UNDERSTANDING THE RHYTHMS OF EMAIL PROCESSING STRATEGIES IN A NETWORK OF KNOWLEDGE WORKERS

By

ASHISH GUPTA

Bachelor of Engineering
Marathwada University
Nanded, India
1998

Master of Science
Oklahoma State University
Stillwater, Oklahoma
2001

Submitted to the Faculty of the
Graduate College of the
Oklahoma State University
in partial fulfillment of
the requirements for
the Degree of
DOCTOR OF PHILOSOPHY
May, 2007

UNDERSTANDING THE RHYTHMS OF EMAIL PROCESSING STRATEGIES IN A NETWORK OF KNOWLEDGE WORKERS

Dissertation Approved:

Dr. Ramesh Sharda, Dissertation Adviser

Dr. Rick L. Wilson

Dr. Marilyn Kletke

Dr. Manjunath Kamath

Dr. A. Gordon Emslie, Dean of the Graduate College

ACKNOWLEDGMENTS

I could not have accomplished this work without the blessings and support of numerous faculties, friends, and family members. I would like to start by thanking my dissertation chair, Dr. Ramesh Sharda, on whom I could always count for the excellent guidance and support from him throughout my years as a doctoral student. I can never thank him enough for being so patient with me and for his contributions and suggestions that tremendously helped shape this study. He has been a great mentor for me!

I would also like to extend my sincere thanks to committee members Dr. Rick Wilson, Dr. Marilyn Kletke and Dr. Manjunath Kamath who not only helped me during different stages of this study but also at various points during my life as a Masters and Doctoral student at OSU.

I would also like to thank my friends Peter Rosen and Han Li for making my stay at OSU memorable.

This was certainly unachievable without the unfaltering support of my Mom and Dad. I thank them for all their lifetime blessings, teachings, love, and support. To my wonderful brother, Abhishek, to my great sister, Shal, and Brother-in-Law, Sudhir, for providing all the encouragement, affections and for helping me through out my life and this program. To my wife, Rupee, for being there for me all the time! I thank her for providing all the unconditional love, and support that kept me going every day. To my lovely daughter Esha, for making every moment of my life worth living and waiting for!

Finally, I would like to thank my Grand mother, who lives in Heaven, and to God.

TABLE OF CONTENTS

Chapter Page

Chapter Page

LIST OF TABLES

Table Page

Table Page

Table Page

LIST OF FIGURES

Figure Page

1. INTRODUCTION

This section first broadly describes various reasons for managing emails. A summary of major problems associated with the use of email at work is presented and the use of email from a paradox perspective is explained. The second subsection describes the research problem that is the focus of this dissertation study. Third subsection discusses the significance of conducting this study.

1.1 Need for Email Management

Email is considered one of the most widely adopted internet-based applications ever built and is used in many business functional concerns. It has now become the most prevalent mode of business communication and information exchange within organizations and has changed the way we spend our life at work. Email, due to its asynchronous nature, provides several advantages over other communication tools such as telephone or instant messaging. We get the much needed flexibility and the latitude needed for correcting any error before responding to a message. Geographical location is no longer a constraint for organizations, as well. Email provides a very cheap and fast means for sharing information and is open to all. Further, one-to-many communication can be accomplished using email without incurring any additional costs. It is known to assist individuals with time-effectiveness and effectiveness (McManus et al. 2002).

But recently, several scientific and anecdotal reports have started to indicate that knowledge workers are spending enormous amounts of their time interacting with emails at work. While this excessive interaction between knowledge worker (email user) and email (technology) has produced several good effects such as increasing productivity and faster information exchange, it has, at the same time, spawned side effects such as email overload, interruptions, technology addiction, attention deficiency, productivity loss, etc.

Today's email office reaches far beyond the limits imposed by the boundaries of traditional organizations and is certainly a lot more intricate due to complex email interactions. And these interactions do not necessarily depend only on simple hierarchies but also on workers' job functions, mutual interests, and collaborative team composition. The work environment using email (Figure 1.1) can be said to comprise several (email) users, email technology and different types of interactions taking place at individual, group, and organizational levels (for example, email interaction in a business-to-business model). This work environment is now marred by problems arising either from email users or technology or both, and (or) from the interaction between email user and technology. Good examples of these problems are (1) email addiction that occurs due to an email user's pathological behavior (Weber 2004), (2) the problem of email content management and archiving that occurs due to deficiencies with the technology (Gupta et al. 2006), and (3) the problem of email interruptions and overload that occurs due to interactions between user and technology, requiring improvement at both ends (Gupta et al. 2004).

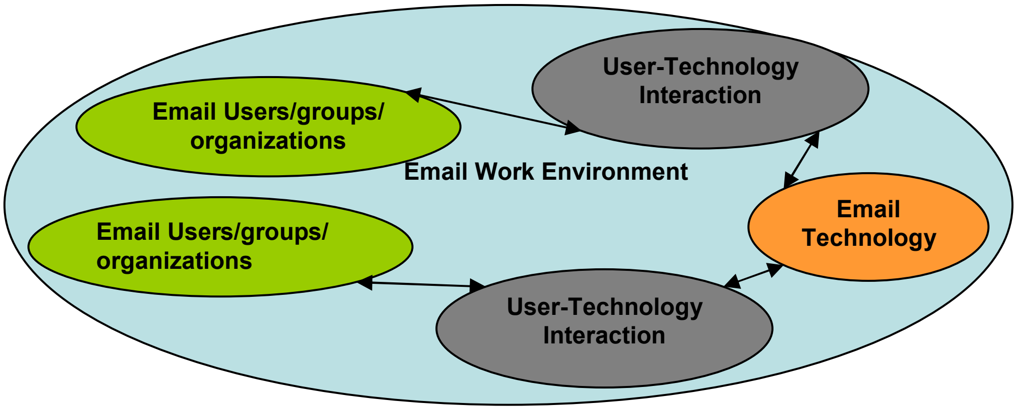


Figure 1.1 Email Work Environment

This dissertation study focuses specifically on two problems: email overload and email use within organizations. Presently, knowledge workers use email processing strategies that are as random and diverse as their daily work requirements and email arrival patterns. This inefficient handling of emails leads to increased overload. We use a multi-dimensional perspective comprising email scheduling strategies, work routines and time to address this problem. We do this by proposing a new way of processing emails that may bring about a change in the timing and the manner in which emails are processed within organizations. The use of various time-based scheduling approaches for processing emails within the organization is investigated. This solution approach can potentially reduce the problems of email overload, interruptions, addiction, etc. and, at the same time, bring more routine and order to the email processing culture within organizations. To understand the rewards and penalties associated with the use of different email processing strategies, we look at various performance measures such as knowledge worker time-effectiveness, value effectiveness, task completion time, and email response time in this study.

As a first step toward finding a solution to the problems and for the purpose of improving productivity, it is important for us to develop a better understanding of the email work environment and its problems. In the next section, we provide a detailed overview of several of these problems.

The following section first discusses the dark side of email and elaborates on some contemporary issues that knowledge workers must deal with while using the current email technology. It presents an analysis of the state-of-the-art email technology and, identifies areas that require further research to ultimately enhance email technology's current capabilities and improve its use at work. We later explain the paradox associated with the use of emails. This overview is not meant to be a literature-driven section, as there is a negligible amount of literature on this topic. It is meant to provide a high-level perspective on the various issues related to the current use of email within organizations. These sections also highlight the areas where academic research on the topic is deficient and where it ought to focus in the near future.

1.1.1 The Problem of Emails

We may start to think that email is a mature technology. With time and the advancement of technology, the way we perform different tasks at work has changed but email technology has not been able to evolve at the same pace (Rohall 2002; Whittaker et al. 2005). Whittaker et al. (2005) reported that today's email client softwares are very similar at the core to clients 15 years ago, except for some additional functionalities such as a graphical interface, drop down menus, attachment facilities, extra storage space, etc. Rohall (2002) refers to them as the "souped-up" cousins of clients we had 30 years ago. We are using email for purposes such as task collaboration, task management, etc. for which it was never designed in the first place (Whittaker and Sidner 1996; Ducheneaut and Bellotti 2001; Whittaker 2005) We are using email so extensively that we have literally started to live in emails (Ducheneaut and Bellotti 2001). As a result, we are starting to see the fault lines. It is time that we recognize these email-related challenges, as they offer opportunities for future research in the area, and increase our research efforts in this direction. We briefly describe the problems (Figure 1.2) that have been reported in the literature.

1. **Email Overload –** Often, knowledge workers feel overwhelmed by the high volume of emails they receive and the unusually high amount of time that they spend in processing them. This large volume results from more communication and work being conducted through emails. The E-Policy Institute (2004) has estimated that the annual rate of email growth is approximately 66 percent. A recent survey of 840 organizations revealed that 47 percent of their workers spend one to two hours and 34 percent spend more than two hours in a given workday on email processing (American Management Association, 2004). All these statistics suggest that the volume of emails is a big contributor to email overload, which is a big problem for knowledge workers. However, it has been reported that not only the quantity but also the quality contributes to the overload (Bellotti et al. 2005). This phenomenon is not so well understood and needs to be further researched. Among several things, it is important to determine "when to process emails," "what emails to process," and "how to process emails."

Also, due to the "anytime" and "anywhere" availability of email systems in the Wi-Fi enabled society, knowledge workers often find themselves addicted to email. They get into the habit of processing emails as soon as they arrive. This irrational behavior leads to frequent interruptions, which is detrimental to the productivity of a knowledge worker. Efficient task scheduling has become another major problem due to the large amount of multitasking taking place in the workplace.

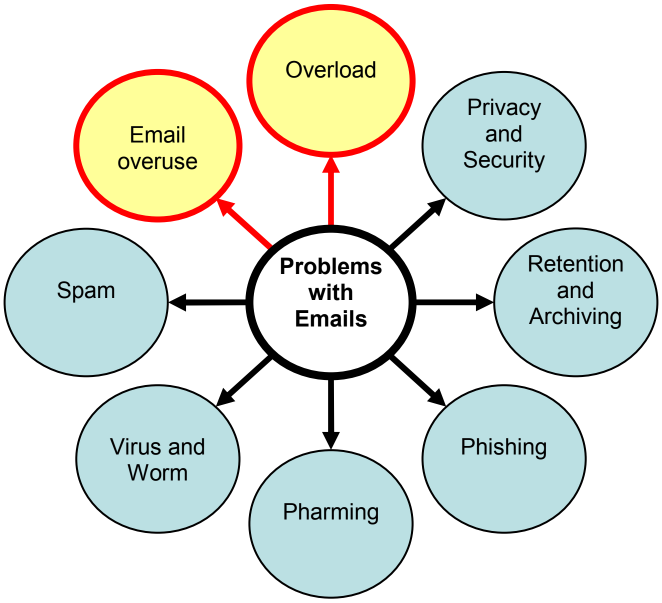


Figure 1.2 Problems with Emails

2. Email Overuse - In addition to these problems, there are several other drawbacks. For example, emails are cheap and open to everyone. There is no cost to the senders other than composing the emails; but receivers, on the other hand, pay a price in terms of time and other respects. This characteristic often results in senders pushing too many emails out and the receivers' inboxes overflowing. Imposition of some sort of monetary charge, similar to the concept of stamps, may be one of several solutions to the problem. However, the impact of such a charge on the overall communication must be studied in detail before implementation. Multicasting is another problem that results due to the openness of email. Anyone can send emails to anyone and several people at a time, contributing further toward recipients' overload. Although it is bliss for email marketers, it is a bane for many and can quickly lead to several problems.

Information replication is another type of problem that results from poor and ineffective use of this technology. Emails have certainly improved our connectedness and elevated our expectations. They have also resulted in people sending multiple reminders (just because they can!) that are often not needed, and queries that are not required. This practice leads to redundant information and multiple interruptions, ultimately causing receiver overload.

Another problem that has been noticed with several email users is addiction to the technology. Osterman Research recently conducted a study to determine how often workers check their emails for new messages when at work. Results confirmed that 67 percent of the workers check continually and 17 percent check a few times each hour. The survey that we conducted for this dissertation study confirmed a similar disturbing pattern in email use, confirming the presence of irrational behavior and attention-deficit disorders in workers dealing with emails. In fact, an NPR story on email compares email processing to smoking a cigarette. It also refers to the Blackberry as “crackberry” (Langfitt 2005).

Emails are also contributing toward the degradation of communication lingua. For example, we have developed numerous short notations and acronyms for sentences (e.g., c u latr, Hw r u, missing salutations, etc.). One of the prime reasons for this development is the phenomenon of email overload. Workers have to process too many emails in a timely manner, which forces them to develop shortcut ways for conveying the messages, which ultimately results in the reduced quality of an email message.

3. Email Retention and Archiving – The recent Sarbanes-Oxley 2002 compliance act has forced managers to focus on several email archiving and retention-related policies. For example, business-related emails can no longer be deleted for a certain period of time, and this requirement has created the need to address storage and caching-related problems. There is also a need to identify more efficient and effective approaches to addressing problems such as integrating email archiving and content management issues, which tends to be a major challenge. This requirement also has implications for research on distributed-database design such as better classification and filtering approaches for extremely large databases. Compliance acts such as HIPAA also pose several legal challenges. It becomes important to restrict the access to archived emails to various employees, as the information contained within emails could be very sensitive and confidential. It is important to do further research to develop better access-control mechanisms.

4. **Email Privacy and Security -** It is important to securely preserve the information held in emails, as we may need it to satisfy auditors. This becomes more important for industries such as the health care industry. A few studies have focused on information privacy in the context of emails (such as Sipior and Ward 1995; Weisband and Reinig 1995; Steven 2004) but clearly more research efforts need to focus on privacy-preserving methods such as link analysis, authentication, and cryptographic methods. Security of important email transactions is also important as organizations such as those operating under the B2B model are required to send encrypted emails to their receivers. However, the economic viability of this option needs to be well understood and studied.

5. **Email Phishing -** The word “phishing” first appeared in a hacker newsletter "2600 Magazine” (1996) and was coined by some crackers attempting to steal accounts

from unsuspecting AOL members. The term really refers to online imposters who use various social engineering and technical subterfuges to steal users' information(- 2006). Webster's Dictionary also provides a more detailed definition: "The practice of luring unsuspecting Internet users to a fake Web site by using authentic-looking email with the real organization's logo, in an attempt to steal passwords, financial or personal information, or introduce a virus attack; the creation of a Web site replica for fooling unsuspecting Internet users into submitting personal or financial information or passwords." (*Webster's New Millennium Dictionary of English, Preview Edition (v 0.9.6)*). Phishing reports received from antiphishing.org revealed that such activities have more than doubled since 2006. More research is needed in the area of fraud and deception detection to better understand the problem.

6. **Pharming, Virus, and Spam -** Pharming is a crimeware that misdirects users to fraudulent sites or proxy servers, typically through DNS hijacking, DNS poisoning or malware (Source: http://www.antiphishing.org/). It has been reported that spam, although increasing, is under control due to effective filters, but a relatively new phenomenon is starting to take place, which is called SPIM (spam in instant messaging). There is scarce research in this area, and substantial efforts need to be focused on it. Also, a virus spread through emails offers several opportunities for research. One promising area for future work could be to study the pace at which modern viruses can spread in networks.

1.1.2 The Paradox of Emails

The use of emails within any organization is very complex, and researchers in the IS discipline have not yet gained deep enough insight, due to lack of research in the area, to completely explain the phenomenon of organizational email processing. We believe that the use of theoretical and conceptual lenses provided by the rich literature on paradox in the organization science field will enable us to achieve two goals: first, additional insights can be gained that can further explain how knowledge workers process emails and second, why there is a need for email management. But before we begin to apply the concept of "paradox" to explain the email processing behavior of knowledge workers, we need to clearly understand what it means. The term "paradox" has been defined in several ways within the organization science literature. For example, Rosen (1994) defines paradox as "dynamic tensions of juxtaposed opposites." In another study, Ford and Backoff (1988) defined paradox as "some 'thing' that is constructed by individuals when oppositional tendencies are brought into recognizable proximity through reflection or interaction." According to Ford and Backoff, "paradoxes may serve as useful conceptual tools that extend our capabilities beyond the limits imposed by formal logic." These definitions have an approximately similar meaning at the core but there is no consensus as to which one is better than the other. We adopt a very neutral stance at this point and, without getting into the analysis of the lingering debate in this area and the process of explaining the nitty-gritty details of these definitions, we start by providing some evidence that can explain the presence of paradox in a knowledge worker's email processing behavior.

Organizations often provide incentives for collaborating within groups and for working toward achieving the overall group or organizational goals. Effectively and efficiently communicating through emails often becomes very important since a majority of collaborative work, deadline accomplishments, and idea exchanges occur through emails. This is a common phenomenon within the communities of practice (Huberman and Hogg 1995). So communicating with workers is encouraged and rewarded on the one hand, but at the same time, it also consumes a worker's time and energy resources. In many cases, the information contained in the email is of high utility for the requesting party but not of as much use for the responding party. Despite this lack of importance, the responding party is inclined, if not forced, to respond in a timely fashion to the email in order to maintain group cohesiveness and overall organizational goals. On the other hand, workers are rewarded for their individual performances, i.e. they are recognized for their individual accomplishment of tasks and goals, which may not always be in line with the group or organizational goals. They can use the time to perform tasks for which they will be directly recognized rather than in helping a colleague accomplish the task for which they may not be directly rewarded. Therefore, not providing a timely response to a particular query which is not of much use to the responding party (the original email's receiver), can be advantageous in the short term. Here, the responding party is marred by a paradox of whether to respond to a sender's email when the response may not be useful to the responder.

Another puzzle associated with the use of email is the issue of timing, i.e., when to respond to an email, if it has to be responded to at all. Speedy responses to emails often give the perception that the person is very efficient, diligent, and always accessible, but at the same time, a knowledge worker's primary task suffers due to the diversion of attention to emails. Assigning a relatively higher priority to email processing motivates workers to constantly check their inbox, which leads to frequent interruptions and possibly reduced performance on primary tasks. Trying to be on top of emails at all times can not only become very taxing to workers but also lead to problems such as attention deficit disorder, stress, lost productivity, and addiction to technology (Hallowell 2005) . On the other hand, providing a slow response to an email can mean retaining more flexibility in terms of when to process email and having more time for reducing the errors in emails. Thus, a slower response not only helps to improve the quality of the email response but also reduces the number of interruptions for the receiver since slower processing does not require constant inbox checking. At the same time, a slow response gives the impression that the worker (email responder) is less efficient. It also increases the response time of emails but decreases the completion time of primary tasks. Complex and opposing forces are in action here, which leads to paradox.

Another example of paradox with the use of emails occurs when knowledge workers prioritize their emails. Prioritizing emails leads to the problem of inclusion and exclusion. In the process of prioritization, what workers essentially do is to satisfy a few recipients on the high priority list and potentially upset a few who are on the lower priority list of email. Recipients are relatively more satisfied if they are included in the higher priority email list and relatively less satisfied if they are excluded from it. Such priority decisions are often based on one’s knowledge and experience, and are not always guaranteed to be hundred percent correct when the volume of information to be processed by a knowledge worker increases, as the knowledge worker may miss important cues or information necessary for correct classification. So whether to prioritize is an issue that often leads to paradox within email processing.

Yet another example of paradox occurs when a conflicting issue is being discussed through email. We know that email is a lean medium and that during an email exchange, due to the text nature of email, information cues can be lost which can completely change the meaning of the entire message. In spite of a knowledge worker's having a complete knowledge of this possibility, heated debates still occur through email, leading to further escalation of conflicts and flame wars. Such topics are better dealt with when discussed in a face-to-face environment. Emails can not only help to build relationships faster but can also contribute to the faster spread of conflict. A knowledge worker often struggles with which issues should be discussed through email, leading to paradox.

These instances and examples confirm the presence of paradox in email processing. We now use a framework presented in (Lewis 2000) in the context of emails to explain the presence of paradox. Figure 1.3 shows the three components of this framework that was developed by Lewis (2000): Tension, reinforcement cycles, and email management.

Figure 1.3 Framework for Email Paradox (Lewis 2000)

1. Tension – Tension results from the high volume of arriving emails, our inability to accomplish timely email processing, overload, burnout, stress, etc. Several other things lead to tension in email processing. For example, expecting to be rewarded sooner rather than later motivates a person to check email frequently in spite of the awareness that checking email only once may reduce the interruption effect or increase overall performance; however, we are unable to delay checking email due to the fear of

missing an opportunity or the eagerness to gain an incentive quickly or to search for more information. Such tension leads to the generation of reinforcing cycles.

2. Reinforcement cycle – A reinforcement cycle occurs when we take an action to cope with the tension and in that process we get more entangled in the reinforcing cycles or the swirl of the paradox (Lewis 2000). Such cycles occurs when we send multiple reminders leading to redundant information and receiver overload, check emails frequently, interrupt ourselves frequently, and stop alerts to avoid interruptions and thus miss important information.

3. Email management – Escaping the power of paradox is difficult because stopping the reinforcing cycles requires counter intuitive reactions (Cameron and Quinn 1988). It often requires a change in the way we do something and, therefore, the need for more research on email management. This can help a knowledge worker escape the paradox.

1.2 The Research Problem

Several types of approaches can be taken to address the problems mentioned above. For example, spam-related problems can be addressed through advanced filter development, security-related problems can be addressed using a more sophisticated algorithm, etc. However, it is not the technology that will provide the solution to problems such as information overload and interruptions; it is improvement in new-email management practices that is needed to solve these problems (Denning 2002). There is a lack of research on email use within organizations, and only a few studies have started to focus on the various strategies related to the use of email such as prioritization and classification of emails, and timing and frequency of email processing. These email-use strategies, if implemented, have the potential to not only improve the way we manage our emails but also make the knowledge workers' overall workday more productive. Although all the problems reported earlier are critical and cannot be overlooked in order to achieve productivity gains, this dissertation study focuses specifically on two of those problems: email overload and email use within organizations. These are the two most widespread problems with organizations that have not received sufficient attention from IS researchers. Nowadays, workers are overwhelmed by the enormous quantities of emails they receive (Denning 2002; Weber 2004; Hallowell 2005; Tassabehji and Vakola 2005; Antone 2006; Gupta et al. 2006; Paul 2006; Smith 2006; Swartz 2006). In order to cope with the increasing number of emails, they continually check for newly arrived emails or focus on processing pending emails, which results in either frequent interruptions (Jackson et al. 2003; Gupta et al. 2004) or addiction to emails (Adam 2002; Hallowell 2005). We previously described several paradoxical reasons for why workers tend to frequently respond to emails even though they know the consequences (see Section 1.1.2). Interruptions are not generally considered good in the organization science and IS literature, and are known to disrupt the routine flow of work (Zellmer-Bruhn 2003). Presently, knowledge workers use email processing strategies that are as random and diverse as their daily work requirements and email arrival patterns. Although the importance of routine has been emphasized in several research studies such as Zellmer-Bruhn (2003), the use of any routine or schedule in the processing of emails is currently lacking. Therefore, senders must wait for responses to their emails without having an idea of when to expect those responses (Jennings 2006) . As a result of this anxiety and because the completion of several other tasks may be dependent upon the information carried by an awaited response, workers check their inboxes more frequently, thereby increasing the number of interruptions. Another related issue is the concept of time management. The 'time' factor has not been studied well in the IS discipline, and this is evident from a recent editorial preface in IRMJ: "Time is a concept that needs to be more developed and integrated into Information Technology research" (Saunders 2002).

This study focuses on all of the above-mentioned dimensions, namely email scheduling strategies, work routines, and time, to address the problem of emails (overload and use). It proposes a new way of processing emails that may bring about a change in the timing and the manner in which emails are processed within organizations. It investigates the use of various time-based scheduling approaches for processing emails within the organization. This solution approach can potentially reduce the problems of email overload, interruptions, addiction, etc. and at the same time, bring more routine and order to the email processing culture within organizations. If workers within a group have a mutual awareness of one another's email processing schedules, the number of daily email interruptions that they must deal with may be reduced. Workers would also have an a priori idea of when they would receive their responses. This information would help them schedule their primary tasks and may lead toward a more productive and disciplined work environment. However, this proposition must be scientifically tested and verified, which is what we set out to achieve in this study.

A few studies have focused on timing and routine issues in email management and have tried to find optimal email strategies using different approaches. For example, Gupta et al. (2006), Yadati (2006) and Greve (2005) recently studied this problem, but from a different perspective. Gupta et al. (2006) modeled email strategies using a queuing theory-driven approach whereas Greve (2005) modeled a knowledge worker's attention span. Yadati (2006) used scheduling approach to determine when a knowledge worker should switch between primary tasks and emails. But, like all research studies, they also have certain limitations. One major limitation is that these studies modeled email interruptions as a time penalty, which suggests that a negative connotation is associated with processing emails within organizations whereas several research studies (such as McManus et al. 2002) reported that an organization derives value through email communication. In addition to interruption, rewards in varying proportion are to be associated with an email. The reward may be either extrinsic or intrinsic and is usually very difficult to measure in terms of one particular measurement unit. For example, an email may provide a piece of information that saves time and, at the same time, may positively influence different emotions such as mood. Time and emotions, as we know, are different constructs and cannot be measured using a single unit. Thus, it is important to understand not just the penalties (due to interruptions) associated with various email processing strategies but also the various accompanying rewards, to realize the true impact of different email processing strategies. To gauge the benefit (or reward) and drawbacks associated with the use of different email processing strategies, we look at knowledge worker time-effectiveness, value- effectiveness, task completion time, and email response time in this study.

Another major limitation of the earlier research is that it has focused on only individual knowledge workers whereas, in reality, an email always represents interaction between two people (a sender and a receiver) or more (such as in a distribution list). It is extremely important to analyze the problem at a higher level such as a group, network, or organization level to gain a deeper and broader understanding of the situation. This study, therefore, adopts a social network perspective to understand the use of various strategies related to the timing and frequency of email processing within a network of knowledge workers by identifying the optimal ones.

Knowledge workers usually belong to several groups or networks (Figure 1.4) and interact with members of these groups in various capacities depending upon the nature of relationships among them. Figure 1.4 shows that a specific worker may have memberships in several groups while involved in various intra- and inter-group email exchanges. As a result, complex email interactions take place within these networks, which makes the problem more challenging since the outcomes become difficult to predict under these circumstances. Focusing on the network level also makes the problem more interesting, as we are able to study several network-level parameters and variables. For example, an individual knowledge worker, who is a part of a bigger network, may be using email processing strategies that reduce his email volume but increases others' volume. Under this scenario, we will be able to study the performance of this particular worker but will not be able to understand the side effects of this policy on the whole network. Another example is that of time-effectiveness and value-effectiveness, which often involve tradeoffs (Ostroff and Schmitt 1993). In the hope of increasing one, the other may be compromised. As such hard-to-anticipate network performance measures can often have negative correlations, it is important to conduct an analysis at the network level and not just at the individual level.

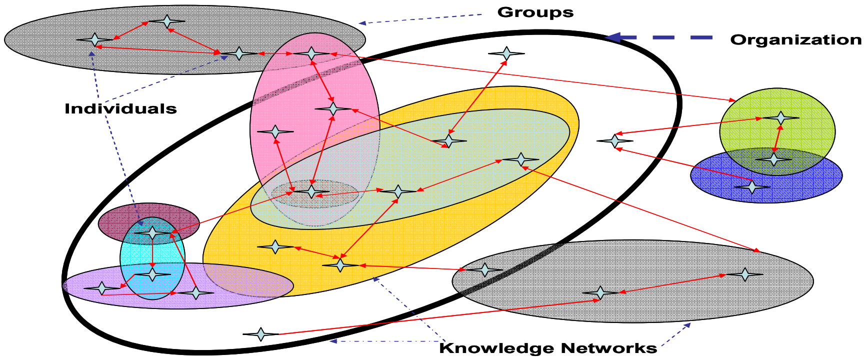


Figure 1.4 Email Networks

These networks can differ in several respects. We consider one such criterion for the purpose of our study: network type (in terms of degree of homogeneity). A network is called "homogeneous" if all the constituting knowledge workers are of the same type and skill and spend almost the same amount of time on emails. We see such types of networks when there is an absence of hierarchies, for example, networks of academics, friends, etc. A heterogeneous network occurs when knowledge workers are of different types and have different email needs. Examples of such networks are more commonly seen in organizations where hierarchies are present. A knowledge worker higher in the hierarchy may be spending more time on email than a worker at a lower hierarchy level. The first research question that we intend to explore is

Q1. How does the performance of various email processing strategies vary under different network types (i.e., homogeneous and heterogeneous)?

Also, earlier research that focused only on individual knowledge workers has shown contradictory results. For example, one stream of studies (Gupta et al. 2004; Greve 2005) showed that processing emails two to four times a day is the best strategy while another study (Venolia 2001) revealed that processing emails once a day is the better policy. Still another study (Jackson et al. 2003) found that processing emails eight times a day is a better policy. These results show a lack of consensus as to which policy works better at the individual level, let alone at the network level. Further, Gupta et al. (2006) noted that the performance of a knowledge worker varies with the email processing strategy in use and follows an inverted C-shape when performance and the number of email hour slots are plotted for an individual worker. However, the nature of this curve remains unexplored in a network setting and when different time-based schedules for email processing are used. Thus, the second research question is

Q2. How does overall network effectiveness change when similar email processing strategies are used throughout the network?

We later develop hypotheses related to each research question. A comparative analysis of these policies from an overload and interruptions point-of-view is conducted to understand their relative impact on the performance of a knowledge worker. Simulation and network-modeling concepts are used to study this problem.

1.3 Significance of the Study

This study has the potential to contribute at a theoretical as well as a practical level in several ways. It will lead to a better understanding of how complex email interactions that occur within a network of knowledge workers can be managed. A broad and deep understanding of email processing within modern work environments is very important and there is hardly any theory to guide it. This study lays the foundation for building and testing an analytical model of email processing that can provide insights into the positive or/and negative impact of various strategies on performance. A rigorous bottom-up approach is taken to look at the phenomenon of email overload and interruptions simultaneously in a network setting. This study could have important practical significance for organizations in which knowledge workers are spending large amounts of time processing email. Developing organization-wide policies to encourage users to check their emails on a scheduled basis rather than continually could save organizations thousands of hours each year. Such schedules can also be implemented by delivering emails to the users' email boxes periodically rather than continuously. It is also conceivable to develop policies that are appropriate for different classes of users.

This study shows that simulation can provide enormous advantages in studying a research problem for which data collection becomes a major challenge due to site unavailability, where field or experimental studies are difficult to conduct, and where human subjects cannot be utilized easily.

1.4 Layout of Dissertation

This study is divided into seven chapters. Chapter one has provided a general overview of the problems of email, a view from a paradox perspective, the research problem, and the objectives of this study. The second chapter discusses the extant literature in the areas of email, email networks, information overload in general and specifically email overload, interruptions, and email processing strategies. The third chapter elaborates on the analytical and the theoretical model developed in this study, and presents the research questions and hypotheses. This section also describes the development of the research design. The fourth chapter describes the research methods used to accomplish the goals of this study: survey, simulation along with parameter evaluation, and the validation process. Chapter five describes the model development and implementation. Chapter six focuses on the discussion of results and summarizes the important findings of the study along with their theoretical and practical implications. Finally, chapter seven talks about limitations and implications for future research.

2. REVIEW OF LITERATURE

This chapter is divided into four sections. The first section provides a comprehensive review of research done on emails and the use of emails in the context of networks. The second section discusses information overload in general and email overload specifically. The third section provides an overview of research on interruptions. This section provides information about the process of interruption and various costs associated with it. Finally, the fourth section briefly discusses the most recent work done on email processing strategies.

2.1 Emails

The IDC predicted that the total volume of emails sent in 2006 would exceed 3.5 billion gigabytes. This, in addition to the numbers cited earlier in Section 1.1.1, highlights the growing importance of emails as an indispensable mode of communication within organizations. There has been a rich tradition of research on emails in other disciplines such as computer science, but email has not been as much of a topic for major research in the IS or MS/OR community. Some of the earlier research that focused on emails mainly looked at the adoption and use of email technology within organizations. For example, one study conducted by Markus (1994) focused on the explanation of the managerial use of emails using several social constructs such as sponsorship, socialization, and control. Another study by Markus (1994) looked at the negative social consequences of emails.

A majority of studies related to emails have looked at the performance variables and are restricted to the individual level. But, since emails involve interactions between a minimum of two people, studies conducted at the individual level are often able to reveal only limited information about complex relationships among workers. A few studies have focused on email communication from a network perspective and used social network analysis to gain knowledge at the individual as well as network levels. Network approaches have proved to be tremendously useful in modeling the information flow within the real-world organizations by making certain simplistic assumptions. Such assumptions make the problem tractable and make it possible to observe a phenomenon that can be as complex as the flow of information within a work environment. Huberman and Adamic (2004) and Wu and Huberman et al. (2004) studied the information flow in groups using network analysis approaches. Several other studies on emails have taken a different viewpoint. Ahuja and Carley (1999) studied the impact of different network structures such as centralization, degree of hierarchy, levels of hierarchy, and different task characteristics such as analyzability and variety on the network performance. Johnson and Faraj (2005) built an entire simulation model of a knowledge network to understand the role of preferential attachment and mutuality in network formation. Some studies have aimed to reduce the overload in a networked environment such as virtual groups, social spaces like UseNet, and email distribution lists. Sharda et al., for example, studied the phenomenon of information overload for group knowledge networks and made several propositions to help to reduce the overload for the entire network (Sharda et al. 1999). Another field study tried to understand how the volume of communication is associated with message complexity in large social spaces (Jones et al. 2004). There is an interesting stream of research focused on understanding the value that an organization derives from communicating in a network (Nasrallah et al. 1998; Nasrallah and Levitt 2001; Nasrallah et al. 2003; Nasrallah 2006), in which the researchers developed a formula for calculating this effectiveness of communication that we discuss in more detail in Chapter 3. But none of these studies have explicitly looked at the time-effectiveness and value-effectiveness of these networks simultaneously, which are the focus of this study.

2.1.1 Lack of Research on Emails

In spite of these research efforts, there is still a paucity of information in this area. A 2004 editorial in *MIS Quarterly* recognizes this lack of research on email and calls for more IS research in order to better understand the problems associated with email (Weber 2004). Recognizing the lack of research in this area, a panel session chaired by Ramesh Sharda and dealing with the technical and managerial issues surmounting email use was hosted at ICIS 2005. The panelists provided perspectives from industry as well as academia and discussed various problems in email management, research methodologies to address these problems, various research opportunities, and an integrative framework for research on email management (Gupta et al. 2006). Also, two special issues on email-based research utilizing different types of methodologies were recently published, which shows that the awareness of this deficiency among IS researchers is being noticed and being addressed as well. One special issue on redesigning and reinventing emails was published in *HCI Journal* (Whittaker et al. 2005), and the other special issue, which focused on mining the Enron corpus for knowledge discovery, was published in

Computational and Mathematical Organization Theory (Carley and Skillicorn 2005). An article in this issue of the *HCI journal* (Ducheneaut and Watts 2005) provided a comprehensive review of state-of-the-art email research and proposed a three metaphor-based taxonomy for categorizing all the previous research on emails: email as a file cabinet, email as a production line, and email as a communication genre. A few research studies have looked at email as a personal management system (Whittaker et al. 2006) and proposed to unify the three functions of task management, personal archiving, and contact management. Other efforts also focused on developing and designing emails for the 21st century (Kerr and Wilcox 2004). Kerr and Wilcox pointed out that functionalities within the email technology have not evolved. They also pointed out three main deficiencies with email technology: lack of context, co-opting of emails, and keeping track of too many emails. A few researchers were quick to notice the problems associated with the use of emails (such as Denning 1982; Whittaker and Sidner 1996). Before describing this in more detail, we provide a brief review of research conducted on information overload, since email overload is one of its manifestations.

