventh Annual Conference on Uncertainty in Artificial Intelligence (UAI-95)*. August 1995.
13. Hobbs, Jerry R., et al. "Interpretation as Abduction." *Proceedings of the 26th Annual Meeting of the Association for Computational Linguistics*. 95–103. 1988.
14. Horvitz, Eric and Finn Jensen, editors. *Uncertainty in Artificial Intelligence: Proceedings of the Twelfth Conference*, San Francisco: Morgan Kaufmann Publishers, August 1996.
15. Josephson, John R. and Susan G. Josephson. *Abductive Inference: Computation, Philosophy, Technology*. Cambridge University Press, 1994.
16. Kanazawa, Keiji. *Reasoning about Time and Probability*. Thesis, Brown University, May 1992.
17. Kjaerulff, Uffe. "dHugin: A computational system for dynamic time-sliced Bayesian networks," *International Journal of Forecasting, Special Issue on Probability Forecasting, to appear* (1995).

18. Ligozat, Gerard. "Weak Representations of Interval Algebras." *Proceedings of the AAAI Conference*. 715–720. 1990.
19. Manna, Zohar and Amir Pnueli. *The Temporal Logic of Reactive and Concurrent Systems: specification*. New York: Springer-Verlag, 1992.
20. McDermott, Drew V. "A Temporal Logic for Reasoning About Processes and Plans," *Cognitive Science*, *6*:101–155 (1982).
21. Nemhauser, George L., et al., editors. *Optimization: Handbooks in Operations Research and Management Science Volume 1*, *1*. North Holland, 1989.
22. Pearl, Judea. *Probabilistic Reasoning in Intelligent Systems: Networks of Plausible Inference*. San Mateo, CA: Morgan Kaufmann, 1988.
23. Peng, Yun and James A. Reggia. *Abductive Inference Models for Diagnostic Problem-Solving*. Springer-Verlag, 1990.
24. Santos, Jr., Eugene. "On the Generation of Alternative Explanations with Implications for Belief Revision." *Proceedings of the Conference on Uncertainty in Artificial Intelligence*. 337–347. 1991.
25. Santos, Jr., Eugene. "On Modeling Time and Uncertainty for Diagnosis Through Linear Constraint Satisfaction." *Proceedings of International Congress on Computer Systems and Applied Mathematics Workshop on Constraint Processing*. 93–106. 1993.
26. Santos, Jr., Eugene. "Unifying Time and Uncertainty for Diagnosis," *Journal of Experimental and Theoretical Artificial Intelligence*, *8*:75–94 (1996).
27. Shafer, G. A. *Mathematical Theory of Evidence*. Princeton University Press, 1979.
28. Shoham, Yoav. *Reasoning About Change: Time and Causation from the Standpoint of Artificial Intelligence*. MIT Press, 1988.
29. Shortliffe, E. H. *Computer-Based Medical Consultation: MYCIN*. Elsevier, 1976.
30. Shortliffe, E. H. and B. G. Buchanan. "A Model of Inexact Reasoning in Medicine," *Mathematical Biosciences*, *23*:351–379 (1975).
31. Suppes, Patrick. *A probabilistic theory of causality*. Acta philosophica Fennica. Fasc. 24., Amsterdam: North-Holland Publishing Co., 1970.
32. Vilain, Marc and Henry Kautz. "Constraint Propagation Algorithms for Temporal Reasoning." *Proceedings of the AAAI Conference*. 377–382. 1986.
33. Young, Joel D. and Eugene Santos, Jr. "Introduction to Temporal Bayesian Networks." *Online Proceedings of the 1996 Midwest Artificial Intelligence and Cognitive Science Conference*, edited by Michael Gasser. URL: "http://www.cs.indiana.edu/event/maics96/Proceedings/Young/maics-96.html", April 1996.
34. Zadeh, L. A. "The Role of Fuzzy Logic in the Management of Uncertainty in Expert Systems," *Fuzzy Sets and Systems*, *11*:199–227 (1983).

Vita

Capt Joel Young was born on September 4, 1968 in Northampton, Massachusetts of Patricia Young and Peter Young. He graduated from Worcester Polytechnic Institute with High Distinction on 19 May 1990. While he was at WPI, Joel was inducted into the Upsilon Pi Epsilon computer science honor society (for which he served as President) as well as the Tau Beta Pi engineering honor society. Joel was also an active member of the Zeta Psi fraternity. Capt Young received his commission through ROTC upon graduation and entered active duty ten months later.

Capt Young was stationed, in August 1991, at Offutt AFB, Nebraska to work at Headquarters Strategic Air Command's Comm-Computer Center where he worked in the Cruise Missile Flight Simulation Section (Cruisers). Joel became chief of the Cruisers in 1993. While at Offutt, HQ SAC stood down and the United States Strategic Command stood up and took its place.

In May 1995 the Cruisers were decommissioned and soon after Capt Young transfered to the Air Force Institute of Technology in order to complete his Master of Science in Computer Science. His research efforts focus on temporal reasoning under uncertainty. While at AFIT, Joel was inducted into the Eta Kappa Nu electrical engineering honor society.

Joel has a wife, Wendy Snow of Salem Massachusetts, to whom he has been married since 1992 as well as a lovely daughter, Alix Elizabeth.

Permanent address: 55 Maple St.
Topsfield MA 01983