Situation Awareness: A Network Centric Approach

Ananya Ojha

Thesis submitted to the Faculty of the

Virginia Polytechnic Institute & State University

in partial fulfillment of the requirements for the degree of

Master of Thesis

in

Computer Engineering

A. Lynn Abbott, Chair

Eunice E. Santos, Co-Chair

Paul Plassmann

September 26, 2008

Blacksburg, Virginia

Keywords: Situation Awareness, Network Centric Operations, Network Characteristics,

Resource Allocation, Decision Making, Information Flow, Information Sharing

Situation Awareness: A Network Centric Approach

Ananya Ojha

(ABSTRACT)

Situation (al) awareness (SA) is critical to analyze, predict and perform tasks effectively in a dynamic environment. Many studies on SA have ignored network dynamism and its effect on SA, focusing on simple environments. Many studies involving the network and SA have refrained from attempting to model information space dynamism (i.e. dynamic scenarios which may have more than one probable outcome). Few studies have identified the need for a flexible, robust and overarching framework which could model both the network and information space dynamisms and provide for analysis of different types of networks (heterogeneous/homogeneous) at multiple scales.

We utilize the NCOPP (Network Centric Operations Performance & Prediction), a uniform framework with "plug-&-play" capabilities to provide analysis and performance prediction of networked information systems. In this work, we demonstrate the flexibility of the NCOPP framework and its ability to model a hierarchical sensor system satisfactorily. We model the network & information space dynamisms using probability and statistics theory (e.g. Bayesian prediction, probability distribution curves). We model the behavior of entities/nodes involved in the process of sharing information to achieve greatly improved situation awareness about a dynamic environment within hierarchical information network systems.

Our behavior model mathematically represents how successful/unsuccessful predictions critically impact the achievement of effective situation awareness. In the behavior model, we tie together the cost of considering predictions which accounts for limited resources and the indirect effect of unsuccessful predictions.

We research and show how the NCOPP framework can model real world networked information systems at different levels of granularity. We leverage the framework's capabilities to perform experiments that not only assist in an objective comparison of distributed information filtering and central data processing paradigms but also provide important insights into the effect of network dynamism on the quality and completeness of information in the system. We demonstrate the ability of incorporating key network information, in the process of achieving SA to improve the performance of the system. We exhibit the improvement in the performance achieved with inclusion of the network characteristics during dynamic allocation of resources. We were able to show that simple hierarchical filtering (via distributed processing) results in significant reduction in the information with regards to "false alarms" when compared to systems employing central information processing. Experimental results show a direct positive impact in the completeness of SA when information sharing in hierarchical systems is supplemented by network delay information.

Overall, we demonstrated the ability of the NCOPP framework to provide meaningful insights into the interactions of key factors involved in operation of networked information systems, with a particular emphasis on SA.

To my parents

Acknowledgments

I would like to thank my adviser Dr. Eunice E. Santos for all the insights and the wisdom she bestowed to me, which made this work possible. I would also like to mention my gratitude for John Korah, a PhD candidate, who has provided me with immense support and often well cooked food.

I would like to thank Dr. Lynn Abbott for his guidance and serving on my committee. I would also like to thank Dr. Paul Plassmann for serving on my committee.

This work was supported in part by Air Force Office of Scientific Research.

Ananya Ojha

Virginia Polytechnic Institute & State University

September 26, 2008

Contents

List of Tables

List of Figures

List of Procedures

Chapter 1

Introduction

This thesis focuses on the intrinsic problem faced when trying to develop a clear and concise picture (i.e. Situation Awareness (SA)) based on the critical and relevant information available. From the discussions to be provided in this chapter, we posit that effective SA cannot be realized without equally effective methodologies and models of the underlying infrastructure or network. The first step is to determine key network characteristics and parameters that have a significant effect on the effectiveness of situation awareness. Once these characteristics are identified, our work will then focus on understanding interactions of these network characteristics with different aspects critical in the process of achieving situation awareness in a dynamic environment.

1.1 Motivation

Situation Awareness spans a diverse group of areas and applications such as plant operations management, emergency response, weather forecasting, homeland security and military operations. In simple terms, SA means "awareness of what is going on around you" [8]. There is an implicit assumption that SA contains the group of relevant and important information needed to understand, analyze and perform tasks effectively in a dynamic environment. Without a doubt, SA is critical even in the most basic of scenarios, e.g. the animals in wild need to be aware of their constantly changing surroundings in order to successfully source their food and stay alive.

In operational terms, SA is required for a wide range of specific purposes. For example, a nuclear power plant operator is not required to have the full knowledge of all objects and scenario within her/his environment (e.g. the birth date of co-workers or their siblings' occupation), but s/he does need the relevant and critical information to successfully operate the plant under controlled conditions. Similarly an airplane pilot or rescue operation personnel would also need SA but for different sets of decisions and goals.

Recent advances in sensor, computing and information technologies, have allowed for rapid increase in the amount of data and information for incorporation into decision making processes in highly dynamic heterogeneous environments. Systems today are capable of generating vast amounts of data and today's pilots, air-traffic controllers, homeland security personnel, plant operators and others must be able to comprehend this vast and often rapidly changing data and environment. They must be able to pick and choose not only the necessary and relevant information but also determine when it is needed. It is easy to understand why more information does not necessarily mean more awareness.

Situation awareness aims at determining the meaning of information regarding environment variables. It serves as the basis of the decision making processes in highly dynamic, heterogeneous environments. Interest in SA began in the mid 1980's and grew rapidly in the 1990's riding on the new growth in technology and the challenges posed by it. In current times, enhancing SA is a major design consideration for those developing operator/observer interfaces, automated systems and programs aimed at training personnel in a wide array of fields including industrial management & operations, air traffic control [9], nuclear industry [24], emergency response [19], public health system [21], weather forecasting, homeland security [25] and many others.

Today, in addition to designing systems that provide the operator with necessary information and capabilities, we must ascertain that the amount of provided information does not overwhelm the operator. We also must know if the system design can support the operator's ability to retrieve the information under dynamic and often critical operational circumstances.

1.2 Formal Definition

Many formal definitions of SA have been developed over the years, many of which are closely tied to the military and aircraft domains and also in more generic [4, 14] domains. SA originated in the domain of aircraft piloting but is now being studied in a variety of domains like weather forecasting, plant management and education.

One of the most acceptable and widely recognized definitions was given by Endsley [5]. Endsley describes SA as "the perception of elements in the environment within a volume of time and space, the comprehension of their meaning and the projection of their status in the near future".

Three critical aspects in Endsley's definition of SA are *perception, comprehension* and *projection.* Jones & Endsley in [9] performed a study aimed at understanding the specific situation awareness requirements for air traffic control. In that study *perception* is the most fundamental level of situation awareness. The authors report that 76% of SA errors by pilots could be attributed to problems in perception of the required information (due to either failures/limitations of the system or problems with the cognitive process).

SA continues beyond perception of the information to integration of different nuggets of information and determining their relevance to personal objectives and the "big picture". This process of integration is referred to as *comprehension* in this study. In this work [9], authors traced 20% of SA errors back to the comprehension problems.

Projection, as the highest level of SA, is the ability to use current events and expert knowledge of scenario dynamics to anticipate future events and their implications. Projection allows operators to plan for scenarios and make timely decisions.

1.3 Achieving SA

Achieving SA involves deriving information from all of our various sources of information. Perception could be received through visual, auditory or tactile sensors. Some of the sources could be explicit (e.g. a warning light) while some could be more subtle (e.g. slight change in pressure levels in a power plant). The system's sensors collect a subset of all information about system's environment and internal system state. Out of this information, operator perceives and interprets some information resulting in SA. It is also important to note that achieving SA is not always a passive process of receiving information from sensors, but one where the observer may be actively involved by controlling which information is displayed or considered. The observer may also be able to direct the system to collect specific information of interest by setting directions and coverage of sensors under its control. Thus the SA is derived from a combination of environment and the sensed information as interpreted and integrated by the individual.

The definition provided by Endsley [5] is widely accepted and used across a number of domains [13, 30, 31]. This definition can be understood as the basis for a number of follow on studies [18, 35, 20]. The model of SA in dynamic decision making processes provided in Fig 1.1, is derived from the theoretical model described by Endsley [7]. Figure 1.1 (derived from Endsley, 1995b [7]) shows what SA, as described in the definition provided in [5], entails.

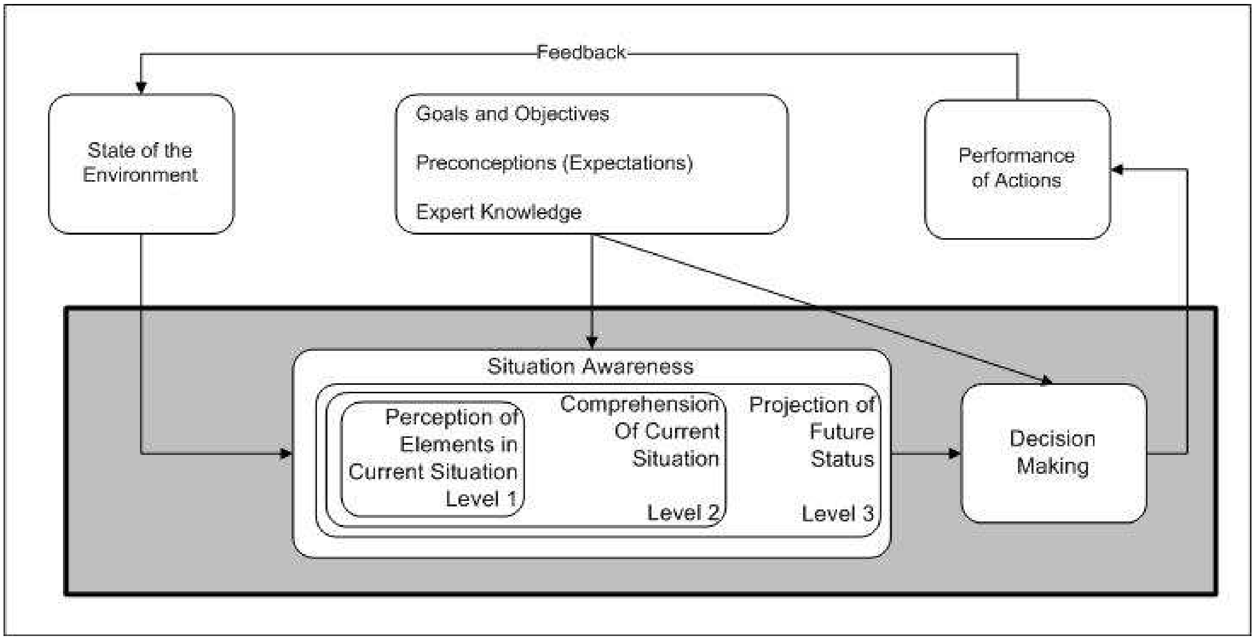


Figure 1.1: Model of SA in Dynamic Decision Making (Adapted From Endsley [7])

Upon a close inspection of the Figure 1.1, however, we notice the absence of a few critical components. We observe that the SA and decision making modules do not consider the network state/characteristics in execution of their tasks. The model shown in Figure 1.1 has been utilized as an underlying model for a variety of scenarios such as maintaining appropriate voltage/load in a power station and changing altitude/direction in commercial aircrafts. These scenarios have negligible network delay as a common factor between them.

However, a number of real-time scenarios cannot guarantee a network with negligible delay. In fact, network characteristics are a critical component of any model that serves to analyze such systems. A few examples are:

- hierarchical networks, involving humans, where information is compartmentalized and the communication follows a strict protocol (e.g. military and rescue/recovery operations)

- mobile/unstable networks where the network characteristics are not static and may vary with time (e.g. MANETs [23])

- systems where network can be dynamically modified (e.g. node addition/deletion) to affect the perception of state of the environment (e.g. rescue/recovery, intelligence gathering operations)

For the above listed system setups, it is imperative to understand the impact of network characteristics on all three levels of SA, namely *perception, comprehension* and *projection*. Also, incorporating network characteristics in decision making processes may potentially be particularly pivotal in improving situation awareness for such scenarios. We thus stress the need for a model that includes the critical component as a feedback of network characteristics to the SA and decision making module. To the best of our knowledge, no such components has been effectively modeled and put to use. We include the *network characteristics feedback* component to the model in Figure 1.1. The altered model is shown in Figure 1.2. It is apparent from the Figure 1.2 that modeling and analyzing the effect of network characteristics on different aspects of performance of systems dedicated toward achieving better information sharing and awareness is an important focus of this thesis.

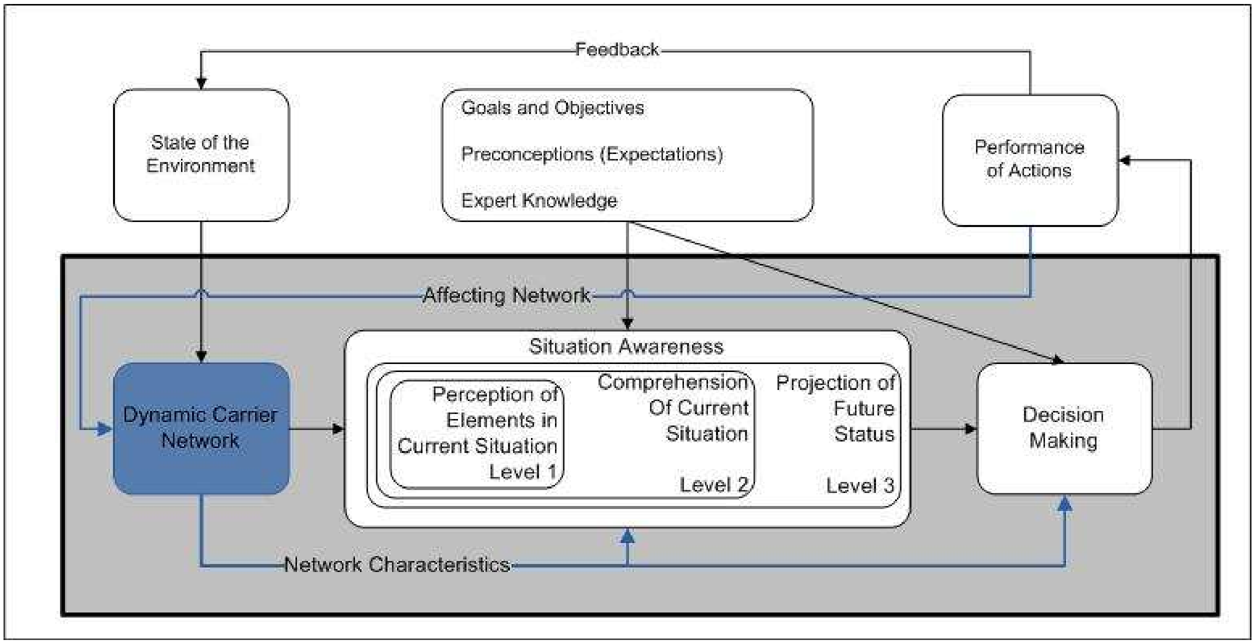


Figure 1.2: Proposed Modification in Endsley's Model [7] of SA

1.4 Previous Work

Previous studies that examined situation awareness and network characteristics together have been very limited in scope. Clearly, this must be one of the critical focal points in order to provide further advance in the field. Discussed here are the three known works that have looked at issues involving situation awareness and network delay. Their strengths and limitations are described subsequently.

Lu in [18] analyzed the expression derived by Walsh in [32], for average SA age of position updates of mobile platforms. Lu extended it to give maximum error estimates at different speeds of the mobile platform at different network delay characteristics. Using the extended formula, Lu was able to arrive at an expression that determined the maximum allowed delay over the network to keep position error estimates within targeted limits. This work included the transmission failure probability for position updates in its analysis.

White in [35] identifies that traditional network characteristics such as end-to-end delay, transmission reliability and throughput are not sufficient to evaluate SA architectures. White looks at position updates from mobile units to establish new measures to evaluate the performance of SA architectures. White proposes *average SA age, average worst SA age, message completion rate, average end-to-end delay and average SA accuracy* as new measures. The measure *average SA accuracy* has a built in assumption that limits units such that they travel in a straight line at a constant speed. White also explains when to apply these measures and how to interpret them in practice.

Manikopoulos in [20] compares trade offs between average delay, packet delivery ratio and resulting *information staleness* for TCP and UDP protocols to deliver packets in a stock update service.

The above mentioned works [18, 32, 35, 20], while seemingly in different areas, could be brought under a common umbrella. These can be described as the studies which considered network processes while examining SA in their respective areas. However, the domain of all of these works was limited to perception of data (position updates, stock prices) which falls into the first level of SA. These works cannot be directly applied to generalized information domains where SA continues beyond perception to comprehension and projection, the three levels of SA defined in [9]. Also, these works do not take temporal/dynamic aspects of scenarios into account. These aspects are discussed in the following section.

1.5 Important Aspects of Real World Scenarios

1.5.1 Temporal Aspect

Temporal dynamics associated with events and perception of time both play an important role in formulation of SA. Previous works [6, 10, 12, 11] have given examples of domains where time is regarded as an important component of SA. Operators decide on attending to parts of a scenario on the basis of not only how far some event is in space but also how soon, in time, it will impact the goals and tasks of the operator. It is easy to see how time is an important aspect of comprehension and projection, levels of SA described in [9], of future events within SA. We utilize concepts of probability theory and statistics to represent the temporal aspect of events, particularly Weibull as the representative probability distribution curve. The Weibull distribution curve is explained in Section 2.2.

1.5.2 Dynamic Aspect

Dynamic aspect of real world situations is another important factor in formulation of SA. This pertains to the fact that situations in the real world are evolving by nature. Their properties and directions change with time and so should the situation awareness of the concerned operator. If the SA does not adapt to the dynamic nature of a situation it may be rendered out-dated. Many studies have emphasized the importance of modeling the dynamics of situations and cognitive processes while modeling SA [1, 29].