2.2 Information Overload

Information overload has been a topic of interdisciplinary research (Wilson 1996) where several disciplines such as accounting(for e.g. Chewning and Harrell 1990; Schick et al. 1990; Tuttle and Burton 1999), marketing (for e.g. Malhotra et al. 1982; Malhotra 1984), management science (for e.g. Chervany and Dickson 1974; Hart and Staveland 1988; Her and Hwang 1989; Xie and Salvendy 2000), and psychology (for e.g. Hancock and Meshkati 1988; Rubio et al. 2004) have made significant contributions. Researchers from the IS area have made several important contributions to this area as well (for e.g. Ackoff 1967; Simpson and Prusak 1995; Grise and Gallupe 1999; Edmunds and Morris 2000; Farhoomand and Drury 2002; Speier 2003; Hall and Walton 2004; Eppler and Mengis 2005). It has been established that information overload occurs when the information to be processed exceeds the information processing capacity of a knowledge worker (Galbraith 1974; Speier et al. 1999).

2.2.1 Email Overload

Email overload is one contributor to the overall information overload of a knowledge worker that may occur due to several things such as pending work, prolonged meeting hours, and time pressure due to deadlines. The problem of email overload was recognized by Denning (1982), who presciently asked, “Who will save the receivers from drowning in the rising tide of information so generated?” The phenomenon of email overload was later reported in a few other studies (Berghal 1997; Ducheneaut and Bellotti 2001; Jackson et al. 2001; Jackson et al. 2003). Email overload remains a burgeoning problem, and Denning revisited the problem of overload that he originally recognized in his 1982 article (2002). Email overload occurs when the cognitive workload created by emails exceeds the mental workload capacity, prohibiting the timely processing of emails. But it is a very complex phenomenon and has not been clearly understood (Bellotti et al. 2005; Gupta et al. 2006). Belloti and colleagues found that it is not only the quantity but also the quality of emails that contributes toward overload. One of the reasons we are experiencing email overload is that we are living in an “Attention Deficit Economy” where too many things are vying for our attention, which leads to problems such as Attention Deficit Trait (ADT), addiction, and stress (Davenport and Beck 2000; Davenport and Beck 2001; Hallowell 2005).

Our major objective in this study is to develop a better understanding of this problem. Many studies have reported the negative impact of emails on productivity (Swartz 2006) but have not quantified it. We will focus on the quantification of the impact of various strategies of email processing on time-effectiveness, value-effectiveness as well as overall effectiveness. However, the work environment of knowledge workers is nowadays marred by constant interruptions due to emails (Ducheneaut and Bellotti 2001). Apart from the loss of productivity mentioned earlier, interruptions often tend to increase information overload, an increase that results in a knowledge worker's feeling of "having too much to do and not enough time to do it"(Perlow 1999). In the context of emails, frequent interruptions aggravate email overload. We will now focus on interruptions before delving into the intricate details of the study.

2.3 Interruptions

Interruptions pose a huge threat to the U.S. economy and have an impact that is bigger than usually anticipated. It has been reported that interruptions consume about 28 percent of the knowledge worker's day, which leads to 28 billion lost hours per year in the United States (Spira and Feintuch 2005). Considering the average cost of $21 per hour as reported by the U.S. Department of Labor Bureau of Labor Statistics in its June 2005 report, this translates into an annual cost of $588 billion to U.S. companies (Spira 2005). Interruptions have been defined in several different ways. For example, according to distraction theory, an interruption is defined as "an externally generated, randomly occurring, discrete event that breaks the continuity of cognitive focus on a primary task" (Corragio 1990). Another definition says that interruptions are incidents or occurrences that impede or delay organizational members' progress on work tasks (Jett and George 2003). In Jett and George's (2003) study, four major types of interruptions were proposed: (1) intrusion, (2) break, (3) distraction, and (4) discrepancy. Interruptions caused by email fall under the category of intrusion. An intrusion is normally viewed from a time management perspective and is defined as an unexpected and unscheduled encounter that interrupts the flow and continuity of an individual's work, thus bringing that work to a temporary halt (Jett and George 2003).

Although some research on interruptions has been done in disciplines other than IS, such as human-computer interaction (HCI), management, and cognitive psychology, it is still an under-researched area from the IS perspective. However, the focus of these studies has been slightly different. For example, related research within the field of HCI has mainly focused on developing interfaces to reduce the cognitive overload and interruptions (for example, see McFarlane 2002). Our intention is not to provide a comprehensive review of all the work that has been done in the area of interruptions. For readers who are interested in the complete overview of research in this area, we suggest a useful online resource that is regularly updated and available at http://www.interruptions.net/. Our focus is only on the studies that are most relevant to the problem that we are studying here. A few authors from IS and other disciplines have started to study the interruptive nature of technologies such as email and instant messaging on primary task performance (for example, Speier et al. 1999; Cutrell et al.

2000; Czerwinski et al. 2000; Speier 2003) and suggested that the intensity of the interruption effect, among several things, depends upon the point at which the primary task gets interrupted and also on the degree of complexity of the primary tasks. They found that interruptions had less of an impact when a task was interrupted earlier in the processing stages. For example, a task interrupted during its planning phase will have a relatively smaller penalty attached to it than a task that is interrupted during later stages, typically called the execution and evaluation phases (Czerwinski et al. 2000).

2.3.1 The Cost and Process of Interruptions

Jackson and colleagues conducted a few studies to understand the cost of email interruptions in organizations (Jackson et al. 2001; Jackson et al. 2003). They found that the overall interruption effect of email is greater than that caused by phone calls, and reported several important parameters on the time lags created due to these interruptions resulting from emails. They videotaped and observed several knowledge workers in a British organization for a certain time period and found that it takes an average of one minute and forty-four seconds to react to a new email by activating the email application. The time needed to switch from a current work medium to the email medium is often referred to as switching time interruption lag (Trafton et al. 2003). A knowledge worker spends extra time to restart a task interrupted by email due to re-immersion. The recovery time due to interruptions caused by email is also referred to as resumption lag (Trafton et al. 2003) or penalty. This penalty has been reported to be about 64 seconds per interruption (Jackson et al. 2001; Jackson et al. 2003). According to these authors, , although this time may appear to be small, the cumulative interruption and resumption lags become large due to the large number of messages arriving everyday. These lags have the potential to increase the non-value-added time of a knowledge worker and decrease the knowledge worker's time-effectiveness. Figure 2.1 describes the process of interruptions graphically and in more detail. When an interruption arrives, a knowledge worker is preempted from a primary task. After spending a small switching time (IL), the worker starts to process the interrupt. Once the processing on the interrupt is over, workers spend a small recall time (RL) before they can resume their previously interrupted task.

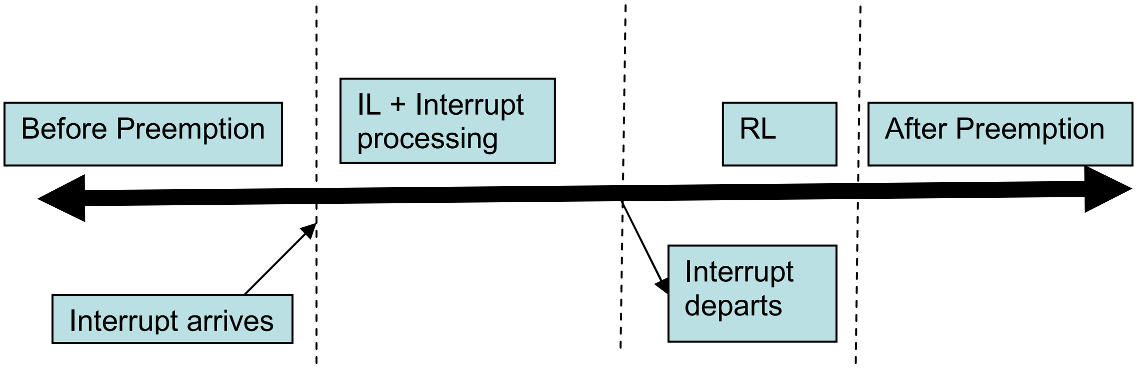


Figure 2.1 The Process of Interruption (Trafton et al. 2003)

Various theories of interruption have been proposed from cognitive psychology, such as consolidation (decay) theory, interference theory, and forgetting as discrimination (for a detailed treatment, please see Neath and Surprenant 2002). A few other theories have also tried to evaluate the effect of these interruptions either quantitatively or qualitatively. For example, the residual memory of a person after a break is considered to be a function of length of break and the performance time immediately after the break (Steedman 1970). Carlson and Rowe (1976) developed a learning-forgetting-learning curve and suggested that the forgetting curve is of exponential form. A study looked at the effects of interrupting a learning process (Bailey 1989) and developed a measure for evaluating the effect of forgetting that occurs due to learning in terms of time lost.

Carlson and Rowe suggested that the time lost due to forgetting is the difference between the total actual performance time and the time predicted by the learning curve. However, the learning-forgetting and relearning curve based on Wright's power function (Wright 1936) has proved to be more time-tested and robust. Ash and Smith-Daniels (1999) applied Wright's power function in a slightly different way to devise and test several types of rules for studying the project-scheduling problem in an interruptive work environment. But there was a slight difference. Rather than using the learning curve, which is primarily downward sloping, they studied proficiency, which is an upward sloping curve. We use these modified versions of formulae presented by Ash and Smith-Daniels (1999) to understand proficiency in this study and provide a more detailed description later in Section 3.

2.4 Prior Research on Email Management Strategies

A few studies have focused on email processing strategies. The study conducted by Jackson suggests that knowledge workers should check email every 45 minutes (approximately eight times a day). However, these studies do not consider the content complexity of the primary task or other factors that may impact choice of timing. Another study reported that processing emails once a day is the best strategy (Venolia 2001). The finding of these studies contradicted the results reported in two recently conducted studies (Greve 2005; Gupta et al. 2006), which suggest that processing emails two to four times a day is the best strategy. However, Greve (2005) and Gupta et al. (2006) have a few differences. Greve (2005) modeled the attention of a knowledge worker as an entity whereas Gupta et al. (2006) modeled email as an entity using queuing theory concepts. Their performance variables were also different.

This dissertation study intends to fill the gap in the literature in the following ways:

1. As recognized earlier in section 2.1, there is a need for more research on emails within the IS discipline. We hope this study answers the call. In addition, none of the studies conducted earlier have looked at the time-effectiveness and value-effectiveness of email communication simultaneously, which is the kernel of this dissertation.
2. Most of the studies conducted within the IS discipline have focused on the individual level rather than the network level due to the sheer difficulty of conducting a network-level study. Earlier studies conducted in this stream of research have all focused on individual performance measures. This study fills this gap by focusing on the network-level characteristics.
3. A majority of studies conducted on interruptions and email management are either empirical or qualitative in nature, with a few exceptions such as Yadati (2006), Greve (2005). Although, we do not claim than one approach is better than the other, we adopt a rigorous analytical and modeling based approach to study interruptions and overload in the email communication domain at the network level. This approach is achieved by developing and adopting (from previous work) a strong theoretical and mathematical base which then tends to serve as a foundation for the network simulation models.

The next section describes the theoretical and mathematical foundations of this dissertation study.

3. RESEARCH MODEL, QUESTIONS, AND HYPOTHESES

In this section, we describe the research methodology in detail. The first section elaborates on the concept of email processing strategies from this study's perspective. These strategies are first described from an individual knowledge worker's perspective and then from a network's point of view. We then present the mathematical and theoretical foundations of the research model along with research questions and hypotheses.

3.1 Email Processing Strategies (EPS)

3.1.1 Individual EPS

According to the Single-Resource theory (Kahneman, 1973), frequently diverting resources such as the attention of a knowledge worker to a secondary task (email) decreases the performance on the primary task. This theory suggests segregating the time during which interruptions and interrupted tasks are given higher priority for processing, thus reducing the interaction between interruptions and interrupted tasks. This strategy is expected to lead to a better performance on primary tasks. Interruption-related literature also confirms that whenever an interruption occurs, switching time as well as recall time is spent before the interrupted task is resumed. As the frequency of interruptions increases, the cumulative resumption and interruption lags increase as well, thereby delaying the completion of the primary tasks.

Thus, controlling the timeframe within which email is allowed to interrupt can reduce the number of interruptions, thereby reducing the cumulative switching and recall time. Such controls also allow for better attention allocation, which is a scarce resource in modern organizations (Davenport and Beck 2000).

To establish such a timeframe, we introduce the notion of “email hour” and “non-email hour.” The overall knowledge work hours in a particular workday can be split into two categories: one during which email is given the highest priority, termed “email-hour,” and the other during which primary tasks are given the highest priority, termed “non-email hour.” All the email processing strategies that we introduce have the same overall email hour length per work day (T_N) for a particular knowledge worker, but they differ in terms of the number of email hour slots (Ω) and length of each email hour slot (τ) within each policy. The overall email hour length (T_N) is a product of the number of email hour slots (Ω) and the length of each email hour slot (τ). Thus,

$$T_N = \Omega \times \tau \tag{1}$$

The value of T_N signifies the total time for which a knowledge worker prioritizes email processing per work day. Variations in the value of T_N also represent different types of knowledge workers depending upon the extent of their email processing requirements. A recent survey (American Management Association, 2004) reported on various time statistics regarding the amount of time a knowledge worker usually spends on email. We use these statistics to classify knowledge workers on the basis of their dependency on email communication. Knowledge workers can broadly be categorized into four types: very high users of email, high users, low users, and very low users. “Very high” users spend an average of four hours per workday processing email (T_N = 4 hrs.), “high” users spend three hours (T_N = 3 hrs.), “low” users spend two hours (T_N = 2 hrs.) and “very low” users spend one hour (T_N = 1 hr.). “Very high” and “high” users of email generally represent workers with a higher need for communicating at work, e.g. executives, CEOs, distribution and marketing managers, sales personnel, programmers, developers, engineers, educators, workers at virtual and geographically dispersed organizations, and those who belong to middle or higher organizational hierarchy. “Low” and “very low” users of email are knowledge workers with fewer communication requirements or those who are typically lower in the organizational hierarchy, e.g. office assistants and analysts. For the purpose of this study, we restrict our focus to two types of knowledge workers: high users of email (T_N = 3 hrs.) and low users of email (T_N = 1 hr.). Different combinations of “ Ω ” and “ τ ” values lead to different EPSs. For a particular type of knowledge worker (either low or high users of email), the different EPSs that we compare have same values of T_N but differ in terms of the values of Ω and τ . Although Table 3.1 summarizes six EPSs, this study focuses on only three of them. C1 lies on one end of the spectrum, and C lies on the opposite end of the spectrum. In other words, EPSs ranging from C1 to C represent the complete range of all processing strategies that can be used by any knowledge worker. Under the C1 policy, knowledge workers process their email in a single batch. Thus, it comprises one email hour slot (Ω = 1) of length 2 hours ($\tau = 2$) for low users of email and 3 hours for high users of email (($\tau = 3$). On the other hand, the C policy represents continual processing of emails, i.e. emails are processed as soon as they arrive. A knowledge worker working on a primary task keeps up with the flow of incoming messages by processing them immediately, as the C policy is adopted. Although the concept of email hours does not hold here, as emails are given priority throughout the day, this policy can have the numbers of email hours as large as the number of arriving emails.

Four other variations of EPSs are considered when the processing of email is scheduled at particular times of the day. In the C2 policy, the entire length of email hours is divided into two time slots ($\Omega = 2$). In the C4 policy, email hours are split into four time slots of equal duration ($\Omega = 4$). C6 has six email hour slots. Processing email every 45 minutes is approximately equivalent to eight email-hour slots. C8 is suggested as the best policy by Jackson et al. (2003). One thing that we will note with all these policies is that as the number of email-hour slots (Ω) increases, the time-length of each slot (τ) decreases, and that ultimately brings an EPS closer to the continual policy (C).

No. of Email-Hour Slots	**Length of Each Email Hour Slot**			**EPS Description**
	Low Users of Email (T_N = 1 hr)	**High Users of Email (T_N = 3 hrs)**		
Ω	τ_L	τ_H	**EPS**	**Processing Emails**
1	1	3	C1	Once a day
2	0.5	1.5	C2	Twice a day
4	0.25	0.75	C4	Thrice a day
6	0.1666	0.5	C6	Four times a day
8	0.1333	0.25	C8	Eight times a day
NA	NA	NA	C	Continuously

Table 3.1 Number of Slots and Time Spent in Processing Email under Low and High EPS

Of course, there are other potential policies for email processing, such as C3, C7, and C9, but they have been excluded from the evaluative procedure adopted in this study. It was necessary to keep the scope of the study to a manageable level and to allow us to focus on policies that differed most from one another. As the value of Ω increases, the value of τ decreases for each policy. Thus, the value of τ approaches zero as Ω approaches infinity.

Our postulation, which we later test, is that the performance of a knowledge worker is a function of one of the two variables, Ω and τ, that can take different values. The value of τ can be known from the chosen value of Ω. We believe that the number of interruptions can be altered by using the proper number of email hour slots.

3.1.2 Network (or Group) EPSs

In the previous section, we discussed EPSs at the individual level. We will now elaborate on the use of EPSs at the social network level, which is the focus of this dissertation. A network is made up of several knowledge workers. We propose a network classification criterion based on the composition of the network for the purpose of our study. Depending upon whether the email processing load of knowledge workers within a network is similar, we classify networks into two types: homogeneous email networks (HEN) and heterogeneous email networks (XHEN). When all the knowledge workers within a network have a similar email processing requirement, it is referred to as a "homogeneous email network" (HEN). Such types of networks can usually be spotted where hierarchies may not be present. These networks are also seen where each knowledge worker contributes almost equally to the information sharing and, thus, has a similar email processing requirement. Examples of such a network include networks involved in brainstorming, idea generation, etc. We may also see such networks in flat hierarchical organizations, clubs, networks of friends, etc. In the other type of network, knowledge workers have different email processing requirements or loads; this network is referred to as a "heterogeneous email network" (XHEN). Such types of networks are more common in organizations than HEN. In such networks certain knowledge workers within the network play a more important role than others or certain workers have a higher need for exchanging emails than other workers. Examples include networks of workers involved in a project. Within a project, we generally have hierarchies where people at different levels have different needs for email processing. Both types of networks can be visualized as comprising several groups or sub-networks. These groups may have overlapping members. We ignore group affiliations in this study and, hence, assume that knowledge workers belong to only one group. They cannot belong to two groups simultaneously. We also assume that these networks are not evolving but have actually attained a stable or an equilibrium state. Thus, all the types of HEN and XHEN that we study have existed for time periods long enough so as not to undergo any drastic fluctuations in their email volume, thereby implying that their group memberships are permanent for the duration of the study. We introduce five new variables 'x', 'y', 'u', 'w' and 'g' to help describe different network configurations that may arise as a result of combinations of different types of networks and different types of EPSs. We define these five variables as follows:

x = {Set of all the different groups or combinations of groups present in the network, order being important}

y = {Set of all the different policies or policy combinations being used in the network}

u = {No. of different policies or policy combinations being used in the network} = {1, 2, 6}

w = {No. of different group or group combinations present in the network, order being important} = {1, 2}

g = {Set of all the possible combinations of x and y i.e. (x, y) present in the network at a particular time}

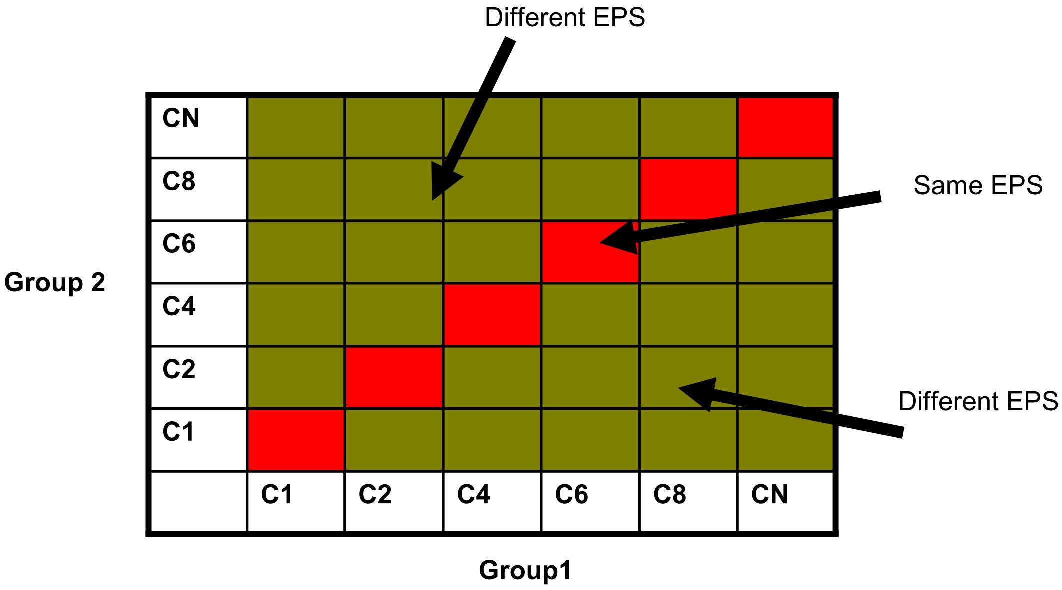


Figure 3.1 EPS between Two Groups of Knowledge Workers

The above variable definitions implies that all the possible combinations of x are represented by {(X_1), (X_2), (X_1, X_2), (X_2, X_1)}and all the possible combinations of y are represented by {(C1), (C2), (C4), (C6), (C8), (CN), (C1-C2), (C1-C4), (C1-C6), (C1-C8), (C1-CN), (C2-C4), (C2-C6), (C2-C8), (C2-CN), (C4-C6), (C4-C8), (C4-CN), (C6-C8), (C6-CN), (C8-CN), (C1,C2, C4, C6, C8, CN)}. We have considered the existence of two groups at the most within a network (Figure 3.1). When the whole network uses one single policy, i.e. u=1, then we have the six EPSs (red-colored cells) that a knowledge worker may choose from. This results in a total of 24 comparisons for both w=1 and w=2. For w= 2, the set has two possible groupings: (X_1, X_2) and (X_2, X_1). As we might notice, the order is very important and needs to be incorporated in the study. As an example, for

(C1-C2) EPS, in one combination X_1 uses C1 and X_2 uses C2, and the other combination that can result is that X_1 uses C2 and X_2 uses C1. There are 30 cells that represent the use of two EPSs in two groups. Similarly, there are 30 comparisons for HEN, which makes the total number of comparisons for groups using two policies equal to 60. Finally, for networks using six policies, we perform three comparisons: two for HEN and one for XHEN. A network that uses six policies is like a "random" policy use and resembles what we presently see in organizations. Under this scenario, each member of the group may use any of the six EPSs. Overall, 87 comparisons, defines the entire set of "g".

Many HEN and XHEN types of networks can be of different dimensions in reality and can also have varying proportions of members within their groups. For instance, in one particular network, only 10 percent of its members may be high users of email, whereas another network may have as high as 90 percent of its members who are high users of emails. In order to restrict our focus and to keep the number of comparisons within a manageable scope, we assume that all the HEN and XHEN are of the same dimension and have an equal number of high users and low users of email.

Also from (1), we can derive the total length of email priority hours for the entire network. For a network comprising "*N*" number of knowledge workers, the total length of email priority hours is given by:

$$\mathrm{T}_{EPS} = \sum_{i=1}^{N} \left(T\right)_i = \sum_{i=1}^{N} \left(\Omega_i \times \tau_i\right). \tag{2}$$

Where, "*i*" represents a particular knowledge worker.

3.2 Theoretical Model

The essence of all the types of EPSs being evaluated lies in how they differ in the timing and length of their email hour slots. Several editorial notes, special journal issues, and other recently published articles have identified “time” as an important factor to consider. For example, the Academy of Management Review had a special issue (Vol. 26, No. 4) about using time as a new research lens. In another editorial preface, there was a call for more time-based research in the Information Technology (IT) discipline (Saunders 2002): “time is a concept that needs to be more fully developed and integrated into IT research.”

Although some preliminary studies have explored the impact that individual EPSs have on knowledge worker performance, it remains to be explored at the network level when complex patterns of email exchanges are involved. Hence, one research question that we set out to investigate in this study is:

Q1. What is the impact of different email processing strategies (EPSs) on the overall network performance?

Time-effectiveness and value-effectiveness are two important measures for understanding the performance of an organization or investigating the process of information exchange through different technologies such as emails within a network. A majority of the time, these terms are mistakenly used simultaneously and interchangeably. As a result, organizations often tend to focus on only one of the measures of systems’ performance, i.e. either effectiveness with respect to time or effectiveness with respect to value attained, but not both. Such a restricted focus does not give us a complete insight into the organizational process of information exchange through emails. For example, it is important not only to be time-effective by meeting the project deadline and not spending time on non-value added activities, but also to be value-effective, which can be achieved by performing project-related activities the correct way. It is difficult, but very important, to achieve both as they involve a tradeoff, and one often comes at the expense of the other (Ostroff and Schmitt 1993). In this study, we want to understand the impact of EPSs not only on time-effectiveness but also on value-effectiveness. Organization science literature has reported distinct and clearly defined definitions of these two terms. Time-effectiveness, often referred to as efficiency, is generally defined as the ratio of output to input. On the other hand, value-effectiveness is derived from the benefit attained from the timely exchange of information contained within an email.

There is a lack of consensus among studies that tried to identify the best email processing strategies. For example, a few studies conducted by a research group proposed that email should be processed no more than every 45 minutes to increase employee productivity in the workplace(Jackson et al. 2003). Another study argued that processing email once per workday is a better policy than continual processing (Venolia 2001), whereas a set of studies by another group revealed that processing email two to four times a day is the optimal policy (Gupta et al. 2006). Besides the lack of general agreement among prior research findings, the policies mentioned above do not represent the entire range of policies that a knowledge worker might be able to use effectively, with respect to time and value, in managing emails and primary tasks.

Also, these studies have focused on individual knowledge workers and have ignored the network aspects. It is not clear how these “strategies” would behave in a group or network setting, and it would be very useful to cluster these strategies on the basis of their time and value effectiveness. Some may be time-effective and value-efficient, time-effective but not value-efficient, efficient but not effective or neither. Figure 3.2 shows such a classification in four quadrants. This classification will serve as a useful tool to help knowledge workers in deciding what policy best suits their working environment.

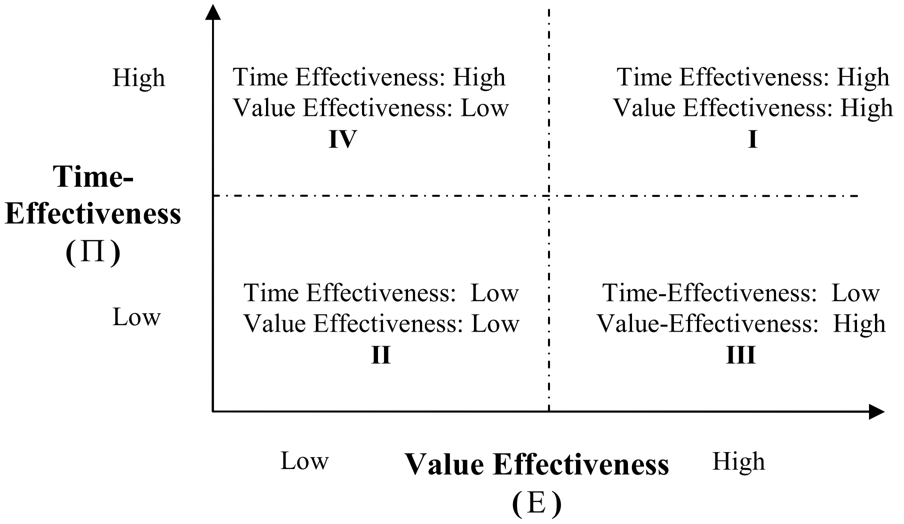


Figure 3.2 Framework for Time-Value Effectiveness of Email

This study analyzes the impact of various email processing strategies on network level performance variables across different levels of network structure. Four performance variables investigated in the study are overall time-effectiveness, overall value-effectiveness, average task completion time and average email response time. The two levels of EPSs that we look at are rhythmic policies vs. arrhythmic policies, and the two levels of network structures are homogeneous networks vs. heterogeneous networks. Figure 3.3 presents the main research model.

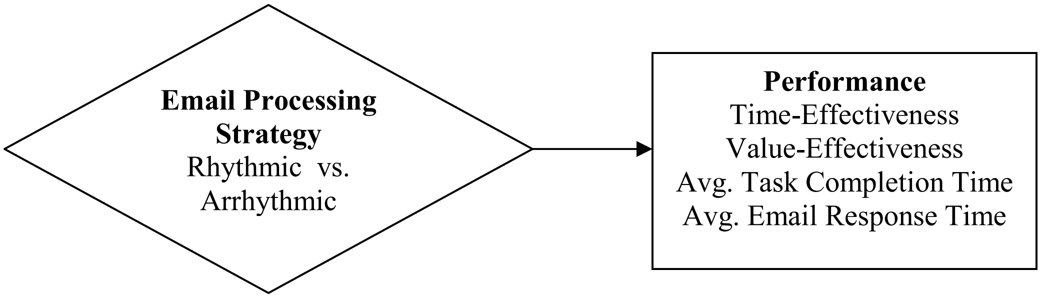


Figure 3.3 Theoretical Research Model

A few studies have emphasized the importance of routines and regularities in our daily work schedules. It is well established that the absence of routines and schedules has a negative impact on performance, but there is a lack of research and understanding of habitual group routines (Gersick and Hackman 1990). Gersick and Hackman suggested that for a group of workers to perform better, it is important not to have dysfunctional time-related group routines. However, the absence of routines leads to an increase in the number of interruptions at the organizational level (Zellmer-Bruhn 2003). One way in which this problem is manifested is through self-interruptions. Workers self interrupt when they randomly check emails. Such interruptions happen when workers use a strategy that has a higher number of email hour slots (Gupta et al. 2006). For example, when workers use a continual strategy (C) to process their email, they interrupt themselves more than if they use a C1 policy. The problems become more serious when a group or network is involved, as in this study. Knowledge workers who do not use a routine policy tend to cause more interruptions to other group members because they send emails randomly. Thus, workers within such networks are unable to develop an expectation of when the new information will arrive. Therefore some rhythm in the processing of email is a better strategy since workers are able to schedule their other primary tasks (Tyler and Tang. 2003). Another study suggested that time-related norms lead to a better understanding of the rhythms of interaction in the workplace (Lawrence et al. 2001). This study suggested that these norms hold their meaning when social exchange of information (such as email) is involved and that having a shared understanding of these norms among the network members leads to better scheduling of tasks and fewer interruptions.

Emails receive higher priority when arrhythmic policies such as a continual EPS is used, thereby leading to quicker email responses, but this quicker response comes as a compromise with primary task completion times. This compromise potentially occurs because primary tasks are allowed to be interrupted as soon as any email arrives, which serves to increase the task completion time. This increase in task completion time implies that a greater amount of time is wasted when arrhythmic policies are used, which will potentially lead to low effectiveness with respect to time. But it also means that arrhythmic policies will result in higher effectiveness with respect to the value that can be derived from a particular email exchange since emails are processed relatively quickly when such strategies are adopted. The utility that an organization may derive with longer response times may be low since the value of the information in the email usually diminishes with time. We believe that this will hold true for homogeneous as well as heterogeneous types of networks. Hence, we propose following hypotheses:

H1a: Rhythmic EPSs will lead to lower value-effectiveness than Arrhythmic EPSs.

H1b: Rhythmic EPSs will lead to higher time-effectiveness than Arrhythmic EPSs.

H1c: Rhythmic EPSs will lead to shorter average task completion times than Arrhythmic EPSs.

H1d: Rhythmic EPSs will lead to higher average email response times than Arrhythmic EPSs.

Further, a qualitative study (Perlow 1999) looked at the effects of the frequency and timing of interruptions on the individual and group productivity of knowledge workers. Although the focus of this study was not specifically on the strategies for processing emails, the findings were that neither perfect a-synchronization nor perfect synchronization between interrupting and interrupted tasks is good for effective time management. This finding leads us to believe that perhaps neither of the extremes is good for knowledge workers' performance and that the optimal policy is somewhere between the nearly synchronous and nearly asynchronous email communication. Figure 3.4 shows the continuum describing various degrees of synchronicity.

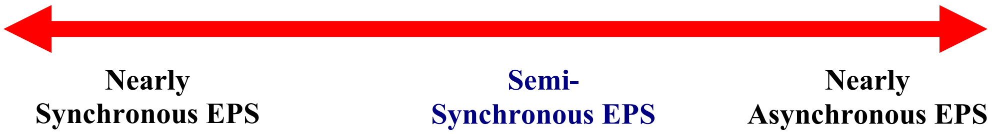


Figure 3.4 Degrees of Synchronicity

A knowledge worker's performance may be analogous to an economic order quantity (Wilson 1934). The analogy that we are drawing from the stochastic inventory total cost model may provide additional insights into the unexplored relationship between the proposed email policies and the performance of knowledge workers in an interruptive work environment. The inventory cost model represents the sum (total cost) of at least two functions, holding cost and setup cost. With the increase in the lot size, the holding cost increases and the setup cost decreases. Similarly, the decrease in a knowledge worker's performance as a result of interruptions can be seen as a function of resulting change in the time required to process email (inclusive of waiting time in inbox, switching time, recall time, etc.) and the increase in the time consumed in processing primary tasks. The response time for email (which is analogous to the setup cost in the inventory model) decreases, but the primary task completion time (which is analogous to the holding cost in the inventory model) increases with the increase in the number of email-hour slots (which is analogous to the lot size). This analogy suggests an inverted U-shaped relationship between the number of email-hour slots and the effect on a knowledge worker's performance.

Such U-shaped relationships can also be hypothesized using the economic theories related to cost functions. These theories also suggest that the nature of the cost function follows a U shape (Varian 1992). There are economies of scale for increasing the output rate only to a certain point. If the output rate is increased beyond this point, then diseconomies of scale arise. If we use the analogy between the output rate from economic cost theories and the number of email-hour slots, then we should arrive at a similar conclusion. Hence, the second research question is as follows:

Q2. How does overall network effectiveness change when similar email processing strategies are used throughout the network?

Hypotheses H3 basically proposes an inverted U-shaped relationship:

H2: The overall network effectiveness (E_o) at extreme levels (with C1 and C strategies) will be lower than that in the middle (with a C4 strategy).

Thus for a network comprising three knowledge workers, we hypothesize that the inverted U-relationship will hold true.

i.e. E_o ((C1, C1), C1)< E_o ((C4, C4), C4) and E_o((C, C), C)< E_o ((C4, C4), C4)

3.3 Analytical Development of Performance Variables

In the context of this study, time-effectiveness is expressed in terms of time and defined as the percent proportion of overall time that is spent on the actual processing of primary tasks and emails per day. Actual task completion time is usually smaller than the total time spent by the knowledge worker on any given day due to several non-value-added components of time, for example due to interruption effects. Since our focus is on evaluating performance measures at the network level, we first need to develop an expression for the calculation of the overall network time-effectiveness (Π).

Time-Effectiveness

Each network can comprise several groups that may or may not have an overlap. Assuming that there is no overlap in the group and that the time-effectiveness of each group is represented by η_g, the relationship between the time-effectiveness of the entire network and the time-effectiveness of constituting groups can be expressed as below:

$$\Pi = \frac{1}{G}\sum_{g=1}^{G}\eta(g) \tag{3}$$

Where,

$\Pi \rightarrow$ Overall network time-effectiveness

$G \rightarrow$ Total or maximum number of groups in the network

$\eta(g) \rightarrow$ Time-effectiveness of g^{th} group. A more detailed meaning of the term "group" (g) is presented in section 3.1.2

Further, group level time-effectiveness for any group can be evaluated using the expression below:

$$\eta(g)=\frac{100}{D}\sum_{d=1}^{D}\left[\left(\sum_{i=1}^{N(g)}T_v(i,d,g)\right)\Big/\left(\sum_{i=1}^{N(g)}T_O(i,d,g)\right)\right] \tag{4}$$

Where,

$N(g)\rightarrow$ Total number of knowledge workers belonging to g^{th} group in the network

$D\rightarrow$ Total number of days for which observation is made

Therefore, from (3) and (4), we get

$$\Pi=\frac{1}{G}\sum_{g=1}^{G}\eta_g=\frac{100}{G\times D}\sum_{g=1}^{G}\sum_{d=1}^{D}\left[\left(\sum_{i=1}^{N(g)}T_v(i,d,g)\right)\Big/\left(\sum_{i=1}^{N(g)}T_O(i,d,g)\right)\right] \tag{5}$$

Where,

$T_O(i,d,g)\rightarrow$ is the total units of time spent (value added + non-value added) by i^{th} knowledge worker belonging to g^{th} group on d^{th} day

$T_v(i,d,g)\rightarrow$ is the actual units of time spent (value-added component of time) on processing primary tasks and emails by i^{th} knowledge worker belonging to g^{th} group on d^{th} day

$T_\psi(m,i,d,g)\rightarrow$ is the overall non-value-added component of $T_O(m,i,d,g)$ and represents the total non-value-added time spent during the completion of m^{th} task by i^{th} knowledge worker belonging to g^{th} group on d^{th} day

$T_v^1(e,i,d,g)\rightarrow$ is a component of $T_v(e,i,d,g)$ and is the total value-added time spent on processing emails until the completion of e^{th} email by i^{th} knowledge worker belonging to g^{th} group on d^{th} day

$T_v^2(m,i,d,g) \rightarrow$ is a component of $T_v(i,d,g)$ and is the total value-added time spent on processing primary tasks until the completion of m^{th} task by i^{th} knowledge worker belonging to g^{th} group on d^{th} day

$T_f(m,i,g,d) \rightarrow$ is a component of $T_O(i,d,g)$ and comprises times spent on processing emails and time spent on interruption lag. It is defined as the time attributed to the forgetting of m^{th} task by i^{th} knowledge worker belonging to g^{th} group on d^{th} day

$T_\psi^1(m,i,g,d) \rightarrow$ is a component of $T_\psi(m,i,d,g)$ and is the total non-value-added time spent due to interruption lag until the completion of i^{th} task by j^{th} knowledge worker belonging to g^{th} group on d^{th} day

$T_\psi^2(m,i,d,g) \rightarrow$ is a component of $T_\psi(m,i,d,g)$ and is the total non-value-added time spent due to recall lag until the completion of m^{th} task by i^{th} knowledge worker belonging to g^{th} group on d^{th} day. It can also be represented as the time required to gain lost proficiency (Ash and Smith-Daniels 1999)

$T_\psi^2 \rightarrow$ is the total resumption time spent by the entire network in D number of days and is the summation of $T_\psi^2(m,i,d,g)$ for D number of days.

The formalization of all the different components of time, especially the non-value-added time, is a crucial step here since the entire expression of time-effectiveness is based on it. We know from above that the total time spent by all the knowledge workers belonging to a network on any given day is the summation of total value-added and non-value-added times spent by the knowledge worker. As described earlier, each of these components has two subcomponents. The value-added time that knowledge workers actually spend on performing the work, i.e., $T_v(i,d,g)$ comprises the time spent on processing emails, i.e. $T_v^1(e,i,d,g)$ and primary tasks, i.e. $T_v^2(m,i,d,g)$. On the other hand, the non-value-added time i.e. $T_\psi(m,i,d,g)$ comprises time accounted for by interruption lag denoted by $T_\psi^1(m,i,g,d)$ and resumption lag denoted by $T_\psi^2(m,i,d,g)$. Although several other overheads can potentially occur when emails are processed, we chose to ignore them in this study as they are more governed by the limitations of the underlying technology than by the inefficiencies surrounding the use of emails at work. Examples of such overhead activities are deleting junk emails and filing, and organizing emails in the inbox.

The work environment of a knowledge worker usually involves some time-based learning. If there is no learning in the job, the work becomes simple enough that it does not involve any forgetting. It is very realistic to consider that a knowledge worker is working in an environment where time-based learning is involved. In such an environment, when a knowledge worker is performing a primary task, an increase in proficiency level occurs when the worker spends an increasing amount of time on the task. However, during the time when an interrupt is being processed, forgetting also occurs and that leads to a loss of proficiency (Carlson and Rowe 1976; Ash and Smith-Daniels 1999).

Only two of our four basic subcomponents of time contribute towards the forgetting of a primary task, namely the time spent on processing emails $T_v^1(e,i,d,g)$ and the interruption lag time $T_\psi^1(m,i,g,d)$. These two subcomponents, when added, give the total time for which forgetting occurred, i.e. $T_f(m,i,g,d)$. The expression below describes it mathematically,

$$T_O(i,d,g) = T_v(e,i,d,g) + T_\psi(m,i,g,d) \qquad (6)$$

$$= \left(T_v^1(e,i,d,g) + T_v^2(m,i,d,g)\right) + \left(T_\psi^1(m,i,g,d) + T_\psi^2(m,i,d,g)\right)$$

$$= T_v^2(m,i,d,g) + \left(T_\psi^1(m,i,g,d) + T_v^1(e,i,d,g)\right) + T_\psi^2(m,i,d,g)$$

$$= T_v^2(m,i,d,g) + T_f(m,i,g,d) + T_\psi^2(m,i,d,g) \qquad (7)$$

Therefore, from (5) and (7), we get

$$\Pi = \frac{100}{G \times D} \sum_{g=1}^{G} \sum_{d=1}^{D} \sum_{i=1}^{N(g)} \left[\left(T_v(i,d,g)\right) \Big/ \left(\sum_{m=1}^{M} \left[T_v^2(m,i,d,g) + T_f(m,i,g,d) + T_\psi^2(m,i,d,g) \right] \right) \right]$$

(8)

M → Total number of tasks performed per day by i^{th} knowledge worker belonging to g^{th} group on d^{th} day. The value of M need not be the same for two knowledge workers.

It is very important to accurately evaluate the time spent on recalling $T_\psi^2(m,i,d,g)$ for calculating the overall network time-effectiveness. As has been found in earlier research studies, time spent on recalling $T_\psi^2(m,i,d,g)$ depends upon the proficiency levels during different timeframes and the learning rate in the environment (Carlson and Rowe 1976; Ash and Smith-Daniels 1999). Using Wright's power function (Wright 1936), Ash and Smith (1999) deduced various formulae for proficiency during the learning, forgetting, and relearning (recall) period. These formulae (equations 9 through 12 below) are reproduced from their work without providing proof and after some very minor modifications to incorporate the network aspects:

$$\mathrm{P}_L(m,i,d,g) = 100\left[1 - \left(T_v^2(m,i,d,g) + 1\right)^b\right] \tag{9}$$

Where

$\mathrm{P}_L(m,i,d,g) \rightarrow$ Proficiency gain at the end of a period before forgetting

$\mathrm{P}_F(m,i,d,g) \rightarrow$ Proficiency level at the end of preemption time (forgetting period)

$\mathrm{P}_R(m,i,d,g) \rightarrow$ Proficiency level at the end of relearning (recall period)

$r_o \rightarrow$ Learning rate of the environment and is a constant that is a characteristic of the intensity with which forgetting occurs in the work environment. A rate of 1 implies that no forgetting takes place at work when interruptions occur, whereas a rate of 0 implies that a knowledge worker forgets everything if interruptions occur. The degree of forgetting is usually higher with a smaller learning-rate value. Usually values of 0.7, 0.8, or 0.9 are encountered.

$b \rightarrow$ Wright's power function exponent $b = \log(r_o)/\log(2)$

$$\mathrm{P}_F(m,i,d,g) = \left(\mathrm{P}_L(m,i,d,g)\right)\left(T_f(m,i,g,d) + 1\right)^b \tag{10}$$

$$\mathrm{P}_R(m,i,d,g) = \mathrm{P}_F(m,i,d,g) + \left[\left[100\left(1 + [1 - r_o]\left[T_f(m,i,d,g) + 1\right]^b\right)\right]\right] \times \left[1 - \left(T_\psi^2(m,i,d,g) + 1\right)^b\right]\Big] \tag{11}$$

Solving the above equation for $T_\psi^2(m,i,d,g)$, we get

$$T_\psi^2(m,i,d,g) = \sqrt[b]{\left\{\left(\mathrm{P}_F(m,i,d,g) - \mathrm{P}_L(m,i,d,g)\right) \Big/ \left[100\left(1 + [1 - r_o]\left[T_f(m,i,d,g) + 1\right]^b\right)\right]\right\} + 1} - 1 \tag{12}$$

Also, the total time spent due to resumption lag is given by the following expression

$$T_{\psi}^{2} = \sum_{g=1}^{G}\sum_{d=1}^{D}\sum_{i=1}^{N(g)}\sum_{m=1}^{M} T_{\psi}^{2}(m,i,d,g)$$

Substituting (12) in the above equation, we get the total value of time lost due to interruptions as

$$= \sum_{g=1}^{G}\sum_{d=1}^{D}\sum_{j=1}^{N(g)}\sum_{i=1}^{M}\left[\sqrt[b]{\left\{\left(p_{Fijdg} - p_{Lijdg}\right)\Big/\left[100\left(1+[1-r]\left[T_{Fijdg}+1\right]^{b}\right)\right]\right\}+1} - 1\right] \tag{13}$$

We can now apply the above proficiency and recall time expressions developed by Ash and Smith-Daniels (1999) to evaluate the overall network time-effectiveness derived earlier in equation (8). Upon substituting equation (12) in equation (8), we get the complete expression for network-level time-effectiveness as shown in equation (14)

$$\Pi = \frac{100}{G \times D}\sum_{g=1}^{G}\sum_{d=1}^{D}\sum_{i=1}^{N(g)}\left\{\left[T_{v}(i,d,g)\right]\Big/\sum_{m=1}^{M}\left[T_{v}^{2}(m,i,d,g) + T_{f}(m,i,g,d)\right.\right. \tag{14}$$

$$\left.\left. + \left[\sqrt[b]{\left\{\left(\mathrm{P}_{F}(m,i,d,g) - \mathrm{P}_{L}(m,i,d,g)\right)\Big/\left[100\left(1+[1-r_{o}]\left[T_{f}(m,i,d,g)+1\right]^{b}\right)\right]\right\}+1} - 1\right]\right]\right\}$$

Value-Effectiveness

We base the formulation of value-effectiveness on the work done initially by Huberman and his colleague (Huberman and Hogg 1995) on communities of practice and later developed by Nasrallah and his colleagues on “Interaction Value Analysis” frameworks (Nasrallah et al. 1998; Nasrallah and Levitt 2001; Nasrallah et al. 2003; Nasrallah 2006) for the mathematical treatment of value-effectiveness in this study. Huberman and Hogg mathematically studied group performance as a function of the interaction that takes place in a community-of-practice setting. This work was later generalized and further developed in several studies by Nasrallah and his colleagues.

Formulations developed by Nasrallah and colleagues can be applied to understanding and evaluating a network's value-effectiveness value. In this model, an email interaction is considered of high value-effectiveness if an organization derives high benefit through the exchange of information contained within an email. In other words, we assume that knowledge workers act rationally by exchanging emails and that they want to maximize the overall organizational value-effectiveness. According to Nasrallah (2003, 2006), the value-effectiveness of a message sent by sender *"i"* to receiver *"j"* on d^{th} day is a function of four factors and can be mathematically described as follows:

$$\xi(i,j,d) = h(i,j,d) \times p(i,j,d) \times s_1(i,j,d) \times s_2(i,j,d) \tag{15}$$

Where,

$h(i,j,d) \rightarrow$ Represents the value generated from an email interaction sent by i^{th} knowledge worker to j^{th} knowledge worker on d^{th} day.

$p(i,j,d) \rightarrow$ Represents the proportion of i^{th} knowledge worker's time spent emailing j^{th} knowledge worker on d^{th} day

$s_1(i,j,d) \rightarrow$ Represents the probability with which email sent by i^{th} knowledge worker to j^{th} knowledge worker on d^{th} day adds value to the organization. Mathematically, Nasrallah (2006) described it as below,

Probability [Email sent by i^{th} knowledge worker to j^{th} knowledge worker on d^{th} day adds value to the organization]

$$= \frac{1}{1 + \mathrm{P}(i,j,d) \times \frac{r_\gamma}{w}} \tag{16}$$

Where, r_γ is the rate at which emails are sent by each worker in the network (interaction-seeking process) and w is the rate at which hints are generated by each member (value-creating process).

$s_2(i, j, d) \rightarrow$ is a function of the load of the knowledge worker and the urgency of the email requested.

Both of these terms are based on queuing theory concepts. “Load” represents a knowledge worker’s utilization and the “urgency” of an email represents the timeframe within which, if it is responded to, provides value to the organization and if it is not, does not provide value. .

Assuming that value-effectiveness is additive in nature, the overall value-effectiveness (E) of the network can than be defined as the mean of the value-effectiveness of its constituting groups spanning the timeframe for which observations are made. Thus, for observations spanning ‘ D ’ number of days, the overall value-effectiveness of the network can be written as

$$\mathrm{E} = \left(\frac{1}{D}\right) \times \sum_{d=1}^{D} \xi(i, j, d) \tag{17}$$

Therefore, for all *i-j* pairs of knowledge workers, we can derive the network-level value-effectiveness function using equations (15) and (17) and after minor subscript modifications,

$$\mathrm{E} = \left(\frac{1}{D}\right) \times \sum_{d=1}^{D} \xi(i, j, d) = \left(\frac{1}{D}\right) \times \sum_{d=1}^{D} \sum_{i} \sum_{j} \left[h(i, j, d) \times p(i, j, d) \times s_1(i, j, d) \times s_2(i, j, d)\right] \tag{18}$$

For a network of size N, $h(i, j, d)$ will be an (N x N) matrix. The values in this matrix show the cardinal values of rankings that each knowledge worker assigns to the other. This interaction value is a function of two variables: differentiation within the network and diversity within the network. Nasrallah (2003) defined differentiation as the ratio of the value of interacting with the favorite versus the value of interacting with the least favorite, and diversity is defined as the number of independent skill types possessed by parties in the network. One of the assumptions that Nasrallah and colleagues made is that differentiation within the network remains the same for all the knowledge workers, which means that the ratio for any two consecutively ranked members within the network is always consistent. While it would be interesting to study a network having a mix of differentiation values, this study includes only those networks that have the same differentiation value through out. Mathematically, the interaction value, as defined in Nasrallah (2003), for a sender- receiver pair can be calculated as follows:

$$h_1(i, j, d) = Differentiation^{((Diversity-(ranking\_of\_j\_by\_i))//(Diversity-1))} \tag{19}$$

Where, the diversity of a network is the highest rank assigned to any knowledge worker within the network.

We need three components to evaluate the interaction values for the entire network: the ranking matrix, the differentiation of the network, and the diversity within the network. The body of the ranking matrix holds the ranking of a knowledge worker *"j"* (represented in columns) by another knowledge worker *"i"* (represented in rows). A value of one signifies the highest preference given by knowledge worker *"i"* to knowledge worker *"j"* whereas a rank assignment of four suggests the lowest preference due to low skill. A network comprising four knowledge workers having different skill-levels derives differing values through email exchanges taking place between each sender-receiver pair. If all the knowledge workers within this network are assigned integer ranks, the diversity value for this network is four since the highest rank assigned to/by a particular knowledge worker is four, with the lowest being one. As shown in row one of Table 3.2, a rank of four assigned by knowledge worker (KW) 1 to himself means that KW1 assigns the least importance to self-addressed emails, whereas KW1 assigns the maximum importance to emails received from KW4. Similarly, we can interpret the preference rankings assigned by knowledge workers 2, 3 and 4. The ranking assigned by KW *"j"* to communication received by KW *"i"* need not be the same as that assigned by KW *"i"* to communication received by KW *"j"*. We see this happening in reality and hence have modeled it in Table 3.2

Table 3.2 Ranking Matrix (Diversity 4)

	KW1	KW2	KW3	KW4
KW1	4	2	3	1
KW2	2	4	3	1
KW3	3	1	4	2
KW4	3	2	1	4

This rank matrix can now be used to calculate the value that a network of knowledge workers derives from any particular pair-wise exchange of emails using equation (19). Assuming a homogeneous differentiation value of twenty and a diversity value of four, we derive the values that the network will derive when knowledge worker *"i"* processes an email sent by knowledge worker *"i"* using the ranking values reported in the body of Table 3.2. These values are reported in Table 3.3. Rows represent the receiver and columns represent the sender.

Table 3.3 Interaction Value (h) for differentiation= 20 and diversity=4

	KW1	KW2	KW3	KW4
KW1	1	7.368063	2.714418	20
KW2	7.368063	1	2.714418	20
KW3	2.714418	20	1	7.368063
KW4	2.714418	7.368063	20	1

However, there are limitations in the approach adopted by Nasrallah and colleagues (2003). For example, their studies do not model the life of an email based on time. Equations (15) and (18) used to evaluate the value-effectiveness in this study only considers "network-wide" heterogeneity in terms of the skill level of knowledge workers. The value derived from a single exchange of email between a pair of knowledge workers can depend, in addition to their skill levels, on two other things:

1) Value of information contained in the email. All the emails sent by a particular sender *"j"* to a particular recipient *"i"* are not going to be of the same value. Some emails sent by *"j"* to "*i"* will be of high value, some of moderate value, and some may be of low value depending upon the information contained in the email. We may even expect to occasionally receive emails of almost negligible value, for example forwarded jokes, from a highly skilled sender who would usually send emails of high value. It is, therefore, necessary to consider the heterogeneity in the value of information carried by the email. Nasrallah's formulation ignored this and considered only the heterogeneity among the skill levels of the workers.
2) Life of the email. The value of information contained in an email does not always remain constant. For example, an email informing the recipient about the start of an event (say, a meeting) at 3:00 pm will provide value to the recipient if it is processed before 3:00 pm but will not provide any value if it is processed after that time. The value that an organization derives from an email exchange occurring between sender

“k” and receiver *“l”* is a function of time. This value remains constant for certain time period before it starts to diminish at a certain rate. Finally, beyond a particular time period, the email’s value becomes zero. All emails follow a life cycle that approximates a sigmoid function but with differing shapes. Although we do not prove the above statement empirically, it is based upon the observation of hundreds of emails over the past one year. We will leave the empirical validation of this assumption for a later study, but the mathematical treatment of this concept is provided in section 4.4.

The life of any email can be said to comprise three phases. During the first phase, the value of the email remains constant. If a desired action is taken that leads to the temporary resolution of the email while it is still in the first phase, this email provides maximum utility, i.e. value to the network. After a time period has elapsed, its value starts to diminish at a particular rate. This rate keeps declining until it approaches zero. This phase is referred to as the “second” phase of an e-mail’s life. If an email is processed during this time, a positive value which is less than the maximum but greater than zero is derived. After this phase, an email reaches its third phase, where any action taken on the email does not provide any value to the network. This phase could potentially happen for several reasons. For example, information contained in the email may not be relevant to the receiver anymore, if the email is processed after the event occurs or if the information becomes obsolete for other reasons. Figure 3.5 describes how the value of an email diminishes with time.

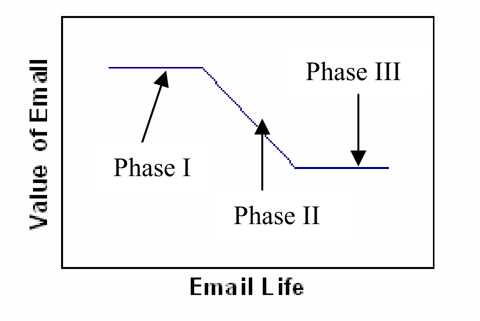


Figure 3.5 Email Value vs. Email Life

Both the above-mentioned variables need to be factored into the value-effectiveness of the email communication. If the value of information contained in an email between an *i-j* pair on d^{th} day is denoted by $h_2(i, j, d)$ and the time-based value, which is a function of the e-mail's life that the network derives after the email is processed on d^{th} day is denoted by $h_3(i, j, d)$, then the actual value-effectiveness can be mathematically expressed as:

$$\varepsilon_{actual} = h_1(i, j, d) \times h_2(i, j, d) \times h_3(i, j, d) \qquad (20)$$

Where, $h_1(i, j, d)$ is the interaction value calculated using Nasrallah's equation (19).

One assumption that is implicit here is that all three values in equation (20) are multiplicative, else the equation will not hold true. Ideally, the effectiveness derived by the organization is always going to be less than or equal to the actual effectiveness (ε_{actual}). If ε_{ideal} represents the value-effectiveness derived by the network if we were to assume that the value of life does not vary with time, then the overall value-effectiveness of the entire network could be calculated using the following expression:

$$\mathrm{E} = \left(\frac{100}{D \times (Total\_no.\_of\_unique\_i - j\_pairs)} \right) \sum_{\forall i,j,d} (\varepsilon_{actual} / \varepsilon_{ideal}) \qquad (21)$$

To summarize, we have combined independent bodies of literature here to develop two different formulations, namely time-effectiveness and value-effectiveness. We first developed our own expression for evaluating time-effectiveness and then integrated it with formulations borrowed from the theory of learning and forgetting. The expression for value-effectiveness was derived based on part of the work done by Nasrallah and colleagues. We borrowed the concept of interaction value from them and contributed two new concepts that are used in evaluating value-effectiveness: the value of information contained in any email and the life of an email. Our contribution is also unique in that we extended these formulations to include the network aspect of the organizations. All the related studies conducted earlier were done primarily at the individual level.

Email Response Time and Task Completion Time

In addition to time and value effectiveness, two other performance variables are being studied here: average task completion time and average email response time. Average task completion time is defined as the average amount of time for which a task remains with the knowledge worker before it is completed. This time includes also includes the entire primary task wait time. Mathematically, it can be expressed as

$$\mathrm{T}_m = \sum_d \sum_g \sum_i \sum_m \left(\mathrm{T}_{m_{out}}(m,i,g,d) - \mathrm{T}_{m_{in}}(m,i,g,d) \right) / \mathrm{M} \tag{22}$$

Where,

$\mathrm{T}_{m_{in}}(m,i,g,d) \rightarrow$ Represents the time at which m^{th} primary task arrives for processing at *"i^{th}"* knowledge worker belonging to "g^{th}" *group on "d^{th}"* day

$\mathrm{T}_{m_{out}}(m,i,g,d) \rightarrow$ Represents the time at which the processing on m^{th} primary task that arrived at *"i^{th}"* knowledge worker belonging to "g^{th}" *group on "d^{th}"* day was completed.

$\mathrm{M} \rightarrow$ Represents the total number of primary tasks processed with the network

Finally, average email response time can be defined as the average amount of time an email remains with the knowledge worker before a desired action on it such as pre-processing, responding, reading, and filing, is completed. It includes the entire inbox time of the email until processing on it concludes. Mathematically, it can be expressed as

$$\mathrm{T}_e = \sum_{d}\sum_{g}\sum_{i-j}\left(\left(\mathrm{T}_{e_{out}}(e,i,j,g,d) - \mathrm{T}_{e_{in}}(e,i,j,g,d)\right)/\mathrm{Z}\right) \tag{23}$$

Where,

$\mathrm{T}_{e_{in}}(i,j,g,d) \rightarrow$ Represents the time at which e^{th} email was exchanged between "*i-j*" knowledge worker pair belonging to "g^{th}" group *on* "d^{th}" day

$\mathrm{T}_{e_{out}}(i,j,g,d) \rightarrow$ Represents the time at which the processing on the same e^{th} email that was exchanged between "*i-j*" knowledge worker pair belonging to "g^{th}" group *on* "d^{th}" day is completed.

$\mathrm{Z} \rightarrow$ Represents the total number of emails processed within the network.

3.4 Research Design

This study seeks to understand the influence of two variables, namely network structure and email processing strategies, on four different performance variables. Table 3.4 lists all the factors and their levels that were used in the study. Network structure differentiates a network on the basis of whether it is homogeneous or heterogeneous. Homogeneous networks can, in turn, be classified into two categories: one where all workers have high email processing requirements and the other where all workers have low email processing requirements. Email processing strategies, as described earlier, also have three levels: rhythmic same, rhythmic different, and arrhythmic. The performance variables are time-effectiveness, value-effectiveness, average task completion time and average email response time. Table 3.5 lists all performance variables.

Table 3.4 Factors Involved in the Study

S. No	IV	Levels
1	Network Stricture	HEN-High
		HEN-Low
		XHEN
2	EPS	R-Same
		R-Diff
		AR

Table 3.5 Performance Variables

S. No	DV
1	Time-effectiveness
2	Value-effectiveness
3	Avg. Task completion time
4	Avg. Email response time

We developed a full factorial design to study all the scenarios that could potentially exist as a result of combinations of all the levels of independent variables. Full factorial designs, although time consuming to build and difficult to execute, give results that are more robust and have greater external validity since these results account for greater variability in the affecting variables. Given an artificial experiment setting, it becomes all the more important to consider all possible combination of factors to compensate for the lost generalizability. Following is the procedure that we used to identify all the strategy and network level combinations devoid of all duplicates and redundant combinations:

A three-step filtering process was employed to identify all the duplicates and redundant combinations. In this study, we analyzed a network of three knowledge workers who not only interact with each other through the exchange of emails but also interact with the world outside the network. Since this outside world can represent a varying and potentially large number of knowledge workers with whom the workers within the network exchange emails at any time of the day, it is logical to assume that the outside world processes emails more or less continually throughout the day due to the lack of control. So our focus is on identifying the complete set of combinations of strategies that could be used by the three knowledge workers within the network. In all, there are forty-five possible combinations. Table 3.6 shows all such policy strategy combinations along with duplicates. Each cell in the table represents a particular policy combination. For example, the cell in row one and column three has the value "C1C4C4," which implies that knowledge workers KW1, KW2 and KW3 use C1, C4, and C4 respectively.

The first step in the filtering process was to remove the duplicate policy-sets from the list in Table 3.6 through visual inspection. After the removal of duplicates, the total number of distinct policy combinations was reduced to twenty-seven. Table 3.7 lists these policy combinations with duplicates having a cross marked against them.

Table 3.6 Complete Set of Policy Combinations

All Policies Same	Two Policies Same					All Policies Different	
C1C1C1	C1C1C	C1C4C4	C4C4C	C1CC	C4CC	C1C4C	C1C1C4
C4C4C4	C1CC1	C1C4C4	C4CC4	C1CC	C4CC	C1CC4	C1C4C1
CCC	C1C1C	C4C1C4	C4C4C	CC1C	CC4C	C4C1C	C1C1C4
	C1CC1	C4C4C1	C4CC4	CCC1	CCC4	C4CC1	C1C4C1
	CC1C1	C4C1C4	CC4C4	CC1C	CC4C	CC1C4	C4C1C1
	CC1C1	C4C4C1	CC4C4	CCC1	CCC4	CC4C1	C4C1C1

The second and third step in the filtration process involved the simultaneous use of two rules, a policy-filter rule and an affiliation-filter rule. The criterion that is evaluated in the policy filter rule is to check whether the policy combinations meet the requirements specified within the definitions of rhythmic-same, rhythmic-different, and arrhythmic policies. The affiliation filter assesses whether the group or network definitions described in the earlier section hold true for a particular policy combination.

Table 3.7 Policy Combinations with Duplicates Identified

S. No	Policy	Duplicates
1	C1-C1-C1	
2	C4-C4-C4	
3	C-C-C	
4	C1-C4-C	
5	C1-C-C4	
6	C4-C1-C	
7	C4-C-C1	
8	C-C1-C4	
9	C-C4-C1	
10	C1-C1-C4	
11	C1-C4-C1	
12	C1-C1-C4	X
13	C1-C4-C1	X
14	C4-C1-C1	X
15	C4-C1-C1	
16	C1-C1-C	
17	C1-C-C1	
18	C1-C1-C	X
19	C1-C-C1	X
20	C-C1-C1	X
21	C-C1-C1	
22	C1-C4-C4	
23	C1-C4-C4	X
24	C4-C1-C4	X
25	C4-C4-C1	X
26	C4-C1-C4	
27	C4-C4-C1	
28	C4-C4-C	
29	C4-C-C4	
30	C4-C4-C	X
31	C4-C-C4	X
32	C-C4-C4	X
33	C-C4-C4	
34	C1-C-C	
35	C1-C-C	X
36	C-C1-C	X
37	C-C-C1	X
38	C-C1-C	
39	C-C-C1	
40	C4-C-C	
41	C4-C-C	X
42	C-C4-C	X
43	C-C-C4	X
44	C-C4-C	
45	C-C-C4	

The last column in Tables 3.8 and 3.9 represents the set of strategy combinations that meet both the filter tests. The last column reports those combinations that pass through all the filters. Table 3.8 represents the selected policy combinations for homogeneous networks of high as well as low type. For example, the policy-set in the row with serial no. 1, i.e. C1-C1-C1, is selected since it passes the policy filter (P) as well as the affiliation filter (PA). This strategy set meets the policy filter test because all the strategies in this combination are rhythmic, whereas it meets the affiliation test because all the strategies within and across groups are same. Similarly, Table 3.9 represents the selected strategy combinations for a heterogeneous type of network. For example, the policy set in the row with serial no. 10, i.e. C1-C1-C4, passes the policy filter since all the policies in this set are rhythmic in nature and it passes the affiliation filter since within-group strategies are the same but across-group strategies are different. Using similar logic, we identify the complete set of strategy policy combinations that meet all the selection criteria.

Table 3.8 Policy and Affiliation filter for HEN- High and Low

		Rhy same		Rhy Different		AR		Selected
S. No	Policy	Policy filter	Affiliation filter	Policy filter	Affiliation filter	Policy filter	Affiliation filter	
1	C1-C1-C1	P	PA					yes
2	C4-C4-C4	P	PA					yes
3	C-C-C					P	PA	yes
4	C1-C4-C					P	PA	yes
5	C1-C-C4					P		
6	C4-C1-C					P		
7	C4-C-C1					P		
8	C-C1-C4					P		
9	C-C4-C1					P		
10	C1-C1-C4			P	PA			yes
11	C1-C4-C1							
12	C4-C1-C1							
13	C1-C1-C			P	PA			yes
14	C1-C-C1							
15	C-C1-C1							
16	C1-C4-C4							
17	C4-C1-C4							
18	C4-C4-C1			P	PA			yes
19	C4-C4-C			P	PA			yes
20	C4-C-C4							
21	C-C4-C4							
22	C1-C-C							
23	C-C1-C							
24	C-C-C1			P	PA			yes
25	C4-C-C							
26	C-C4-C							
27	C-C-C4			P	PA			yes

Table 3.10 (a, b, c) summarizes all the selected policy combinations that passed the three filters, the duplication filter (conducted through visual inspection), the policy filter, and the affiliation filter, for each of the three types of networks. These combinations serve as different scenarios for comparative purposes in this study. Once,

we have identified all the scenarios that need to be studied, it becomes easier to classify them into different factors. We identified ten distinct scenarios for HEN-Low, ten for HEN-High and eighteen for XHEN.

Table 3.9 Policy and Affiliation Filters for XHEN

		Rhy same		Rhy Different		AR		
S. No	**Policy**	**Policy filter**	**Affiliation filter**	**Policy filter**	**Affiliation filter**	**Policy filter**	**Affiliation filter**	**Selected**
1	**C1-C1-C1**	P	**PA**					**yes**
2	**C4-C4-C4**	P	**PA**					**yes**
3	**C-C-C**					P	**PA**	**yes**
4	**C1-C4-C**					P	**PA**	**yes**
5	**C1-C-C4**					P	**PA**	**yes**
6	**C4-C1-C**					P		
7	**C4-C-C1**					P	**PA**	**yes**
8	**C-C1-C4**					P		
9	**C-C4-C1**					P		
10	**C1-C1-C4**			P	**PA**			**yes**
11	**C1-C4-C1**						**PA**	**yes**
12	**C4-C1-C1**							
13	**C1-C1-C**			P	**PA**			**yes**
14	**C1-C-C1**						**PA**	**yes**
15	**C-C1-C1**							
16	**C1-C4-C4**							
17	**C4-C1-C4**						**PA**	**yes**
18	**C4-C4-C1**			P	**PA**			**yes**
19	**C4-C4-C**			P	**PA**			**yes**
20	**C4-C-C4**						**PA**	**yes**
21	**C-C4-C4**							
22	**C1-C-C**							
23	**C-C1-C**						**PA**	**yes**
24	**C-C-C1**			P	**PA**			**yes**
25	**C4-C-C**						**PA**	**yes**
26	**C-C4-C**							
27	**C-C-C4**			P	**PA**			**yes**

Table 3.10 (a) Policies for HEN: Low Type of Network

Scenario	**Network Structure**	**Total E-hrs spent by**				**EPS**	**Email Policy used**			
		Group 1		**Group 2**	**OW**		**Group 1**		**Group 2**	**OW**
		KW1	**KW2**	**KW3**	**KW4**		**KW1**	**KW2**	**KW3**	**KW4**
1	HEN	1	1	1	1	Rhy-Same	C1	C1	C1	C
2	HEN	1	1	1	1	Rhy-Same	C4	C4	C4	C
3	HEN	1	1	1	1	Rhy-Differerent	C1	C1	C4	C
4	HEN	1	1	1	1	Rhy-Differerent	C1	C1	C	C
5	HEN	1	1	1	1	Rhy-Differerent	C4	C4	C1	C
6	HEN	1	1	1	1	Rhy-Differerent	C4	C4	C	C
7	HEN	1	1	1	1	Rhy-Differerent	C	C	C1	C
8	HEN	1	1	1	1	Rhy-Differerent	C	C	C4	C
9	HEN	1	1	1	1	Arrhythmic	C1	C4	C	C

Table 3.10 (b) Policies for HEN: High Type of Network

Scenario	Network Structure	Total E-hrs spent by				EPS	Email Policy used			
		Group 1		Group 2	OW		Group 1		Group 2	OW
		KW1	KW2	KW3	KW4		KW1	KW2	KW3	KW4
1	HEN	3	3	3	3	Rhy-Same	C1	C1	C1	C
2	HEN	3	3	3	3	Rhy-Same	C4	C4	C4	C
3	HEN	3	3	3	3	Rhy-Differerent	C1	C1	C4	C
4	HEN	3	3	3	3	Rhy-Differerent	C1	C1	C	C
5	HEN	3	3	3	3	Rhy-Differerent	C4	C4	C1	C
6	HEN	3	3	3	3	Rhy-Differerent	C4	C4	C	C
7	HEN	3	3	3	3	Rhy-Differerent	C	C	C1	C
8	HEN	3	3	3	3	Rhy-Differerent	C	C	C4	C
9	HEN	3	3	3	3	Arrhythmic	C1	C4	C	C
10	HEN	3	3	3	3	Arrhythmic	C	C	C	C

Table 3.10 (c) Policies for XHEN Type of Network

Scenario	Network Structure	Total E-hrs spent by				EPS	Email Policy used			
		Group 1		Group 2	OW		Group 1		Group 2	OW
		KW1	KW2	KW3	KW4		KW1	KW2	KW3	KW4
1	XHEN	3	3	1	1	Rhy-Same	C1	C1	C1	C
2	XHEN	3	3	1	1	Rhy-Same	C4	C4	C4	C
3	XHEN	3	3	1	1	Rhy-Differerent	C1	C1	C4	C
4	XHEN	3	3	1	1	Rhy-Differerent	C1	C1	C	C
5	XHEN	3	3	1	1	Rhy-Differerent	C4	C4	C1	C
6	XHEN	3	3	1	1	Rhy-Differerent	C4	C4	C	C
7	XHEN	3	3	1	1	Rhy-Differerent	C	C	C1	C
8	XHEN	3	3	1	1	Rhy-Differerent	C	C	C4	C
9	XHEN	3	3	1	1	Arrhythmic	C1	C4	C	C
10	XHEN	3	3	1	1	Arrhythmic	C1	C	C4	C
11	XHEN	3	3	1	1	Arrhythmic	C4	C	C1	C
12	XHEN	3	3	1	1	Arrhythmic	C1	C4	C1	C
13	XHEN	3	3	1	1	Arrhythmic	C1	C	C1	C
14	XHEN	3	3	1	1	Arrhythmic	C4	C1	C4	C
15	XHEN	3	3	1	1	Arrhythmic	C4	C	C4	C
16	XHEN	3	3	1	1	Arrhythmic	C	C1	C	C
17	XHEN	3	3	1	1	Arrhythmic	C4	C	C	C
18	XHEN	3	3	1	1	Arrhythmic	C	C	C	C

Figure 3.7 describes the full factorial design for this study. Network structure and email processing strategies (EPSs) are the two main factors in the study. Network structure has two levels, homogeneous and heterogeneous networks. Homogeneous networks can be further classified into two categories on the basis of email load, high or low. Email processing strategies also have two types of levels, rhythmic and arrhythmic. Rhythmic EPSs can again be of two types, the same or different. Thus, we have nine different pairs: homogeneous-high type of network that uses a same-rhythmic policy, a

homogeneous-high type of network that uses a different rhythmic policy, a homogeneous-high type of network that uses an arrhythmic policy, a homogeneous-low type of network that uses a same-rhythmic policy, a homogeneous-low type of network that uses a different-rhythmic policy, a homogeneous-low type of network that uses an arrhythmic policy, a heterogeneous network that uses a same-rhythmic policy, a heterogeneous network that uses a different-rhythmic policy, and a heterogeneous network that uses an arrhythmic policy. All the thirty-eight scenarios identified in Table 3.10(a through c) fall under one of these nine pairs.