Dynamic aspects of scenarios can be modeled in a simple *if-then-else* rule framework which provides for multiple possible directions, hence outcomes, of a scenario. But such a framework must be *a*) able to represent incomplete knowledge *b*) provide analysis at different scales *c*) mathematically robust and *d*) easy to update with new knowledge. In our work, we use Bayesian Knowledge Bases (BKBs) which subsume Bayesian networks, are able to represent incomplete knowledge and could be easily updated with new incoming knowledge. The framework for BKBs was proposed by Santos and Santos in [28]. This framework is further explained in Section 2.1.

1.5.3 Our Focus

We study the efficacy of the model proposed in Figure 1.2 in scenarios where:

- the magnitude of delay on the transmission of information, from sensors to operator over a network, reaches the order of duration between two consecutive events in a dynamic scenario
- the sensors are working in harsh environments which may induce, in sensors, a temporary loss of transmission ability
- link delay, a network characteristic, becomes a dynamic variable
- situations and event have respectively, dynamic and temporal aspects (described above) associated with them

- such conditions may arise in the domains of *homeland security, public health monitoring, emergency response, weather forecasting* and *nuclear power plant maintenance.*

The above issues we chose to investigate introduce us to a new and interesting problem. The networked systems we are interested in, are composed of heterogeneity in nodes, links and communication technologies. We must aim to design a model which is scalable and mathematically robust. Since different domains may have different definitions of SA and may have different criterion to perform analysis, our model must be theoretically flexible and overarching in its nature. With these concerns we come across a active field of research in military domain known as Network Centric Operations (NCO).

1.6 Paradigm of NCO

Network Centric Operations is the concept which proposes that the application of information age concepts to speed communications and increase situational awareness through networking improves both efficiency and effectiveness of operations in both military and civilian domain [36]. NCO relies on computerized systems and communication technology to provide a shared awareness of the information/operation space. NCO was conceptualized to encourage collaboration by allowing greater flow of information across this space. This is aimed at achieving a state where acquisition of data, processing it into information and providing the information to person or system that needs it, can be achieved in significantly less time. *leveraging a network to maximize situation awareness* is seen to be at the heart of this domain.

The emphasis in NCO domain has been on developing components based either on network (e.g. *network security, hierarchical routing, data aggregation*) [33, 16, 37] or information space alone (e.g. *information display, information quality, threat analysis, target tracking*) [17, 3, 22, 2].

Since these networks are frequently employed in adaptive and dynamic environments, it is imperative to access the infrastructure, pinpoint weaknesses and suggest remedies before deployment and even during operations. Very few theoretical models have been proposed to analyze and estimate the reliability and robustness of the networks in the context of situation awareness. The concept of *Network Centric Operations Performance & Prediction* (NCOPP), developed by Santos [26], addresses the issue of being able to model NCO networks into interacting decomposable components. This framework promises to be an overarching, robust and flexible framework which provides for study of different components of networked information systems and a common platform where advances in different domains associated with such systems could be brought together and their impact on the performance (e.g. SA) of these complex heterogeneous systems studied. Such a model allows use of "plug-&-play" components through defining functional interactions based on specifics. Santos argues that NCOPP would allow for realistic performance prediction and analysis of NCO networks spanning a variety of metrics, scales and techniques. We modeled our system along the lines of the NCOPP framework. In Chapter 2 we provide description of the NCOPP framework. In Chapter 3 we discuss our design of networked information systems based on the NCOPP framework.

1.7 Problem

The NCOPP framework accommodates the heterogeneity of nodes and communication technologies allowing us to address an overarching problem of being able to gain insights into important interactions of various factors involved in NCO operations. These interactions could be within the network space or between the network and the information spaces. In this work, we attempt to get insights into the importance of incorporating network characteristics/dynamism in achieving SA and decision making process affecting the network. We also model all three levels of SA in the system. This work provides a foundation for further validation of the NCOPP framework, and establishes its viability as a theoretical framework in network performance prediction and analysis domain.

In order to provide critical and timely results, we focus our study on hierarchical sensor networks where information is compartmentalized and information flow follows a strict protocol. We look at the effectiveness of information filtering over allowing the apex operator in the hierarchy (or the "root") to process all the incoming information. We are also interested in understanding how, in such dynamic environments, incorporating the network characteristics, while interpreting the information coming in from various sensors or lack of the same, is vital to building a better SA and effective management of resources while deploying/directing sensors to collect information of interest.

1.8 Contribution of this Thesis

We establish the ability of the NCOPP framework to model networked information systems together with network and information space dynamisms. Our focus is toward designing the framework and employing critical high level (coarse-grained) behavior analysis to obtain intuitive insights to evaluate the correctness of our model.

In this thesis, we model hierarchical networked information gathering systems (in a logical "tree" structure). We model the information flow instead of data packet level transmission over the links and represent the path followed by the information between two logical nodes in the network using a direct link. Here, we abstract network characteristics and defer modeling network behavior in fine detail for future works and hence we do not compare our results with data from real systems to evaluate our framework. We focus more on modeling the dynamism in information flow and the dilemma involved in decision making processes for information sharing and resource allocation.

Since little or no work has been done toward designing and establishing such all-inclusive theoretical frameworks which allow us to get significant insights into performance prediction and analysis of scalable networked information systems, we conduct experiments that could investigate critical behaviors of such systems and provide intuitive and meaningful insights.

We note that the results and definitions presented in this thesis can also be found in [27].

1.9 Outline

In this thesis we focus on validating the NCOPP framework which accommodates heterogeneity and allows us to address the problem of being able to gain insights into critical interactions in NCO networks. Here we try to gain insights into the interactions of dynamic information and network spaces. We look to examine the importance of incorporating network characteristics in achieving SA and decision making processes in systems where both network and information spaces have dynamic properties.

In Chapter 2, we provide background information on modeling temporal aspects of the system (both in information and network spaces) and formally discuss the concept of NCOPP as proposed by Santos in [26]. In Chapter 3, we discuss how real-world systems can be encapsulated into the NCOPP framework and its components/submodels. We also elaborate on the design specifics of our system in particular. Experiments and results are discussed in Chapter 4. Chapter 5 summarizes the contribution of this thesis and provides concluding remarks.

As noted before, the results and definitions presented in this thesis can also be found in [27].

Chapter 2

Background Information

2.1 Bayesian Knowledge Bases (BKBs)

Santos [28] provided the framework for BKBs as a flexible, intuitive and semantically sound knowledge representation. It unifies the *if-then* style rules with probability theory. We use BKBs to represent uncertainty and temporal dynamism of events and scenarios under observation of the system. Also, BKBs can easily be updated and maintained with new incoming information. To compare BKBs against Bayesian network, Santos and Santos in [28] point out to the following properties of real world scenarios/environments:

- In the real world, complete information about one's environment is typically unattainable
- Certain conditional probabilities may not exist or are not meaningful in the target

domain.

Thus, any representation must have the ability to recognize and accommodate incompleteness as it occurs. Bayesian networks, as opposed to BKBs, require complete information of dependencies and interaction among various independent variables which may not be feasible. We discuss the theoretical framework of BKBs in the following section.

2.1.1 Basic Definitions and Model

A Bayesian knowledge-base (BKB) represents objects, world states and the relationships between them using a directed graph. The graph consists of nodes which denote various random variable instantiations while the edges represent conditional (in)dependencies between them.

BKB

Let $\Re$ denote the real numbers, $\Re^+$ denote the non-negative reals, and Φ denote the empty set.

Definition 1 (Def 1 from [28]): A correlation graph $G = (I \bigcup S, E)$ *is a directed graph such that* $I \bigcap S = \Phi$ *and* $E \subseteq \{I \times S\} \bigcup \{S \times I\}$*. Furthermore, for all* $a \in S$*,* (a, b) *and* $(a, \acute{b})$ *are in* E *if and only if* $b = \acute{b}$*.* $\{I \bigcup S\}$ *are the nodes of* G *and* E *are the edges of* G*. A node in* I *is called an instantiation-node (I-node) and a node in* S *is called a support-node (S-node).*

I-nodes represent various possible instantiations of random variables (RVs). In other words, I-nodes represent the events occurring in the information space. S-nodes, on the other hand,

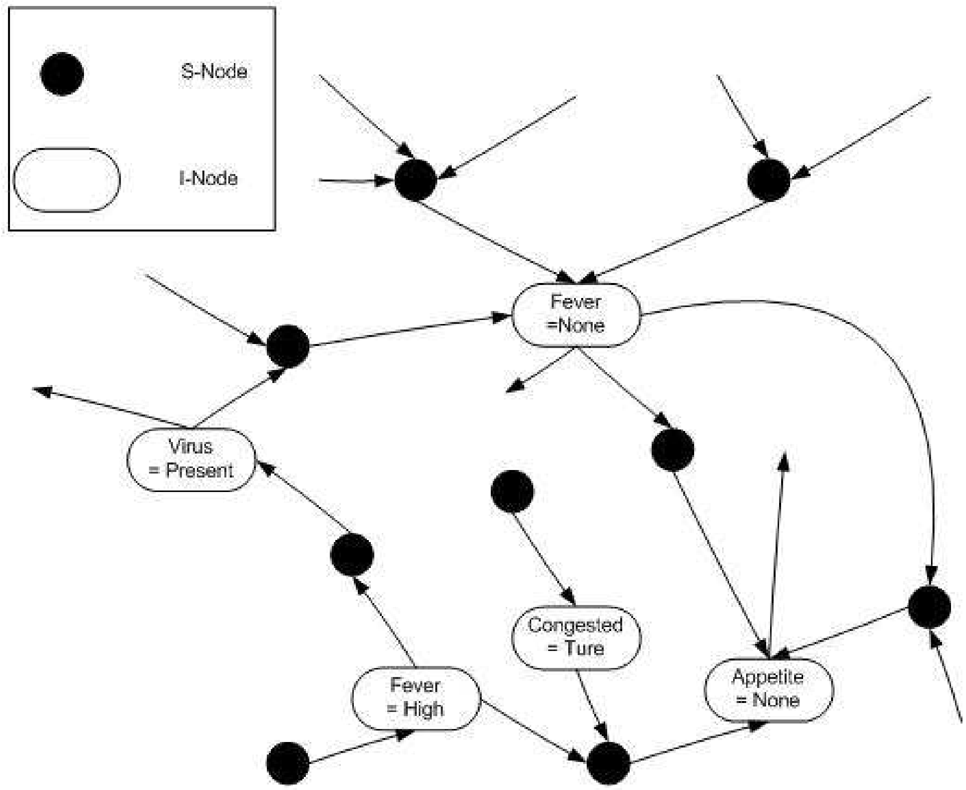


Figure 2.1: Example of a Correlation Graph Fragment (Adapted From [28])

explicitly represent pre-conditions/(in)dependencies between the I-nodes. In Figure 2.1, note that each S-node has at most a single outgoing edge.

Let a be any node in $I \bigcup S$. $PRED_G(a) = \{b|(b,a) \in E\}$ are the *immediate predecessors of* a in graph G. $DESC_G(a) = \{b|(a,b) \in E\}$ are the *immediate descendants of* a in graph G.

Let π be a partition on I. Each cell in the partition π denotes the set of I-nodes (events/instantiations) which are mutually exclusive instantiations of a single RV. In BKB, we can represent any random variable with discrete and multiple instantiations of the possible states that particular variable might attain. In Figure 2.1, one cell in π would be $\{fever = none, fever = high\}$

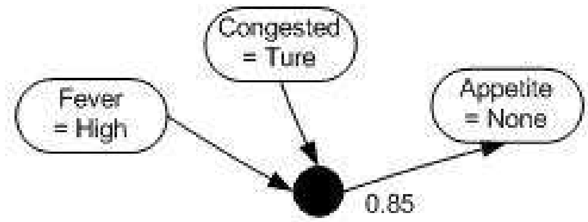


Figure 2.2: Head and Tail. The RV instance 'Appetite= None' is the head and the I-node RV instances 'Fever= High' and 'Congested= True' constitute the tail of the S-node. (Adapted From [28])

Figure 2.3: Priors. The S-node Denotes the Prior Probability $P(Fever = \ High) = 0.4$ (Adapted From [28])

which are two instantiations for the RV $fever$. $|.|$ denotes cardinality.

Definition 2(Def 2 from [28]): G is said to respect π if

- for any S-node $b \in S$, the predecessor I-nodes of b, $PRED_G(b)$, assigns at most one instantiation of each RV, and

- for any two distinct S-nodes b_1 and b_2 in S such that $DESC_G(b_1) = DESC_G(b_2)$, there exists an I-node in $PRED_G(b_1)$ whose RV instantiation contradicts an I-node in $PRED_G(b_2)$. Furthermore, the b_1 and b_2 are said to be mutually exclusive.

Intuitively, direct conditional dependency between the single immediate I-node, i.e. head of the S-node (referred here to as descendants), and the tail (immediate I-node predecessors) is represented by a S-node (see Figure 2.2). 'Appetite = None ' (a RV instance) represents the head and, the I-node RV instances 'Fever = High ' and 'Congested = True ' form the tail of the S-node. The value associated with the S-node in Figure 2.2 represents the conditional probability P(Appetite = None — Fever = High, Congested = True) = 0.85. Priors are denoted by S-nodes without inputs, as shown in Figure 2.3 where the S-node denotes the prior probability P(Fever = High) = 0.4. We later describe, how temporal properties of events in the information space is incorporated in the BKB representation.

Definition 3(Def 3 from [28]): A Bayesian knowledge-base K is a 3-tuple (G, w, π) where $G = (I \bigcup S, E)$ is a correlation-graph, w is a function from S to $[0, 1]$, pi is a partition on I, and G respects π. Furthermore, for each $a \in S$,$w(a)$ is the weight of a.

2.1.2 Probabilistic Reasoning

A BKB, in our system, is used frequently to obtain answers for the following questions, "*What is the probability that a particular I-node a_i will get instantiated?*". We know from Definition 1 that a S-node supports only one I-node. Thus,

$$P(a_i) = \begin{cases} 1 & if\ a_i \in Evidence \\ 0 & if\ a_j \in Evidence, (a_i, a_j) \in \pi_a \\ Max\{P(a_i|b_i) \times P(b_i) : (b_i, a_i) \in E\} & otherwise \end{cases}$$

where $(a_i, a_j) \in \pi_a$, means a_i, a_j are mutually exclusive. Also, for any S-node b,

$$P(b) = \begin{cases} \prod\{P(c_i) : (c_i, b) \in E\} & if\ PRED_G(b_i) \neq \Phi \\ 1 & otherwise \end{cases}$$

2.2 Temporal Dynamics

In our experiment, to include temporal dynamics of events and scenarios, in addition to the weight $w(a)$, we also include $\lambda(a)$ and $k(a)$ as parameters associated with the S-node a. These two parameters control the shape of probability distribution curve, which determines the temporal dynamics of I-node b, such that $(a, b) \in E$, once the preconditions in $PRED_G(a)$ are satisfied. We are using Weibull distribution [34] as a representation for possible probability distribution curves because its shape can be easily controlled by two variables and it is mathematically easy to work with. A Weibull distribution is given by the following equation.

$$f(x; k, \lambda) = \frac{k}{\lambda} \left(\frac{x}{\lambda}\right)^{k-1} e^{-(t/\lambda)^k} \tag{2.1}$$

for $x > 0$ and $f(x; k, \lambda) = 0$ for $x \leq 0$, where $k > 2$ is the shape parameter and $\lambda > 0$ is the scale parameter of the distribution. Example of Weibull distribution curve for different values of λ and k are shown in Figure 2.4. It is important to notice here that

$$\int_0^x f(x; k, \lambda) = 1 - e^{-(t/\lambda)^k} \tag{2.2}$$

The probability for the instantiation of an I-node b at a time step $t > t_1$, where t_1 is the

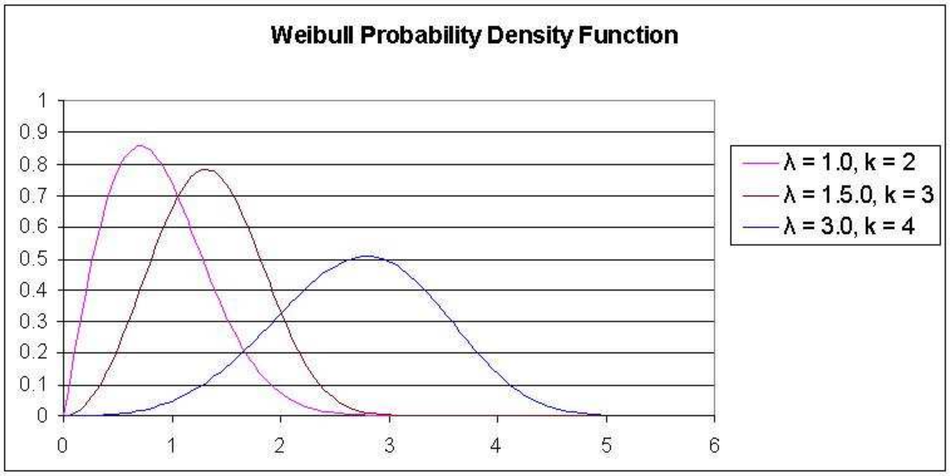


Figure 2.4: Weibull Probability Density Function

time step at which all $PRED_G(a)$ are satisfied for a S-node a, is given by

$$P_{(b,a)}(t; k, \lambda) = w(a) \times f(t - t_1; k(a), \lambda(a)) \tag{2.3}$$

We use the Weibull distribution [34] to model the interval between successive failures of the transmission ability between a sensor node and its relay node authority. We also employ this distribution to model the variability in network delay characteristics over logical links between various nodes in the networked sensor system.

It is important to note here that Weibull distribution is a representative distribution curve and could be substituted by any other probability distribution curve to suit the requirements of different domains or scenarios where this framework is employed to model the system.

2.3 NCOPP Framework

The Network Centric Operations Performance & Prediction (NCOPP) [26] framework is designed to be capable of analyzing and predicting performance quickly within a dynamic environment. Added strength of such a framework is its ability to use theoretical "plug-&-play" components by defining functional interactions based on specifics of the component.

The key components that are required to express and analyze the network are identified below. The interaction of the components is shown in Figure 2.5 This framework can model the network at multiple scales and is able to represent the adaptive and dynamic nature of the NCO/NCW networks.