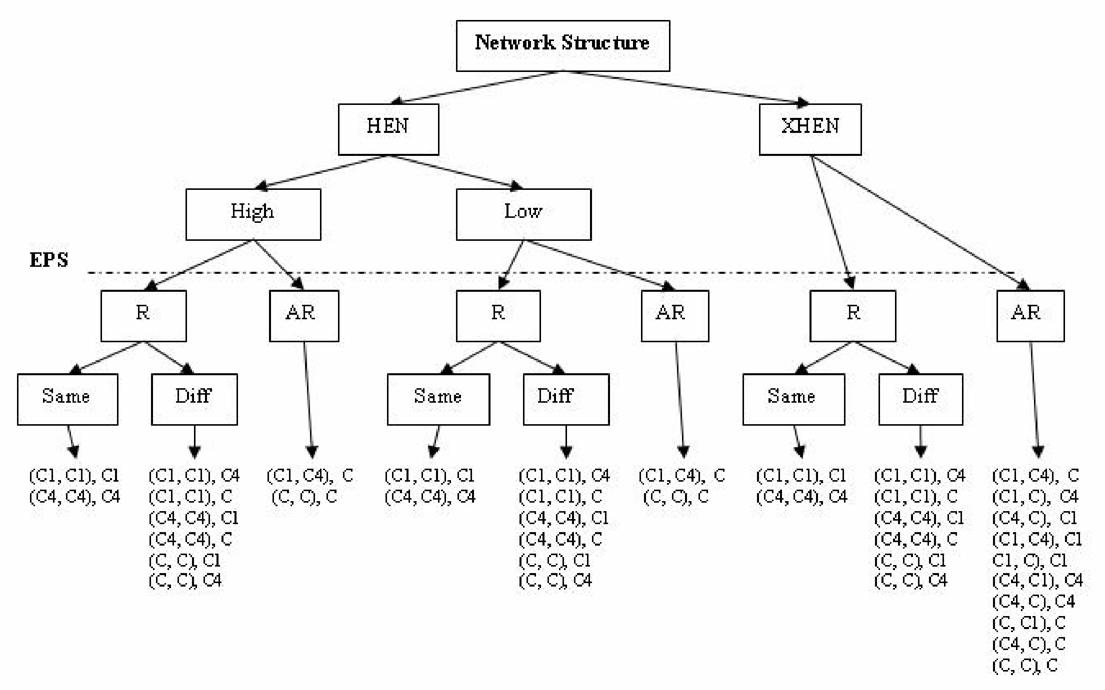


Figure 3.6 Full Factorial Design
(Total number of scenarios/factors=38)

Next section discusses the method used for developing the model.

4. MODEL DEVELOPMENT

4.1 Data Collection and Parameter Validation

We developed a survey to understand how emails and primary tasks are processed by different types of knowledge workers in different types of work environments. In particular, the survey helped us in gaining a better understanding of the following issues:

(1) Email overload- Addressed the question of how much time knowledge workers spend on emails.

(2) Interruptive nature of email- Helped us identify the frequency with which people process emails, and are therefore interrupted, at work.

(3) The different types of emails received

(4) The patterns of email arrival in different work environments

(5) Email priorities – Helped us identify different criteria that workers use in prioritizing their email processing

Only working professionals who used email at work were invited to complete the survey. For this purpose, students at a large Midwestern university provided help in recruiting the subjects. Each student established contact with two working professionals who were later requested to fill out the questionnaire. Since a majority of students were full time employees, they did not have any difficulty in recruiting two professionals. The students were later awarded with ten points as extra credit.

The survey was facilitated electronically and was divided into four sections: In the first section, general questions helped us understand the nature of the respondents' work environment. The second section focused on questions related to the amount of time spent on email and the frequency of email processing. The third section focused on collecting statistics related to the number of incoming and outgoing emails. Finally, the fourth section included questions to help us identify the criteria that respondents used for categorizing and prioritizing their emails. See Appendix for a copy of the complete survey.

The data from this survey were later analyzed to collect different statistics, to identify different patterns related to email use at work, and to mathematically derive the parameters to drive the simulation model. Since these parameters were based on data collected from the real world, the issue of questioning the validity of input parameters for the simulation model becomes less critical in this study. We now describe these initial parameters and explain the process of deriving the parameters needed to develop the simulation model. In this process, we will also describe some of the assumptions that were made as well as the procedure used to validate these parameters. We will also discuss different classification schemes, for example the basis for the urgency level, the phases of email processing, content, action needed, etc., that were used to categorize emails. This classification scheme not only helped us come up with the taxonomy for emails but also understand characteristics that are specific to particular email types.

First, based on data collected from the survey, we classified all the emails into three main categories depending upon their level of urgency. The level of urgency of an email is defined as an indicator of estimated time-length beyond which the value of the email for the recipient (or the organization) starts to diminish at a particular rate if appropriate action is not taken to resolve that email. Depending upon this time-length, we could broadly categorize all the emails that knowledge worker reported to receive into three categories: emails with high urgency, moderate urgency and low urgency. Emails with different urgency levels also enter various phases of their life at different points of time. For instance, emails with high urgency have a short phase I and the rate at which the value of the email diminishes during phase II is very high. Such emails enter phase III relatively quickly. On the other hand, emails with moderate urgency have a relatively longer phase I and their value drops at a relatively smaller rate during phase II. Such emails take relatively longer to enter phase III. Finally, emails with low urgency will have the longest phase I and the slowest rate of drop in their value while in phase II. These emails take longest to enter phase III of their life due to a minimal drop in their value with respect to time.

The survey revealed that approximately fifty-seven percent of all emails belonged to the high-urgency category and required a near-immediate action of some sort such as responding, reading, pre-processing, filing, etc. Approximately seventeen percent were found to belong to the moderate urgency category, while the remaining twenty-six percent belonged to the low urgency category. Table 4.1 summarizes this information.

Table 4.1 Email Categorization Based on Urgency

Urgency Level of Email (*U*)	Percentage of Emails	Cumulative Percent Values
High Urgency	57	57
Moderate Urgency	17	74
Low Urgency	26	100

Before we go further, we would like to elaborate on the mathematical formalisms that were developed to derive various parameters for the simulation model. These formulae, along with the notations that were used, are presented below:

$p_C \rightarrow$ Proportion of a knowledge worker's overall email time spent on newly created emails per working day.

$p_A \rightarrow$ Proportion of a knowledge worker's overall email time spent on arriving emails per working day

Analysis of the survey data made it evident that any email that a knowledge worker processed could be classified based on whether it was a new email or part of an ongoing email thread. In other words, it could be an email that was originated by a knowledge worker with no prior link to any other email or it could be a response to an "arriving" email i.e. sent by some other knowledge worker within or outside of the network. Using the survey data, we found that of all the emails that knowledge workers processed on any given day, approximately thirty percent were newly created ones i.e. were not generated as a response to an existing email but were actually new emails just being originated. The remaining seventy percent of emails were arriving emails i.e. emails that somebody else sent and that the recipient was required to act on, such as reading, responding, etc. Table 4.2 summarizes this information collected from the survey.

Table 4.2 Proportions of Time Spent on New and Arriving Emails

p_C	0.3
p_A	0.7

Since this study was limited to three knowledge workers and their interaction with the outside world, all *i* and *j* subscripts used below could possibly take four values: 1, 2, and 3 represent the three knowledge workers, and 4 represents the outside world. The outside world essentially comprised all the senders who were outside of the knowledge network and could essentially be treated as the 4th knowledge worker.

$\left(T_N\right)_i \rightarrow$ Total time spent by i^{th} knowledge worker on emails per working day. This study modeled two types of knowledge workers: high users of emails and low users of emails. The survey suggested that high users typically spent 180 min on email and low users approximately 60 minutes on emails per day. For the XHEN type of working environment, knowledge workers, $\left(T_N\right)_1 = \left(T_N\right)_2 = 180$ minutes per day whereas, $\left(T_N\right)_3 = 60$ minutes per day. For the HEN-High type of network, $\left(T_N\right)_1 = \left(T_N\right)_2 = \left(T_N\right)_3 = 180$ minutes per day and for the HEN-Low type of network, $\left(T_N\right)_1 = \left(T_N\right)_2 = \left(T_N\right)_3 = 60$ minutes per day.

$\left(T_C\right)_i \rightarrow$ Total time spent by i^{th} knowledge worker on newly created emails per working day

$\left(T_A\right)_i \rightarrow$ Total time spent by i^{th} knowledge worker on arriving emails per working day

The values of $\left(T_C\right)_i$ and $\left(T_A\right)_i$ depend upon the proportion of overall email time spent on newly created and arriving emails respectively. In other words, their values are constrained by the overall email time that knowledge workers usually spend within each type of network. Mathematically, $\left(T_C\right)_i$ and $\left(T_A\right)_i$ can be quantified as below:

$$\left(T_C\right)_i = p_C \times \left(T_N\right)_i \tag{24}$$

$$\left(T_A\right)_i = p_A \times \left(T_N\right)_i \tag{25}$$

Using the above equations and the data presented in Table 4.2, we can derive $\left(T_C\right)_i$ and $\left(T_A\right)_i$ values for all three types of networks: HEN-High, HEN-Low and XHEN. Table 4.3 presents this information. For the first three knowledge workers, the values of $\left(T_C\right)_i$ and $\left(T_A\right)_i$ for each type of network add up to $\left(T_N\right)_i$. However, for the fourth knowledge worker, these values are higher since the fourth knowledge worker is essentially a representative of a collection of several knowledge workers outside of the first three networks. Knowledge workers belonging to the HEN-High type of network spend more time on emails than those belonging to HEN-Low networks. In XHEN networks, we see a mix of high and low users of email. Knowledge workers 1 and 2 are high users of email whereas worker 3 is a low user of email.

Table 4.3 $\left(T_C\right)_i$ and $\left(T_A\right)_i$ values for HEN-High, HEN-Low and XHEN Network Types

Knowledge Worker	Time for HEN-High in minutes		Time for HEN-Low in minutes		Time for XHEN in minutes	
i	$\left(T_C\right)_i$	$\left(T_A\right)_i$	$\left(T_C\right)_i$	$\left(T_A\right)_i$	$\left(T_C\right)_i$	$\left(T_A\right)_i$
1	54.0	126.0	18.0	42.0	54.0	126.0
2	54.0	126.0	18.0	42.0	54.0	126.0
3	54.0	126.0	18.0	42.0	18.0	42.0
4	113.4	340.2	37.8	113.4	88.2	264.6

$p_{ij} \rightarrow$ Proportion of email time that i^{th} knowledge worker spends with j^{th} knowledge worker. Thus, p_{12} represents the proportion of time that knowledge worker 1 spends on email transactions taking place with 2. Technically, $p_{12} \neq p_{21}$ since the amount of time that knowledge worker 1 spends on emails from 2 need not be same as the amount of time that 2 spends on emails from 1. This also mimics what we might possibly see in real world situations. The values of p_{ij} will depend upon the overall time that the knowledge worker spends on emails i.e. $\left(T_N\right)_i$. Different types of networks will have different values of p_{ij}. Table 4.4 (a) represents the p_{ij} matrix for HEN-High and HEN-Low, and table 4.4 (b) represents the p_{ij} matrix for XHEN networks.

Table 4.4 (a) Time Proportion Matrix p_{ij} for HEN-High and HEN-Low

	Sender (*j*)			
Receiver (*i*)	**1**	**2**	**3**	**4**
1	0.000	0.350	0.350	0.300
2	0.350	0.000	0.350	0.300
3	0.350	0.350	0.000	0.300
4	0.333	0.333	0.333	0.000

Table 4.4 (b) Time Proportion Matrix p_{ij} for XHEN

	Sender (*j*)			
Receiver (*i*)	**1**	**2**	**3**	**4**
1	0.00000	0.52500	0.17500	0.30000
2	0.52500	0.00000	0.17500	0.30000
3	0.35000	0.35000	0.00000	0.30000
4	0.42860	0.42856	0.14286	0.00000

$q_{ij} \rightarrow$ Total time i^{th} knowledge worker spends on newly created emails sent to j^{th} per day. For a group of four knowledge workers, q_{ij} will be a 4x4 matrix. Mathematically, the relationship between q_{ij} and p_{ij} can be described as follows:

$$q_{ij} = p_{ij} \times \left(T_C\right)_i \tag{26}$$

Equation (26) can be used to derive a q_{ij} matrix for all three types of networks being studied here. Table 4.5 (a through c) shows the derived values of q_{ij} for different network types. We will later use a q_{ij} matrix to derive the rate matrix emails belonging to different categories.

Table 4.5 (a) q_{ij} Matrix for Time Spent on Newly Created Emails in HEN-High Networks

	Sender (*j*)			
Receiver (*i*)	**1**	**2**	**3**	**4**
1	0.0	18.9	18.9	16.2
2	18.9	0.0	18.9	16.2
3	18.9	18.9	0.0	16.2
4	37.8	37.8	37.8	0.0

Table 4.5 (b) q_{ij} Matrix for Time Spent on Newly Created Emails in HEN-Low Networks

	Sender (*j*)			
Receiver (*i*)	**1**	**2**	**3**	**4**
1	0.0	6.3	6.3	5.4
2	6.3	0.0	6.3	5.4
3	6.3	6.3	0.0	5.4
4	12.6	12.6	12.6	0.0

Table 4.5 (c) q_{ij} Matrix for Time Spent on Newly Created Emails in XHEN Networks

	Sender (*j*)			
Receiver (*i*)	**1**	**2**	**3**	**4**
1	0.0	28.35	9.45	16.2
2	28.35	0.0	9.45	16.2
3	6.3	6.3	0.0	5.4
4	37.8	37.8	12.6	0.0

$q'_{ij} \rightarrow$ Total time i^{th} knowledge worker spends on arriving emails sent to j^{th} per day. For a group of four knowledge workers, this will be a 4x4 matrix. It is a function of p_{ij} and $\left(T_C\right)_i$ and can be mathematically described as their product.

$$q'_{ij} = p_{ij} \times \left(T_C\right)_i \tag{27}$$

In the same way we calculated the values of q_{ij}, we derive the values of q'_{ij} using equation (27) and Table 4.5(b). Table 4.6 (a through c) presents those derived values. The q'_{ij} matrix can not be used directly to derive the overall arrival-rate matrix for emails, since the time values in this matrix represent the combined time spent on emails in their 2nd and 3rd processing stages. However, we will use the q'_{ij} matrix for validation purposes later on.

Table 4.6 (a) q'_{ij} Matrix for Time Spent on Arriving Emails in HEN-High Networks

	Sender (*j*)			
Receiver (*i*)	**1**	**2**	**3**	**4**
1	0.0	44.1	44.1	37.8
2	44.1	0.0	44.1	37.8
3	44.1	44.1	0.0	37.8
4	113.4	113.4	113.4	0.0

Table 4.6 (b) q'_{ij} Matrix for Time Spent on Arriving Emails in HEN-Low Networks

	Sender (*j*)			
Receiver (*i*)	**1**	**2**	**3**	**4**
1	0.0	14.7	14.7	12.6
2	14.7	0.0	14.7	12.6
3	14.7	14.7	0.0	12.6
4	37.8	37.8	37.8	0.0

Table 4.6 (c) q'_{ij} Matrix for Time Spent on Arriving Emails in XHEN Networks

	Sender (*j*)			
Receiver (*i*)	**1**	**2**	**3**	**4**
1	0.00	66.15	22.05	37.8
2	66.15	0.00	22.05	37.8
3	14.7	14.7	0.00	12.6
4	113.4	113.4	37.8	0.00

$s_{kl} \rightarrow$ Time taken by an email of k^{th} type while in its l^{th} processing phase

All the emails can be broadly classified into two main types, ones that elicit a response from the recipient and those that don't require a response from the recipient. Emails that require a response can further be classified into two types based upon the time it takes to process them: complex emails and simple emails. Complex emails require a relatively longer time to process, whereas simple emails require a short time to process. Figure 4.1 shows the entire classification scheme used to distinguish one type of email from the other. In this study we modeled the third level of the hierarchy and the categories of emails are represented by different values of the *'k'* subscript. C stands for complex emails, S stands for simple emails, and F stands for FYI emails. Thus *'k'* can potentially have any of the three values, C (*k*=1), S (*k*=2), and F (*k*=3).

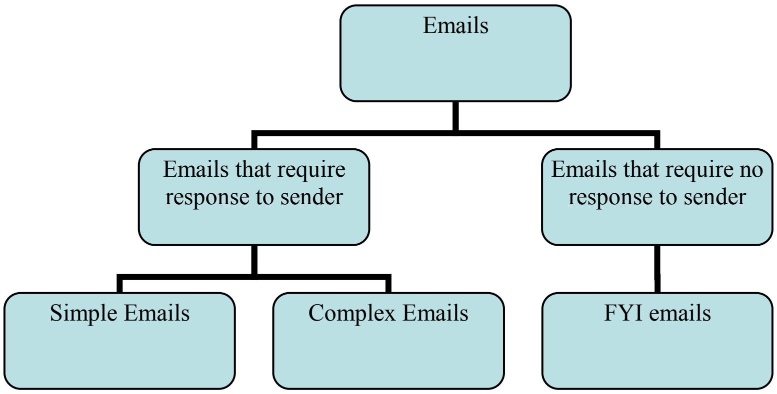


Fig 4.1 Hierarchy of Email Types

Emails that require a response by the receiver go through three phases of processing before they are resolved and are represented by the subscript *'l'*. An email during its first processing stage (*l*=1) has been processed by the sender and sent to the receiver. As soon as this email reaches the receiver, it enters its second processing phase (*l*=2). After a wait of a certain time period, the receiver will begin processing this email.

This processing may involve reading the email, creating the response, and sending it back to the original receiver. As soon as this email response reaches the original sender, the email enters its third stage of processing (*l*=3). During this last stage of processing, the original sender will read the email, extract the necessary information and file it away in the inbox. The processing time for an email in the different phases is different. The survey data revealed that processing an email usually takes longer in the second phase than in the first or third processing phases. The emails that do not elicit any response from the receiver have a short message thread life as they undergo only two stages of processing. In phase one, such emails are created by the sender and sent to the receiver. In the second phase, the receiver extracts the information contained in the email and takes the necessary action to resolve it. This may include deleting the email or filing it away but it does not require the receiver to generate a response to be sent to the original sender. Such emails are usually FYI emails, notification email, CC emails, listserv emails, etc. It usually takes a relatively small time to process such emails. Figure 4.2 (a, b) explains the sequential processing phases of different types of emails and also how they differ from each other.

Table 4.7 describes the processing time taken by emails of different types during their different processing stages. A majority of the statistics reported in this table are based on approximations from the data collected through the survey. As we might notice here, the second stage of processing usually takes longer for emails that require a response, i.e. email types with *k*=C and *k*=S.

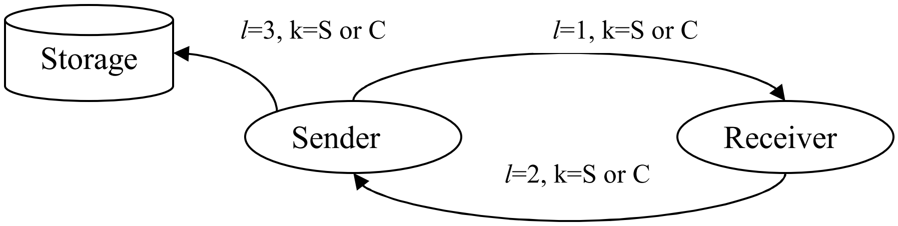


Fig 4.2 (a) Email Processing Phases that Require Response

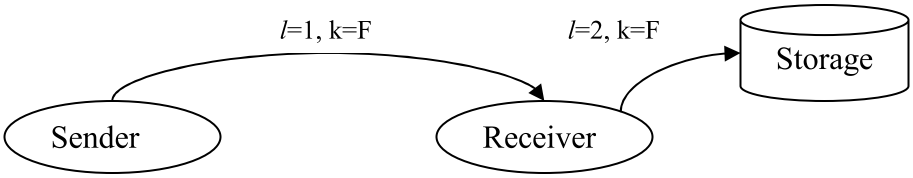


Fig 4.2 (b) Email Processing Phases that Do Not Require Response

Table 4.7 Processing Time Matrix (s_{kl})

	Time taken in different processing Stages (l) in min		
Email Type (k)	l=1	l=2	l=3
k=C	4	6	4
k=S	2	3	2
k=F	2	2	-

$p_k \rightarrow$ Represents the proportion of time that any knowledge worker typically spends on k^{th} type of email on any working day

We found that knowledge workers typically spent forty percent of their overall email time on processing complex emails, another forty percent on processing simple emails, and the remaining twenty percent of overall email processing time on dealing with FYI emails. This was found to be true irrespective of the type of network to which they belonged. Table 4.8 summarizes the collected statistics.

Table 4.8 Proportion of Email Time (p_k) Spent on Different Types of Emails

	Email Type (*k*)		
Proportion of Email Time Spent (p_k)	C	S	F
	0.4	0.4	0.2

$s_{ikl} \rightarrow$Represents the time taken by i^{th} knowledge worker to process k^{th} type of email during l^{th} processing stage

Table 4.9 represents the three-dimensional s_{ikl} matrix, which has been represented in the form of a two dimensional matrix for reading purposes. Each subscript represents an axis of the matrix. In this study, we have assumed that all the knowledge workers consume stochastically the same amount of time to process a particular type of email. This explains the reason for the time values' being repeated in rows for different values of *i*.

Table 4.9 s_{ikl} Matrix

Workers	Time taken in different processing Stages (*l*) in min by different types of emails							
	Complex Emails			Simple Emails			FYI Emails	
i	*k*=C, *l*=1	*k*=C, *l*=2	*k*=C, *l*=3	*k*=S, *l*=1	*k*=S, *l*=2	*k*=S, *l*=3	*k*=F, *l*=1	*k*=F, *l*=2
1	4	6	4	2	3	2	2	2
2	4	6	4	2	3	2	2	2
3	4	6	4	2	3	2	2	2
4	4	6	4	2	3	2	2	2

The q_{ij} matrices developed in Table 4.5 (a through c) report the total time that any particular receiver spends on newly created emails. From these matrices, we need to first calculate the total time that every knowledge worker spends on each type of email during its various processing stages. These split-time values can then be used to calculate the arrival-rate matrix needed for simulation models to be developed later on. We now describe the process of deriving these arrival-rate matrices for all three types of networks.

$\left(v_{ijk}\right)_l$ → Represents the total time spent by i^{th} knowledge worker on emails of k^{th} type, in l^{th} processing stage and sent to j^{th} knowledge worker on any given day. This is actually a certain proportion of q_{ij} and can be derived using the formula presented below:

$$\left(v_{ijk}\right)_l = p_k \times q_{ij} \tag{28}$$

For newly created emails, equation (28) becomes $\left(v_{ijk}\right)_{l=1} = p_k \times q_{ij}$. Using the data from Table 4.8 and Table 4.5(a), we derive the $\left(v_{ijk}\right)_{l=1}$ matrix and present it in Table 4.10 (a through c). This will give us the split time for newly created emails. Though it is a three dimensional matrix, we present it in a two-dimensional format without losing any details.

Table 4.10 (a) $\left(v_{ijk}\right)_{l=1}$ Matrix for Newly Created Emails for HEN-High

	j=1			*j*=2			*j*=3			*j*=4		
i	*k*=1	*k*=2	*k*=3	*k*=1	*k*=2	*k*=3	*k*=1	*k*=2	*k*=3	*k*=1	*k*=2	*k*=3
1	0.00	0.00	0.00	7.56	7.56	3.78	7.56	7.56	3.78	6.48	6.48	3.24
2	7.56	7.56	3.78	0.00	0.00	0.00	7.56	7.56	3.78	6.48	6.48	3.24
3	7.56	7.56	3.78	7.56	7.56	3.78	0.00	0.00	0.00	6.48	6.48	3.24
4	15.12	15.12	7.56	15.12	15.12	7.56	15.12	15.12	7.56	0.00	0.00	0.00

Table 4.10 (b) $\left(v_{ijk}\right)_{l=1}$ Matrix for Newly Created Emails for HEN-Low

	j=1			*j*=2			*j*=3			*j*=4		
i	*k*=1	*k*=2	*k*=3	*k*=1	*k*=2	*k*=3	*k*=1	*k*=2	*k*=3	*k*=1	*k*=2	*k*=3
1	0.00	0.00	0.00	2.52	2.52	1.26	2.52	2.52	1.26	2.16	2.16	1.08
2	2.52	2.52	1.26	0.00	0.00	0.00	2.52	2.52	1.26	2.16	2.16	1.08
3	2.52	2.52	1.26	2.52	2.52	1.26	0.00	0.00	0.00	2.16	2.16	1.08
4	5.04	5.04	2.52	5.04	5.04	2.52	5.04	5.04	2.52	0.00	0.00	0.00

Table 4.10 (c) $\left(v_{ijk}\right)_{l=1}$ Matrix for Newly Created Emails for XHEN

	j=1			*j*=2			*j*=3			*j*=4		
i	*k*=1	*k*=2	*K*=3	*k*=1	*k*=2	*k*=3	*k*=1	*k*=2	*k*=3	*k*=1	*k*=2	*k*=3
1	0.00	0.00	0.00	11.34	11.34	5.67	3.78	3.78	1.89	6.48	6.48	3.24
2	11.34	11.34	5.67	0.00	0.00	0.00	3.78	3.78	1.89	6.48	6.48	3.24
3	2.52	2.52	1.26	2.52	2.52	1.26	0.00	0.00	0.00	2.16	2.16	1.08
4	15.12	15.12	7.56	15.12	15.12	7.56	5.04	5.04	2.52	0.00	0.00	0.00

Now we can derive the individual email rate creation matrix. Let, $\left(r_{ijk}\right)_l \rightarrow$ represent the rate at which emails are created by i^{th} knowledge worker of k^{th} type, in l^{th} processing stage and sent to j^{th} knowledge worker on any given day. These split email creation rates are the ratio of total time spent per category of email to the number of emails created in that category and can be calculated using the following formula described in equation (29). Table 4.11 (a through c) describes $\left(r_{ijk}\right)_{l=1}$ values for all three networks:

$$\left(r_{ijk}\right)_l = \frac{\left(v_{ijk}\right)_{l=1}}{\left(s_{ij}\right)_{l=1}} \tag{29}$$

Table 4.11 (a) $\left(r_{ijk}\right)_{l=1}$ Split-Rate Matrix for Newly Created Emails for HEN-High

	j=1			*j*=2			*j*=3			*j*=4		
i	*k*=1	*k*=2	*k*=3	*k*=1	*k*=2	*k*=3	*k*=1	*k*=2	*K*=3	*k*=1	*k*=2	*K*=3
1	0.00	0.00	0.00	1.89	3.78	1.89	1.89	3.78	1.89	1.62	3.24	1.62
2	1.89	3.78	1.89	0.00	0.00	0.00	1.89	3.78	1.89	1.62	3.24	1.62
3	1.89	3.78	1.89	1.89	3.78	1.89	0.00	0.00	0.00	1.62	3.24	1.62
4	3.78	7.56	3.78	3.78	7.56	3.78	3.78	7.56	3.78	0.00	0.00	0.00

Table 4.11 (b) $\left(r_{ijk}\right)_{l=1}$ Split-Rate Matrix for Newly Created Emails for HEN-Low

	j=1			*j*=2			*j*=3			*j*=4		
i	*k*=1	*K*=2	*k*=3	*k*=1	*k*=2	*k*=3	*k*=1	*k*=2	*k*=3	*k*=1	*k*=2	*k*=3
1	0.00	0.00	0.00	0.63	1.26	0.63	0.63	1.26	0.63	0.54	1.08	0.54
2	0.63	1.26	0.63	0.00	0.00	0.00	0.63	1.26	0.63	0.54	1.08	0.54
3	0.63	1.26	0.63	0.63	1.26	0.63	0.00	0.00	0.00	0.54	1.08	0.54
4	1.26	2.52	1.26	1.26	2.52	1.26	1.26	2.52	1.26	0.00	0.00	0.00

Table 4.11 (c) $\left(r_{ijk}\right)_{l=1}$ Split-Rate Matrix for Newly Created Emails for XHEN

	j=1			*j*=2			*j*=3			*j*=4		
i	*k*=1	*k*=2	*k*=3	*k*=1	*k*=2	*k*=3	*k*=1	*k*=2	*k*=3	*k*=1	*k*=2	*K*=3
1	0.00	0.00	0.00	2.835	5.67	2.835	0.945	1.89	0.945	1.62	3.24	1.62
2	2.835	5.67	2.835	0.00	0.00	0.00	0.945	1.89	0.945	1.62	3.24	1.62
3	0.63	1.26	0.63	0.63	1.26	0.63	0.00	0.00	0.00	0.54	1.08	0.54
4	3.78	7.56	3.78	3.78	7.56	3.78	1.26	2.52	1.26	0.00	0.00	0.00

Table 4.12 Cumulative % Split-Rate Matrix for Newly Created Emails for HEN-High, HEN-Low and XHEN

	j=1			*j*=2			*j*=3			*j*=4		
i	*k*=1	*k*=2	*k*=3	*k*=1	*K*=2	*k*=3	*k*=1	*k*=2	*K*=3	*k*=1	*k*=2	*K*=3
1	0.00	0.00	0.00	0.25	0.75	1.00	0.25	0.75	1.00	0.25	0.75	1.00
2	0.25	0.75	1.00	0.00	0.00	0.00	0.25	0.75	1.00	0.25	0.75	1.00
3	0.25	0.75	1.00	0.25	0.75	1.00	0.00	0.00	0.00	0.25	0.75	1.00
4	0.25	0.75	1.00	0.25	0.75	1.00	0.25	0.75	1.00	0.00	0.00	0.00

Since the simulation model requires the cumulative percentages of split creation rates, those values are presented in Table 4.12. The cumulative percentage creation rates turned out to be the same for all three types of networks.

Thus, the unified rate can be calculated by adding the arrival rates of all types of emails for any particular *i-j* pair. Thus if $\left(R_{ij}\right)_l$ represents the unified rate at which emails in their l^{th} stage are processed by *i* and sent to *j*, it can be computed using following equation:

$$\left(R_{ij}\right)_l = \sum\left(r_{ijk}\right)_l = \sum\frac{\left(v_{ijk}\right)_l}{\left(s_{ij}\right)_l} \quad \text{Where, } l = 1 \text{or } 2 \text{ or } 3 \tag{30}$$

Although unified rates can be calculated for all stages of emails using the above formula, here it is used to calculate the unified arrival rates of only new emails i.e. for *l*=1. Thus for *l*=1, equation (30) reduces to $\left(R_{ij}\right)_{l=1} = \sum\left(r_{ijk}\right)_{l=1} = \sum\frac{\left(v_{ijk}\right)_{l=1}}{\left(s_{ij}\right)_{l=1}}$. Table 4.13 (a through c) shows the results $\left(R_{ij}\right)_{l=1}$ calculated using the above equation for all three networks.

Table 4.13(a) $\left(R_{ij}\right)_{l=1}$ Unified Rate Matrix for HEN-High

	J				
i	1	2	3	4	$\left(R_i\right)_C$
1	0.00	7.56	7.56	6.48	**21.60**
2	7.56	0.00	7.56	6.48	**21.60**
3	7.56	7.56	0.00	6.48	**21.60**
4	15.12	15.12	15.12	0.00	**45.36**
$\left(R_{j=i}\right)_C$	30.24	30.24	30.24	19.44	

Table 4.13(b) $\left(R_{ij}\right)_{l=1}$ Unified Rate Matrix for HEN-Low

	J				
i	1	2	3	4	$\left(R_i\right)_C$
1	0.00	2.52	2.52	2.16	**7.20**
2	2.52	0.00	2.52	2.16	**7.20**
3	2.52	2.52	0.00	2.16	**7.20**
4	5.04	5.04	5.04	0.00	**15.12**
$\left(R_j\right)_C$	**10.08**	**10.08**	**10.08**	**6.48**	

Table 4.13(c) $\left(R_{ij}\right)_{l=1}$ Unified Rate Matrix for XHEN

	J				
i	1	2	3	4	$\left(R_i\right)_C$
1	0.00	11.34	3.78	6.48	**21.60**
2	11.34	0.00	3.78	6.48	**21.60**
3	2.52	2.52	0.00	2.16	**7.20**
4	15.12	15.12	5.04	0.00	**35.28**
$\left(R_j\right)_C$	**28.98**	**28.98**	**12.60**	**15.12**	

Now we can calculate the unified email arrival rates using the formula described below:

$$\left(R_i\right)_A = \left(R_{j=i}\right)_C + \left(R_i\right)_C - \sum_{j=1}^{4} \left(r_i\right)_{l=1,k=3} \tag{31}$$

Where,

$\left(R_i\right)_A$ → Represents the rate at which emails arrive at i^{ths} knowledge worker

$\left(R_{j=i}\right)_C$ → Represents the rate at which emails are received by the same knowledge worker. Hence, *j=i.* This is the column sum calculated in Table 4.13 (a through c).

$\left(R_i\right)_C$ → Represents the rate at which newly created emails are sent by i^{th} knowledge worker. These are the row sum values calculated in Table 4.13 (a through c)

Table 4.14 represents the $\left(R_i\right)_A$ matrix for all three types of networks.

Table 4.14 $\left(R_i\right)_A$ Matrix: Total Email Arrival Rates

i	HEN-High	HEN-Low	XHEN
	$\left(R_i\right)_A$	$\left(R_i\right)_A$	$\left(R_i\right)_A$
1	**46.44**	**15.48**	**45.18**
2	**46.44**	**15.48**	**45.18**
3	**46.44**	**15.48**	**18.00**
4	**53.46**	**17.82**	**41.58**

4.2 Validation Process: The Reverse Method

The method that we have illustrated just now describes the process of calculating email creation and arrival rates. It is undeniably a fairly complex procedure and before we take the next step, it is important to make sure that this method is doing what we intend it to do. For this purpose, we need to verify that the email creation and arrival rates calculated for each pair of knowledge workers are actually correct and can lead us to the same numbers if a reverse process is adopted. In the method described earlier, we did not use the q'_{ij} matrix derived earlier in Table 4.6 (a through c). It is used now for the validation purpose described in detail below. Figure 4.3 describes different parameters were evaluated along with the procedure for validating these parameters. The main outcomes of the above mentioned method was the three rate matrices that we deduced: the split rate matrix for newly created emails, the unified rate matrix for newly created emails and the unified rate matrix for arriving emails. These outcome matrices should yield a matrix similar to q'_{ij} if a reverse process is adopted. This matrix is q''_{ij} and the difference between the two matrices (Δ) is calculated. For the validity of this process to be established, this difference should be minimal. This method is referred to as the reverse method.

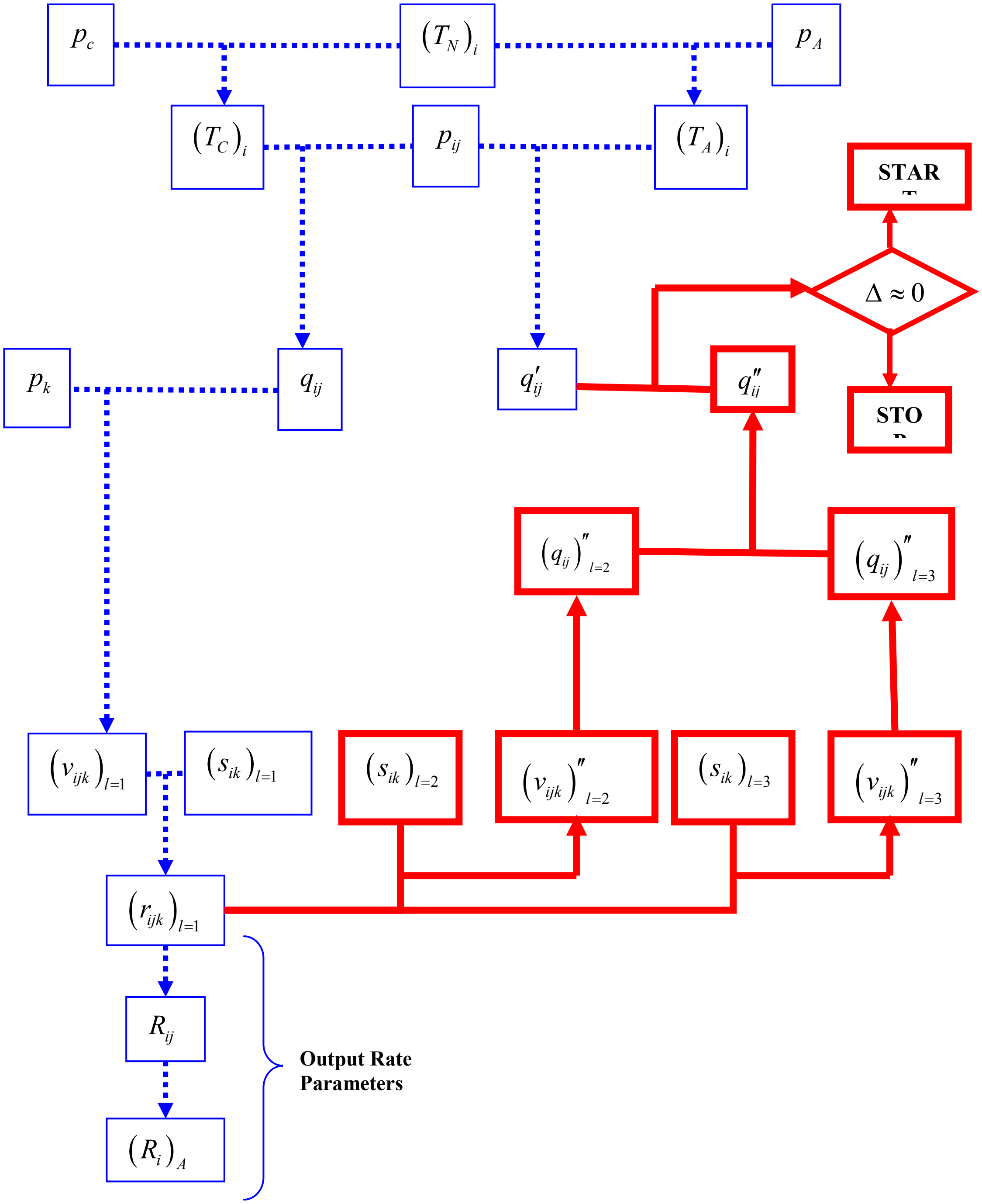


Figure 4.3 The Process of Parameter Calculation (Forward Method) and Parameter Validation (Reverse Method)

Following is the four-step process used for this purpose:

Step 1: From the split email creation rate matrix $\left(r_{ijk}\right)_{l=1}$, calculate the split-time matrix for arriving emails in their second (l=2) and third (l=3) processing stages. The formula for calculating this is:

$$\left(v_{ijk}\right)''_{l} = \left(r_{ijk}\right)_{l-1} \times \left(s_{ik}\right)_{l} \qquad \text{where, } l = 2 \text{ and } 3 \tag{32}$$

Step 2: Using the split email time matrix for arriving email derived in step 1, calculate the time that knowledge worker spends on "arriving" emails that are in their second (l=2) and third (l=3) processing stages. Mathematically, this time matrix can be evaluated using the formula described below:

$$\left(q''_{ij}\right)_{l} = \sum_{\forall k} \left(v''_{k}\right)_{ij,l} \qquad \text{where, } l = 2 \text{ and } 3, \quad k = 1, 2, 3 \tag{33}$$

Step 3: Calculate the combined time matrix for emails in their second (l=2) and third (l=3) processing stages.

$$q''_{ij} = \left(q''_{ij}\right)_{l=2} + \left(q''_{ij}\right)_{l=3} \tag{34}$$

Step 4: Calculate the absolute value of the difference between the combined time matrix for arriving emails calculated through both the forward and the reverse methods.

$$\Delta_{i} = \sum_{\forall j} \Delta_{ij} = \left| q'_{ij} - q''_{ij} \right| \tag{35}$$

A minimal error confirms the validity of the forward method used to derive various rate parameters.

4.3 Implementation of the Reverse Method

We will now show the implementation of this method for different types of networks, starting with HEN-High networks. Table 4.15 (a and b) shows the split-time matrix for arriving emails in phases 2 and 3 of their processing for HEN-High networks

Table 4.15 (a) $\left(v_{ijk}\right)''_{l=2}$ Split-Time Matrix for Arriving Emails in Phase 2 for HEN-High

	j=1			*j*=2			*j*=3			*j*=4		
I	*k*=1	*k*=2	*k*=3	*k*=1	*k*=2	*k*=3	*k*=1	*k*=2	*k*=3	*k*=1	*k*=2	*k*=3
1	0.00	0.00	0.00	11.34	11.34	3.78	11.34	11.34	3.78	9.72	9.72	3.24
2	11.34	11.34	3.78	0.00	0.00	0.00	11.34	11.34	3.78	9.72	9.72	3.24
3	11.34	11.34	3.78	11.34	11.34	3.78	0.00	0.00	0.00	9.72	9.72	3.24
4	22.68	22.68	7.56	22.68	22.68	7.56	22.68	22.68	7.56	0.00	0.00	0.00

Table 4.15 (b) $\left(v_{ijk}\right)''_{l=3}$ Split-Time Matrix for Arriving Emails in Phase 3 for HEN-High

	j=1		*j*=2		*j*=3		*j*=4	
i	*k*=1	*k*=2	*k*=1	*k*=2	*k*=1	*k*=2	*k*=1	*k*=2
1	0.00	0.00	7.56	7.56	7.56	7.56	6.48	6.48
2	7.56	7.56	0.00	0.00	7.56	7.56	6.48	6.48
3	7.56	7.56	7.56	7.56	0.00	0.00	6.48	6.48
4	15.12	15.12	15.12	15.12	15.12	15.12	0.00	0.00

In the second step, we calculate the time matrix for emails in their second (l=2) and third (l=3) processing stages for HEN-High networks. Table 4.16 (a and b) shows those values.

Table 4.16 (a) $\left(q''_{ij}\right)_{l=2}$ Matrix for Time Spent on Phase 2 Arriving Emails in HEN-High

	Sender (*j*)			
Receiver (*i*)	**1**	**2**	**3**	**4**
1	0.00	26.46	26.46	22.68
2	26.46	0.00	26.46	22.68
3	26.46	26.46	0.00	22.68
4	52.92	52.92	52.92	0.00

Table 4.16 (b) $\left(q''_{ij}\right)_{l=3}$ Matrix for Time Spent on Phase 3 Arriving Emails in HEN-High

	Sender (*j*)			
Receiver (*i*)	**1**	**2**	**3**	**4**
1	0.00	15.12	15.12	12.96
2	15.12	0.00	15.12	12.96
3	15.12	15.12	0.00	12.96
4	30.24	30.24	30.24	0.00

The third step involves calculating the combined time spent on emails in the 2nd and 3rd stages. We obtain this by adding the values presented in Table 4.16(a) and Table 4.16(b). The summated values are given in Table 4.17.

Table 4.17 q''_{ij} Matrix for Time Spent on Arriving Emails in HEN-High

	Sender (*j*)			
Receiver (*i*)	**1**	**2**	**3**	**4**
1	0.00	41.58	41.58	35.64
2	41.58	0.00	41.58	35.64
3	41.58	41.58	0.00	35.64
4	83.16	83.16	83.16	0.00

Finally, we evaluate the difference (Δ_i) between the q''_{ij} matrix obtained through this reverse process and the q'_{ij} matrix obtained earlier through the forward process and present it in Table 4.18. Although these values are non-zero, they are very small and can be considered well within the tolerance limits. Such minor deviations from zero can be easily accounted for by the probabilistic distributions that are used in simulation models.

We will now show the implementation of this method for HEN-Low networks. Table 4.19 (a and b) shows the split-time matrix for arriving emails in phases 2 and 3 of their processing for HEN-Low networks.

Table 4.18 Δ_i Values for HEN-High

i	Δ_i
1	7.2
2	7.2
3	7.2

Table 4.19 (a) $\left(v_{ijk}\right)''_{l=2}$ Split-Time Matrix for Arriving Emails in Phase 2 for HEN-Low

	j=1			*j*=2			*j*=3			*j*=4		
I	*k*=1	*k*=2	*k*=3	*k*=1	*k*=2	*k*=3	*k*=1	*k*=2	*k*=3	*k*=1	*k*=2	*K*=3
1	0.00	0.00	0.00	3.78	3.78	1.26	3.78	3.78	1.26	3.24	3.24	1.08
2	3.78	3.78	1.26	0.00	0.00	0.00	3.78	3.78	1.26	3.24	3.24	1.08
3	3.78	3.78	1.26	3.78	3.78	1.26	0.00	0.00	0.00	3.24	3.24	1.08
4	7.56	7.56	2.52	7.56	7.56	2.52	7.56	7.56	2.52	0.00	0.00	0.00

Table 4.19 (b) $\left(v_{ijk}\right)''_{l=3}$ Split-Time Matrix for Arriving Emails in Phase 3 for HEN-Low

	j=1		*j*=2		*j*=3		*j*=4	
i	*k*=1	*k*=2	*k*=1	*K*=2	*k*=1	*k*=2	*k*=1	*k*=2
1	0.00	0.00	2.52	2.52	2.52	2.52	2.16	2.16
2	2.52	2.52	0.00	0.00	2.52	2.52	2.16	2.16
3	2.52	2.52	2.52	2.52	0.00	0.00	2.16	2.16
4	5.04	5.04	5.04	5.04	5.04	5.04	0.00	0.00

In the second step, we calculate the time matrix for emails in their second (l=2) and third (l=3) processing stages for HEN-Low networks. Table 4.20 (a and b) shows those values.

Table 4.20 (a) $\left(q''_{ij}\right)_{l=2}$ Matrix for Time Spent on Phase 2 Arriving Emails in HEN-Low

	Sender (*j*)			
Receiver (*i*)	**1**	**2**	**3**	**4**
1	0.00	8.82	8.82	7.56
2	8.82	0.00	8.82	7.56
3	8.82	8.82	0.00	7.56
4	17.64	17.64	17.64	0.00

Table 4.20 (b) $\left(q''_{ij}\right)_{l=3}$ Matrix for Time Spent on Phase 3 Arriving Emails in HEN-Low

	Sender (*j*)			
Receiver (*i*)	**1**	**2**	**3**	**4**
1	0.00	5.04	5.04	4.32
2	5.04	0.00	5.04	4.32
3	5.04	5.04	0.00	4.32
4	10.08	10.08	10.08	0.00

The third step involves the calculation of combined time spent on emails in the 2nd and 3rd stages. We obtain this by adding the values presented in Table 4.20(a) and Table 4.20(b); the summated values are shown in Table 4.21.

Table 4.21 q''_{ij} Matrix for Time Spent on Arriving Emails in HEN-Low

	Sender (*j*)			
Receiver (*i*)	**1**	**2**	**3**	**4**
1	0.00	13.86	13.86	11.88
2	13.86	0.00	13.86	11.88
3	13.86	13.86	0.00	11.88
4	27.72	27.72	27.72	0.00

Finally, we evaluate the difference (Δ_i) between the q''_{ij} matrix obtained through this reverse process and the q'_{ij} matrix obtained earlier through the forward process and present it in Table 4.22. Again, these deviations are almost negligible for the HEN-Low type of networks.

Table 4.22 Δ_i Values for HEN-Low

i	Δ_i
1	2.4
2	2.4
3	2.4

We will now show the implementation of the reverse validation method for XHEN networks. Table 4.23 (a and b) shows the split-time matrix for arriving emails in phases 2 and 3 of their processing for XHEN networks.

Table 4.23 (a) $\left(v_{ijk}\right)''_{l=2}$ Split-Time Matrix for Arriving Emails in Phase 2 for XHEN

	j=1			*j*=2			*j*=3			*j*=4		
I	*k*=1	*k*=2	*k*=3	*k*=1	*k*=2	*k*=3	*k*=1	*k*=2	*k*=3	*k*=1	*k*=2	*k*=3
1	0.00	0.00	0.00	17.01	17.01	5.67	5.67	5.67	1.89	9.72	9.72	3.24
2	17.01	17.01	5.67	0.00	0.00	0.00	5.67	5.67	1.89	9.72	9.72	3.24
3	3.78	3.78	1.26	3.78	3.78	1.26	0.00	0.00	0.00	3.24	3.24	1.08
4	22.68	22.68	7.56	22.68	22.68	7.56	7.56	7.56	2.52	0.00	0.00	0.00

Table 4.23 (b) $\left(v_{ijk}\right)''_{l=3}$ Split-Time Matrix for Arriving Emails in Phase 3 for XHEN

	j=1		*j*=2		*j*=3		*j*=4	
i	*k*=1	*k*=2	*k*=1	*k*=2	*k*=1	*k*=2	*k*=1	*k*=2
1	0.00	0.00	11.34	11.34	3.78	3.78	6.48	6.48
2	11.34	11.34	0.00	0.00	3.78	3.78	6.48	6.48
3	2.52	2.52	2.52	2.52	0.00	0.00	2.16	2.16
4	15.12	15.12	15.12	15.12	5.04	5.04	0.00	0.00

In the second step, we calculate the time matrix for emails in their second (*l*=2) and third (*l*=3) processing stages for HEN-High networks. Table 4.24 (a and b) shows those values.

Table 4.24 (a) $\left(q''_{ij}\right)_{l=2}$ Matrix for Time Spent on Phase 2 Arriving Emails in XHEN

	Sender (*j*)			
Receiver (*i*)	**1**	**2**	**3**	**4**
1	0.00	39.69	13.23	22.68
2	39.69	0.00	13.23	22.68
3	8.82	8.82	0.00	7.56
4	52.92	52.92	17.64	0.00

Table 4.24 (b) $\left(q''_{ij}\right)_{l=3}$ Matrix for Time Spent on Phase 3 Arriving Emails in XHEN

	Sender (*j*)			
Receiver (*i*)	**1**	**2**	**3**	**4**
1	0.00	22.68	7.56	12.96
2	22.68	0.00	7.56	12.96
3	5.04	5.04	0.00	4.32
4	30.24	30.24	10.08	0.00

The third step involves the calculation of the combined time spent on emails in the 2nd and 3rd stages. We obtain this by adding the values presented in Table 4.24(a) and Table 4.24(b) and present the summated values in Table 4.25.

Table 4.25 q''_{ij} Matrix for Time Spent on Arriving Emails in XHEN

	Sender (*j*)			
Receiver (*i*)	**1**	**2**	**3**	**4**
1	0.00	62.37	20.79	35.64
2	62.37	0.00	20.79	35.64
3	13.86	13.86	0.00	11.88
4	83.16	83.16	27.72	0.00

Finally, we evaluate the difference (Δ_i) between the q''_{ij} matrix obtained through this reverse process and the q'_{ij} matrix obtained earlier through the forward process and present it in Table 4.26. Again, these deviations are almost negligible. Since for all the three types of networks, the error was almost negligible, we can conclude that the different rate parameters developed using the forward method are valid.

Table 4.26 Δ_i Values for XHEN

i	Δ_i
1	7.2
2	7.2
3	2.4

4.4 Other Parameters

Section 3.4 above conceptually described the value of an email and how it diminishes with time (Fig. 3.6). However, the section did not contain a mathematical treatment of this concept and nor a description of how different parameters related to it can be calculated. This topic is more appropriate for this section. The survey conducted by us helped identify three major categories of emails based on their urgency level (Table 4.1): emails with high urgency, emails with moderate urgency, and emails with low urgency. Emails with high urgency demand a quick response from the recipient, as their value to the organization drops very quickly. On the other extreme are the emails with low urgency. Such emails require a rather slow response and usually have a longer life time. Between the two extremes lie emails with moderate urgency. Figure 4.4 explains this more explicitly. The shape of these curves resembles that of a sigmoid curve.

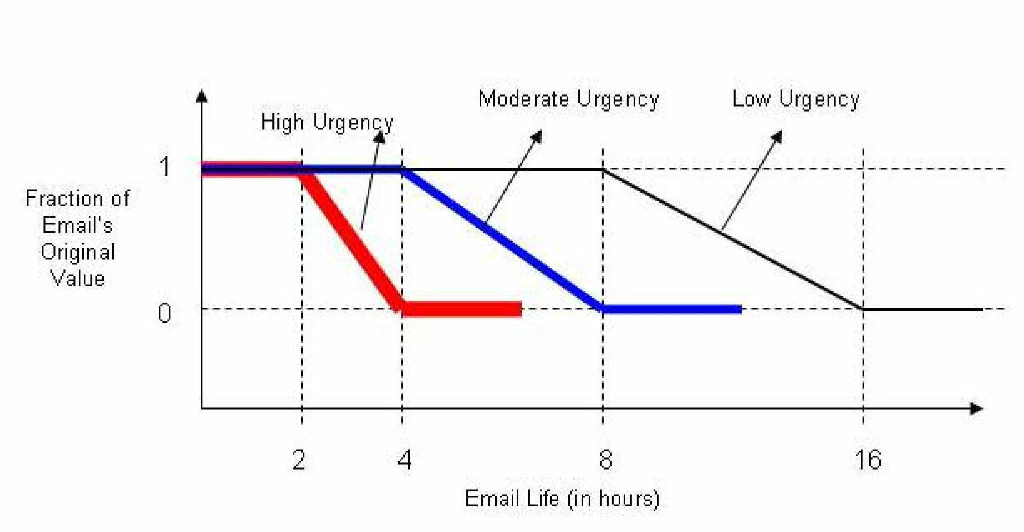


Figure 4.4 Email Value vs. Email Life for High, Moderate and Low Urgency Emails

If emails belonging to the high urgency category are responded to within two hours, than organizations derive maximum value from the information contained in the email. If instead they are processed after two to four hours, the value depends on a negative linear function. Using simple co-ordinate geometry concepts, we can easily find that this is the equation of a straight line with a slope of -1/2. The fraction of value derived at any time ‘t’ (in hours) can be computed using equation (36) for emails belonging to the high urgency category. However, if the wait time of emails exceeds 4 hours than a zero value is derived.

$$(h_3)_{High} = 2 - \left(\frac{1}{2}\right) \times \frac{t}{60} \qquad \text{Where, t is in hours} \qquad (36)$$

If emails of moderate urgency are processed any time within four hours of their arrival, the organization derives maximal value for their processing. However, if the wait time exceeds four hours but is less than eight hours, the value derived is calculated using a simple linear equation. Using the two test points on this line, namely (4, 1) and (8, 0), we can easily determine the slope of this curve to be -1/4 with a positive intercept of 2. Equation (37) is used to determine the value of email during this time period. If any email belonging to this category is processed after 8 hours of inbox waiting time, the organization derives a zero value.

$$\left(h_3\right)_{Mod} = 2 - \left(1/4\right) \times t/60 \tag{37}$$

Finally, if emails belong to the low urgency category, they can tolerate a longer waiting period before their value starts to diminish. Their value starts to decline at a constant rate only if the wait period exceeds eight hours. The linear function that governs this rate of fall has a slope of -1/8 with a positive intercept of 2. This linear relationship holds valid until the time period of sixteen hours. If emails are processed after a sixteen-hour wait period, a value of zero is derived. Equation (38) below describes this linear function:

$$\left(h_3\right)_{Low} = 2 - \left(1/4\right) \times t/60 \tag{38}$$

Also, since we assume a learning rate of 70 % for the knowledge work environment that we modeled in this study, Wright’s power function exponent becomes, -0.1549.

Finally we used the statistics provided by Jackson et al. (2003) for modeling the switching time: a triangular distribution with a minimum, mode, and maximum of 1, 1.5 and 2 minutes respectively.

5. MODEL IMPLEMENTATION

5.1 Schematic Diagram

The network that was modeled comprises three knowledge workers interacting with the outside world. Each knowledge worker within the network was capable of sending or receiving emails to or from anyone within the network. Figure 5.1 describes the flow of email with this network. The outside world collectively represents a group of knowledge workers not belonging to the network and processing emails at random times of the day. Since these outside knowledge workers can be geographically located anywhere in the world and in any time zone, they process emails even at night. On the other hand, knowledge workers belonging to the network are capable of processing emails only during their regular office hours, which also implies that they do not exchange emails during the night hours. Further, each email undergoes a cycle of processing before it gets resolved. The length and nature of this cycle depends upon the type of email i.e. whether the email requires a response or not.

The processing cycle is relatively small if the email is of type 3, such as FYI emails, informative emails, CC emails, etc. Type 3 emails do not require a response to the original sender and tend to be resolved once they have been processed by the receiver. This cycle will be relatively long if the email is a type 1 or 2 (simple or complex email) that requires the receiver to respond to the original sender. For such emails, the cycle ends after the response from the receiver has been processed by the email's original sender. This also marks the resolution of the email.

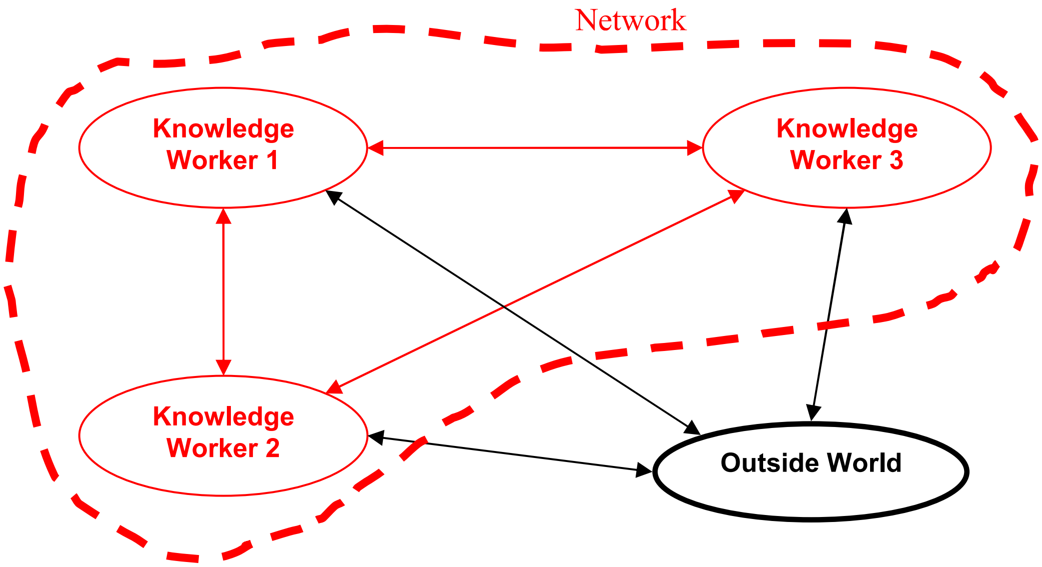


Figure 5.1 Email Flow within a Network

One thing to be noted here is that the focus of the model is on email resolution and not issue resolution. An email resolution does not guarantee the resolution of the issue discussed in the email, when it undergoes one cycle of processing. An issue being discussed over email often requires more than one cycle to be resolved, but such situations are not modeled in this study.

Figure 5.2 shows how email processing strategies are used to manage interruptions within any work environment. A new email created by the sender reaches the inbox of the recipient, but the processing of this email does not begin until several conditions have been met. The first is whether email priority hours are in progress or not. If email priority hours are not in-progress and the recipient is also busy with primary-task processing, the email will wait until the arrival of next scheduled email hour or the completion of the primary task, whichever occurs first.

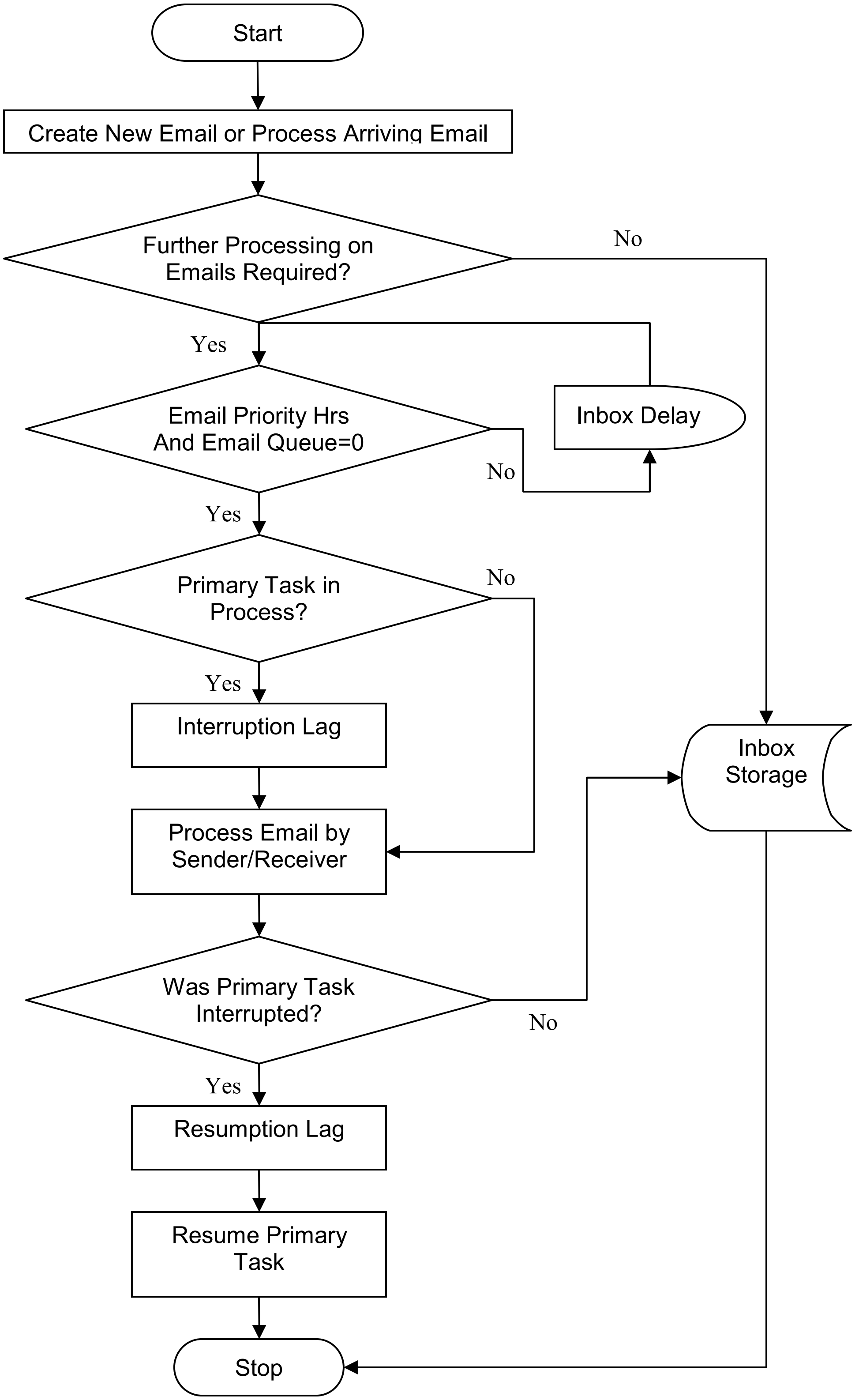


Figure 5.2 Use of Email Processing Strategies to Manage Interruptions at Work

If email priority hours are in progress and other emails are already waiting, there will also be a delay in the processing of this new email. Once the queue of all pending email clears up, the knowledge worker may begin processing the new email. If the email queue is zero and email priority hours are underway, any processing of the primary task will be interrupted, interruption lag will occur due to the switching involved, and processing on the new email will begin. Resumption lag time will be added before any work on the primary task is resumed. The fate of this email depends on its type. An email requiring a response will go through a similar pattern but this time at the sender's end. In this event, the email not requiring a response will be stored in the inbox.

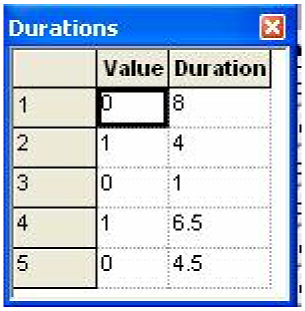
5.2 Implementation of the Simulation Model

Due to the time-length and the nature of the policies being compared, it was extremely difficult to conduct this study as an experimental or field study with enough control. Hence, a simulation-based computer experiment was chosen to study this problem. Simulation has often served as a very useful tool for theory enquiry and development (Di Paolo 2000; Peschl 2001) and theory development and investigation (Hans-Joachim 2001) and can be used to conduct virtual experiments (Winsberg 2003). Several researchers have described simulation as a way of doing "thought experiments" and as a technique that can give surprising "emerging" results due to "large-scale interactions of local agents which are often difficult to anticipate" (Axlerod 2003).

Following is a description of the entire architecture of the simulation model, which was built using Arena 8.01, to implement email exchanges within a network. In all, thirty-eight different network models were created to implement thirty-eight different policy-sets. Of these, ten models were for HEN-High networks, ten were for HEN-Low networks and the remaining eighteen were for XHEN networks. The major components of the model are described in the following subsections.

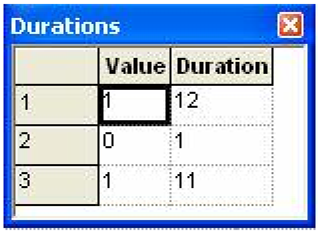
5.2.1 Implementation of Primary Task Processing

This section focuses primarily on explaining how the processing of the primary task was modeled. All the knowledge workers that were modeled worked for more than 8 hours per work day. Fig 5.3(a) shows the schedule of a typical knowledge worker. The column titled “value” shows the capacity for the “duration” of time mentioned in the corresponding cell of the second column. The unit of duration is hours. A capacity of 0 implies that a knowledge worker (KW) is unavailable, whereas a capacity of 1 implies that a knowledge worker is available to either process primary tasks or email. The outside world was modeled as a 4th knowledge worker in the simulation model. Figure 5.3 (b) shows that this particular KW is available throughout the day except during the lunch hour, but there is a difference. The 4th KW only processes email and not primary tasks. We are not interested in collecting the output parameters of this KW since our focus is on the performance of the network.



	Value	Duration
1	0	8
2	1	4
3	0	1
4	1	6.5
5	0	4.5

Figure 5.3 (a) Schedule of KW 1, 2 and 3

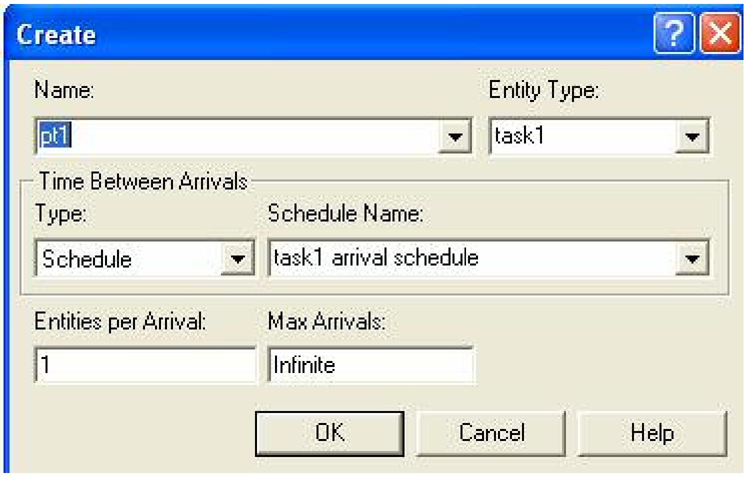


	Value	Duration
1	1	12
2	0	1
3	1	11

Figure 5.3(b) Schedule of KW 4

A “create” module from basic Arena processes was used to generate primary tasks (Fig. 5.4). Each knowledge worker works on his or her own task only and not that of some other knowledge worker; i.e. knowledge workers do not send their primary tasks to other knowledge workers. Although primary tasks are non-collaborative, emails, by their very nature, are collaborative. As soon as processing on a particular primary task is finished, it leaves the system after various statistics have been collected on it.

Since the 4th knowledge worker does not process any primary tasks, task arrival schedules were created for only three knowledge workers. These schedules are stochastic in nature. The “Schedule” module was used to create these schedules. Figure 5.5 (a) shows the use of this “schedule” block for creating the primary task arrival schedule for the 1st knowledge worker. After defining the time units, “hours” in this case, we set the hourly exponential arrival rate by populating the “duration” box. For example, primary tasks do not arrive at night. So an arrival distribution for the first 8 hours (starting from midnight) was set to an exponential with zero rate. An arrival rate of Expo (2) was set for the time between 8:00am and 9:00am. Fig 5.5(b) shows how hourly rates are defined within the schedule block. HEN-High networks, where all knowledge workers spend stochastically the same amount of time on primary tasks, have an arrival schedule that looks similar to the one shown in Fig. 5.6 (a). For HEN-Low networks, primary task schedules are shown in Fig. 5.6 (b). However, for a XHEN type of network, knowledge workers 1 and 2 used the arrival schedule presented in Fig 5.6 (a), whereas knowledge worker 3 used the schedule in Fig. 5.6 (b). All the arrival rates for primary tasks are exponential in nature.