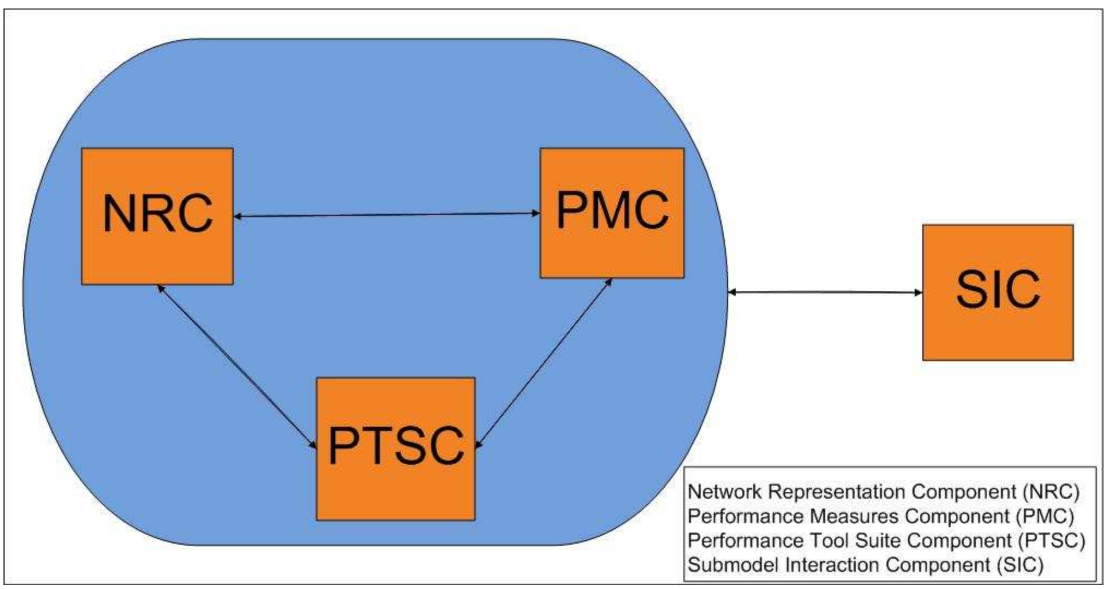


Figure 2.5: Key Components of NCOPP and Their Interaction

2.3.1 Network Representation Component (NRC)

NRC formally represents and defines the static snapshot of the system's network and information space. Using graph-theoretic notations for NRC allows us to formally and concisely represent the state of the system at multiple resolutions.

In the NRC representation of network space, nodes are represented as vertices and edges represent abstracted path followed by the information between any two nodes (or sub-networks). Nodes and edges have labels associated with them which help in determining the specific type of any node or edge. Weights associated with each entity in NRC would represent the value of its metrics at a particular time. In the information space, the vertices represent events occurring in the information space or preconditions which need to be satisfied for any event to be triggered. Edges represent the (in)dependencies between events/pre-conditions.

Formal Definition(from [26]): *Given a network N, NRC will construct $G_N(V, E, W, L)$ where the vertex set V represents the set of nodes in the network, the edge set E represents "hard" lines as communications (wired telephones, computer networks, etc). In the information space, given a knowledge base G, NRC will construct $G_N(V, E, W, L)$ where the vertex set V represents the set of events, conditions in the knowledge base, the edge set E represents (in) dependencies between events and conditions. Also, the sets W and L represent the collection of weights and labels, respectively, on nodes and edges.*

2.3.2 Performance Measures Component (PMC)

PMC provides mappings which allow the framework to represent dynamism of network/information space in such systems. Using PMC, NRC matrices can be extrapolated to predict/project the future states either of the environment in which the system is deployed or the system's infrastructure itself for any specific time step or change in time for a known time step.

Measures in PMC are required to be mathematically defined based on the metrics represented in NRC model. Example of dynamic measurements that could be defined using PMC include node termination/deactivation, faultiness of node/edge, movement and data reception.

Formal Definition(from [26]): *Let $G(V, E, W, L)$ be the graph (NRC) submodel and $\mathcal{G}$ is the collection of all subgraphs of G, the PMC submodel consists of a collection of functions F, where if f is a function in F then $f : \mathcal{G} \times \mathcal{T} \rightarrow \mathcal{G}$ where*

- *given a new time t, and a subgraph G_x that contains the input vertices, edges, weights, and labels of f, then $f(G_x, t) = G_y$ denotes the changes in the subgraph of NRC at time t on subgraph G_x based on function f,*

- *given a time t, a change in time τ, and a subgraph G_x that contains the input vertices, edges, weights, and labels of f, then $f(G_x, t + \tau) = G_y$ denotes the changes in the subgraph of NRC after τ time steps have elapsed on subgraph G_x based on function f,*

- *to determine the effect of time across multiple measures, we define the composition function $h : \mathcal{G} \times \mathcal{G} \rightarrow \mathcal{G}$ for the composition of two output subgraphs in order to obtain* *a more complete understanding of the graph at the new time. Composition for subgraphs at the same time step can be performed to obtain a more complete specification.*

In the information space the PMC is utilized to predict future events and provide a probability for a scenario/event-chain to reach an end state successfully.

2.3.3 Performance Tool Suite Component (PTSC)

PTSC module is a basket of functions/criterion which provide performance measures of the current system's matrices as defined in NRC or future matrices which could be obtained using NRC and PMC together.

Example of performance measures could include throughput, average node connectivity, aggregate signal strength or user defined functions to measure the performance of system in the information space. PTSC using only NRC metrics will provide performance of the state of the network at a known time step. PTSC applied on NRC along with PMC provides the measure of performance of future states of the system.

Formal Definition(from [26]): *PTSC is a collection of functions F, where*

- *if f is a function in F then $f : \mathcal{G} \rightarrow \Re$ where $\Re$ denotes the set of reals.*
- *F is closed under composition of metrics.*

2.3.4 Submodel Interaction Component (SIC)

SIC is another critical component of the NCOPP framework which provides a qualitative context to the quantitative performance measures of the current or future states of the system provided by PTSC. SIC component utilizes the information provided by NRC and PMC component about the current and future state of the system along with the PTSC measures to detect developing performance bottlenecks and provides suggestions for refinement in the interaction/behavior of system components to improve the performance.

Chapter 3

Modeling Real Systems in NCOPP

In Section 2.3 we provided the formal definitions of the NCOPP and its components. In this chapter we discuss and elaborate how complex real-world systems can be encapsulated and defined into NCOPP and its components/submodels. We note that results and definitions provided here can also be found in [27].

3.1 The Real World System

Most networked information systems which may involve/represent human operation control/decision making are structured in a hierarchical structure. Hierarchical structures allow for efficient flow of instructions and swift synchronization of operations which is imperative in networked operations involving human elements. Thus, in this thesis we concentrate on modeling the flow of information and information sharing aimed at improving situation awareness in hierarchical networks. We focus our study on understanding the impact of key network characteristics on situation awareness and how this understanding could be leveraged to further improve situation awareness.

Hierarchical networks are unique in a way that information flow is compartmentalized and more restricted. Nodes have limited view of the world and as we move up the hierarchy, view of the real world consolidates and is more refined. A sample representation of hierarchical networks that we are considering in our study is shown in Figure 3.1.

Hierarchy can also be defined as the representation of interaction among information processing/decision making nodes at the lowest level of hierarchy. In such a representation, a fusion node at levels other than the lowest level would represent the interaction among a set of nodes at the lowest level. Following are the definitions of basic elements in our real-world system.

- **Knowledge Base:** We define a knowledge base as a collection of Bayesian Knowledge Bases (BKBs) (defined in Section 2.1) with each BKB capturing the expert knowledge of a possible dynamic scenario. Different BKBs can be understood as disjoint partitions of a single global BKB. These BKBs are defined to capture previous experiences and expert knowledge about various scenarios we intend to monitor.

- **Event:** An event is the fundamental unit of the information space and has temporal and spatial characteristics. Events represent significant changes/patterns in a situation depicting "important occurrences". With respect to BKBs, a true instantiation of any

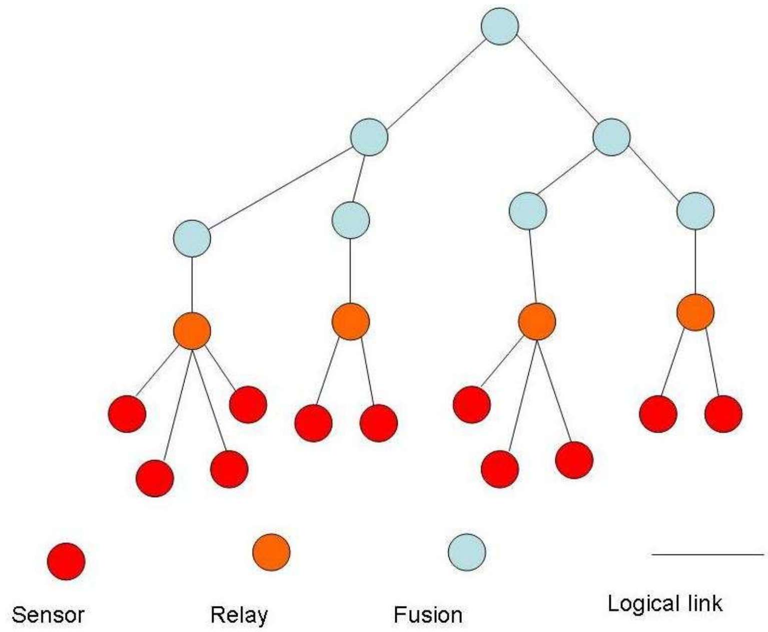


Figure 3.1: A Representative Hierarchical Network

I-node defined in Knowledge Base is referred to as "Event". In our system, only the true instantiations can be detected by sensors deployed in the field. False instantiations of any I-node can only be inferred. Each event has specific properties which lead to detection of the event by a sensor unit.

- **Event Chain/Situation:** An event chain is the instantiation of a disjoint BKB in knowledge base. Instantiation here refers to activation of one or more I-nodes as True. An event chain is composed of active S-nodes (representing achieved preconditions) and instantiated I-nodes when supported by one S-node. Event chain, at a time step t is regarded:
  - *Active*: When instantiation of at least one I-node is pending in future.

  - *Expired*: When no new instantiations of I-nodes are pending.
  - *Successful*: If the chain has instantiated at least one leaf I-node from the disjointed BKB as True.
  - *Failed*: When the chain expires without instantiating at least one leaf I-node from the disjoint BKB as True. These are also referred to as *false alarms*.

- **Sensor:** A sensor (node capable of wireless communication) is deployed in the field with ability to detect events with certain properties. Such properties could be *thermal, radio, visual, nuclear, motion, etc.* Sensors have limited transmission range. Sensors is the element of logical network where perception (SA at level 1) begins.

- **Relay Node:** A relay node or base is the nodal point for collection of information from a set of sensors deployed in a particular region. Base receives information, i.e. detected events, from sensors assigned to it and deployed in the region. A relay node also has the ability to deploy more sensors or switch-on backup sensors if advised from an authority (explained in 3.2).

- **Fusion Node:** Information collected at a relay node is transmitted to a hierarchy of fusion nodes, responsible to provide situation awareness for operations, where further processing on the collected information takes place. A fusion node may be collecting information from more than one relay nodes deployed in same or different regions.

- **Link:** A link is the logical representation of the path followed by data-packets when information travels between any two logical nodes. Each data packet traveling on this link, experiences a variable delay. This delay represents aggregate queuing, transmission and retransmission delay. Initially, we have also assumed in-order delivery of information on a link.

- **Dynamic Conditions:** In a dynamic environment, sensor nodes may move in/out of the relay node's range and thus experience temporary loss of communication. In harsh environments, sensor nodes may also transmit less frequently or in bursts in order to avoid detection. In periods, when a sensor is experiencing loss of contact with the relay node, it may continue to gather evidences, which are transmitted to the relay node when the contact resumes. We capture these two conditions in our system.

Described below in detail are different modules of the NCOPP framework demonstrating how real-world systems can be encapsulated into NCOPP and its submodels.

3.2 NRC

NRC submodel is necessary for static representation of the network/system state. Here in addition to physical network, we have to represent state of the environment under observation. We associate labels with entities in the system to represent their state and behavior.

- **Network Space:** The Network space is represented as a graph $G(V, E)$. $V = F_G \bigcup B_G \bigcup S_G$ represents the vertex set and $E = F_E \bigcup B_E \bigcup S_E$ represents the edge set where:

- F_G, B_G, S_G are the set of fusion, relay and sensor nodes respectively,
- $F_E = \{(a, b) \in F_G \times F_G : a, b \in F_G, a = Authority(b)\}$,
- $B_E = \{(a, b) \in F_G \times B_G : a \in F_G, b \in B_G, a = Authority(b)\}$ and
- $S_E = \{(a, b) \in B_G \times S_G : a \in B_G, b \in S_G, a = Authority(b)\}$

Different components of NRC and the respective associated labels are enumerated below:

Sensor Node:

- *Sensing Ability:* Properties of an event which the sensor may detect
- *Authority:* Relay Node with which the sensor communicates
- *Failure Characteristics:* Average duration d for which the sensor stays in and out of communication range of relay node. Also, (λ, k) which determine the distribution around d in Equation 2.1.
- *Observation Zone:* Zone in which an occurring event may be noticed, provided necessary sensing ability

Relay Node:

- *Authority:* Fusion Node to which the relay node communicates all collected data from sensors.
- *Observation Zone:* Zone of operations for sensor nodes associated with the relay node

Fusion Node:

- *Authority:* Fusion Node to which this node communicates important developments and processed information.
- *Hierarchy level:* Hop distance away from the apex node of the hierarchical network. Hierarchy level implicitly determines a lot of important behavior characteristics of the fusion node.

Link:

- *Delay Characteristics:* (λ, k), which determine the probability distribution curve in Equation 2.1.
- *Present Delay:* Delay last suffered by a data packet/information when transmitted over the link.

- **Information Space:** In the information space, the current state is represented by:

  Active Chains: List of event chains or situations, instantiated form of disjoint BKBs, which will progress further or have a pending I-node instantiation in future.

  Successful Chains: List of event chains or situations which eventually result in activation of a leaf I-node, from the BKB, as True.

  Failed Chains: In other words, "*false alarms*". Event chains that could not activate a leaf I-node as True.

  Events: Each I-node, activated as True, has position coordinates and a time of instantiation associated with it. Every event also has one or more properties that enable its detection by a sensor node having the respective sensing ability.

Fusion Nodes: Each fusion node maintains a *events record* that contains various events observed in the information space and communicated to it by other nodes. These events are essentially I-nodes instantiated as True. Each fusion node also maintains a record of events from different developing situations, which it deems important to monitor the environment for. These are called predictions and are not yet registered in *events record*. An example of *events record* is shown in Figure 3.2

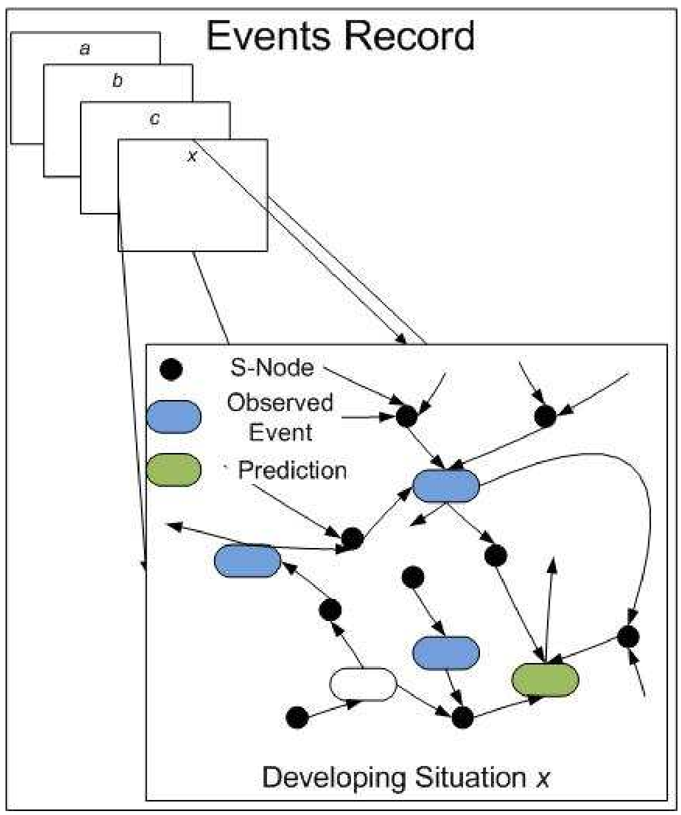


Figure 3.2: Diagram Showing *events record* & Predictions at a Fusion Node

3.3 PMC

In this study we focus on event prediction and situation assessment as the central ideas guiding the PMC submodel design. We use the PMC module to allow fusion nodes to assess the development of situation and predict events which can occur in near future. The block diagram explaining PMC module is shown in Figure 3.3.