Name: pt1

Entity Type: task1

Time Between Arrivals

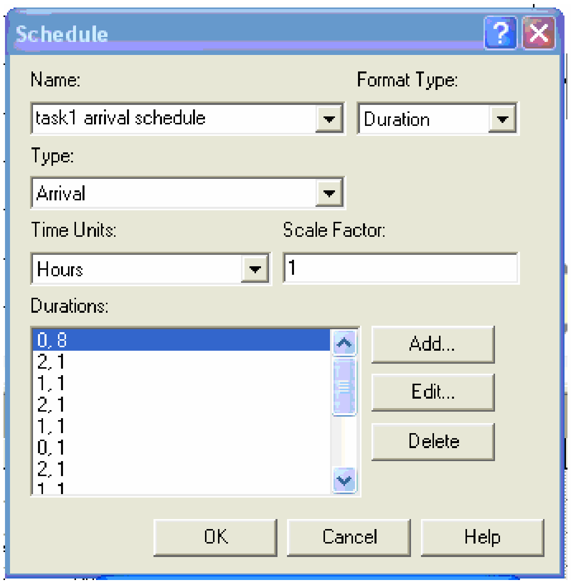
Type: Schedule

Schedule Name: task1 arrival schedule

Entities per Arrival: 1

Max Arrivals: Infinite

Figure 5.4 Creating Primary Task Entities using “Create” Module



Name: task1 arrival schedule

Format Type: Duration

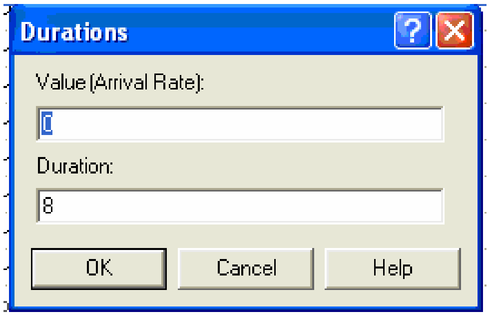
Type: Arrival

Time Units: Hours

Scale Factor: 1

Durations:
0, 8
2, 1
1, 1
2, 1
1, 1
0, 1
2, 1
1, 1

Figure 5.5(a) Schedule Block



Value (Arrival Rate): 0

Duration: 8

Figure 5.5 (b) Setting Arrival Rates

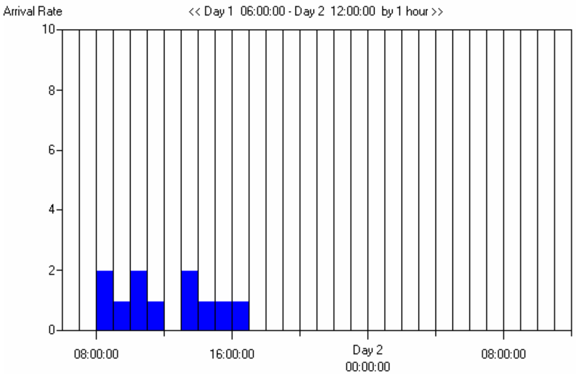


Figure 5.6(a) Task 1 and 2 Arrival Rate

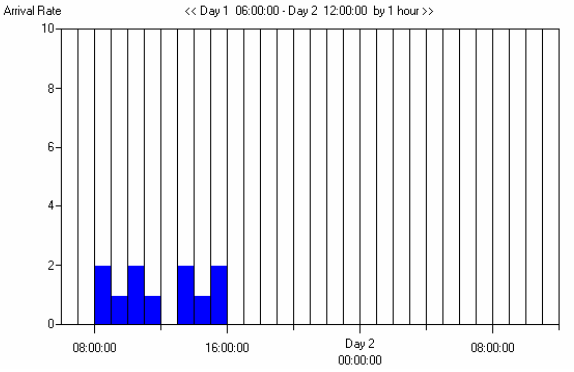
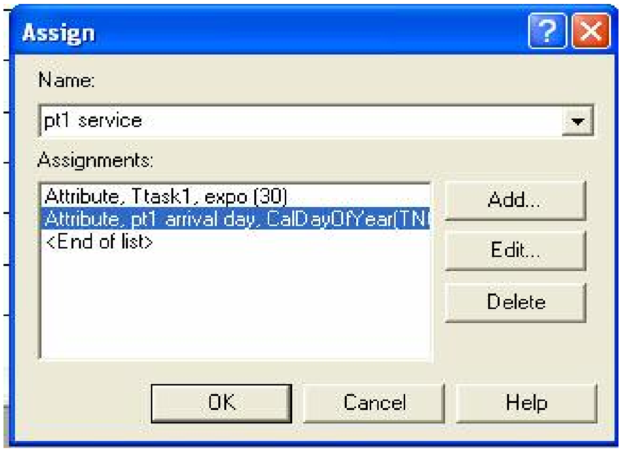


Figure 5.6(b) Task 3 Arrival Rate

As soon as a new primary task is created, it is stamped with two attributes. An "Assign" block is used for this purpose. Attribute "Ttask1" informs the simulation model about its service time, which has been set to Expo (30) with time units being minutes. Another attribute tracks the day of its arrival using the "CalDayofYear( )" function. The argument of this function is set to "TNOW." Fig 5.7 shows the use of the "Assign" block for this purpose. Another "Assign" block kept track of the time a new primary task entered the system. This entry time was later used to calculate the task completion time when the task was about to exit the system.

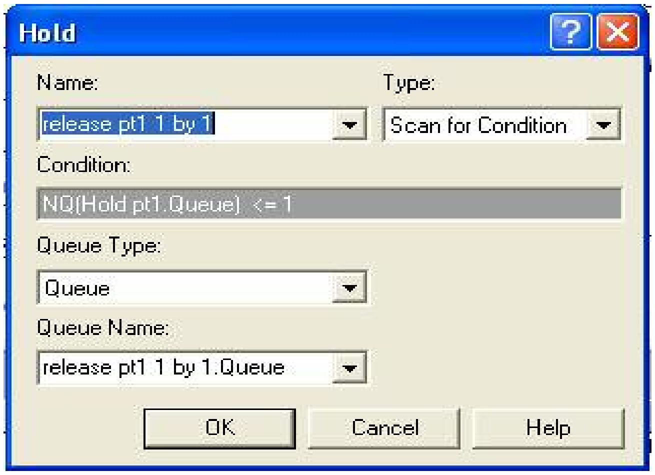

Assign

Name: pt1 service

Assignments:

Attribute, Ttask1, expo (30)
Attribute, pt1 arrival day, CalDayOfYear(TN
<End of list>

Figure 5.7 Assign Block to set Primary Task Service Time and Track Arrival Day


Hold

Name: release pt1 1 by 1

Type: Scan for Condition

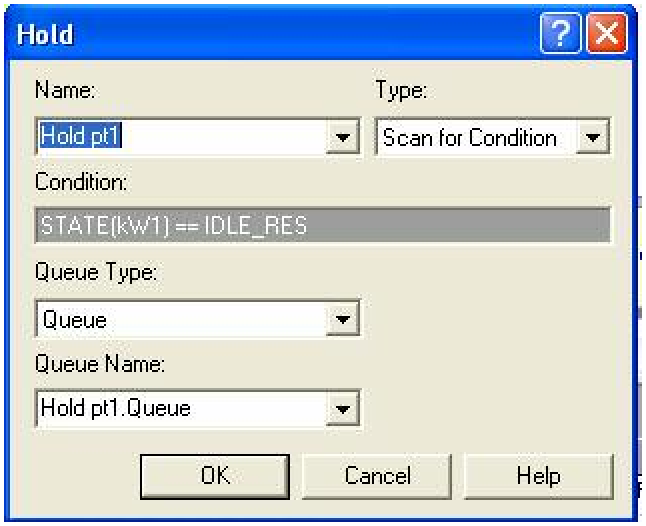
Condition: NQ(Hold pt1.Queue) <= 1

Queue Type: Queue

Queue Name: release pt1 1 by 1.Queue

Figure 5.8 Hold Block for Releasing Tasks One by One

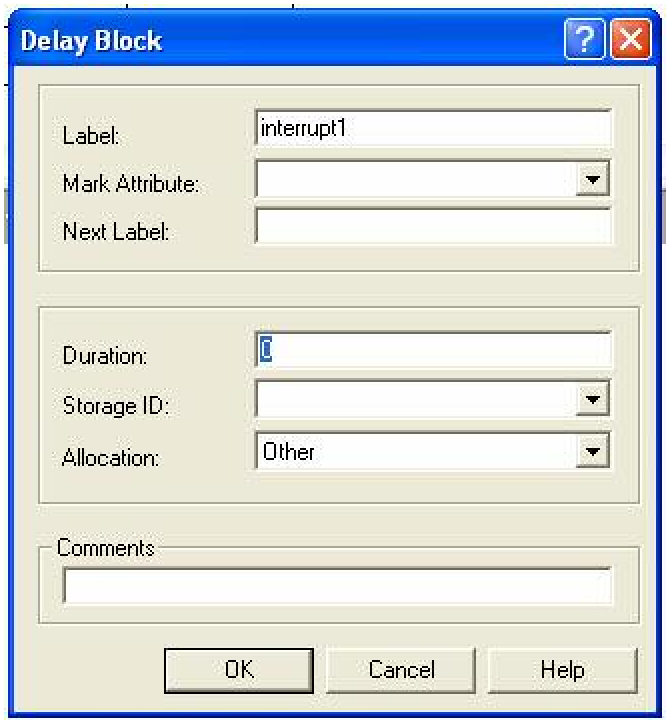
Each primary task, after passing through the assign block, reaches two “Hold” blocks. These “Hold” blocks hold the entity till a certain condition has been met or a signal has been received. The first “Hold” block shown in Fig 5.8 allows the primary tasks to be released to the knowledge worker one by one. The second “Hold” block allows a new task to be released only if all the previous tasks have been processed by the knowledge worker. Fig. 5.9 describes the condition to implement this release. Once a task is released by both “Hold” blocks, it is ready for processing by the knowledge worker. This resource is seized as soon as a task arrives in the “Process” block and is released upon the completion of the primary task, if no interruption occurs during this time. The time spent on the task while it is undergoing processing is recorded as value-added time.



Name: Hold pt1
Type: Scan for Condition
Condition: STATE(kW1) == IDLE_RES
Queue Type: Queue
Queue Name: Hold pt1.Queue

Figure 5.9 Hold Block for Releasing Tasks when Knowledge Worker is Idle

However, if an interruption occurs while the knowledge worker is processing the task and an email priority hour is in progress, the knowledge worker is preempted and is diverted to processing emails. The remaining time to be spent on the primary task is recorded in an internal variable referred to as RT1/2/3 depending upon which resource was preempted, knowledge worker 1, 2, or 3. Once the preempted resource is released, the knowledge worker begins processing the remaining primary tasks. However, a task that has been previously interrupted is handled quite differently within the simulation model. An interrupted task, instead of being taken to its original processing location (i.e. “Process” block), is taken to a “Delay” block, as shown in Fig. 5.10.


Delay Block

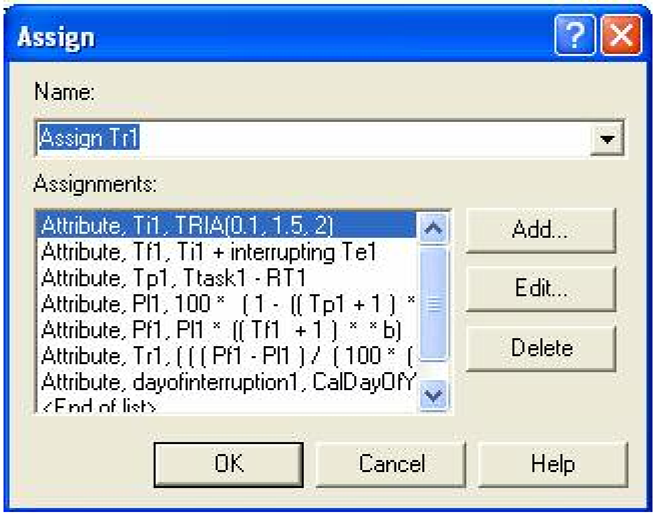
Label: interrupt1
Mark Attribute:
Next Label:

Duration: 0
Storage ID:
Allocation: Other

Comments

Figure 5.10 Delay Block Where a Previously Interrupted Task Enters

The “Delay” block then passes the interrupted primary task to a new ‘Assign’ block (Fig 5.11). This is where the previously stored value of remaining time (RT1/2/3) is used to evaluate switching time and recall time using formulae described in Section 3. All the values thus calculated are then stamped on each interrupted tasks by declaring it an attribute of the entity. The task then undergoes a switching time represented by ‘Ti1’ (or 2 or 3) in the model. This is also based on an expression defined in this assign block. Before knowledge workers can resume work on a previously interrupted primary task, they have to spend some time recalling the part of the task that was previously accomplished. This time is represented as ‘Tr1’ (or 2 or 3) in the model. The switching time and recall time both preoccupy the resource but this time is considered non-value-added time in the model.



Name:

Assign Tr1

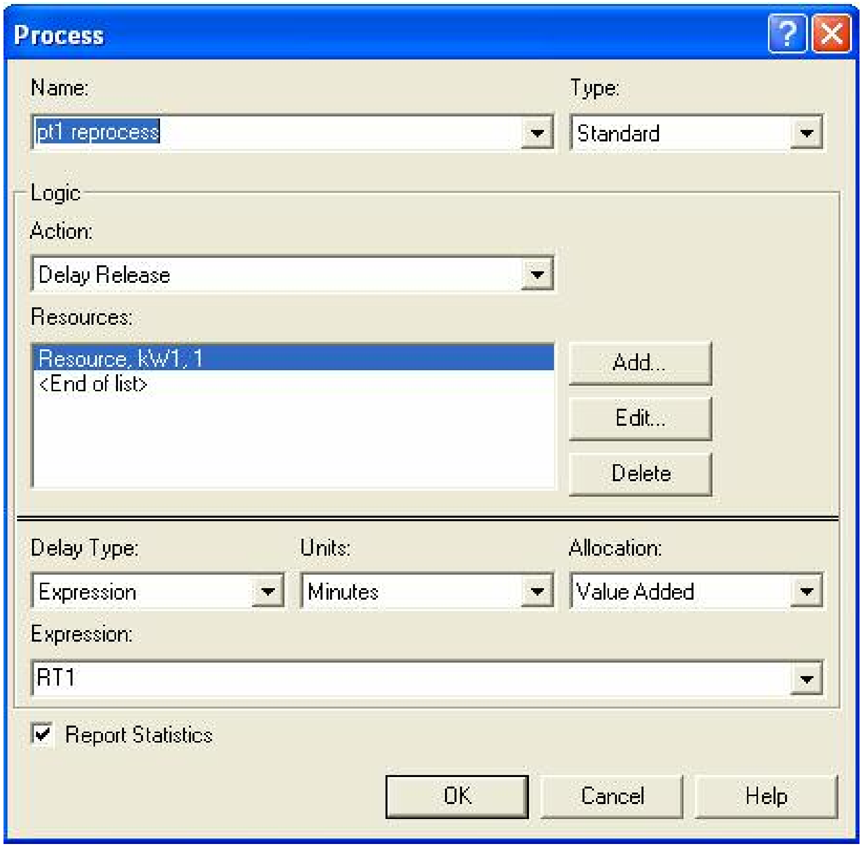
Assignments:

```
Attribute, Ti1, TRIA(0.1, 1.5, 2)
Attribute, Tf1, Ti1 + interrupting Te1
Attribute, Tp1, Ttask1 - RT1
Attribute, Pl1, 100 *  ( 1 - (( Tp1 + 1 ) *
Attribute, Pf1, Pl1 * (( Tf1  + 1 ) *  * b)
Attribute, Tr1, ( ( ( Pf1 - Pl1 ) /  ( 100 *  (
Attribute, dayofinterruption1, CalDayOfY
<End of list>
```

Figure 5.11 Assign Block to Evaluate Switching Time and Recall Time

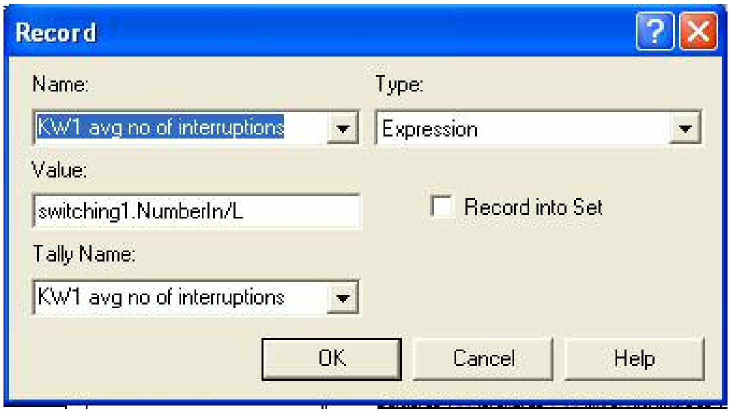
The tasks then enter a new “Process” block where the resource is seized once again so that work on the task can be started. However for this time, the resource is occupied only for the duration of the remaining task time. This time is also treated as value-added time by the system. Fig. 5.12 shows the new “Process” block where the interrupted task is processed for the duration of time stored in variable “RT1.” A task can potentially be interrupted several times while undergoing processing. The cycle is repeated each time an interruption occurs. The value stored in the RT1 variable is updated and the new entity is again taken to the “Delay” block. From this point onwards, the processing begins very much in the manner just described.

Finally, various statistics are calculated and collected using the “Record” module. Several such modules captured different statistics to be used later on. An example of one such “Record” module is shown in Fig. 5.13, where we use it to evaluate the average number of interruptions that occur per day. Also data for each entity was recorded using the “Read/Write” module from the advanced process template within Arena. Eventually, the entity exits the system after its exit time has been recorded.



Name: pt1 reprocess
Type: Standard
Action: Delay Release
Resources: Resource, KW1, 1
<End of list>
Delay Type: Expression
Units: Minutes
Allocation: Value Added
Expression: RT1
☑ Report Statistics

Figure 5.12 Processing of an Interrupted Task at the New ‘Process’ Block



Name: KW1 avg no of interruptions
Type: Expression
Value: switching1.NumberIn/L
☐ Record into Set
Tally Name: KW1 avg no of interruptions

Figure 5.13 Record Module used to Calculate Average Number of Interruptions

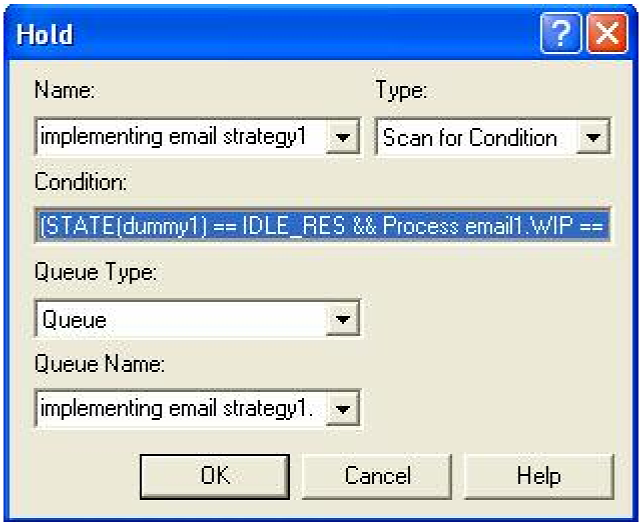
5.2.2 Implementation of Email Processing Strategies

This section explains the logic used to implement various email processing strategies (EPS) within different types of networks along with how emails were created in the model. To implement EPS for each knowledge worker, we created a pair of dummy resources labeled “dummy” and “anti dummy.” Thus, for four knowledge workers, we built four such pairs. These resources were built to have contradictory work schedules, i.e. if “dummy” is working, then “anti dummy” is not available. The “dummy” resource is active only during email-priority hours, whereas the “anti dummy” resource remains active only during non-email-priority hours. These resources do not perform any activity such as process primary tasks or emails. So whenever they are available, their state is idle since they are not being utilized anywhere. This feature helped us identify when to switch priority between emails and primary tasks. When the “dummy” resource is idle, email priority hours are in progress and priority is given to processing email. On the other hand, if the “anti dummy” is idle, non email-priority hours are underway and the priority switches to primary task processing.

Fig. 5.14 shows how the “Hold” block was used for modeling EPS for the 1st knowledge worker. Any email that arrives at this “Hold” block is held in an internal queue until the following condition is not met.

(STATE (dummy) == IDLE_RES && Email_Process.WIP == 0 && switching.WIP == 0 && relearning.WIP == 0)

(STATE (anti dummy1) == IDLE_RES && Primary_Task_Process.WIP== 0 && Primary_Task Reprocess.WIP== 0 && NQ (Hold Primary_Task_Process.Queue) == 0 && switching.WIP == 0 && relearning.WIP == 0)

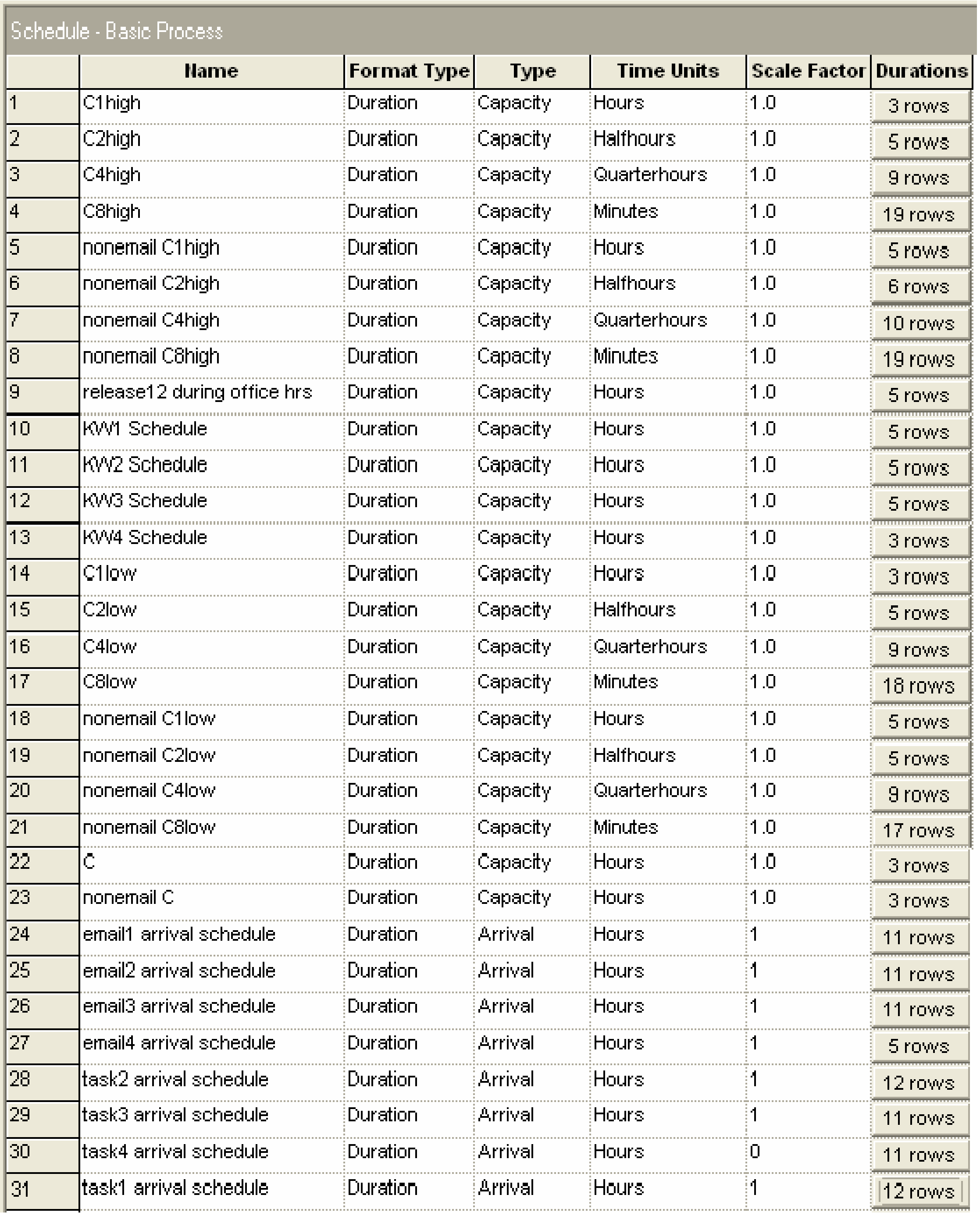


Name: implementing email strategy1
Type: Scan for Condition
Condition: (STATE(dummy1) == IDLE_RES && Process email1.WIP ==
Queue Type: Queue
Queue Name: implementing email strategy1.

Figure 5.14 Hold Block for Implementing EPS

The above condition comprises two components and determines when an email is released to the knowledge worker for processing. The first part of the condition ensures that emails are released from the “Hold” block as soon as they arrive, provided the knowledge worker is not engaged in processing other emails, referred to as work in progress (WIP). This part of the condition also continuously verifies whether the “dummy” agent is idle or not. If the state of “dummy” is idle and the WIP inventory of emails is zero, emails are released or else they wait. The second part of the condition determines how emails are released to the knowledge worker during non email-priority hours. An email is released from the “Hold” block during non priority hours only if a) the knowledge worker is not processing any primary task currently, b) the knowledge worker is not processing any previously interrupted primary task, and c) no primary task is currently held in any other queue.

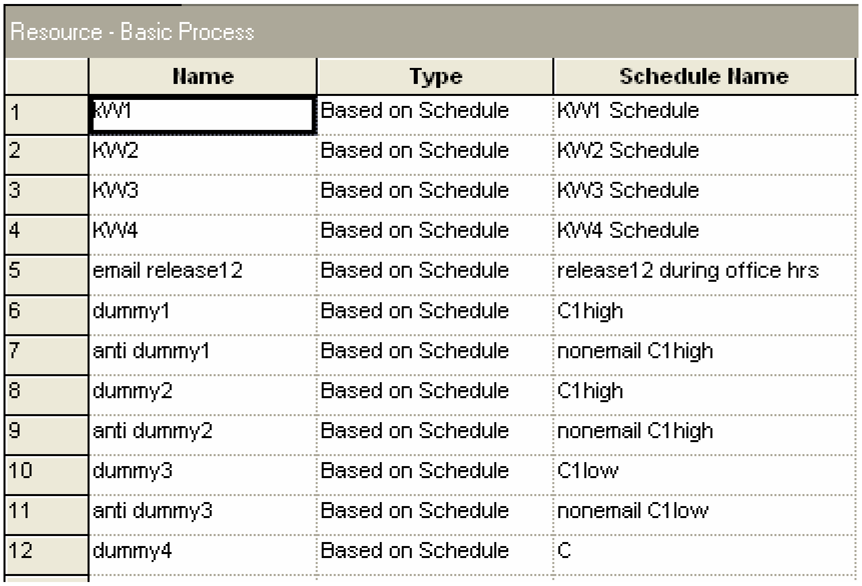
Figure 5.15 shows various schedules developed in the simulation model. Thus if the “C1 High” schedule is being used by “dummy1,” then the “non-email C1 High” schedule will be used by the corresponding “anti dummy 1.” The fourth column describes the type of a particular schedule, whether it is a capacity type or an arrival type. Schedules that were used to create entities were declared an “arrival” type, whereas all the schedules that were used to determine the availability of a particular resource were declared the “capacity” type. The fifth column describes the time units for these schedules. Knowledge workers can later pick and choose from these resources depending upon which scenario is being evaluated. For example, Fig. 5.16 shows the schedules for a XHEN network where ((C1, C1), C1) policy set is being used by the knowledge workers. The sixth column (labeled “duration”) specifies the time periods during which a resource utilizing that specific schedule is available. Although we developed schedules for several policies, only the schedules for policies that this study focuses on are reported here. Figure 5.17 (a through e) reports the schedules that were developed for C1 (High and low users of email), C4 (high and low users of email) and C policy (indifferent for high and low users of email). A value of “0” in these schedules signifies that the resource using that schedule is unavailable, whereas a value of “1” signifies that a resource is available.



Schedule - Basic Process

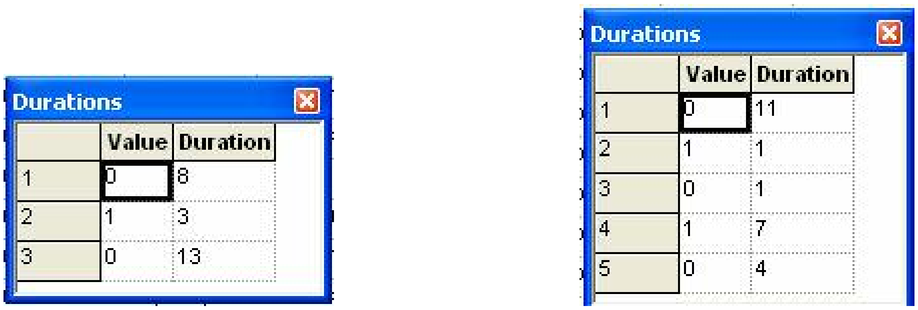
	Name	Format Type	Type	Time Units	Scale Factor	Durations
1	C1high	Duration	Capacity	Hours	1.0	3 rows
2	C2high	Duration	Capacity	Halfhours	1.0	5 rows
3	C4high	Duration	Capacity	Quarterhours	1.0	9 rows
4	C8high	Duration	Capacity	Minutes	1.0	19 rows
5	nonemail C1high	Duration	Capacity	Hours	1.0	5 rows
6	nonemail C2high	Duration	Capacity	Halfhours	1.0	6 rows
7	nonemail C4high	Duration	Capacity	Quarterhours	1.0	10 rows
8	nonemail C8high	Duration	Capacity	Minutes	1.0	19 rows
9	release12 during office hrs	Duration	Capacity	Hours	1.0	5 rows
10	KW1 Schedule	Duration	Capacity	Hours	1.0	5 rows
11	KW2 Schedule	Duration	Capacity	Hours	1.0	5 rows
12	KW3 Schedule	Duration	Capacity	Hours	1.0	5 rows
13	KW4 Schedule	Duration	Capacity	Hours	1.0	3 rows
14	C1low	Duration	Capacity	Hours	1.0	3 rows
15	C2low	Duration	Capacity	Halfhours	1.0	5 rows
16	C4low	Duration	Capacity	Quarterhours	1.0	9 rows
17	C8low	Duration	Capacity	Minutes	1.0	18 rows
18	nonemail C1low	Duration	Capacity	Hours	1.0	5 rows
19	nonemail C2low	Duration	Capacity	Halfhours	1.0	5 rows
20	nonemail C4low	Duration	Capacity	Quarterhours	1.0	9 rows
21	nonemail C8low	Duration	Capacity	Minutes	1.0	17 rows
22	C	Duration	Capacity	Hours	1.0	3 rows
23	nonemail C	Duration	Capacity	Hours	1.0	3 rows
24	email1 arrival schedule	Duration	Arrival	Hours	1	11 rows
25	email2 arrival schedule	Duration	Arrival	Hours	1	11 rows
26	email3 arrival schedule	Duration	Arrival	Hours	1	11 rows
27	email4 arrival schedule	Duration	Arrival	Hours	1	5 rows
28	task2 arrival schedule	Duration	Arrival	Hours	1	12 rows
29	task3 arrival schedule	Duration	Arrival	Hours	1	11 rows
30	task4 arrival schedule	Duration	Arrival	Hours	0	11 rows
31	task1 arrival schedule	Duration	Arrival	Hours	1	12 rows

Figure 5.15 Schedules Implemented in the Simulation Model


Resource - Basic Process

	Name	Type	Schedule Name
1	KW1	Based on Schedule	KW1 Schedule
2	KW2	Based on Schedule	KW2 Schedule
3	KW3	Based on Schedule	KW3 Schedule
4	KW4	Based on Schedule	KW4 Schedule
5	email release12	Based on Schedule	release12 during office hrs
6	dummy1	Based on Schedule	C1high
7	anti dummy1	Based on Schedule	nonemail C1high
8	dummy2	Based on Schedule	C1high
9	anti dummy2	Based on Schedule	nonemail C1high
10	dummy3	Based on Schedule	C1low
11	anti dummy3	Based on Schedule	nonemail C1low
12	dummy4	Based on Schedule	C

Figure 5.16 Resources and their Respective Schedules

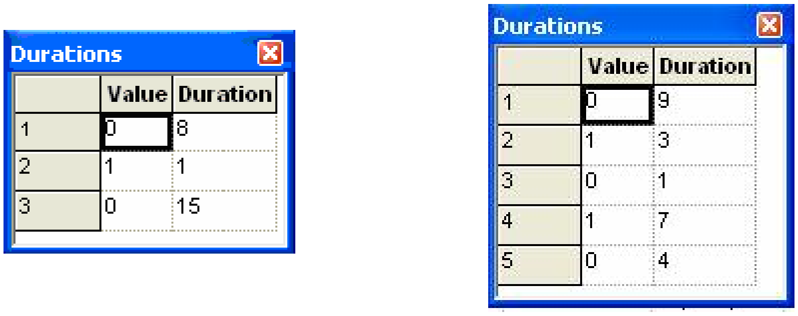

Durations

	Value	Duration
1	0	8
2	1	3
3	0	13

Durations

	Value	Duration
1	0	11
2	1	1
3	0	1
4	1	7
5	0	4

Figure 5.17(a) Schedule for C1High and Non-email C1High (duration in hours)

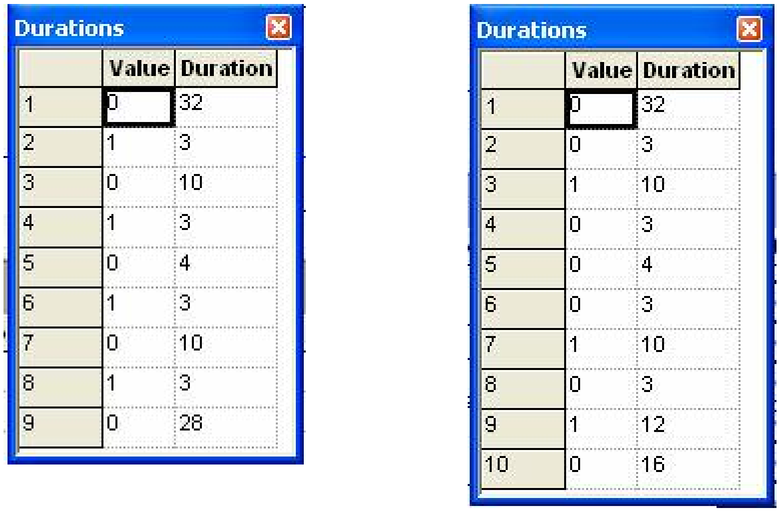

Durations

	Value	Duration
1	0	8
2	1	1
3	0	15

Durations

	Value	Duration
1	0	9
2	1	3
3	0	1
4	1	7
5	0	4

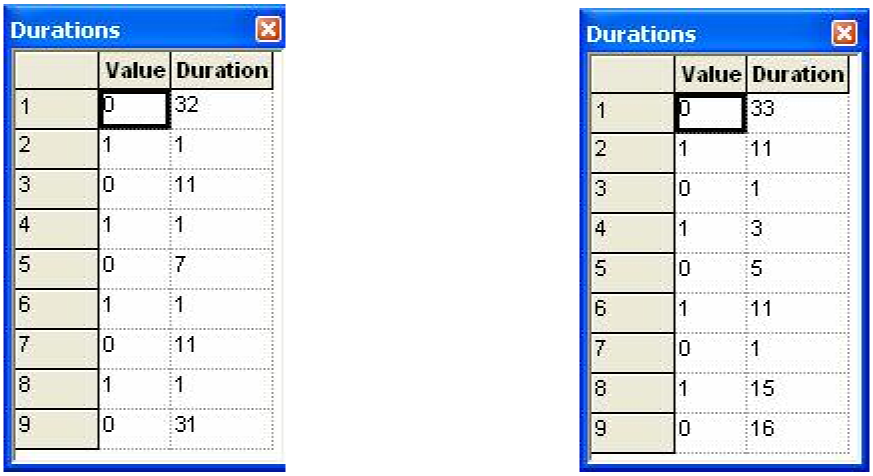
Figure 5.17(b) Schedule for C1Low and Non-email C1Low (duration in hours)



	Value	Duration
1	0	32
2	1	3
3	0	10
4	1	3
5	0	4
6	1	3
7	0	10
8	1	3
9	0	28

	Value	Duration
1	0	32
2	0	3
3	1	10
4	0	3
5	0	4
6	0	3
7	1	10
8	0	3
9	1	12
10	0	16

Figure 5.17(c) Schedule for C4High and Non-email C4High (duration in quarter hours)



	Value	Duration
1	0	32
2	1	1
3	0	11
4	1	1
5	0	7
6	1	1
7	0	11
8	1	1
9	0	31

	Value	Duration
1	0	33
2	1	11
3	0	1
4	1	3
5	0	5
6	1	11
7	0	1
8	1	15
9	0	16

Figure 5.17(d) Schedule for C4 Low and Non-email C4 Low (duration in quarter hours)

	Value	Duration
1	1	12
2	0	1
3	1	11

	Value	Duration
1	0	8
2	0	12
3	0	4

Figure 5.17(e) Schedule for C and Non-email C (duration in hours)

5.2.3 Implementation of Email Flow in the Network

Emails were created in the same way primary tasks were created using the "Create" module. Four different "Create" modules were used, one for each of the four knowledge workers. Fig. 5.18 shows the flow of emails in the initial part of the model where they are created and the original value is assigned to each email. We assumed that creating emails is a need-driven process, i.e. a knowledge worker will originate an email when it is needed and therefore creating one does not result in the knowledge worker's interruption and hence no penalty occurs when emails are created. In other words, the model assumes interruptions only for the receiver and no interruptions for the sender. Creating an email can occur when the knowledge worker takes a natural break from working on primary tasks, so processing emails during this time does not cause any interruption. Fig. 5.19 shows one such "Create" module that was used to create emails for the 1st knowledge worker.