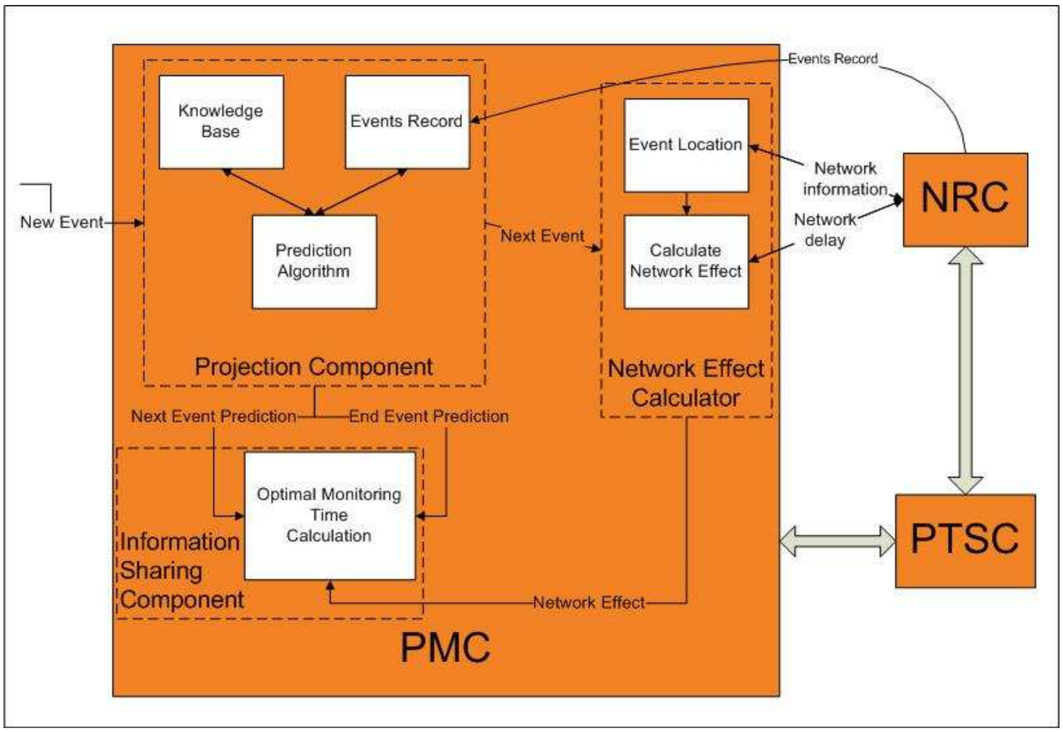


Figure 3.3: Flow Chart Explaining PMC Module

3.3.1 Understanding the PMC

At a fusion node, when a new event not already present in its *events record* is reported by another node, it is added to *events record*. Using the knowledge base(KB) broader understanding of the situation i.e. probability of possible end events and possible future events are inferred. For each possible future event, zone of occurrence is deduced. Taking into consideration the development of the scenario as a whole, significance of possible next event, network characteristics of the subnet under the fusion node's command in the respective zone and time-table for sending the information about the developing situation upstream/downstream in the hierarchy is calculated. The background theory for the calculation is explained in the following section.

A important note must be made here: *In our model, ability of a fusion node to predict a future event is limited by the evidence of new events it receives from other nodes.* Thus in a hierarchy, fusion nodes at the lower level can control the *filtering* of information by restricting the sharing of information, about newly occurred events, with the fusion nodes higher in the hierarchy.

3.3.2 Optimizing Monitoring Time Calculation

We design the framework that may represent the dilemma behind information sharing decisions in a hierarchical network. The inherent dilemma for a fusion node, of "*What evidences must be shared with fusion nodes higher/lower in the hierarchy*" and "*When*", arises from the temporal nature of event progression and dynamism of the scenario (explained in Section 1.5). There are always multiple directions a scenario may take. In other cases, the scenario may not progress altogether (e.g. false alarms).

Motivation

Following points must be considered to better understand the source of the dilemma.

- We assume that the time of instantiation of an I-node follows a probability distribution curve (e.g. Weibull distribution equation 2.1), once the respective pre-conditions or S-node is satisfied.
- Failure of a S-node a to instantiate an I-node b can only be inferred and not detected. $\int_t^\infty P_{(b,a)}$ described in Equation 2.3 can be used as the pointer to arrive at this inference.
- The motivation to predict future events arises from the confidence that a fusion node collects when the fusion node receives information about occurrence of an event, it was predicting *will happen.* This confidence can be represented numerically and henceforth, referred to as reward.
- Motivation to reject or degrade the possibility of a future event as a *false alarm* comes from the cost a fusion node incurs per unit time for predicting a future event.
- Cost and reward may vary both with the level of fusion node in the hierarchy and with the closeness of the predicted event to the end events of a scenario.

- Prior probability of instantiation of the predicted event may also be taken into account.

It can be inferred that by modifying the cost and reward structure, desired information sharing behavior of the fusion nodes in a hierarchical structure can be achieved. For example, if the rewards increase and the cost decreases as we move up the hierarchy; fusion nodes in the hierarchy would be motivated to undertake most of the information processing higher in the hierarchy and vice-versa. Also, if for events which are closer to the end stage in a scenario rewards are higher as we move up the hierarchy when compared to events farther from the end stages, fusion nodes will be encouraged to undertake processing higher in the hierarchy.

In other words, the entire spectrum of the information sharing behavior (from central processing to standalone & scattered processing) can be covered by varying the cost and reward structure in our framework. This helps our framework remain flexible to accommodate different information sharing behavior of different domains. Most common information networks have the following features of information sharing behavior:

- Nodes higher in the hierarchy are more important and hence there is a higher premium attached to their attention.

- As scenario progresses toward end stages commander/officers higher in the hierarchy get increasingly involved in monitoring the situation.

To capture this behavior, we employ the following cost-reward structure in our system:

- we increase the cost monotonically as we move up the hierarchy and there is no differentiation according to the closeness of the predicted event to end events in the scenario.
- We increase the rewards monotonically as we move up the hierarchy.
- The reward increases monotonically as the closeness of predicted event to the end event in the scenario increases.
- We also scale up the rewards depending on the probability of scenario progressing to completion as well.

Mathematical Formulation

We consider a hierarchy with i levels and root node of the hierarchy at level 1. Let us assume that all preconditions required by a S-node a are satisfied at $t = 0$. Labels associated with the S-node a are $w(a), \lambda(a)$ and $k(a)$. For a I-node b, such that $(a, b) \in E$:

- Minimum remaining possibility beyond which the possibility of instantiation of an I-node becomes negligible is denoted as κ
- Time step after which the possibility of instantiation of b is less than κ is given by $T = \left\lceil \lambda(a) \times (ln\frac{1}{\kappa})^{(1/k(a))} \right\rceil$, as derived from Equation 2.2

- In our system, when a fusion node at level j shares relevant evidence with fusion node at level $j-1$, the cost is incurred only at level $j-1$. The cost at level j starts incurring again when fusion node at level $j-1$ stops predicting b and communicates the same to fusion node at level j.

- Minimum fraction of time T which a prediction must spend at higher level in the hierarchy to induce a fusion node to inform its superior in the hierarchy be denoted as η

- Using Equation 2.3 it is easy to show that probability of instantiation of I-node b between time step t_1 and t_2 can be given as:

$$P_{(b,a)}(t_1, t_2; k, \lambda) = w(a) \times (e^{-(t_1/\lambda)^k} - e^{-(t_2/\lambda)^k}) \tag{3.1}$$

- Let us assume the per unit time cost of predicting instantiation of an I-node be c_j at level j.

- Let the $r_j = w(a) \times R_j$, be the reward awarded if the information of instantiation of b reaches the fusion node, which is incurring the cost for prediction of b, at level j.

- Let the time step (*a variable*) at which a fusion node in hierarchy level j shares the relevant evidences (preconditions to satisfy a and predict b) with fusion node at level $j-1$ be u_j. It is necessary to note that u_1 does not exists.

- Let the time step (*a variable*) at which a fusion node in hierarchy level j stops predicting b (due to the dilemma as discussed before in this section) be l_j.

- Expected reward collection by a fusion node at level j can be given as:

  $fr_j = [P_{(b,a)}(u_{j+1}, u_j; k, \lambda) + P_{(b,a)}(l_{j-1}, l_j; k, \lambda)] \times r_j$, $\forall j \in (2, i)$ and

  $fr_1 = P_{(b,a)}(u_2, l_1; k, \lambda) \times r_1$

- Expected cost incurred by a fusion node at level j can be given as:

  $fc_j = ([u_j - u_{j+1}] + [l_j - l_{j-1}]) \times c_j$, $\forall j \in (2, i)$

  and $fc_1 = (l_1 - u_2) \times c_1$

We formulate the *optimizing monitoring time calculation* as a maximizing problem where we:

$$\textbf{Maximize} \sum_{j=1}^{i}(fr_j) - \sum_{j=1}^{i}(fc_j) \tag{3.2}$$

where,

- $u_j \leq \ u_{j-1} \forall j \in (2, i)$
- $u_2 \leq l_1$
- $l_{j-1} \leq l_j \forall j \in (2, i)$
- $l_i = T$ and $u_{i+1} = 0$

We use Lindo API, a standard non-linear solver from Lindo Systems Inc. Chicago, IL. [15], with multiple-starts option to solve this non linear optimization problem. Also if in the solution for any j, $l_{j-1} \leq u_j + \eta \times T$, the fusion node at level j decides not to share the information with the fusion node at level $j - 1$.

3.4 PTSC

As our focus is to gain insights into the effects of network dynamism on performance of the system that includes network and information space dynamism; we look at SA which is affected by the interplay between the network and information spaces. There can be many varied and self defined measures of SA. Different domains where SA is required may have their own standards and procedure to measure SA. We define a sample measure, calculation of which is defined in the following paragraph, called **General Awareness Factor(GAF)**.

General Awareness Factor: This measure computes the ratio of events (true instantiations of I-nodes) from successful scenarios which are flagged by any fusion node when it deems the scenario critical. The accumulation of all such flagged events is called **General Awareness Picture (GAP)**. Let:

- e_s denote the events flagged by all fusion nodes from successful chains.
- e_f denote the events flagged by all fusion nodes from *false alarms.*
- T_s denote the set of all events from successful chains instantiated in the information space.
- $|.|$ represent the cardinality of a set

$$GAF = \frac{|e_s|}{|T_s|} \tag{3.3}$$

Criteria for deeming a scenario critical: Each fusion node i, depending on its position in the hierarchy, is assigned a threshold confidence ψ_i. When probability of any possible end event from a developing scenario exceeds ψ_i, the fusion node is obligated to flag all observed events (in its *events record*) for the corresponding scenario.

Signal to Noise Ratio (SNR): SNR is the measure defined to contrast the efficiency of non-filtering based approach with a approach stressing on filtering of information to gain SA in distributed hierarchical systems. In simple words, SNR quantifies the clutter in awareness.

$$SNR = \frac{|e_s|}{|e_s| + |e_f|} \tag{3.4}$$

3.5 SIC

We model a representative SIC component which is responsible for refinement and suggesting corrective measures to improve the performance of the NCO network. SIC in our model is an instrument for dynamic resource allocation driven by requirements of the information space and supported by characteristics of the network space. Resource allocation (RA) is frequently employed in domains needing SA (e.g. weather forecasting, homeland security and flight testing). RA is a tool which the operator/fusion node/system utilizes to improve the perception of the environment under supervision. RA could translate into deploying new sensors, switching on backup sensors or redistribution of sensors. Figure 3.4 below represents the SIC module in our system.

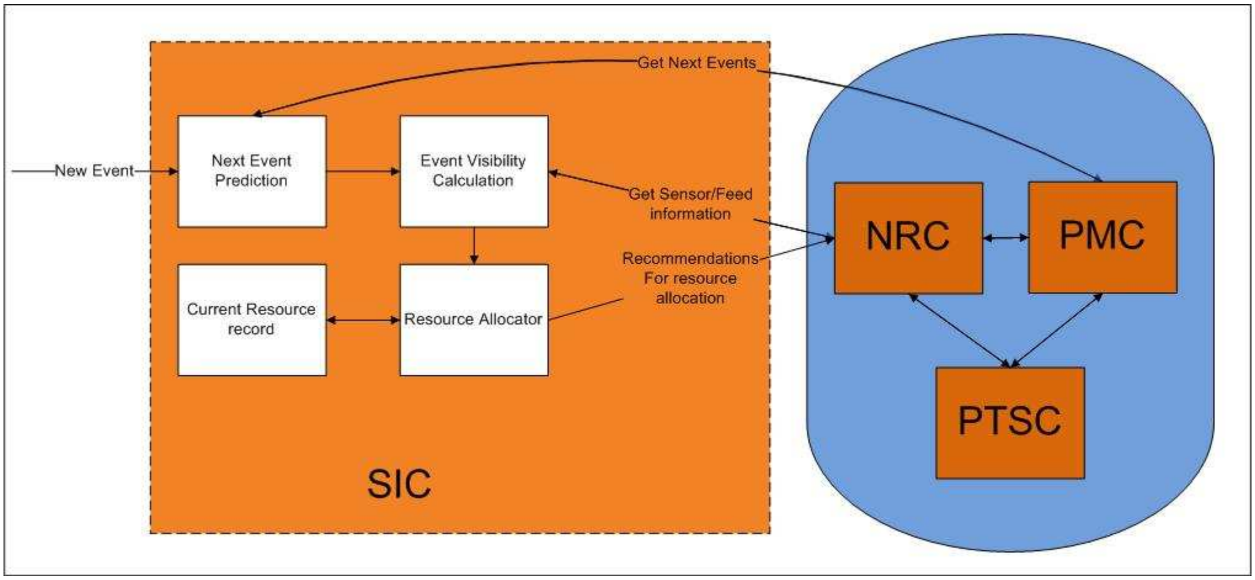


Figure 3.4: Flow Chart Explaining SIC Module

3.5.1 Resource Allocation

In our model, basic elements of resource allocation are explained below:

- **Observability:** Observability here is measured as the ability of network to report occurrence of an event with least failures. In our model, loss of transmission ability and sensor moving out of relay nodes communication range could cause failures. For us satisfactory observability is one where upper bound on probability that no sensor is able to transfer information about the event is less than certain threshold, e.g. 0.30.

- **Feed:** Necessity for a information "feed" arises when a fusion node from one branch of hierarchy is interested in information from the information space monitored by another branch. In simple words, a fusion node requests information from a relay node about an event as and when it occurs if *a*) the supervision zone of a relay node is same as that deduced for the predicted event *b*) the relay node has, under its authority, at least a sensor capable of observing the properties exhibited by the predicted event (if the event were to take place) *c*) the relay node does not fall in the logical network under the fusion node's authority.

Expiration: The feed expires when a) the event occurs and is reported as requested to the respective fusion node and b) if the probability of instantiation of the predicted event is less than κ (described in Section 3.3.2) in future.

- **Special Sensors (SS):** Special Sensors differ from sensors only in terms of re usability, as explained in Section 3.1. In simple words, special sensors can be redistributed in the network and are able to alter their sensing capabilities as requirements dictate whereas normal sensors are permanently assigned certain sensing capabilities and a relay node as their authority.

  – SS is deployed in the field with the objective of observing a certain predicted event.

  – SS is deployed/redistributed only when existing infrastructure is unable to observe the predicted event or satisfy observability requirements.

  – SS is equipped with at least one sensing ability to observe the properties exhibited by the predicted event at the time of allocation/redistribution .

  – When deployed, a SS is assigned the objective to report occurrence of a certain predicted event .

- SS only reports events which are present in its objectives.
- SS can be redistributed only when its objective list is empty.
- Sensing abilities of a SS can only be altered during redistribution
- Objectives could be revised/updated/added by the relay node upon special requests by fusion nodes.
- If a feed is established relying on a SS already under a relay node's authority, the corresponding event is added to objectives of the SS.

Expiration: Objectives assigned to SS expire in the manner identical to the one described above for a feed.

3.5.2 Understanding the SIC

For each new event predicted by the PMC module in order of decreasing criticality, SIC module

- computes the present visibility/observability if the event were to take place in the information space.
- allocates resources, if supervision zone of at least one relay node is same as the zone deduced for the predicted event, observability is not as desired and resources are available, in the following order:

  - A feed based on a normal sensor
  - A feed based on a special sensor
  - Redistribution of a SS, if objective list of the SS is empty
  - Allocation of a new SS

- Returns to previous step if the observability is not acceptable.

Here, we must stress that this level of modeling is not binding for other systems. One can choose as simplified or as convoluted model as per requirements and perform studies using this framework. In the following chapter, we discuss the motivation and results of the experiments performed with the networked sensor system modeled using the NCOPP framework.

Chapter 4

Testing and Evaluation of the NCOPP framework

Testing is a critical component of research in any framework for it to be established as a benchmark and to encourage its adaption in the relevant fields. Since there has been very little work focused at such overarching frameworks in the domain of networked information systems, it is even more essential for us to undertake experiments which provide critical intuitive validations to establish the NCOPP framework's ability to model heterogeneity in such complex systems. There are several inherent difficulties in the validation of frameworks such as NCOPP, e.g. initial steps to model the framework require a balance to be maintained between increasing complexity of the model and capturing realism. The model also needs to be left flexible enough to accommodate requirements of different domains. Since initial models employ abstraction to reduce complexity, these models are not primed to be validated by comparing data from real systems. Thus, initial experiments require critical and less complex experiments to establish the model's correctness.

To evaluate NCOPP and establish its ability to provide insights into interactions and effects of various factors (in the network and information spaces) on the overall performance of the system, we focus on a few simple yet critically important interactions between network and information space factors. Our experiments serve to reinforce the NCOPP framework's ability to model, represent and imitate the behavior of real systems. Before asserting that this framework can be utilized to get meaningful insights into prediction and performance analysis of networked information systems, we need to ascertain if this framework can indeed replicate known behavior and provide intuitively correct results for familiar interactions of different factors in the information and network space. In other words, we need to successfully explore the "known" with this framework before we rely on this framework to provide insights when we venture into the "unknown". By "unknown" we refer to complex interactions which are theoretically difficult to keep track of and model. We, however, reiterate that we are not modeling the network at the fine grain level but rater abstracting network characteristics without simulating data packet level interactions among the nodes in the network. These in and of themselves are the critical research questions, an d as such we leave the complete modeling of network for future work and thus, we are not aiming to match and validate the results of our simulation experiments with any type of real world data.

For our first step in the "known", we cross examine the need to filter the information by comparing advantages of a hierarchical sensor system (say A) employing information filtering against a system (say B) where no information filtering is undertaken at lower levels i.e. all computations and processing of information is done centrally at the apex node. The apparent advantage of employing system B against system A is the inherent simplicity in defining and modeling the interactions among various fusion nodes (agents) where as system A incorporates the complexity of information filtering and sharing among various agents involved. Also, system B will have, by definition, all information centrally available and thus awareness is more complete. On the other hand, the obvious benefit of system A is its scalability as the amount of incoming information increases. System B is bound to face critical computational barriers (such as limited processing power and memory), as the data collected and generated by such networked information systems increase; thus rendering system B largely infeasible. While mentioned above are all easily apparent advantages, another advantage of system A over system B is its ability to filter out the inconsequential information and thus, improving the quality of SA in the system. In essence, with our first experimental setup, we intend to explore the capability of our framework to provide the insight into the quality of SA to compare the information sharing strategies of systems A and B.

Our motivation to model both network and information space dynamisms collectively comes from our opinion that increasing network space dynamism has a deteriorating effect on the information sharing ability and thus the SA in such systems. We take the next step into exploring this intuition and question on whether network space dynamism has an apparent effect on the SA to find out if our framework is able to test and provide this intuition. To the best of our knowledge, this intuition within an overarching framework has been largely unexplored. We use our first experimental setup to also explore this critical question.

Once we evaluate effects of network dynamism on SA, we further run tests on the framework to get insights into the level of improvement achieved in SA when key network characteristics (e.g. network delay and node failure characteristics) are intuitively incorporated in the information sharing and resource allocation processes respectively. With these experiments we further expect to test the ability of the NCOPP framework to model interactions between the network and information spaces. We look to establish the capability and robustness of the framework to provide significant insights when unexplored interactions between key factors are predicted and analyzed. We note that discussions and results can also be found in [27].

4.1 Experiment I

In order to make the case for both hierarchical filtering and the cost/reward structure to model information sharing (described in Section 3.3.2) as our first step, we contrast the two approaches:

a) a hierarchical system of information processing and gathering, and

b) a centrally processed information gathering system,

in terms of superiority of the information quality in the two systems when they are employed in dynamic environments. We also use this experimental setup to confirm the effect of increasing network dynamism on the qualitative and/or quantitative performance of such systems.

4.1.1 Hypothesis 1

Since very few investigations have been undertaken previously to design such frameworks it is critical to meticulously undertake experiments to test even the most basic evaluations to affirm the framework's suitability to model and predict/analyze the performance of such complex systems. We first test the following critical hypothesis:

> *Information filtering at different levels in the hierarchy improves the quality of information in the general awareness picture (GAP).*

4.1.2 Our Approaches

To test our hypothesis and confirm both the proper functioning of different modules designed in the framework and the ability of the framework to replicate known behavior and insights, we compare the hierarchical filtering approach(referred to as Static) with our baseline approach (referred to as Primitive) which represents the central information processing system. The macroscopic description of these two approaches is provided below.

Primitive: This approach represents centrally processed distributed information gathering systems.

Procedure 1 Procedure Followed in Primitive Approach

for all event i **do**

  for all sensor j such that j records i **do**

    j routes i to the apex node k

  end for

  k adds i to update $Records(k)$

  k updates future event predictions

  k updates the corresponding scenario as (in)active

end for

At the apex fusion node k

for all Active scenario $l \in Records(k)$ **do**

  if $EndConfidence(l, k) \geq 20\%$ **then**

    Flag all knowledge and recorded events for the scenario l

    Reflect the flagged information in GAP

  end if

end for

where, a) $Records(k)$ refers to the event record register at fusion node k and b) $EndConfidence(l, k)$ refers to the probability of instantiating an end event if the scenario l progresses, according to the understanding of node k.

Static: This approach represents hierarchical distributed information gathering systems following information filtering paradigm.

Procedure 2 Procedure Followed in Static Approach

for all event i **do**

  for all sensor j which records i **do**

    j routes i to a fusion node k

  end for

  for all fusion node k such that k receives i **do**

    k adds i to update $Records(k)$

    k updates future event predictions

    k updates the corresponding scenario as (in)active

  end for

end for

At the each fusion node k:

for all prediction $m \in Predictions(k)$ **do**

  Share prediction m up/down the hierarchy following

  the schedule obtained by calculations in Section 3.3.2

end for

for all Active scenario $l \in Records(k)$ **do**

  if $EndConfidence(l, k) \geq 20\% \times Level(k)$ **then**

    Flag all knowledge and recorded events for the scenario l

    Reflect the flagged information in GAP

  end if

end for

where, a) $Predictions(k)$ refers to the predicted future events at fusion node k, b) $Level(k)$ refers to the tier of hierarchy to which the fusion node k belongs and c) $Level(apex\ node)$ is one.

4.1.3 Experimental Setup

Below, we describe the different aspects of our experimental setup.

Computing Resources:

- We conducted our experiments in a Unix environment on a system with 2.2 GHz dual core Centrino processor and 2GB RAM.
- We used Python4.4 to develop the discrete time event simulator and the NCOPP framework for our experiments.
- We are also using Lindo API to perform optimization calculations.

We used Python since it has in-built functions which help in quick development of the experimental setup. It can also successfully interface with the C language where intensive calculations and repetitive routines could be performed, thus allowing us to reduce the running time of our experiments.

Physical Network: At the physical level, we make the following choices while designing the network structure:

- The hierarchical network comprised of an apex node at level 1, with two fusion nodes at level 2 under its command.
- Each fusion node at level 2 has four fusion nodes under its command at level 3. Each fusion node at level 3 receives information from two relay nodes.
- Each relay node is randomly placed in one of the nine possible distinct zones, and it has authority over two sensor nodes.
- A sensor node is capable of detecting two independent (out of possible ten) properties which events, occurring in its supervision zone, may exhibit.

These choices are aimed at making our system large enough to represent a real networked information system. Three tier systems are particularly common today. The choice of nine zones for 16 relay nodes in total allows with high probability that at least one relay node in each zone. Also the choice of allowing only 2 distinct sensing capabilities to a sensor ensures that our system is dealing with added uncertainty and less redundancy while predicting future events.

Network dynamism in physical environment:

- Each sensor node temporarily looses transmission ability or moves out of coverage of the relay node once every X time steps with a 90% probability. We vary this duration from sixty to ten in steps of twenty to simulate the increasing effect of network dynamism. We refer to this duration as the *average duration between failures*.

- Once out of the coverage area of relay node or losing the transmission ability, the sensor stays out of the coverage area or without transmission ability for an average of ten time steps.

- Information transmission over communication links on an average takes three time steps.

The setup detailed above was chosen in order to cover a broad range of realistic unstable networks by varying the *average duration between failures*. Also the information transmission delay captures the delay experienced in systems where human operators are involved.

Information space:

- In our view, a normal human operator can multi task up to at least 3 different scenarios. since we have 8 fusion nodes at the lowest tier of our hierarchy, we run a discrete time event simulation with 30 active scenarios at a time.

- Running time for our simulation is 1000 time steps in order to allow for the effect of randomness in different variables to average out and provide a clearer picture.

- Since most common scenarios take a maximum of five stages to completion and about 6-10 different possible events. Thus, the scenarios we run are an instantiation of a sample 5-tier BKB which comprised of 8 I-nodes and 10 S-nodes with one end state.

- Each event occurring in the information space (here nine zones) exhibit two of the possible ten independent properties. This allows for 90 possible types of events which we believe provides enough coverage for different events in any type of scenario.

- Instantiation of an I-node (as a event in information space), when a corresponding S-node is activated, occurs after an average delay of 40, 50, 60, 70 or 80 time steps.

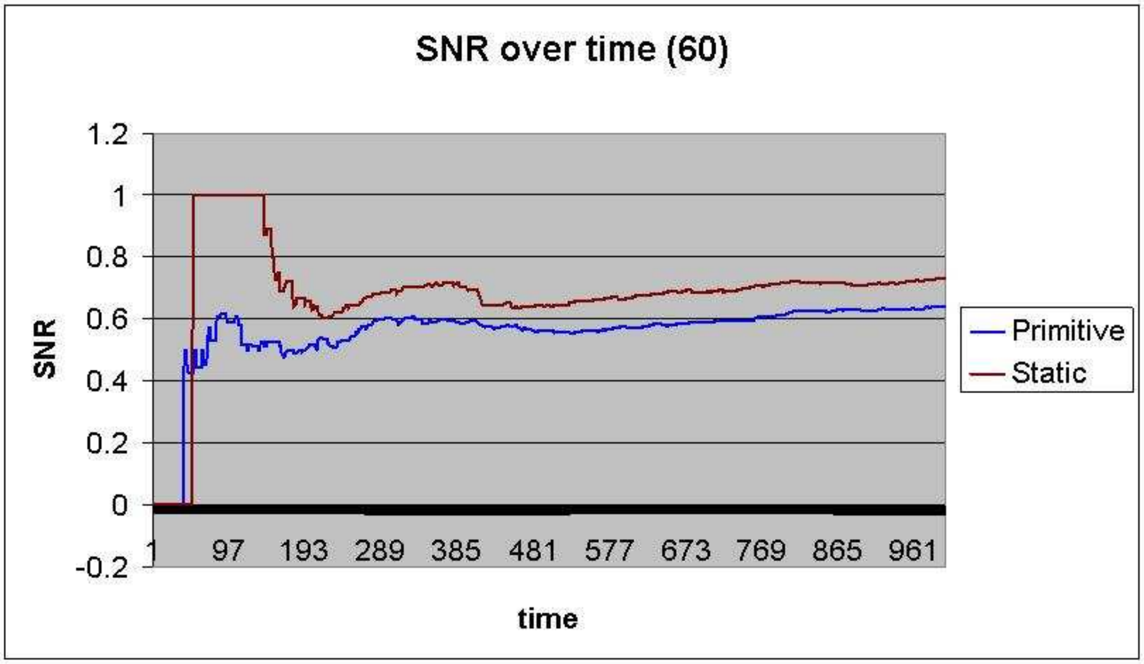
In addition to the above design choices we also note that:

- All variable trends follow Weibull probability distribution curve. We chose Weibull distribution since it can easily be modified to have different shapes and mathematically it was easier to formulate its optimization problem.

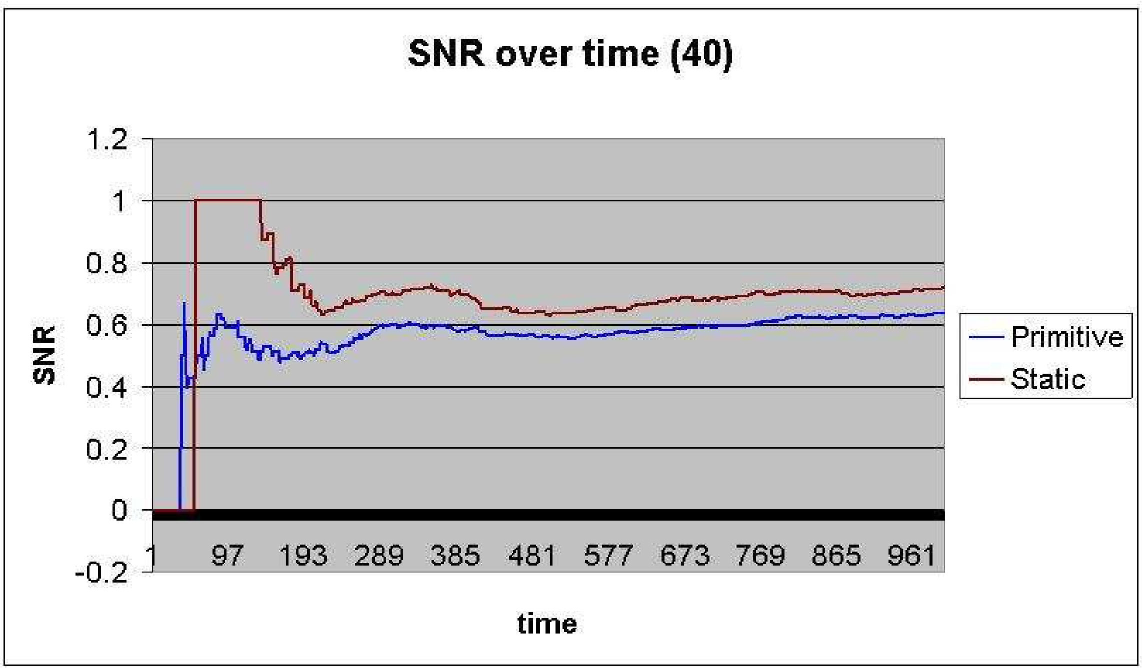
- All quantities (numbers) mentioned here can be configured to suit particular needs in any domain. Our choices are based on a setup that creates commonly occurring systems.

4.1.4 Results

In this experiment we are asking questions from our framework to provide insights into the difference in quality of situation awareness of two systems following different paradigms of information sharing (here centrally processed and hierarchical filtering). We are interested in investigating the impact of increasing network dynamism on performance of networked information systems.

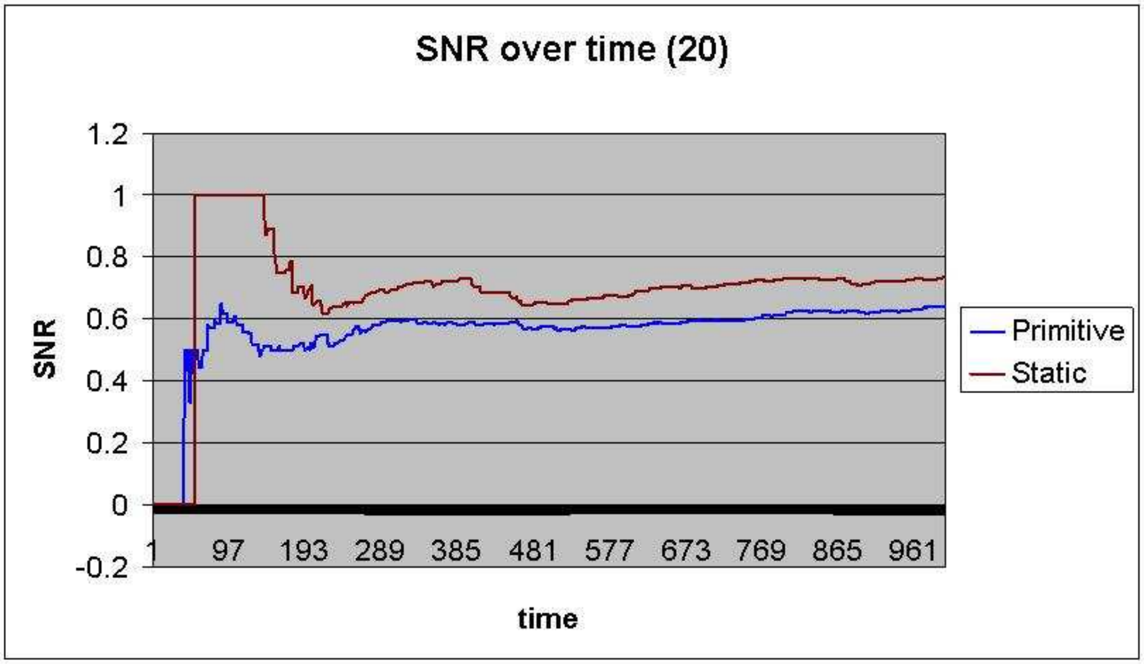


(a) Average Duration Between Failure = 60

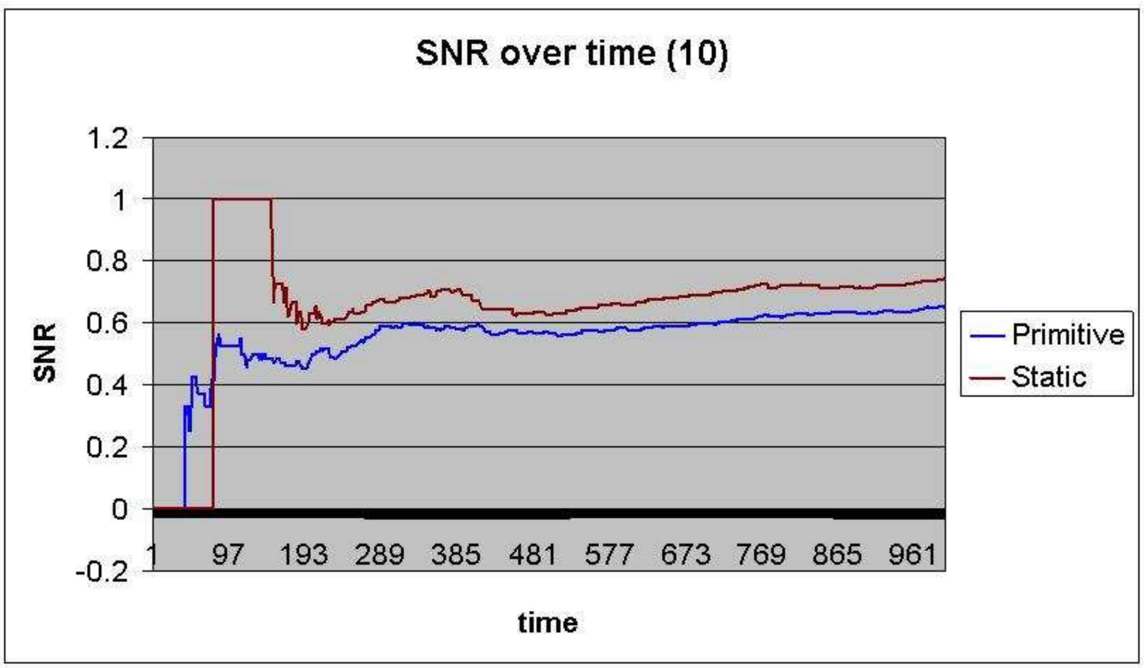


(b) Average Duration Between Failure = 40

Figure 4.1: Accumulative SNR Plot I



(a) Average Duration Between Failure = 20



(b) Average Duration Between Failure = 10

Figure 4.2: Accumulative SNR Plot II

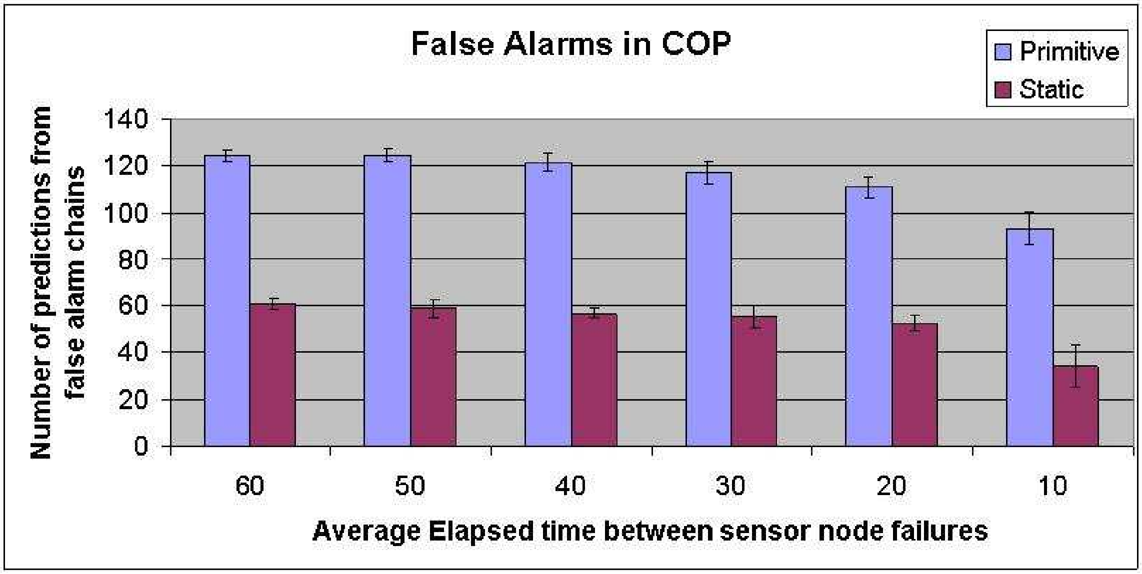


Figure 4.3: Comparison on Amount of Clutter in Two Systems with Increasing Dynamism

In our experiments we measure the quality of SA as SNR at each time step using Equation 3.4. According to our expectations, the quality of information in the system following hierarchical filtering was consistently better than in the system which delegated all processing to the apex fusion node. We ran simulations for different dynamic conditions and the accumulated SNR plots over time are shown in Figure 4.1 and 4.2. It is difficult to see the effect of network dynamism in quantitative terms in the SNR plots (Figure 4.1 and 4.2). We use the plot of information from *false alarms* in the GAP over the period of simulation in Figure 4.3. Using these two figures it is easy to demonstrate the effect of information filtering and network dynamism on SA. The **important take-away** from these results (Figure 4.1, 4.2 and 4.3) are:

- There is a consistent better quality of information (SNR) in the system following information filtering under dynamic conditions.