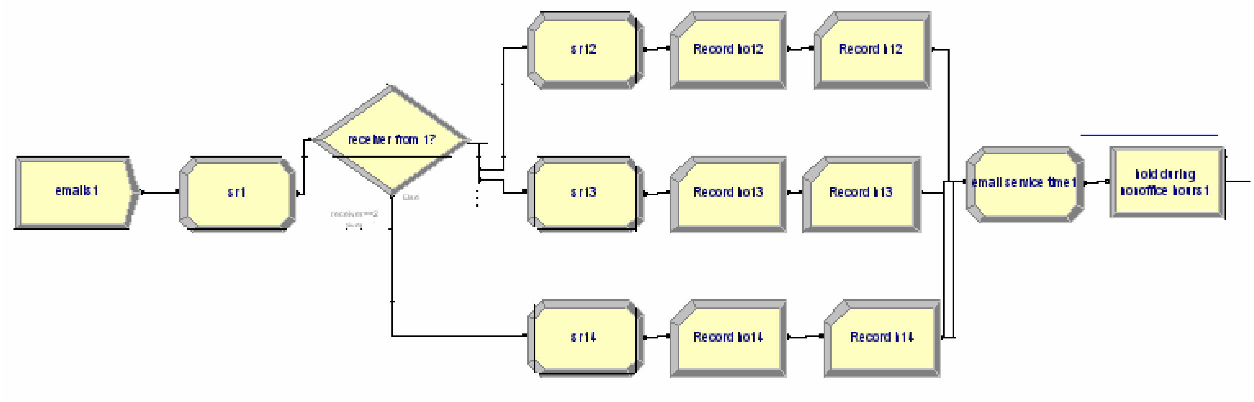
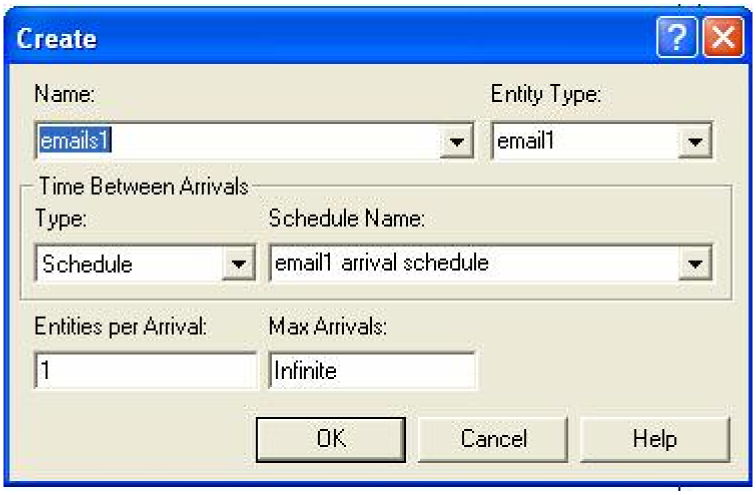


Figure 5.18 Email Creation and Initial Value-Effectiveness Calculation

As described in the previous section, a separate schedule was created for emails being originated at each knowledge worker's end. The "Create" block in Fig. 5.19 shows the “email1 arrival schedule” for emails created by knowledge worker 1. These schedules use exponential hourly creation rates. Fig. 5.20 (a and b) shows these creation rates. Emails are created only during the day time in Fig 5.20 (a) since the three knowledge workers belonging to the network do not work at night, but emails are created by the outside world throughout the day as shown in Fig. 5.20 (b), since knowledge workers outside the network may be located in different time zones around the world. Each of these schedules starts at midnight and is repeated at the beginning of another day. Knowledge workers are not available to process emails during lunch hours.



Name: emails1
Entity Type: email1
Time Between Arrivals
Type: Schedule
Schedule Name: email1 arrival schedule
Entities per Arrival: 1
Max Arrivals: Infinite

Figure 5.19 Create Block for Emails

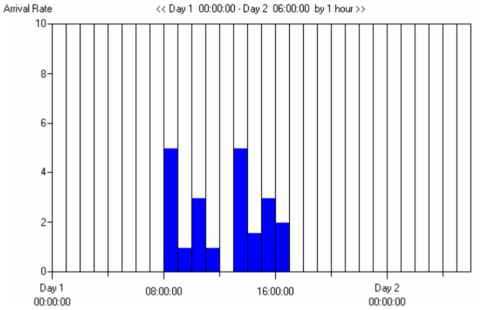


Figure 5.20 (a) Arrival Schedule for KW 1/2/3

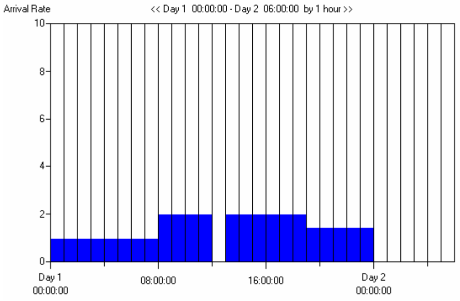
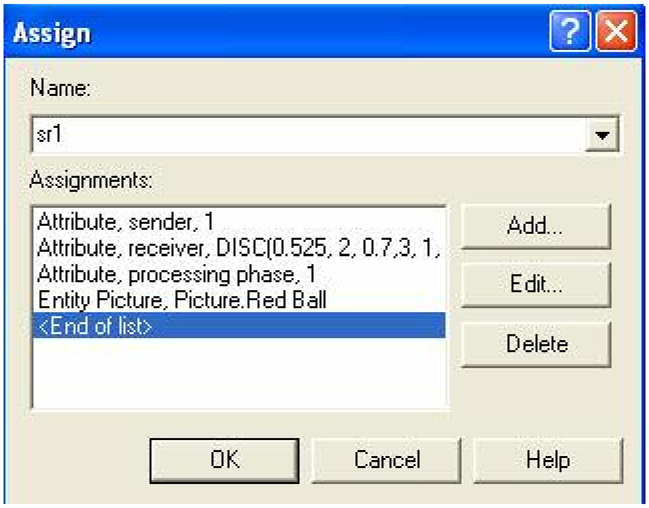


Figure 5.20 (b) Arrival Schedule for Outside World (KW 4)

Soon after the emails are created, they arrive at the “Assign” block labeled “sr1,” where several different attributes are assigned to them. Fig. 5.21 shows the “Assign” block for knowledge worker 1. Each email arriving at this block is stamped with an attribute that holds the name of the original sender. The second attribute uses a discrete probability function to determine its receiver. For example, in the XHEN network, the probability that KW 1 sends a new email to KW 2 is 0.525, to KW 3 is 0.175 and to somebody in the outside world (i.e. KW 4) is 0.3. These probabilities are based on the values derived in section 4.2. Each email is also stamped with information about its potential receiver. Although knowledge workers often send emails to themselves in the real world, knowledge workers in the model do not send emails to themselves. Emails used for self-reminders, a to-do list, etc. can easily be avoided using online calendars and other electronic gadgets. The third attribute tags the information about the current processing stage of the email.



Name:
sr1
Assignments:
Attribute, sender, 1
Attribute, receiver, DISC(0.525, 2, 0.7,3, 1,
Attribute, processing phase, 1
Entity Picture, Picture.Red Ball
<End of list>

Figure 5.21 Assign Block for Stamping Sender and Receiver Attributes

Once an email is stamped with these attributes, it reaches a “Decision” block where a check is made on each email to determine the receiver of that email. This “Decision” block, based upon the receiver information, routes the email to one of the three branches shown in Fig. 5.18. Each branch handles emails heading out to one of the receivers. Another “Assign” block determines what percentage of emails for a particular sender-receiver pair are of type 1, type 2 and type 3. The type of each email is then assigned as an attribute of each email that passes through. Also, the initial value-effectiveness of an email is recorded. The three branches converge at a point from where all the emails enter a “Decision” block as shown in Fig. 5.22. This “Decision” block checks on the processing phase of an email. If the processing phase is “1,” i.e. new email, than the email is routed to the top branch where processing begins one by one, but if the processing phase is “2” or “3,” emails are routed to the lower branch for their next phase of processing. Emails are released by the “Hold” block implementing EPS and then reach the “Preempt” block. As soon as emails arrive at this block, the knowledge worker is preempted from processing primary tasks and the remaining time is stored in an internal variable referred to as “RT.” Emails are then processed at the next “Process” module. The time that an email spends here comes from a two-dimensional matrix and depends upon the type of email and the current processing phase of the email. This time is treated as value-added time by the system. Once the processing on the email in the part shown in Fig 5.22 is finished, it is taken to a “Branch” block similar to the one shown in Fig. 5.23. This block segregates the email traffic pertaining to each sender/receiver pair and transfers it to any one of the three Branch blocks. For example, for an email sent by KW 1 (sender), the possible recipients are KW2 or KW3, or KW4. The first “Branch” block branches out the emails into any one of the three categories: (KW1-KW2), (KW1-KW3), (KW1-KW4). The next destination of emails is determined by the conditions implemented in these three “Branch” blocks, which also perform the routing function.

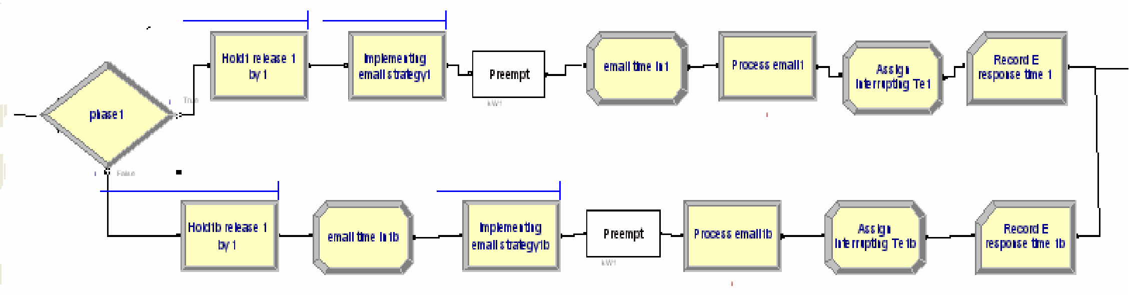
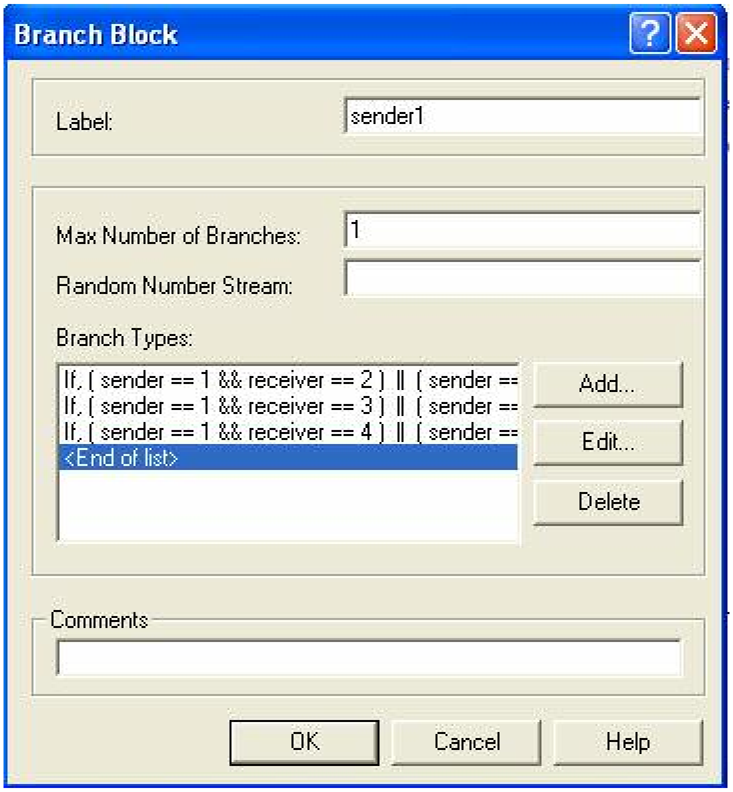


Figure 5.22 Flows of New and Arriving Emails



Label: sender1

Max Number of Branches: 1

Random Number Stream:

Branch Types:

If, (sender == 1 && receiver == 2) || (sender ==
If, (sender == 1 && receiver == 3) || (sender ==
If, (sender == 1 && receiver == 4) || (sender ==
<End of list>

Comments

Figure 5.23 Branch Block Splitting the Emails According to Receiver

Once the email in a particular phase has been processed, the attribute storing the phase information is updated. For example, once the 1st phase of processing on an email is over, the attribute storing the current phase information of the email is updated to 2 before another knowledge worker begins to work on it. If the email was in 2nd phase, it is updated to phase 3. However, only emails of type 1 and 2 are able to reach the 3rd phase. Once an email of type 3 reaches the 2nd stage of processing, the "Branch" block detects that and exits the email out of the system. Before that happens, the email is routed to a part of the model where several email statistics such as time spent by the email in the system, final value-effectiveness, etc. are evaluated. One such component of the model is shown in Fig. 5.24. For each knowledge worker within the network, six such components were created.

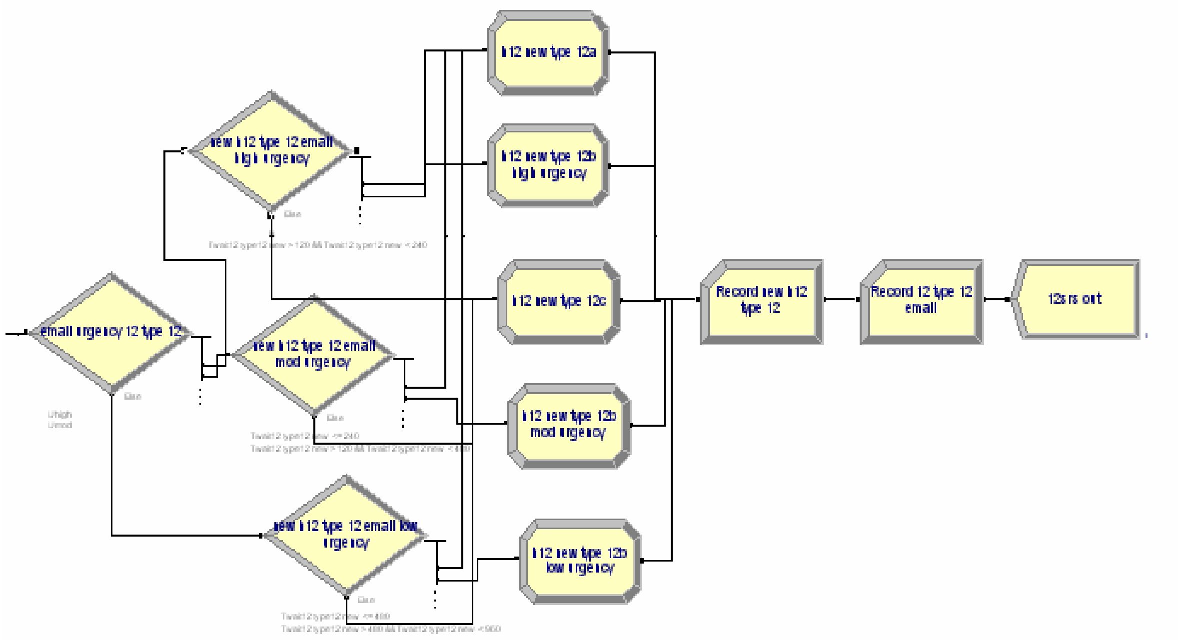


Figure 5.24 Calculation of Final Value-Effectiveness and Other Statistics

5.3 Model Validation

The model went through a rigorous testing and debugging phase. Several strategies were employed to ensure the internal validity of these models. Although, it is difficult to reproduce a complete list of all the validation activities, examples of some of them follow:

1. Several test schedules were created for both types of emails as well as primary task arrivals to ensure that the model remained stable under different load conditions. Different components of the model were tested separately. For example, new email and primary task arrivals were turned on while the new task and email arrival schedules of all other knowledge workers were turned down to make to make sure that the model was behaving in the manner it should. While we did this, all the other three knowledge workers remained active and worked as usual.
2. Initial testing revealed that the model was not calculating recall time correctly. Much testing, revealed that the advanced panel in Arena was not calculating the required expressions. So we performed this calculation alternatively through the use of an "Assign" block.
3. We tested the model to make sure that for each type of entity, the number in was equal to the number out.
4. We checked to make sure that email and task arrival schedules were actually implemented correctly. One of the methods for checking implementation was to calculate the overall utilization of the resource and then calculate the split-utilizations (i.e. utilizations for email and primary tasks) of the same resource to make sure that the proportion of time spent on each type of entity was actually what we intended it to be.

5. To validate the parameters and other entity creation rates, we turned off the primary task arrivals and kept only the flow of email on at first with the scale factor of 1.0. Then we turned off the email arrival rate and kept the arrivals of primary tasks on at a scale factor of 1.0. We checked the utilization in both cases and added them up. This gives the utilization within an environment where there are no interruptions. Next we ran the simulation with both the creation rates on. This gave us the utilization for the scenario where interruptions occur i.e. a penalty was paid in terms of time. Subtracting the two gave us the increase in utilization that occurred due to interruptions. Multiplying this difference by the number of hours for which a knowledge worker works per day gave us the total time lost per day due to interruptions.

6. Different internal and external queues at "Process" and "Hold" blocks were inspected to identify which one was making the entire system unstable.

7. We compared the individual output statistics of all four knowledge workers to figure out whether any one was behaving oddly.

5.4 Warm-up Time, Run Length and Number of Replications

Warm-up time for the simulation model was determined externally using Welch's graphical method (Welch 1983). We used email response time statistics of a HEN-High network using the ((C, C), C) policy configuration, i.e. scenario # 10. We ran the simulation model for this particular scenario for 5 replications and found the average email response time across replications for each day. To smooth the curve, moving average for 15 data points was taken. Fig. 5.25 shows the plot of the moving average of email response time versus the number of days.

The data did not show signs of very large variability, so after eye-balling the graph carefully, we noticed that on or around the 30th day, the curve started to become stable. Hence, 30 days was chosen as the warm-up time for all simulation models. All the data generated during the first 30 days of the simulation run was therefore deleted.

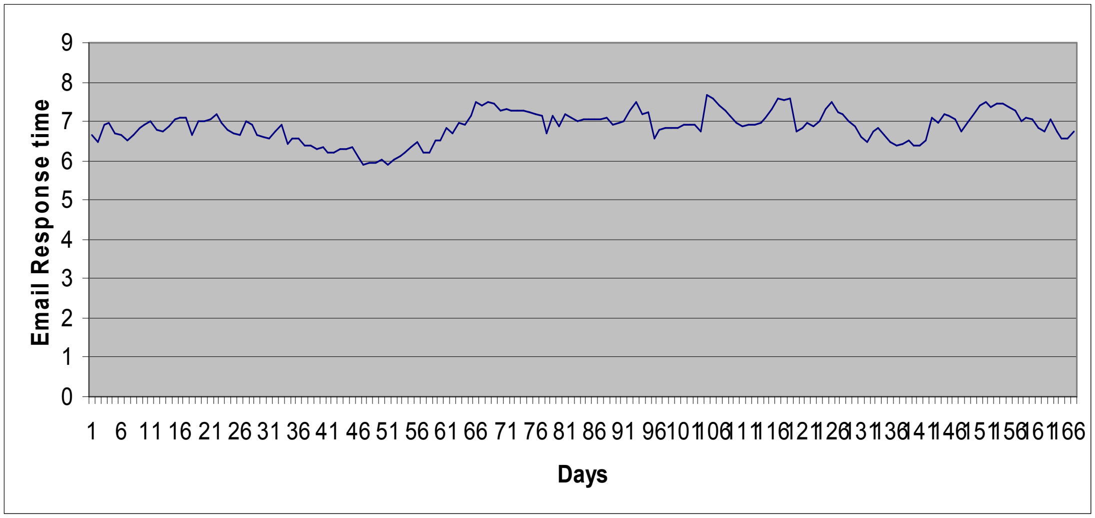


Figure 5.25 Plot of Moving Average of Email Response Time versus Number of Days

Next we needed to know the time duration for which to run each simulation model. We ran the same scenario for four different run lengths: 150 days, 210 days, 270 days and 390 days and recorded the output statistics on the average task completion time. We did not find a large variation in the output statistics with the increase in the run length. For example, we found that the average task completion time for KW 2 with run lengths of 150 days, 210 days, 270 days and 390 days turned out to be 194.5 min, 196.7 min, 198.52 min, and 200.97 min respectively. Hence, in order to be efficient and at the same time collect enough data, we chose 210 days as the run length for all our simulation models.

Finally, we needed to calculate the number of replications for which each simulation model would run. Since all the parameters in the model are probability distributions, it is not a good practice to use data from just one replication. In order to calculate the appropriate number of replications, we applied the replication truncation method using the relative precision approach (Chung 2003). We used the desired relative precision value of 9% as reported in Law and Kelton (2000). To begin, we chose a random value, 20, for the number of replications and used the same scenario to collect the output statistics on email response time for each replication using the replication length of 210 days with a warm-up time of 30 days. The data collected was used to evaluate the relative precision obtained using the formula reported in Chung (2003) as follows:

$$\text{Relative Precision} = \frac{\left(t_{1-\alpha/2,n-1} \times std.dev\right)}{\sqrt{n} \times \overset{=}{x}}$$

(40)

Where, $t_{1-\alpha/2,n-1}$ is t distribution for n-1 degrees of freedom and $1-\alpha/2$

$std.dev$ is the standard deviation for replication mean

n is the number of replications

$\overset{=}{x}$ is the mean of the replication mean

Table 5.1 summarizes the method. The relative precision in the output statistics was found to be 2.7%, which is acceptable. Hence, our assumed value for “n” is acceptable and each simulation model be run for 20 replications.

Table 5.1 Replication Truncation Method

Replication Truncation Method (Using relative precision Approach)	
Run length for each replication	210 days
Warm up time	30 days
Desired relative precision (As reported in Law & Kelton (2000))	0.09
No of replications needed, n	20
Degree of freedom (d.f)	19
t-value for alpha=0.025, d.f=19	2.09302405
square root of n	4.472135955
Replication mean for email response time statistics (in minutes)	
xbar1	10.128
xbar2	10.019
xbar3	10.948
xbar4	10.54
xbar5	9.6672
xbar6	9.9642
xbar7	10.882
xbar8	11.248
xbar9	10.224
xbar10	11.383
xbar11	10.507
xbar12	9.4622
xbar13	9.7674
xbar14	10.543
xbar15	10.531
xbar16	11.275
xbar17	9.5085
xbar18	9.5396
xbar19	10.647
xbar20	10.056
$\overset{=}{x}$	**10.342005**
Std. deviation of the replication mean	0.605072074
Relative precision attained	**0.027381774**

6. RESULTS AND DISCUSSION

Data collected from the simulation experiments were analyzed using multivariate analysis of variance to understand the differences between performances of various policy groups. We begin by explaining the results pertaining to the first research question, which focuses on the impact of different email policies on various performance measures across different network types.

Tables 6.1 and 6.2 show the results of multivariate tests conducted using the MANOVA technique. At the significance level of 0.05, all interaction effects were significant for all performance variables except value-effectiveness, which implies that email policies had significantly different effects on time-effectiveness, task completion time, and email response time across different types of networks but not on value-effectiveness.

Hypothesis H1a was supported at $\alpha=0.05$. The difference between the impact of arrhythmic email policy and rhythmic email policy on value-effectiveness was statistically significant for HEN as well as XHEN networks. Figure 6.1 explains these differences graphically and gives us a sense of the directionality of the difference. For HEN networks, arrhythmic policies provided an average value-effectiveness of 97.11 percent, while rhythmic policies provided a value-effectiveness of 93.85 percent. For XHEN networks, this difference was relatively small. In such networks, arrhythmic policies provided an average value effectiveness of 95.54 percent, whereas rhythmic policies were 93.26 percent effective in terms of value to the organization.

Table 6.1 Tests of Between-Subjects Effects For All Factors

Source	Dependent Variable	Type III Sum of Squares	df	Mean Square	F	Sig.
Corrected Model	Value- effectiveness	1144.580(b)	3	381.527	36.294	.000
	Time effectiveness	3222.972(c)	3	1074.324	28.780	.000
	Avg. Task completion time	31356.752(d)	3	10452.251	18.017	.000
	Email Response Time	43182.821(e)	3	14394.274	87.644	.000
Intercept	Value-effectiveness	5366445.953	1	5366445.953	510503.7	.000
	Time-effectiveness	4777620.768	1	4777620.768	127987.0	.000
	Avg. Task completion time	12360361.31	1	12360361.31	21306.10	.000
	Email Response Time	455700.543	1	455700.543	2774.678	.000
Network Type	Value-effectiveness	172.589	1	172.589	16.418	.000
	Time-effectiveness	50.634	1	50.634	1.356	.245
	Avg. Task completion time	827.166	1	827.166	1.426	.233
	Email Response Time	3265.441	1	3265.441	19.883	.000
Policy Type	Value-effectiveness	1144.257	1	1144.257	108.852	.000
	Time-effectiveness	2628.678	1	2628.678	70.419	.000
	Avg. Task completion time	19288.720	1	19288.720	33.249	.000
	Email Response Time	14727.607	1	14727.607	89.674	.000
Network Type* Policy Type	Value-effectiveness	35.079	1	35.079	3.337	.068
	Time-effectiveness	548.429	1	548.429	14.692	.000
	Avg. Task completion time	3303.067	1	3303.067	5.694	.017
	Email Response Time	11682.772	1	11682.772	71.134	.000
Error	Value-effectiveness	7947.118	756	10.512		
	Time-effectiveness	28220.672	756	37.329		
	Avg. Task completion time	438580.079	756	580.132		
	Email Response Time	124162.022	756	164.235		
Total	Value-effectiveness	6798367.497	760			
	Time-effectiveness	6252331.721	760			
	Avg. Task completion time	15721480.81	760			
	Email Response Time	885068.843	760			
Corrected Total	Value-effectiveness	9091.698	759			
	Time-effectiveness	31443.644	759			
	Avg. Task completion time	469936.831	759			
	Email Response Time	167344.842	759			

a Computed using alpha = .05
b R Squared = .126 (Adjusted R Squared = .122)
c R Squared = .102 (Adjusted R Squared = .099)
d R Squared = .067 (Adjusted R Squared = .063)
e R Squared = .258 (Adjusted R Squared = .255)

Table 6.2 Multivariate Tests(c)

Effect		Value	F	Hypothesis df	Error df	Sig.
Intercept	Pillai's Trace	1.000	989777.365(b)	4.000	753.000	.000
	Wilks' Lambda	.000	989777.365(b)	4.000	753.000	.000
	Hotelling's Trace	5257.7	989777.365(b)	4.000	753.000	.000
	Roy's Largest Root	5257.7	989777.365(b)	4.000	753.000	.000
Network Type	Pillai's Trace	.160	35.755(b)	4.000	753.000	.000
	Wilks' Lambda	.840	35.755(b)	4.000	753.000	.000
	Hotelling's Trace	.190	35.755(b)	4.000	753.000	.000
	Roy's Largest Root	.190	35.755(b)	4.000	753.000	.000
Policy Type	Pillai's Trace	.136	29.511(b)	4.000	753.000	.000
	Wilks' Lambda	.864	29.511(b)	4.000	753.000	.000
	Hotelling's Trace	.157	29.511(b)	4.000	753.000	.000
	Roy's Largest Root	.157	29.511(b)	4.000	753.000	.000
Network Type * Policy Type	Pillai's Trace	.125	27.003(b)	4.000	753.000	.000
	Wilks' Lambda	.875	27.003(b)	4.000	753.000	.000
	Hotelling's Trace	.143	27.003(b)	4.000	753.000	.000
	Roy's Largest Root	.143	27.003(b)	4.000	753.000	.000

a Computed using alpha = .05
b Exact statistic
c Design: Intercept+ Network Type+ Policy Type + Network Type* Policy Type

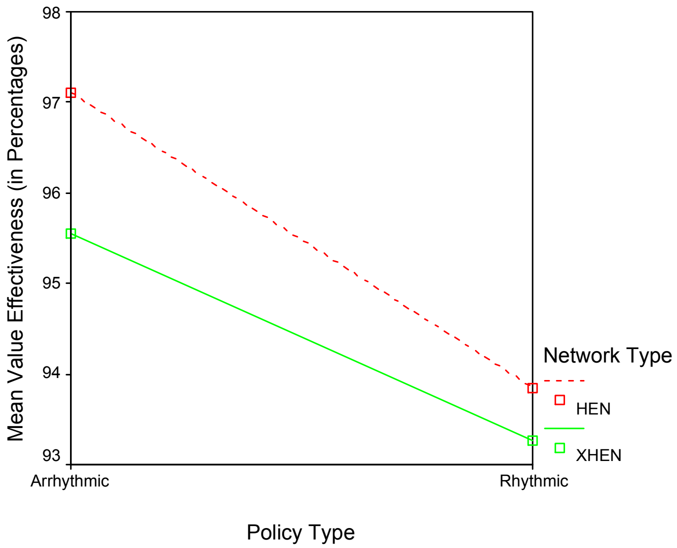


Figure 6.1 Impact of Policy Type on Value-Effectiveness across Different Networks

This trend can be easily explained. The value that an organization derives depends on how quickly an email is processed after it has arrived in an inbox. Since the value of an email diminishes with the passage of time, a higher cumulative value is derived from an email that is resolved quickly, whereas a lower value is derived from an email that has to wait for a relatively longer period of time. A network that uses any arrhythmic policy typically has a larger number of knowledge workers using continuous email processing strategies, which also implies that such emails are resolved quickly leading to a higher value-effectiveness.

Another reason such a pattern is exhibited with the use of different email policies is the presence of variety in the email policies being used by knowledge workers. Networks employing arrhythmic polices have a relatively higher variety of email policies than those using rhythmic policies. Arrhythmic policies are different not only within groups but also across groups. For example, a network using an arrhythmic policy having a ((C1, C4), C) configuration has policies that are different within the group as well as different across groups. It exhibits more variety. On the other hand, networks employing rhythmic policies exhibit a lower degree of variety. Rhythmic-same policies have almost negligible variety since email processing strategies are the same both within and across the groups belonging to the network. Rhythmic–different policies have a relatively higher degree of variety than rhythmic-same but always less than any arrhythmic policy. For example, a network using a ((C1, C1), C1) set of policies has almost the same email schedules both within and across the groups, whereas a network using ((C1, C1), C4) has the same schedule within the group but different across groups.

There is a strong relationship between the degree of variety in email policies used within any network and email resolution time. Although email arrival rates and service times have stochastic characteristics, knowledge workers follow specific policy schedules such as C1, C4, C, etc. when it comes to processing them. The greater the degree of variety of email policies, the higher the probability that emails will have to wait for a short time before they get processed. For example, in a network using a ((C1, C1), C1) set of policies, if an email remains unattended after email-priority hours are over, it will have to wait till the next day before it gets resolved. This potentially increases the resolution time, leading to a decrease in the value being generated for the network. However, if a ((C1, C4), C) set of policies is used, then there is a very high probability that an email sent by a worker with a C1 schedule will be responded to by a worker using a C schedule on the same day. This leads to a reduction in the email wait time and thereby an increase in the value-effectiveness.

We found strong support for hypothesis H1b at $\alpha=0.05$; i.e. rhythmic policies lead to time-effectiveness that is statistically different from and higher than the time-effectiveness of the arrhythmic policies. The pattern exhibited for time-effectiveness is opposite to what we saw for value-effectiveness, as shown in Fig 6.2. Arrhythmic policies provided an average time-effectiveness of 86.23 percent in HEN networks, whereas rhythmic policies provided a time effectiveness of 92.35 percent. A similar trend was observed in XHEN networks; the difference was slightly smaller but still significant. In these networks, arrhythmic policies provided a time-effectiveness of 88.73 percent, whereas rhythmic policies were 91 percent time effective. These results are not surprising. In arrhythmic policies, emails are answered quickly due to the presence of a larger number of workers using continuous policies. As a result, the number of interruptions that take place becomes large as well. Thus the cumulative effect of all the interruptions leads to a greater wastage of time. For example, in a network employing a ((C, C), C) set of policies, emails are processed as soon as they arrive, leading to frequent interruptions of the knowledge workers in their primary tasks. This reduces the time effectiveness in this policy. Fig. 6.3 (a, b, and c) shows that the value of time effectiveness for HEN-High networks is 71.55 percent, for HEN-Low networks is 90.14 percent and for XHEN networks is 77.12 percent. In a network employing a ((C1, C1), C1) set of policies, emails are processed in batches and for this reason, usually the interruptions are fewer.

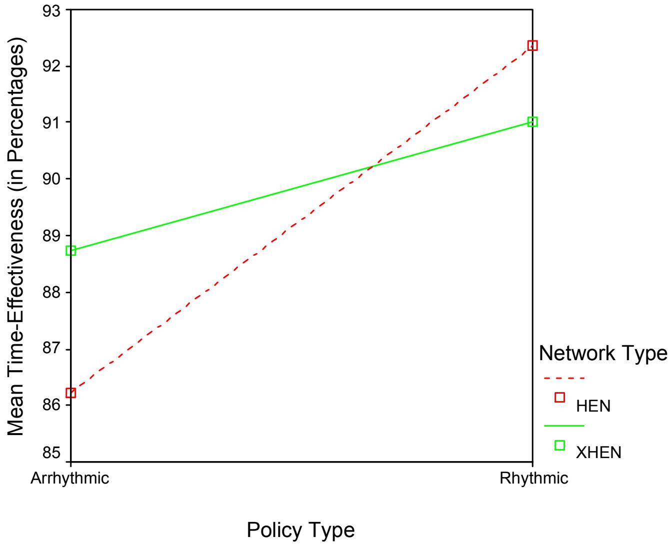


Figure 6.2 Impact of Policy Type on Time-Effectiveness across Different Networks

The above finding is further confirmed by Fig. 6.3 (a, b, and c). For a ((C1, C1), C1) policy set, time-effectiveness is 97.27 percent for HEN-High networks, 99.22 percent for HEN-Low networks, and 97.8 percent for XHEN networks. This explains why rhythmic policies are more effective time-wise than arrhythmic polices. Fig. 6.3 (a, b and c) shows the performance of each policy belonging to rhythmic as well as arrhythmic categories with respect to value and time effectiveness for different types of networks: HEN-High, HEN-Low and XHEN.

However, an important observation to be noted here is the contrary patterns exhibited by value and time effectiveness curves for all types of networks in Fig. 6.3 (a, b, and c). As we move to the right, the value-effectiveness increases but the time-effectiveness decreases. Thus the policies found on the extremes either provide high time-effectiveness and low value-effectiveness or they provide low time-effectiveness and high value-effectiveness. The policies in the middle provide moderate levels of both value and time-effectiveness.