- Surprisingly, there is little or no effect on quality of SA (SNR) as the network dynamism is increased.

- In Figure 4.3, it is evident that the apex node will be overwhelmed by *false alarms* in case of no filtering.

- Hierarchical filtering retains significantly less percentage of false alarms and thus allows more and effective focus on critical situations.

The idea of distributed information processing and gathering hinges on the concept of quick local response and scalability. Besides these obvious advantages of distributed systems; hierarchical networks may come under scrutiny from suggestions advocating central processing of all data resulting in more complete and qualitatively better forecasting and awareness. We, while acknowledging more complete awareness in centrally processed systems, successfully exhibit that the deterioration in quality of the awareness needs to be considered as we move from distributed toward central information processing. Using this framework we can arrive at a comfort zone between the two extremes balancing the trade-off between completeness and quality of awareness. Our system with variable cost and reward structure allows for a smooth movement between the two extremes as explained in Section 3.3.2. While SNR of SA does not get affected by increasing network dynamism, we have successfully exhibited the perceptible effect of increasing network dynamism on the SA in a networked information system. This provides for the motivation to sustain or even improve SA in a system as network dynamism increases.

4.2 Experiment II

Having established the effect of network dynamism on the SA, we strive to understand the effect of accounting for network characteristics, on SA in a very basic/intuitive way. We argue that, *accommodating for network delay in the process of achieving SA* would improve the general awareness factor (GAF), as the dynamism in the network increases.

4.2.1 Hypothesis 2

Continuing the testing and evaluation of the NCOPP framework from the basic level, we aim to incorporate network delay (a critical network characteristic) in the process of information sharing in its very basic form and get a meaningful insight into its effect on the GAF factor and thus the SA.

> *Intuitively incorporating network delay characteristics while information sharing/filtering may improve the GAF factor.*

4.2.2 Our Approaches

In the hierarchical information filtering systems, which achieve a superior SNR as compared to the centrally processed systems, we are interested in employing network characteristics information to improve the GAF factor as network dynamism increases. Here we compare two approaches, Static and Dynamic. The Static approach, which acts as our baseline approach, remains same as described in Section 4.1.2. We explain the dynamic approach and the differences it has with the static approach above where, *a*) $SendingDown(m, t, k)$ refers to a condition check to find if k needs to send down/degrade the prediction m at time t according to the schedule obtained from calculation in Section 3.3.2, *b*) $PathExists(k, m)$ refers to all possible logical paths through which information about occurrence of event m will reach the fusion node k, *c*) $Delay(p)$ refers to the current network delay on the logical network path p and *d*) $DelaySending(m, k, Mindelay)$ is the routine which accommodates minimum delay information to defer degradation of event prediction.

Dynamic: This approach represents hierarchical distributed information gathering systems following information filtering paradigm and accounting for network delay characteristics while sharing information. Difference between dynamic and static approaches is shown in Procedure 3.

In the Dynamic approach the fusion node, before sending down(degrading) a prediction according to the time table, checks to find the minimum delay on the path from field to itself, if such a path exists. The path must exist between a sensor node, in the zone of predicted event capable of noticing the event, and the fusion node. The fusion node accordingly delays relegating the prediction to fusion nodes in the lower hierarchy. We expect, this step, though very basic, will improve perception and could improve the GAF as a result.

Experimental Setup for this experiment is same as defined in Section 4.1.3

Procedure 3 Procedural Differences in Static and Dynamic Approaches

IN STATIC APPROACH

for all fusion node k and time step t **do**

  for all prediction $m \in Predictions(k)$ **do**

    Share prediction m up/down the hierarchy following the

    schedule obtained from calculations in Section 3.3.2

  end for

end for

IN DYNAMIC APPROACH

for all fusion node k and time step t **do**

  for all prediction $m \in Predictions(k)$ **do**

    Share prediction m up the hierarchy following the

    schedule obtained by calculations in Section 3.3.2

    if $SendingDown(m, t, k)$ **then**

      Mindelay $\leftarrow \infty$

      for all path $p \in PathExists(k, m)$ **do**

        MinDelay $\leftarrow$ Min(MinDelay,$Delay(p)$)

        $DelaySending(m, k, MinDelay)$

      end for

    end if

  end for

end for

4.2.3 Results

In this experiment we recorded the GAF factor at each time step for the two approaches. We also recorded the total number of predictions from successful chains at different levels of network dynamism.

Important take away comments from these experiments are as follows:

- The dynamic approach presents better results and a more complete GAP in comparison to the static approach as shown in Figure 4.4, 4.5.
- We expected a steady improvement in the GAF as network dynamism increases, this was seen in all cases except when *average duration between failure* is 20 (Figure 4.5(a)) where the trend did not continue strongly. This could be attributed to comparatively excessive failures of sensors during this particular simulation of the dynamic approach since these failures are driven by probability and are not certain.
- However in Figure 4.5(b) and 4.4(b), we observe a decisive advantage of using network delay characteristics while sharing information in a hierarchical network.
- This also helps in recognizing the impact this basic way of incorporating network delay in information sharing has on the completeness of GAP in environments with high degree of instability.
- We included Figure 4.6 to demonstrate that the dynamic approach has increased resistance as compared to the static approach against the deteriorating effect of network

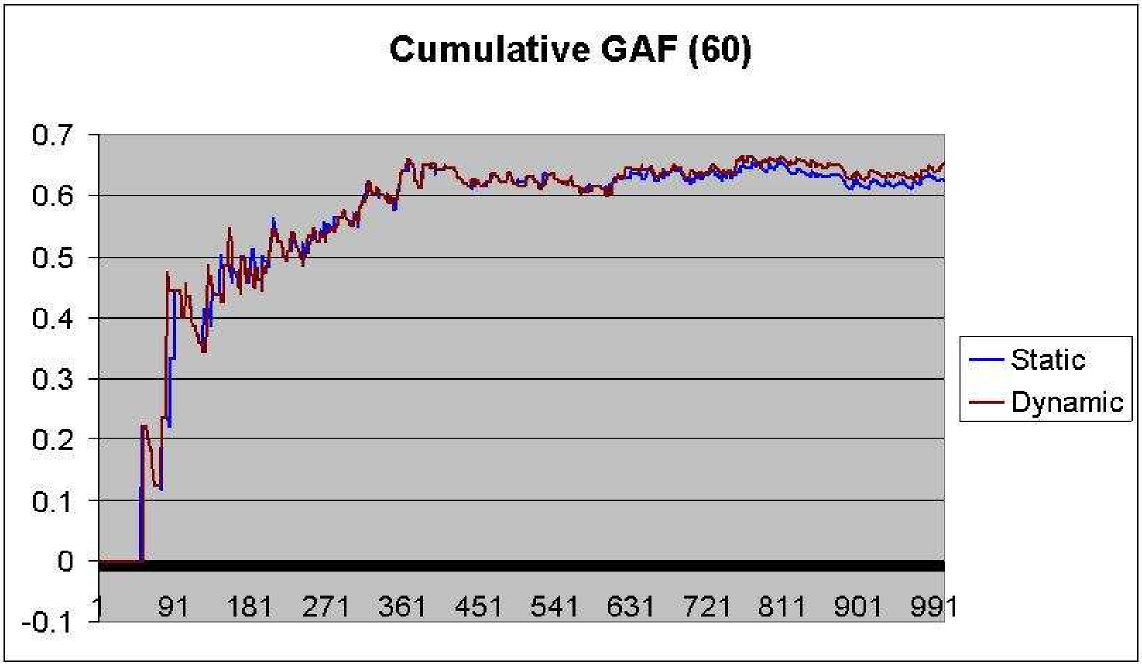
dynamism.

- Comparing the upper range of static approach with the lower range of dynamic approach at different levels of increasing dynamism in Table 4.1, we clearly see that there is no overlap of the two ranges (except in case where elapsed time between sensor node failure is 20 time steps) and the pattern clearly points at the advantage of including network delay characteristic in the information sharing process.

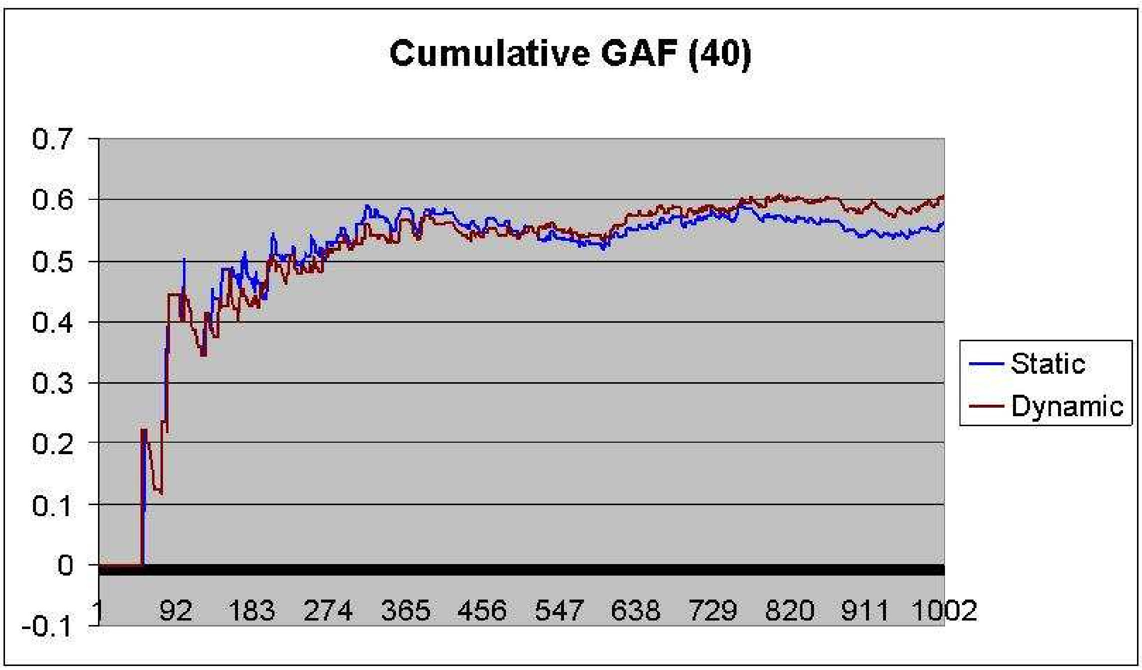
Here we demonstrated the ability of an intuitive and very basic approach which is accounting for network delay while sharing information, to improve the performance measurably. The framework allows us to get insights in the impact of technique adoptions/changes (no matter how small) on the performance of a system; and these results encourage us to look at the impact on SA when resource allocation is complemented with a critical network characteristics i.e. network delay. Here we have experimented with one network characteristic. There are numerous unexplored factors (including other network characteristics) and effects of such factors on SA could be studied/investigated in different combinations using NCOPP.

Table 4.1: Comparison of Lower(L) and Upper(U) Ranges for the Two Approaches.

	60	50	40	30	20	10
Static U-range	0.737	0.712	0.714	0.687	0.624	0.472
Dynamic L-range	0.737	0.731	0.716	0.690	0.618	0.522
Dynamic L - Static U	0	0.019	0.002	0.003	-0.006	0.05

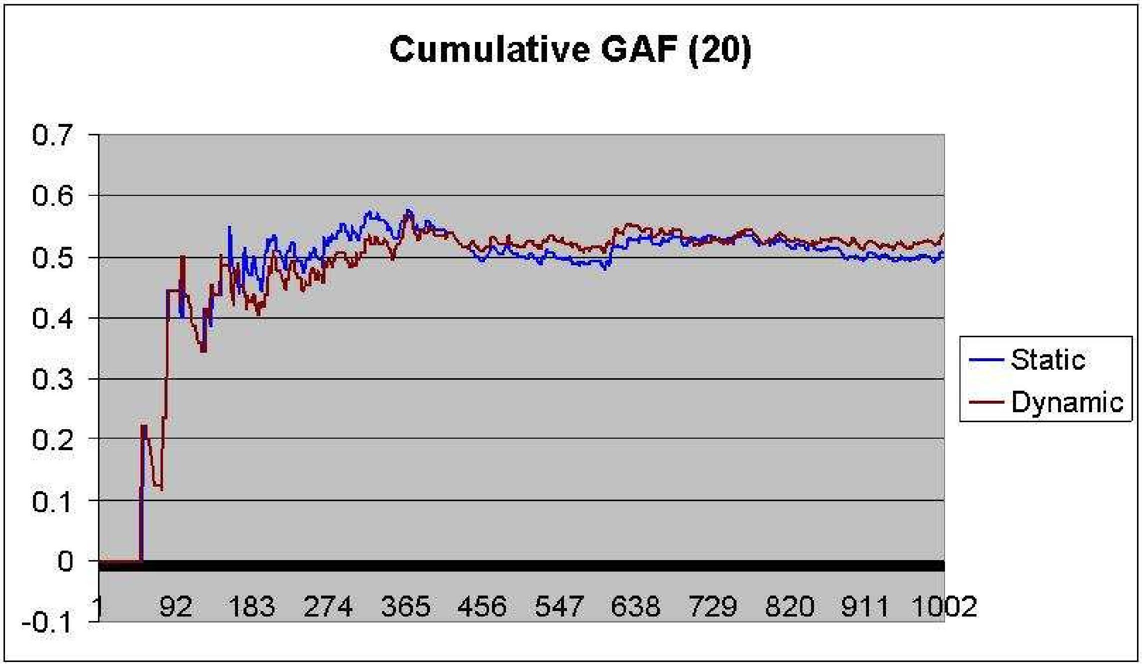


(a) Average duration between failure = 60

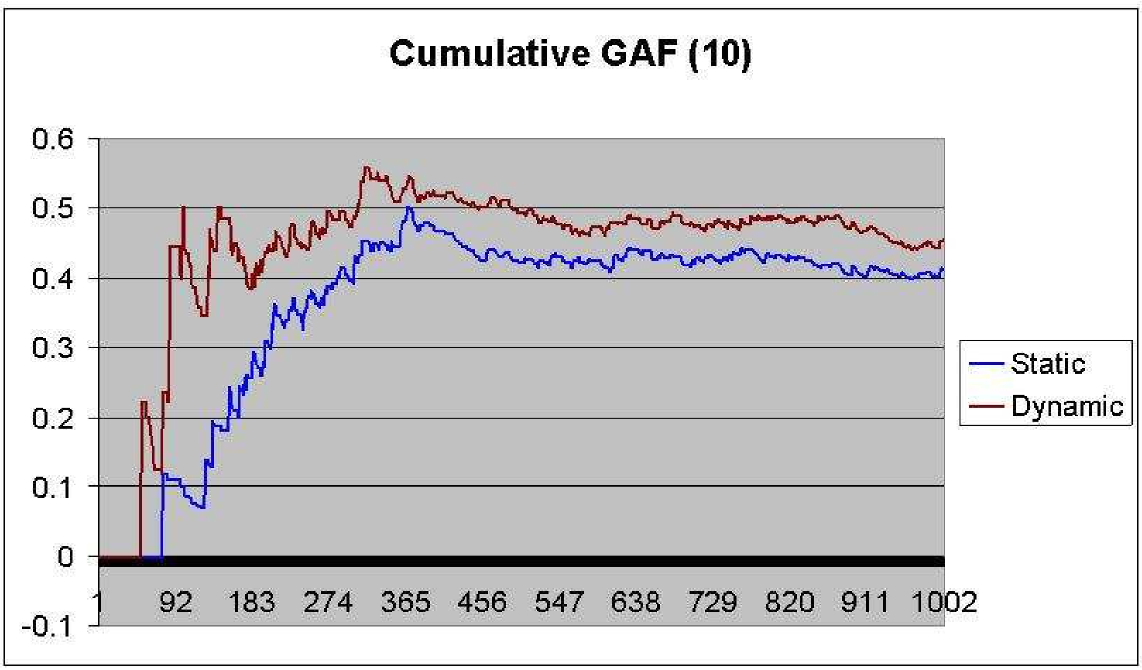


(b) Average duration between failure = 40

Figure 4.4: General Awareness Factor Plot I



(a) Average duration between failure = 20



(b) Average duration between failure = 10

Figure 4.5: General Awareness Factor Plot II

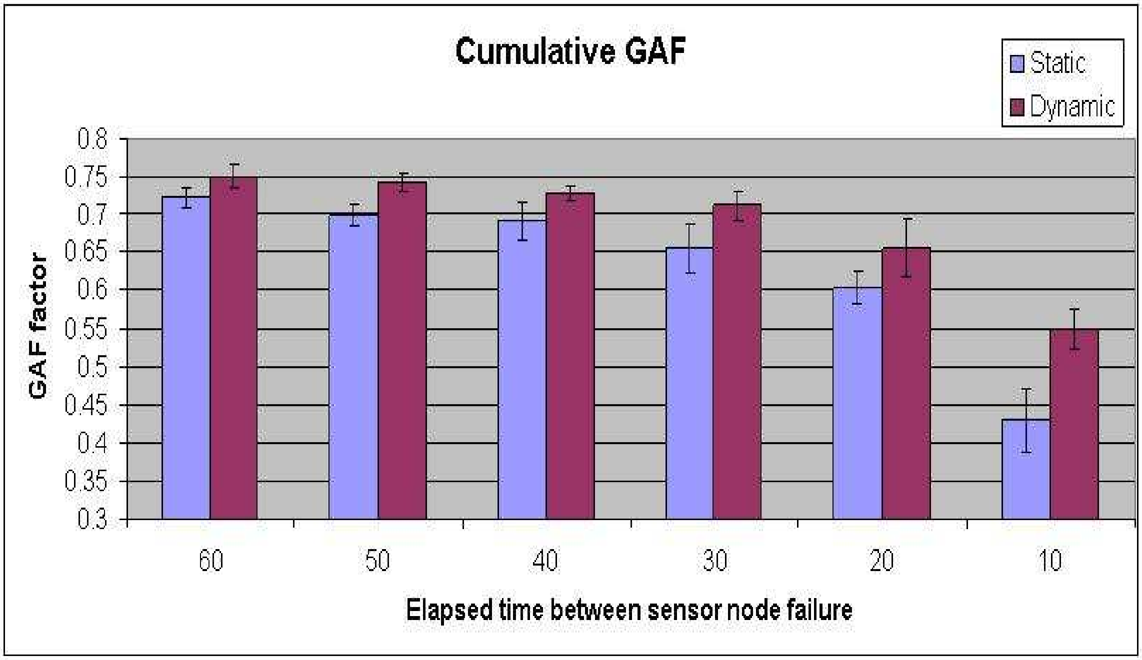


Figure 4.6: Comparison on Amount of Useful Information with Increasing Dynamism

4.3 Experiment III

Having successfully demonstrated the importance of incorporating network delay characteristics while sharing information, we focus on SA affecting network structure through dynamic resource allocation. Here the impact of supporting resource allocation decisions with network characteristics (e.g. sensor failure probability) and how it impacts perception is studied.

4.3.1 Hypothesis 3

Testing the ability of the NCOPP framework to accommodate increasing complexity and explore if limited resources guided by the state of both information and network space could provide better results than unlimited resources guided solely by the status of the information space. We compare resource allocation guided by requirements/state of information space only (referred to as *Proactive-Base*) with resource allocation guided by state of information space but also restricted by limited resource availability (referred to as *Proactive-Limited*). Hence, our hypothesis:

> *Taking sensor node failure characteristics into account while making decisions for dynamic resource allocation will improve the completeness of resulting situation awareness or GAF.*

Both approaches, however, are extended from the dynamic approach described in Section 4.2.2 allowing us to increase the complexity in a controlled manner using our framework.

4.3.2 Our Approaches

Description of both approaches is given below. In the Figure 4.7, we explain the steps followed by both approaches for resource allocation. These steps are in addition to the steps followed in the dynamic approach. The differences in the approaches are explained later.

Proactive-Base: This approach represents hierarchical distributed information gathering systems following information filtering paradigm and accounting for network characteristics while sharing information.

- This approach however does not take network characteristics into consideration while allocating resources in real time.

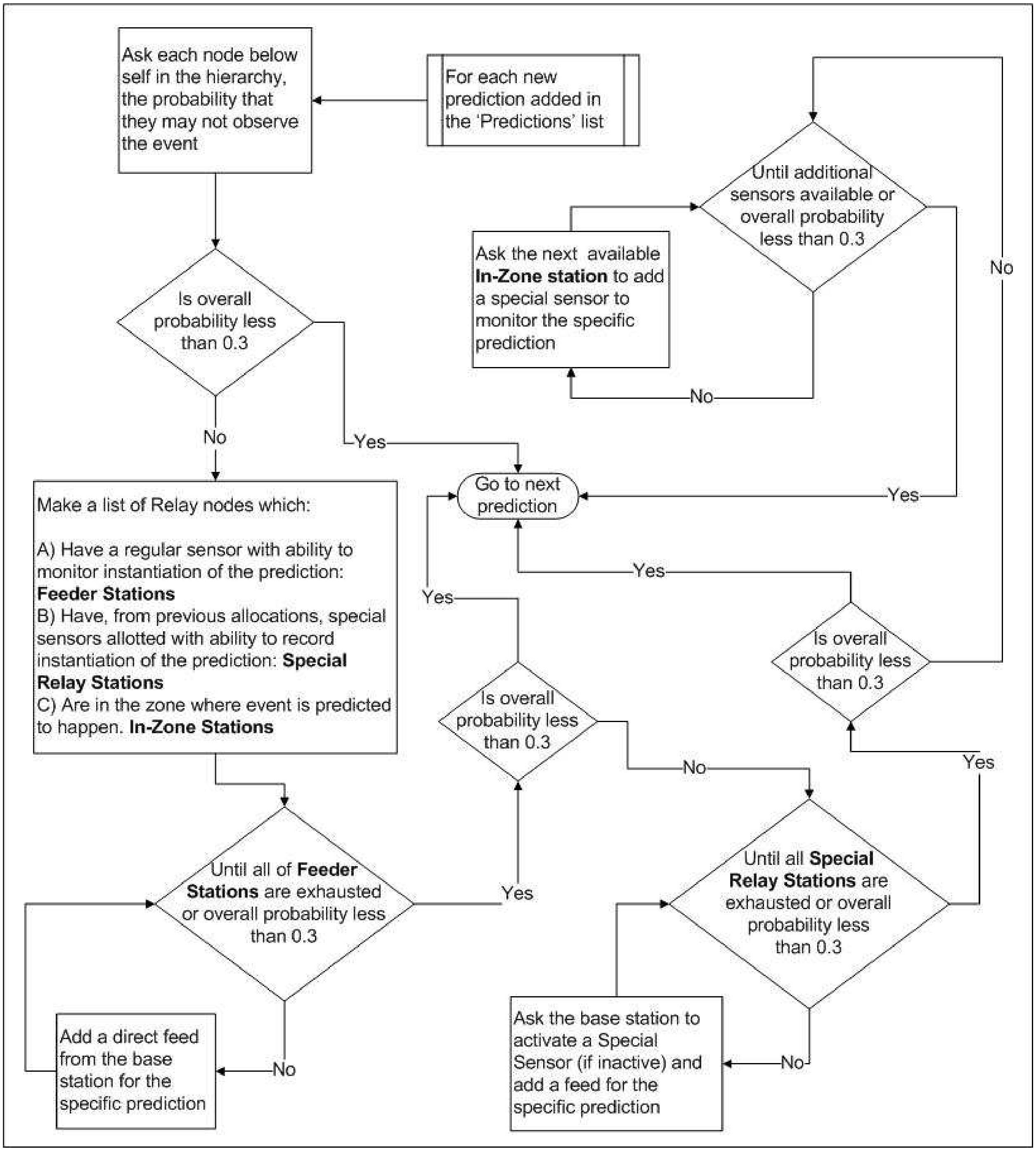


Figure 4.7: Flow Chart Explaining Resource Allocation Procedure

- This approach has no upper bound on the number of resources it could deploy.
- Overall probability that a particular prediction may not be observed goes to zero if at least one sensor node can monitor the event and the information can be delivered to the concerned fusion node.

Proactive-Limited: This approach represents hierarchical distributed information gathering systems following the/an information filtering paradigm and accounting for network characteristics while sharing information.

- This approach however does take network characteristics into consideration while allocating resources in real time.
- This approach has specific upper bounds on the number of resources it could deploy.
- For a fusion node predicting an event, the probability of not being able to observe the event decreases by the factor of the sensor node failure probability for each sensor that can detect instantiation of the prediction and evidence can be relayed to the fusion node.

4.3.3 Experimental Setup

Our experimental setup is different from that defined in Section 4.1.3 in the following ways:

- In this approach we allocated three relay nodes instead of two for each fusion node at level 3 since this increases the possibility of having at least one relay node in each of the 9 zones.
- Here, each sensor carries only one sensing ability to detect events occurring in its supervision zone. This ensures that the deployed network is not already saturated, and dynamic resource allocation is an essential requirement.
- In the *proactive-limited* approach the upper bounds on available resources (special sensors for deployment) are 8,16,24 and 32.
- All other properties of the system remain unchanged from the system described in Section 4.1.3.

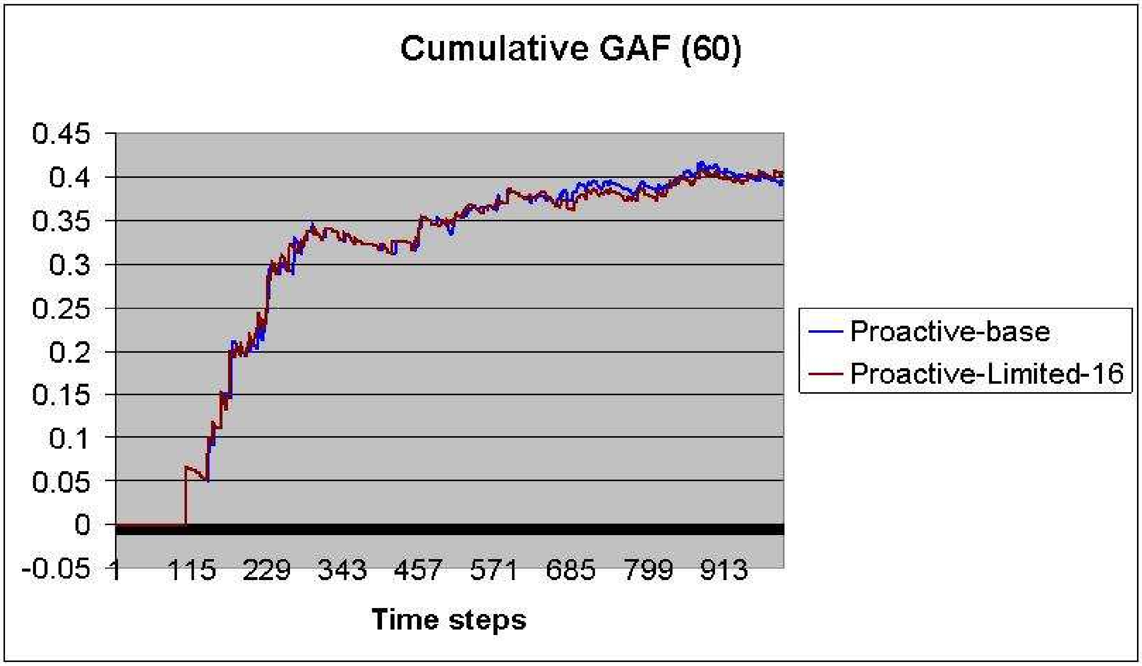
4.3.4 Results

In this experiment, the *proactive-limited* approach provides better results and a more complete GAP in comparison to *proactive-base* approach as shown in Figure 4.8 and 4.9. Out of the four upper bounds we chose for the *proactive-limited* approach, we only show the approach which performed better than the *proactive-base* approach with the minimum available resources. The explanation for this is that our *proactive-limited* approach is a heuristics based approach which does not always guarantee the optimum solution. However we demonstrate here that even a basic and intuitive approach can produce better results when network characteristics are taken into account while allocating resources.

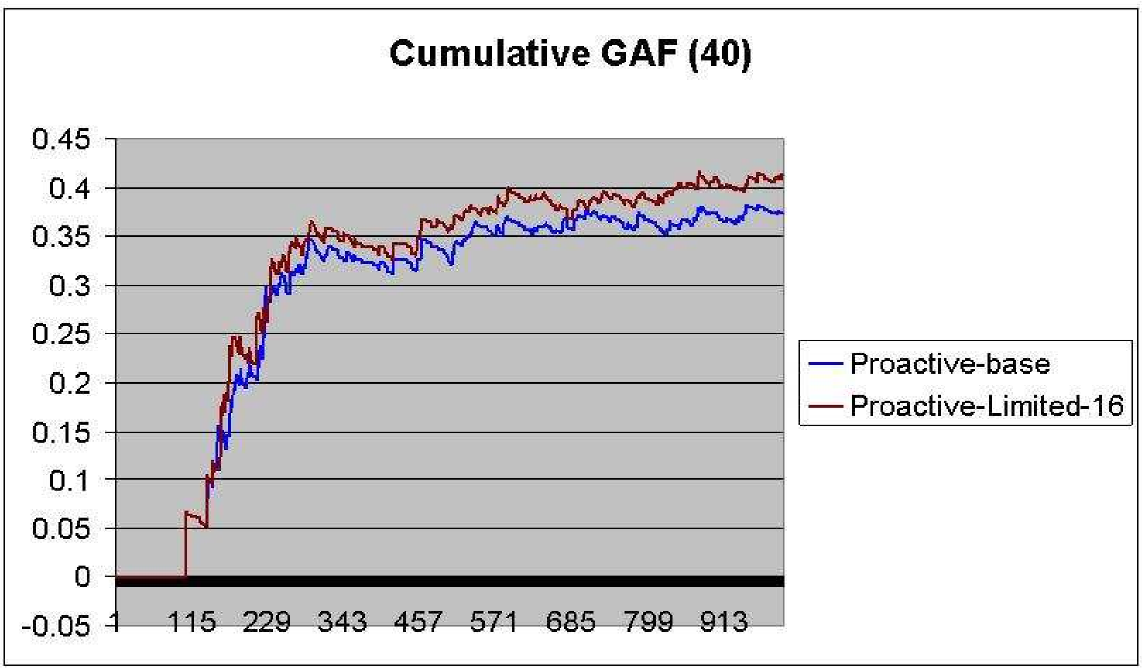
We identify a steady improvement in the GAF factor as the network dynamism increases, and in Figure 4.9(b) and Figure4.8(b), we observe a decisive advantage of utilizing network characteristics while allocating resources in a hierarchical sensor network.

If we see Figure 4.10, we can clearly observe a pattern which suggests that our heuristic based resource allocation scheme consistently achieves maximum GAF factor with limited resources when guided by sensor node failure probability. We can also observe a bell shaped pattern in the performance which suggests that not all resource allocation scheme may work best at different levels of network dynamism. It also implies that different resource allocation scheme must be studied to find *what is the most appropriate resource allocation scheme at different degrees of network dynamism* and *why*.

Here we demonstrated the ability of the approach, which accounts for sensor node failure in the network while allocating resources to improve perception of developing situations, in improving the GAF factor. The framework allows us to gather insights into how different strategies of resource allocation could improve the performance of the system.

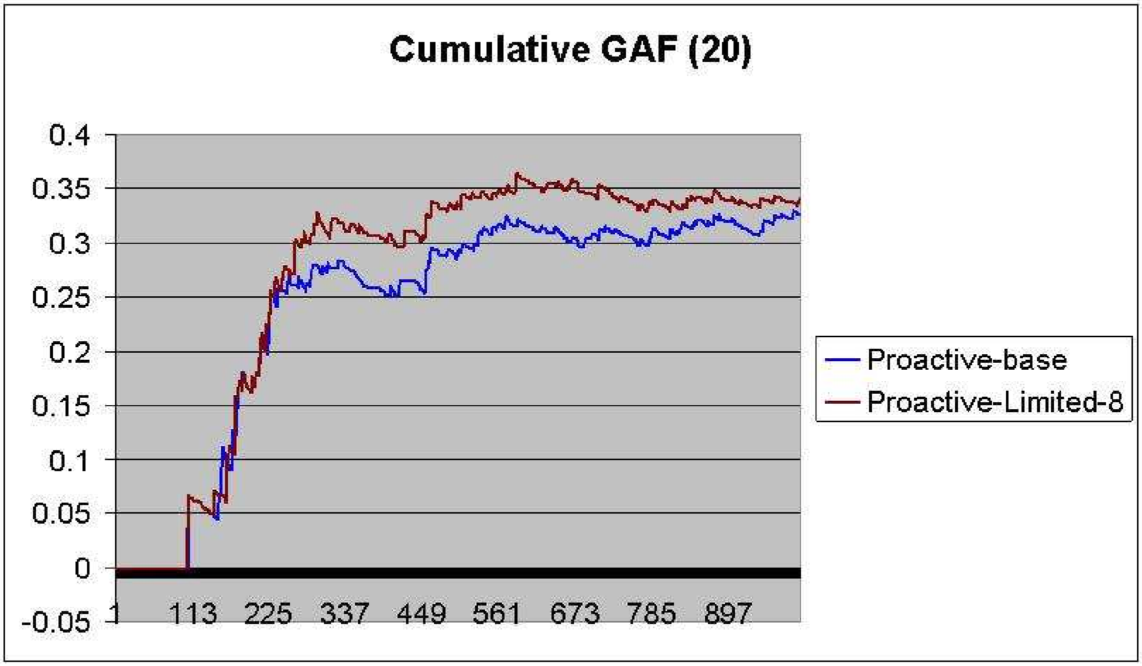


(a) Average duration between failure = 60

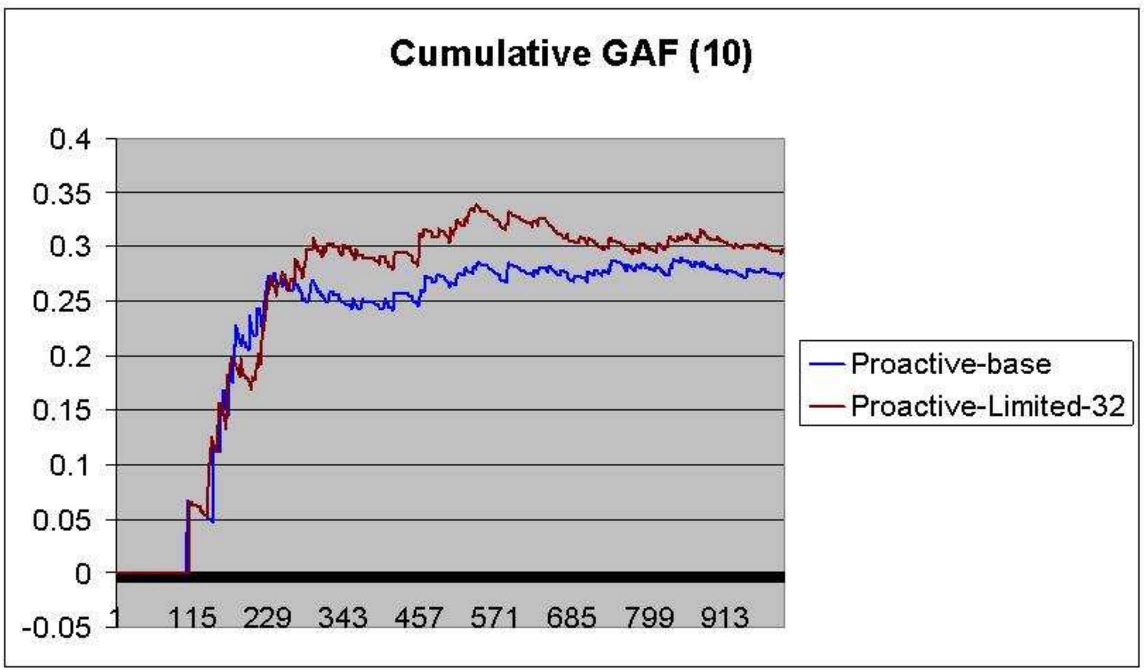


(b) Average duration between failure = 40

Figure 4.8: General Awareness Factor with Resource Allocation I



(a) Average duration between failure = 20



(b) Average duration between failure = 10

Figure 4.9: General Awareness Factor with Resource Allocation II

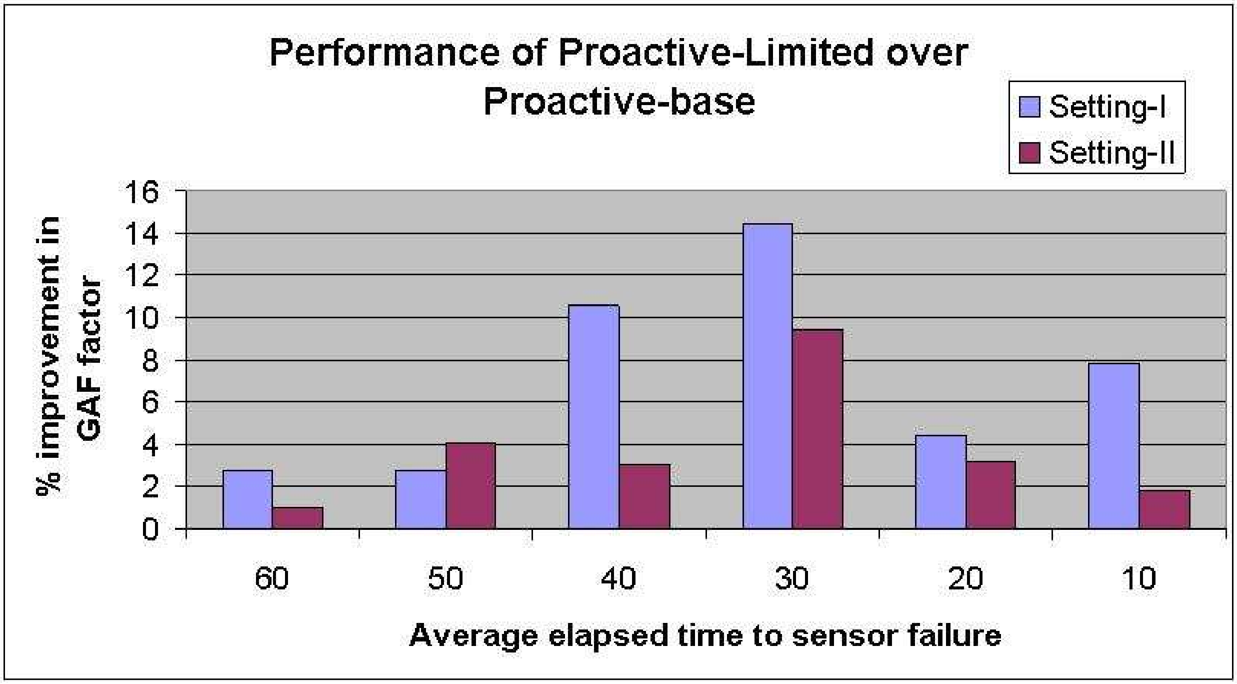


Figure 4.10: Bar Graph Comparison of Percentage Improvement in GAF Factor

4.4 Summary

Insufficient study has been done toward overarching frameworks to model networked information systems and hence there has been limited study of a multitude of network characteristics which effect the performance of networked information systems. We also believe that at the heart of the complex behavior of the network lie basic network characteristics, such as path delay and node failure characteristics. And these components/factors have not been critically studied to understand the behavior of large networked information systems. As such under NCOPP, this capability can be achieved by modeling such systems and performing experiments to obtain meaningful insights before analyzing the added complex interaction. We studied impact of network delay and node failure characteristics on performance of the network. We achieved insights into the ability of the most simple and intuitive methods of utilizing basic network characteristics to improve performance.

In this chapter we asked key questions regarding our framework to ensure that our basic model and concept are sound; also that different submodels are able to capture reality reasonably (with limited complexity in hierarchical networks). We have a "plug-&-play" framework and thus flexible to accommodate networked information systems from a variety of domains. This work now paves the way for possibly modeling the network at a finer level, extending this model to accommodate mesh/irregular networks and increasing the complexity of information sharing and resource allocation strategies.

The results of this work are particularly promising as they illustrate that any networked information system can be encapsulated into the NCOPP framework and its submodels; and meaningful insights (spanning coarse to fine grained) into the interaction of various external/internal factors and their effect on the performance of the network can be extracted. Using our framework, these insights can be further utilized to present recommendations for structural or policy modifications to improve the performance of such systems. This work further strengthens the potential of this overarching framework to be further refined and studied in networked information systems across different domains.

Chapter 5

Conclusion and Future Work

5.1 Conclusion

Situation awareness is a major performance consideration in networked information systems. The widely accepted model of achieving SA, provided by Endsley [7], does not consider the network as an important part of that model. A multitude of scenarios (enumerated in Section 1.3) emphasize that network characteristics along with a variety of factors play a very important role in determining the quality and completeness of situation awareness and there is a need for network and its characteristics to be incorporated in the Endsley's model.

However, current methods have refrained from studying SA with a uniform framework that can robustly model these various factors and aid in prediction and performance analysis of any system dedicated to build and maintain situation awareness. We design and evaluate

Network Centric Operation Performance & Prediction (NCOPP) [26], an overarching framework with the abilities of using “plug-&-play” modules, accommodating heterogeneity and providing real time prediction and performance analysis. This framework allows us to model network and information space dynamisms and complex decision making processes which reflect dilemma of information sharing in such systems.

In this thesis, we designed the framework with which a system’s approach toward information sharing (between the extremes of completely distributed and completely centralized processing) can be studied in light of balancing a trade-off between competing performance measures of networked information systems (completeness of awareness and quality of general awareness picture (GAP) in this work). We modeled a representative 3-tier hierarchical fusion network (a simple tree structure) which relied further on relay and fusion nodes to collect information from a dynamic environment. We studied the impact of incorporating key network characteristics (network delay and sensor node failure probability) in important decision making processes (information sharing and resource allocation), on the improvement in SA (a measure of information sharing ability) of the system.

We showed an average reduction of 47% induced by hierarchical filtering in the false-alarm information over central processing. This framework will also help in predicting and analyzing the performance of a networked information system with respect to understanding the interactions and effects of various factors at play in network and information spaces. We demonstrated the deteriorating effect of network space dynamism on the performance of the hierarchical networked system which comes across dynamic scenarios and temporal properties of events in the information space. Our experiments with the plain approach of incorporating the knowledge of network delay while information sharing, have shown a 10% improvement in the GAF factor under increasingly dynamic network conditions. We also registered an improvement of 5% in the GAF factor when sensor node failure characteristics are considered in the process of allocating resources to improve the observability of scenarios. This must encourage future research in the direction of utilizing network information in a sophisticated way, to further improve the quality and completeness of situation awareness.

Lastly, we were able to design and evaluate a framework that can robustly accommodate heterogeneity of nodes and links in a network. This framework has the flexibility to provide insights into the interactions between various factors in the information and network spaces. This framework allows for predicting and analyzing the effectiveness of networks and different employable strategies, in providing and maintaining a reliable situation awareness. This is required for successful operations under varying physical environments and logical scenarios.

The results underline the potential of this overarching NCOPP framework to define any networked information system into its submodels, predict, analyze and provide substantive understanding of the effect of various factors (individually or in a combination) on the performance of such systems.

5.2 Future Work

We focused on coarse grained analysis of hierarchical networked information systems by abstracting key network characteristics. To focus on establishing the ability of the NCOPP framework to model such systems, we reduced the focus and coarsely modeled some aspects of such system. Following are some of the possible future works which could be investigated to further study and validate the modeling capability of this framework:

- Incorporate learning from incoming new information
- Modeling network at the level of data packets and including routing protocols to provide comparison of data from real time systems
- Modeling mesh or irregular networked information systems using this framework

These above mentioned works would provide new challenges and avenues to further investigate the capability of theoretical overarching frameworks (especially NCOPP) in modeling real world systems and providing important insights into behavior of such scalable networked information systems.

Bibliography

[1] M. J. Adams, Y. J. Tenney, and R. W. Pew. Situation awareness and the cognitive management of complex systems. volume 37 of *Human Factors: The Journal of the Human Factors and Ergonomics Society*, pages 85–104. Human Factors and Ergonomics Society, 1995.

[2] D. Akselrod, A. Sinha, T. Kirubarajan, M. Farooq, and Z. J. Ding. Network-centric multisensor-multitarget tracking testbed based on peer-to-peer communication. In *Proceedings of 2006 Canadian Conference on Electrical and Computer Engineering*, pages 46–51. Ottawa Congress Centre, Ottawa, Canada, 2006.

[3] L. Chen, X. Yin, and W. Chen. Analyzing the effectiveness of network centric operating system in information domain. *Systems Engineering and Electronics*, 26(7):918–923, 2004.

[4] C. Dominguez. Can SA be defined? Situation Awareness: Papers and annotated bibliography. Technical Report AL/CF-TR-1994-0085, Armstrong Laboratory, Ohio, June 1996.

[5] M. R. Endsley. Design and evaluation for situation awareness enhancement. In *Proceedings of the Human Factors Society 32nd Annual Meeting*, pages 97–101. Human Factors Society, 1988.

[6] M. R. Endsley. A survey of situation awareness requirements in air-to-air combat fighters. *International Journal of Aviation Psychology*, 3(2):157–168, 1993.

[7] M. R. Endsley. A taxonomy of situation awareness errors. volume 3 of *Human Factors in Aviation Operations*, pages 287–292. Ashgate Publishing, Limited, 1995.

[8] M. R. Endsley. Theoretical underpinnings of situation awareness: A critical review. In M. R. Endsley and Garland D. J., editors, *Situation Awareness Analysis and Measurement*. Lawrence Erlbaum Associates, Mahwah, NJ, 2000.

[9] M. R. Endsley and D. G. Jones. Situation Awareness in air traffic control. In *Advances in Applied Ergonomics*, pages 394–397. USA Publishing, Lafayette, IN, 1996.

[10] M. R. Endsley and E. O. Kiris. Situation awareness in FAA airway facilities maintenance control centers (MCC): Final report. Technical report, Texas Tech University, Lubbock. TX, August 1994.

[11] M. R. Endsley and M. M. Robertson. Team situation awareness in aviation maintenance. In *Proceedings of the Human Factors and Ergonomics Society Annual Meeting*, pages 1077–1081. Human Factors and Ergonomics Society, 1996.

[12] M. R. Endsley and M. D. Rodgers. Situation Awareness information requirements analysis for en route air traffic control. In *Proceedings of the Human Factors and Ergonomics Society 38th Annual Meeting*, pages 71–75. Human Factors Society, 1994.

[13] K. A. Ericsson, N. Charness, P. J. Feltovich, and R. R. Hoffman. *The Cambridge Handbook of Expertise and Expert Performance*. Cambridge University Press, 2006.

[14] M. L. Fracker. A theory of situation assessment: Implications for measuring situation awareness. In *Proceedings of the Human Factors Society 32nd Annual Meeting*, pages 102–106. Human Factors Society, 1988.

[15] Lindo Systems Inc. *LINDO API 5.0, User Manual*. 2006.

[16] Y. Lin, L. Wang, Y. Chen, J. Zhong, Z. Chen, and T. Tong. A distributed data-centric clustering hierarchical routing algorithm for sensor networks. *Acta Electronica Sinica*, 32(11):1801–1805, 2004.

[17] J. Loomis, R. Porter, A. Hittle, C. Desai, and R. White. Net-centric collaboration and situational awareness with an advanced user-defined operational picture (UDOP). In *Proceedings of 2008 International Symposium on Collaborative Technologies and Systems (CTS '08)*, pages 275–284. Irvine, California, USA, 2008.

[18] N. H. Lu. Characterization of situation awareness performance. volume 2 of *Proceedings of Military Communications Conference, 1996.*, pages 632–636. McLean, VA, USA, 1996.

[19] R. Maciejewski, K. Y. Sung, D. King-Smith, K. Ostmo, N. Klosterman, A. K. Mikkilineni, D. S. Ebert, E. J. Delp, and T. F. Collins. Situational Awareness and visual analytics for emergency response and training. In *Proceedings of 2008 IEEE Conference on Technologies for Homeland Security*, pages 252–256. Waltham, MA, USA, 2008.

[20] C. N. Manikopoulos and J. L. Ucles. Average Information Staleness (AIS) as a system measure of performance. In *Proceedings of Third IEEE Symposium on Computers and Communications. ISCC'98*, pages 478–482. Athens, Greece, 1998.

[21] P. Mirhaji, Y. F. Michea, J. Zhang, and S. W. Casscells. Situational Awareness in public health preparedness settings. volume 5778 of *Proceedings of the SPIE - The International Society for Optical Engineering*, pages 81–91. Orlando, FL, USA, 2005.

[22] M. G. Oxenham. Enhancing situation awareness for air defence via automated threat analysis. In *Proceedings of the Sixth International Conference on Information Fusion*, pages 870–877. Univ. New Mexico, Gallup, NM, 2003.

[23] K. T. Ozan and G. Ferrari. *Ad-Hoc Wireless Networks: A Communication-Theoretic Perspective*. John Wiley and Sons, 2006.

[24] J. Patrick and S. Belton. Situation awareness in the nuclear industry. *Nuclear Engineering International*, 48(582):36–40, 2003.

[25] T. S. Raghu, R. Ramesh, and A. B. Whinston. Addressing the homeland security problem: a collaborative decision-making framework. *Journal of the American Society for Information Science and Technology*, 56(3):310–324, 2005.

[26] E. E. Santos. A framework for assessing and predicting network loads and performance for network-centric operations and warfare. volume 6578 of *Defense Transformation and Net-Centric Systems 2007*, pages 65780T–1–9. Orlando, FL, USA, 2007.

[27] E. E. Santos, A. Ojha, and J. Korah. Net-Centric operations evaluation using NCOPP framework. Technical Report LCID-08-201, LCID, Virginia Tech, Blacksburg, VA, September 2008.

[28] E. Santos Jr. and E. S. Santos. A framework for building knowledge-bases under uncertainty. *Journal of Experimental and Theoretical Artificial Intelligence*, 11(2):265–286, 1999.

[29] N. B. Sarter and D. D. Woods. Situation Awareness: A critical but ill-defined phenomenon. *International Journal of Aviation Psychology*, 1(1):45–57, 1991.

[30] E. Shahbazian, G. Rogova, and P. Valin. *Data Fusion for Situation Monitoring, Incident Detection, Alert And Response Management*. IOS Press, 2005.

[31] D. A. Vincenzi, M. Mouloua, and P. A. Hancock. *Human Performance, Situation Awareness and Automation: Current Research and Trends*. Routledge, 2004.

[32] P. Walsh. Brigade task force XXI performance. Technical Report IPR Briefing to CECOM, ITT AC/D, October 1995.

[33] H. Wang, X. Liu, J. Lai, and Y. Liang. Network security situation awareness based on heterogeneous multi-sensor data fusion and neural network. In *Proceedings of the Second*

International Multi-Symposiums on Computer and Computational Sciences, pages 352–359. IEEE Computer Society Washington, DC, USA, 2007.

[34] W. Weibull. A statistical distribution function of wide applicability. *Journal of Applied Mechanics*, 18:293–297, 1951.

[35] D. A. White, L. Young, and G. R. Kelsch. Modeling and simulation of situational awareness in the tactical internet. volume 3 of *Proceedings of Military Communications Conference, 1996.*, pages 872–876. Boston, Massachusetts, 1998.

[36] C. Wilson. Network centric operations: Background and oversight issues for congress. Technical Report ADA466624, Library of Congress Washington DC Congressional Research Service, Washington DC, March 2007.

[37] B. Zhou, L. H. Ngoh, B. S. Lee, and C. P. Fu. A hierarchical scheme for data aggregation in sensor network. volume 2 of *Proceedings of 2004 12th IEEE International Conference on Networks*, pages 525–529. Singapore, 2004.