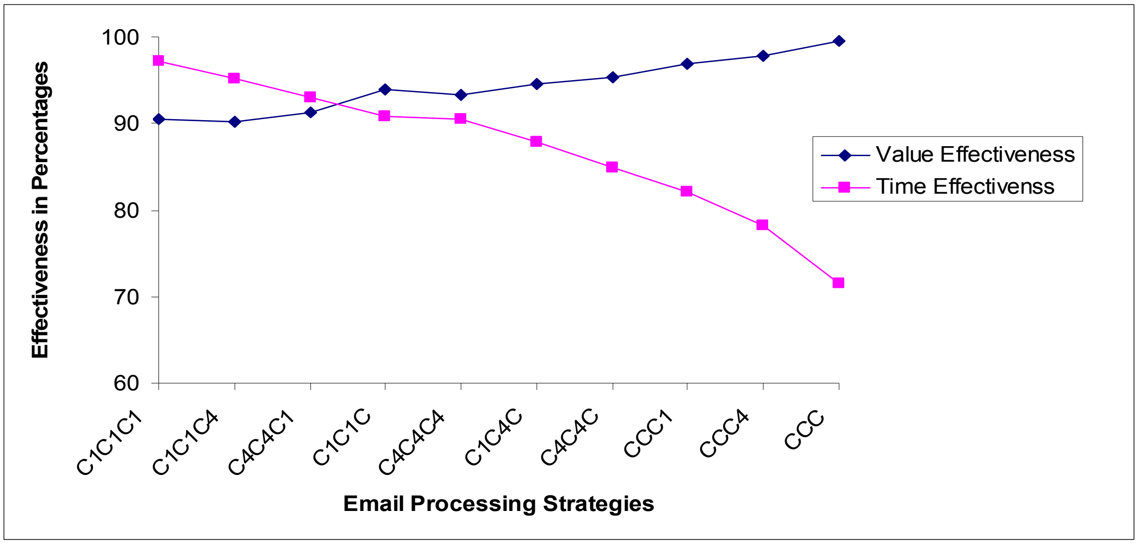


Figure 6.3(a) Value Effectiveness and Time Effectiveness for HEN-High Network

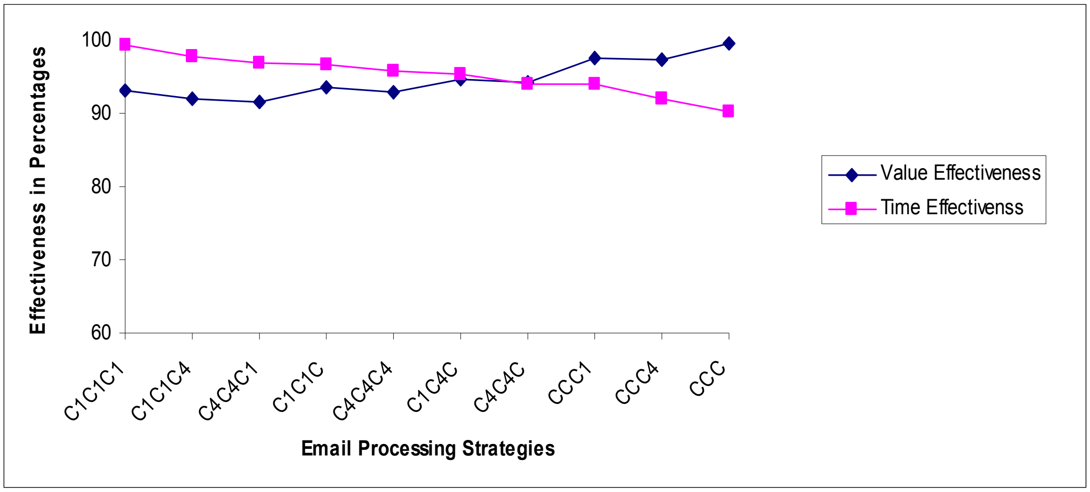


Figure 6.3(b) Value Effectiveness and Time Effectiveness for HEN-Low Network

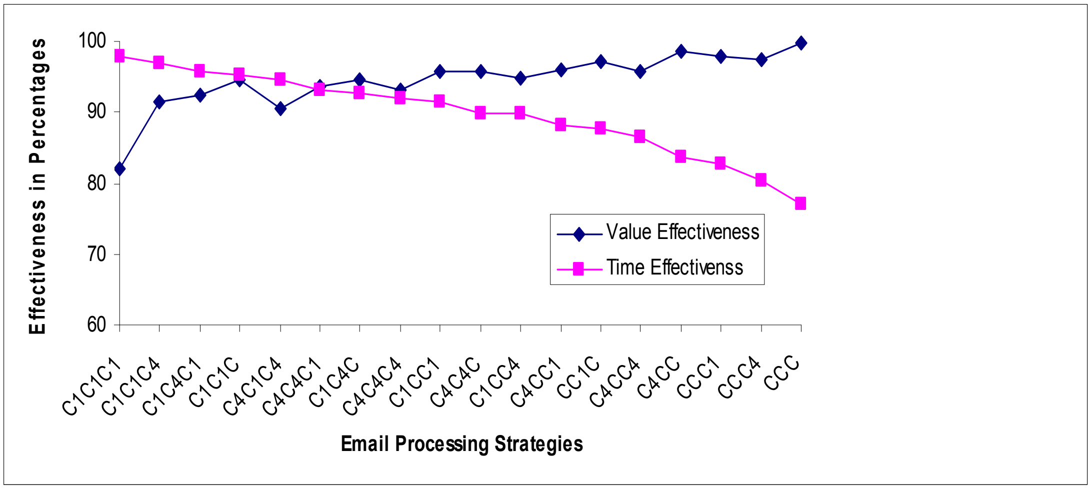


Figure 6.3(c) Value Effectiveness and Time Effectiveness for XHEN Network

We found strong support for hypothesis H1c at α=0.05; i.e. there were statistically significant differences in task completion times between arrhythmic and rhythmic policies. However, the directionality needs to be explained. Fig 6.4 shows that in HEN networks, the average task completion time encountered with arrhythmic policies is 150.96 min whereas with rhythmic policies, it was 134.86 min. In XHEN networks, average task completion time was 148.8 min for arrhythmic policies but 141.9 min for rhythmic policies. This implies that task completion time on an average was lower for rhythmic policies in comparison to arrhythmic policies. These results are consistent with our expectations. In rhythmic policies, more emails are processed using scheduled policies leading to fewer interruptions; hence, a knowledge worker gets more time to catch up on task processing. On the other hand, email processing lags behind. This is evidenced by the task completion statistics reported in Fig. 6.5 (a and b). We see relatively smaller task times in several rhythmic policies.

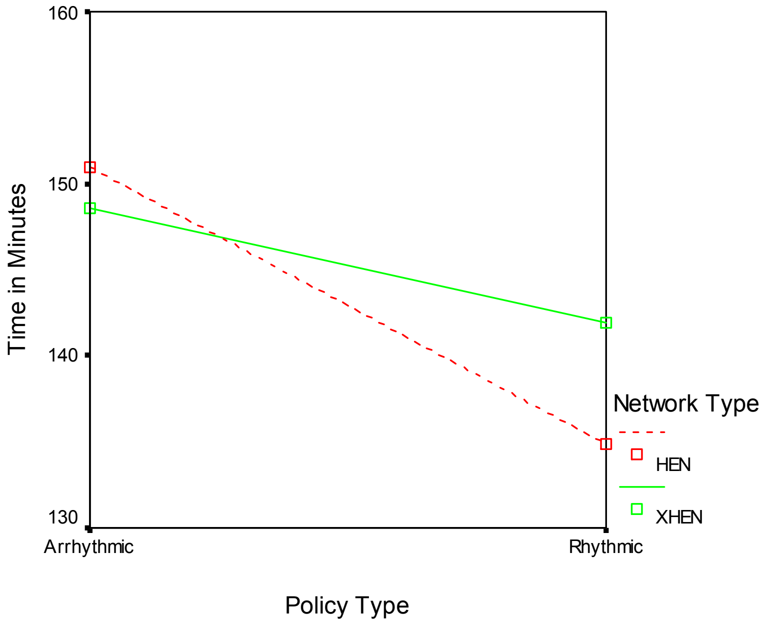


Figure 6.4 Impact of Policy Type on Task Completion Time across Different Networks

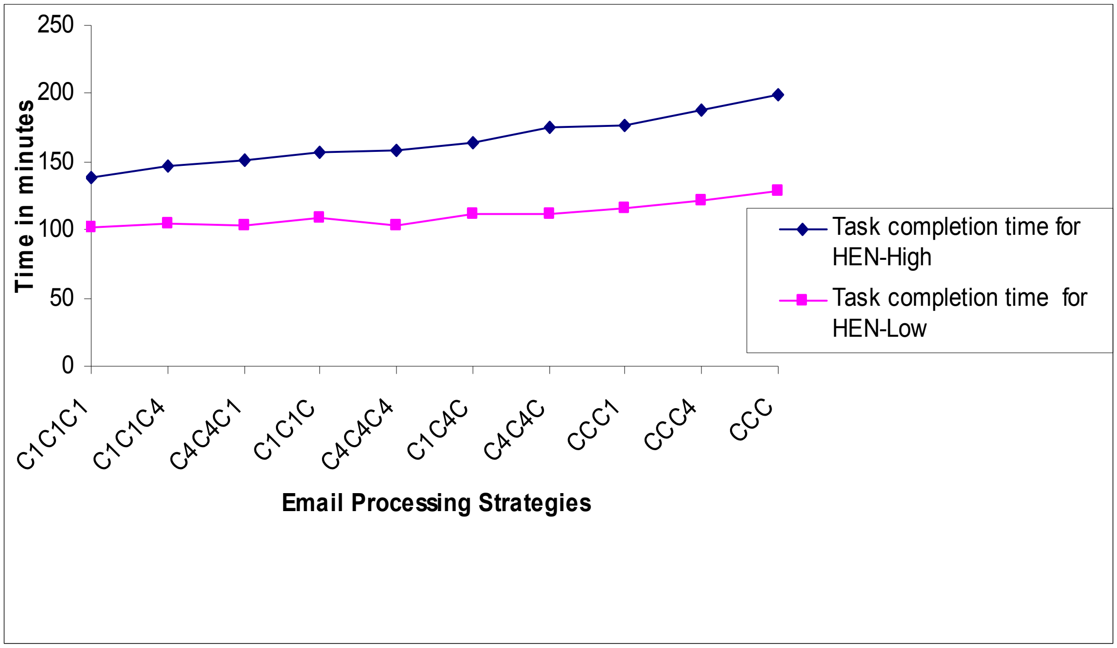


Figure 6.5(a) Average Task Completion Time (in minutes) for HEN-High and HEN-Low Networks

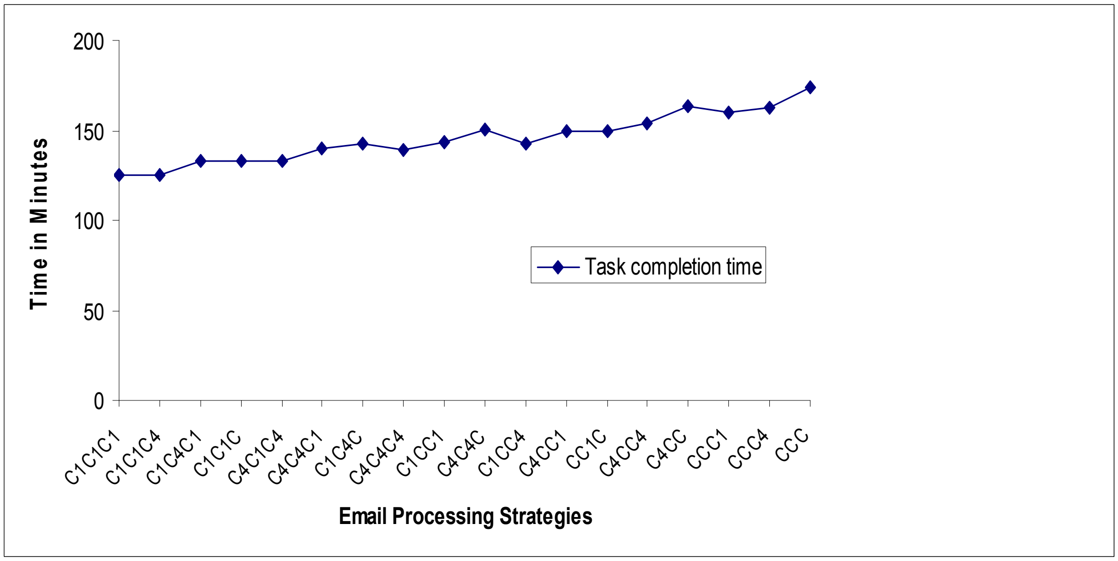


Figure 6.5(b) Average Task Completion Time (in minutes) for XHEN Network

For example, a ((C1, C1), C1) policy has a task completion time of 138.05 minutes in HEN-High networks, 101 minutes in HEN-Low networks, and 124 minutes in

XHEN networks. On the other hand, for an arrhythmic policy such as ((C, C,) C), we see higher task completion times: 198.95 minutes in HEN-High networks, 129 minutes in HEN-Low networks, and 173 minutes in XHEN networks.

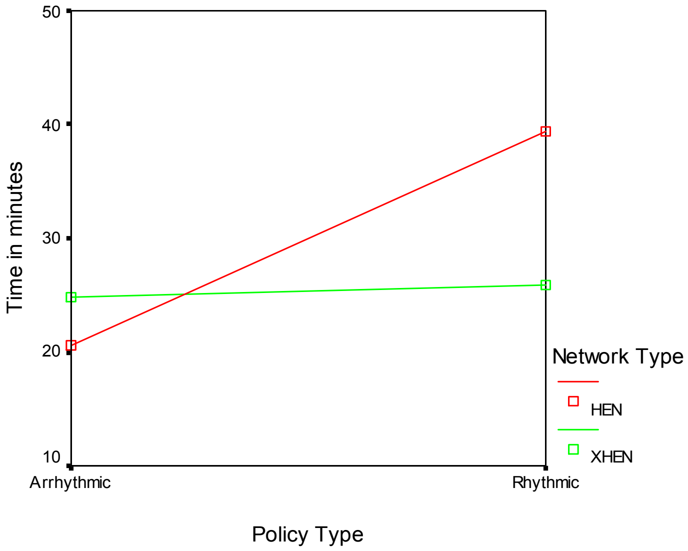


Figure 6.6 Impact of Policy Type on Email Response Time across Different Networks

Finally, we also found support for H1d hypothesis at α=0.05. The average email response time for various types of rhythmic policies was significantly higher than the times for arrhythmic policies. Fig 6.6 also suggests this directionality. For example, in HEN networks, knowledge workers took on average 20.6 minutes to respond to an email under an arrhythmic policy whereas they took 39.41 minutes to respond to an email if a rhythmic policy was used. The main reason for such a pattern is again the number of interruptions that occur in each type of policy. For HEN networks, the average number of interruptions that occurred with the use of arrhythmic policies was 36.19 per day while the number was approx. 22 per day with the use of any rhythmic policy. With XHEN networks, this difference was approx. 7 interruptions per day. As the number of interruptions increases, work on primary tasks starts to lag but more and more emails are processed quickly. This leads to a reduction in the amount of time that they have to wait, resulting in shorter email response time and Fig. 6.7 (a and b) further clarifies this. We see a decrease in the email response time as we proceed towards the right in all three types of networks. As expected, response time is minimum with the use of ((C, C), C) but is maximum at the other extreme, i.e. with the use of a ((C1, C1), C1) policy. Fig 6.7 (b) shows a greater variability in email response time for different policies. This difference is mainly due to policies not being arranged in proper sequence.

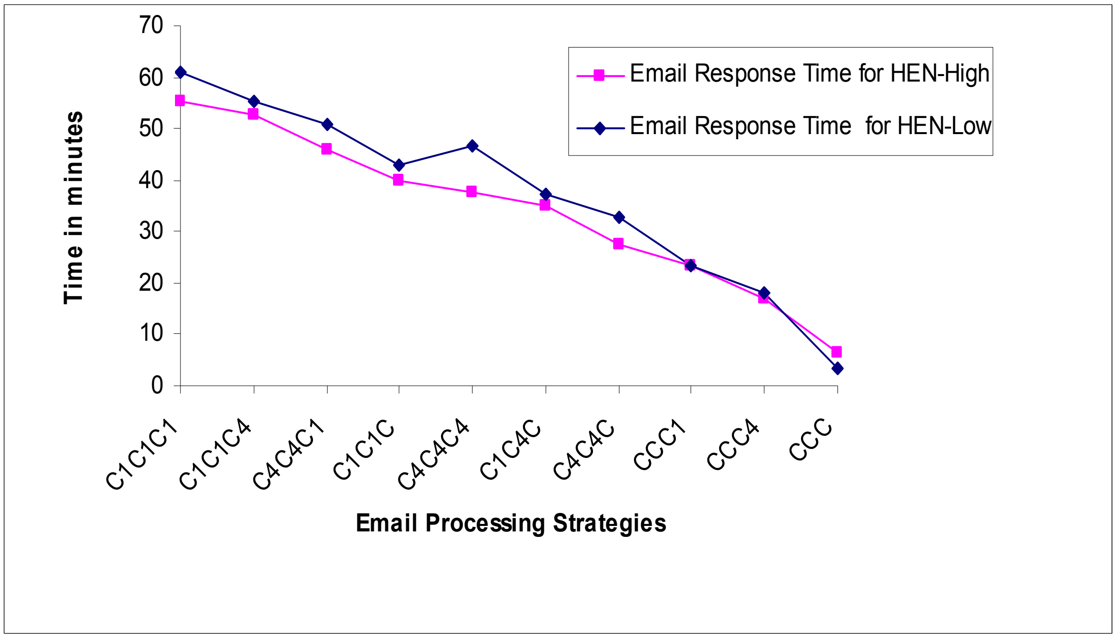


Figure 6.7(a) Average Email Response Time (in minutes) for HEN-High and HEN-Low Networks

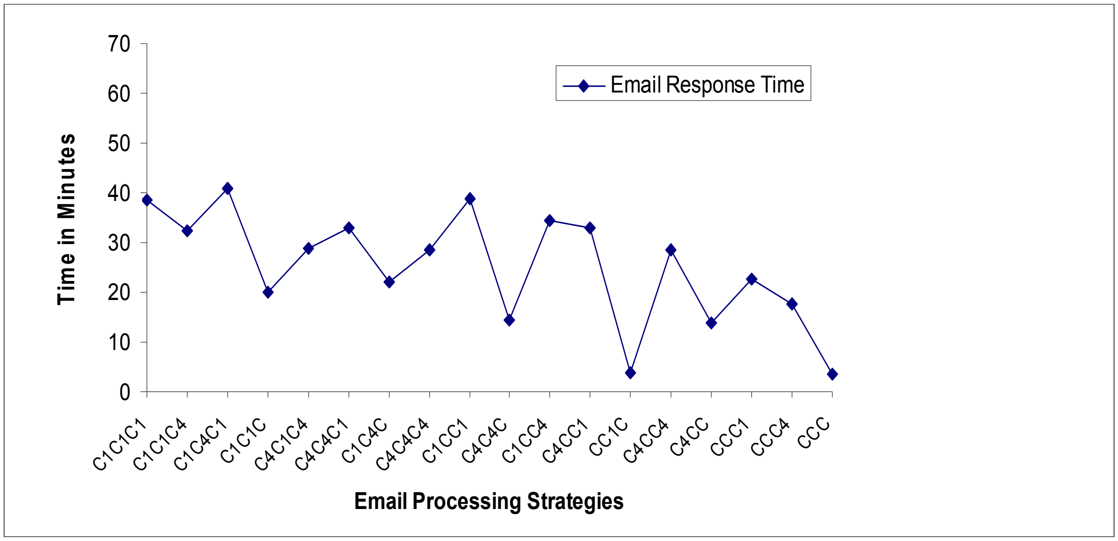


Figure 6.7(b) Average Email Response Time (in minutes) for XHEN Networks

We conducted a separate ANOVA to test hypothesis H2 since this hypothesis had only one dependent variable and required setting up the design differently. Table 6.3 shows the results of univariate between-subject tests conducted to investigate if significant differences existed between ((C1, C1), C1), ((C4, C4), C4), and ((C, C), C). Statistically significant differences existed between the performances of each pure network email processing strategy with respect to overall network effectiveness at α=0.05. However, hypothesis H2 is related more to the directionality of these differences.

Table 6.3 Tests of Between-Subjects Effects (Time and Value Effectiveness Equally Important)

Dependent Variable: Overall Effectiveness (Time and Value effectiveness Equally Important)

Source	Type III Sum of Squares	Df	Mean Square	F	Sig.
Corrected Model	451.439(a)	2	225.719	30.276	.000
Intercept	1517313.691	1	1517313.691	203516.331	.000
EPS	451.439	2	225.719	30.276	.000
Error	1319.621	177	7.455		
Total	1519084.752	180			
Corrected Total	1771.060	179			

(a) R Squared = .255 (Adjusted R Squared = .246)

Figure 6.8 shows the performance of the email policies mentioned above on the overall effectiveness when value-effectiveness and time-effectiveness are equally important for knowledge workers. For this scenario, the overall effectiveness was reported to be 93.23 percent with the use of a ((C1, C1), C1) policy, 92.59 percent with a ((C4, C4), C4) policy, and 89.6 percent with a ((C, C), C) policy. This implies that although E_o((C, C), C)< E_o ((C4, C4), C4) was true, the validity of E_o ((C4, C4), C4)> E_o ((C1, C1), C1) could not be confirmed. Hence, we found only partial support for hypothesis H2. The most plausible explanation for this comes from the rate of increase in value-effectiveness and rate of decrease in time-effectiveness with the increase in the number of email priority hour slots. When the entire network uses a C1 policy, the total number of email priority hours is the smallest across all policies. This number becomes higher when the entire network uses a C4 policy and becomes extremely large when a C policy is adopted. When the number of email-hour slots increases, the rate of increase in value-effectiveness is rather slow but the rate of decrease in time effectiveness is quite substantial. When a C1 policy is replaced by a C4 policy, value-effectiveness increases by 3.87 percent and when a C4 is replaced by a C policy, value-effectiveness increases by 7.2 percent. However, when a C1 policy is replaced by a C4, time-effectiveness drops by 5.15 percent but when a C4 is replaced by a C, this drop is 13.15 percent due to the substantial increase in the number of interruptions. Since the overall effectiveness comprises value and time effectiveness that are equally weighted in hypothesis H3, the total decrease in the overall effectiveness will simply be the sum of drop or gain exhibited by time and value effectiveness. The overall effectiveness that occurred when a C4 policy was replaced by a C was primarily due to the larger drop in time effectiveness.

This confirms a part of the hypothesis to be true. However, the overall effectiveness drops when a C1 is replaced by a C4 due to time-effectiveness dominating over value-effectiveness, which suggests that the other part of the hypothesis does not hold true. Further investigation showed that hypothesis H2 was supported in the case of XHEN networks but only partially supported for HEN-High and HEN-Low types of networks.

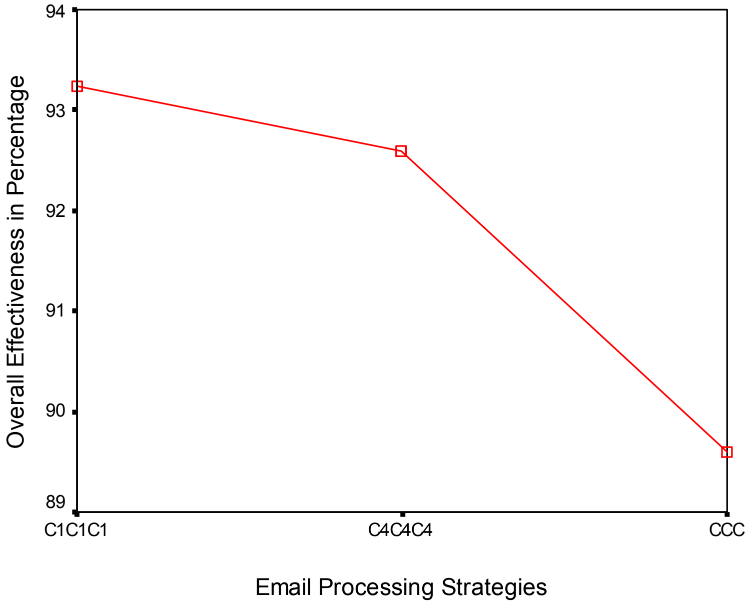


Figure 6.8 Overall Effectiveness (Time and Value Effectiveness Equally Important) vs. Email processing Strategies

We also analyzed the scenarios of different values being assigned to time or value effectiveness. For a scenario where value-effectiveness is three times as important as time-effectiveness, we found that when a C1 policy is replaced by a C4, value-effectiveness went up by 2.9 percent but time-effectiveness went down by 1.62 percent. When a C4 policy was replaced by a C policy, value-effectiveness went up by 5.4 percent but time-effectiveness went down by 3.28 percent. Thus the cumulative effectiveness continuously increased with the increase in the number of email priority hour slots. This explains why the curve has a positive slope as we move from C1 to C4 and from C4 to C. Figure 6.9 also explains this graphically. The average overall effectiveness attained is 90.9 percent in a network where every knowledge worker uses a C1 policy, 92.51 percent in a network using a C4 policy and 94.6 percent in a network utilizing a C policy.

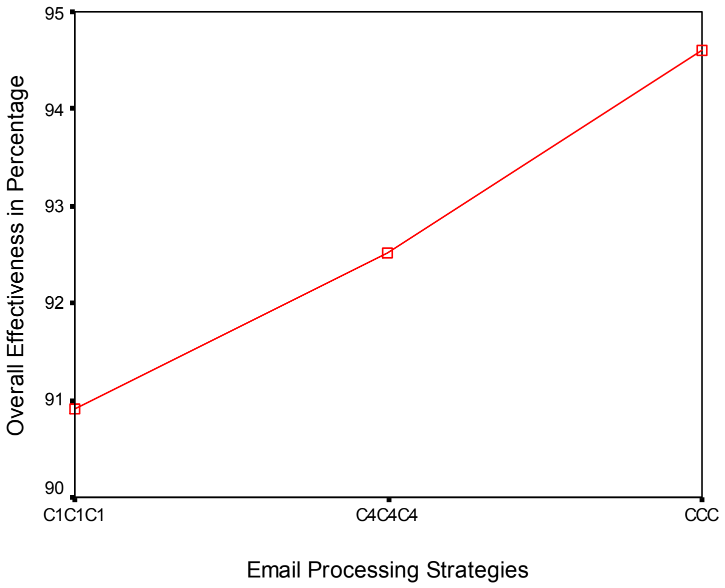


Figure 6.9 Overall Effectiveness (Value-Effectiveness More Important than Time-Effectiveness) vs. Email processing Strategies

We also analyzed the situation where time is given a higher importance than value-effectiveness. The assumption here is that time-effectiveness is three times as important as value-effectiveness. When a C policy was replaced by a C4, we observed a large gain in time effectiveness (9.86 percent) at the cost of a very small loss in value effectiveness (1.8 percent). Again, when a C4 was replaced by a C1, the increase in value-effectiveness was far larger (3.86 percent) than the drop in time-effectiveness (0.96 percent). The cumulative values of time and value effectiveness suggest that overall effectiveness decreases with the decrease in the number of email priority hour slots (i.e. from C to C4 and from C4 to C1) when time effectiveness is given more importance.

Figure 6.10 summarizes the overall effectiveness gained in such scenarios where time effectiveness is more important. The average overall effectiveness is 95.57 percent in a network using only a C1 policy, 92.67 percent in a network using a C4 policy and 84.6 percent in a network using a C policy.

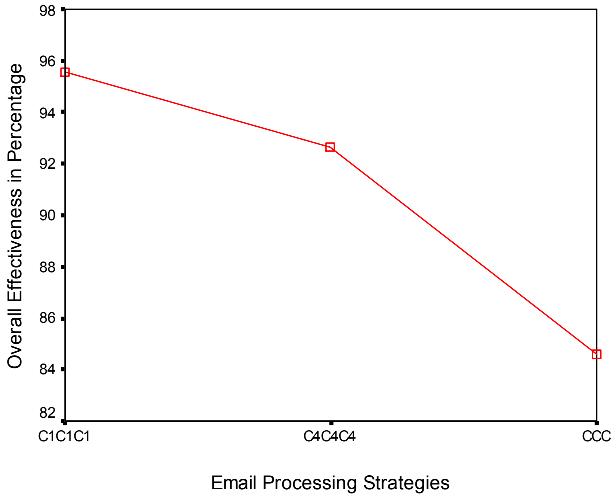


Figure 6.10 Overall Effectiveness (Time-Effectiveness More Important than Value-Effectiveness) vs. Email processing Strategies

6.1 Summary and (Theoretical and Practical) Implications

Input parameters for the simulation model were derived using the data collected from the survey. All the scenarios were implemented in thirty-eight different simulation models. Data collected from simulation output variables were analyzed to test proposed relationships. Support was found for most hypotheses. Hypothesis H2 was only partially supported. Table 6.4 summarizes the results of the hypotheses testing. Findings suggest that for homogeneous as well as heterogeneous networks, rhythmic email processing strategies lead to lower value-effectiveness but higher time-effectiveness. On the other hand, email response times are generally higher with rhythmic policies when compared with arrhythmic policies. Task completion times are usually lower with rhythmic policies.

Table 6.4 Summary of Hypotheses Testing

Hypotheses	**Description**	**Findings (**at α=0.05**)**
H1a	Rhythmic EPSs will lead to lower value-effectiveness than Arrhythmic EPSs.	Supported
H1b	Rhythmic EPSs will lead to higher time-effectiveness than Arrhythmic EPSs.	Supported
H1c	Rhythmic EPSs will lead to shorter average task completion times than Arrhythmic EPSs.	Supported
H1d	Rhythmic EPSs will lead to higher average email response times than Arrhythmic EPSs.	Supported
H2	The overall network effectiveness (E_o) at extreme levels (with C1 and C strategies) will be lower than that in the middle (with a C4 strategy)	Partially supported

These results are of immense practical significance especially for large organizations. If an entire organization uses C policy than each knowledge worker stand to loose, on an average, 28 min per day, given that each workday comprises of approximately 10 hours. This is roughly 4.67 percent of a 10-hour work day. Instead of using C policy, if the organization chooses to use C4 policy through out than a loss of only 12 minutes per day occurs, which is approximately 2 percent of a 10-hour work day. Time lost as a result of interruptions is almost negligible if C1 policy is used (approximately 0.6 percent of a 10-hour workday). Although these time savings appear to be not of as much significance when looked from an individual knowledge worker's perspective but they cumulate into a significantly bigger number when seen from an entire organization's perspective. For example, a small organization having 100 employees could potentially loose up to 2333 work hours with C policy, 1000 work hours with C4 policy, and 305 hours with C1 policy due to interruptions. Assuming an hourly wage of $50, this translates into an annual loss of approx. $600,000 with C policy, $ 260,000 with C4 policy and $80,000 with C1 policy. For a large organization having 10,000 employees, use of C policy could lead to a loss of 233,333 work hours translating into an annual loss of approx. $60 million. With C4 policy, about 100,000 work hours are lost resulting in the loss of approx. $26 million per year whereas, C1 policy leads to approx. 300,000 lost hours, which is equivalent to roughly $8 million per year assuming the wage rate of $50 per hour. However, this saving comes at a cost of approximately 7 to 11 For example, replacement of C policy by C4 policy results in approximately 7 percent reduction in value-effectiveness whereas when a C4 policy is replaced by C policy, value-effectiveness reduces by 3.8 percent In this study, we have looked at the time based value of emails.

When it comes to the overall effectiveness of email communication, for organizations that attach more value to time effectiveness, a C policy outperforms all other policies but if value is more important to an organization, a C1 policy performs better than all other policies. On the other hand, for an organization that attaches equal importance to time and value, a C1 performed better than a C4 and a C4 performed better than a C policy.

Finally, we could rank all policies on the basis of their performance with respect to time and value effectiveness for homogeneous as well as heterogeneous networks. This could serve as a useful guide for knowledge workers in choosing the policy that fits best in their work environment. Figures 6.11 and 6.12 categorize all the email processing strategies for HEN and XHEN networks for this purpose. The markers on the horizontal and vertical lines have been set up at the mean values. All the policies have been divided into four categories. The first quadrant represents a high value-effectiveness and low time-effectiveness group, the second quadrant represents a high value-effectiveness and high time-effectiveness group, the third quadrant represents a low value-effectiveness and high time-effectiveness group, and the fourth quadrant represents a low value-effectiveness and low time-effectiveness group. As is evident from these diagrams, a ((C, C), C) policy provides the highest value-effectiveness but the lowest time-effectiveness, whereas, ((C1, C1), C1) provides the highest time-effectiveness but the lowest value-effectiveness. Policies scattered around the medians are able to strike a better balance between time and value effectiveness. For example, use of a ((C1, C4), C) policy leads to a time-effectiveness of approx. 93 percent and a value-effectiveness of approx. 95 percent.

Figure 6.11 Value Effectiveness vs. Time Effectiveness for HEN Networks

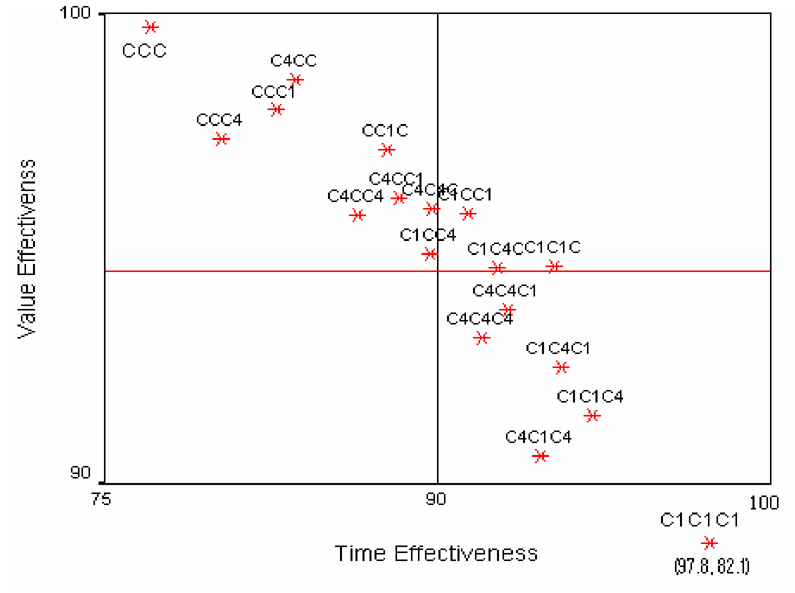


Figure 6.12 Value Effectiveness vs. Time Effectiveness for XHEN Networks

7. LIMITATIONS AND IMPLICATIONS FOR FUTURE RESEARCH

Like any other research, this study has certain limitations. We used simulation methodology, which has some known drawbacks. First, it is a method for conducting computer experiments where we are not able to use real subjects and therefore individual differences between subjects can not be directly accounted for. We address this problem to a certain extent by using probability distributions, which may account for variability in input variables. Simulation also serves to reduce the external validity of the work environment.

Another issue with simulation is that it can generate large amounts of data for analysis purposes. Sample sizes can be reduced by running the simulation for shorter time periods, but this leads to the problem of insufficient half width, which can be corrected by running the simulation for longer time periods. Hence, the solution to one problem creates another problem. We used a slightly different approach to overcome this problem. Rather than using individual values of outcome variables for statistical analysis, we used replication averages. It is very difficult to conduct this study using empirical methods such as experiments, field studies, etc. due to lack of control over treatments, subject attrition, change taking place during the experiment, etc. Further, in any empirical study, we are always limited by the constraints of time and availability of subjects. This study specifically required intensive monitoring of time and hence would have become expensive due to the need for video cameras, additional requirements of manpower for monitoring knowledge workers for extended time periods, etc.

Simulation can easily overcome these limitations. Another advantage is that a large number of treatments can be evaluated using simulation. We can not really give a long questionnaire to real subjects.

We made several assumptions in the modeling process and these also become limitations of this study. We studied work environments where all primary tasks had the same learning rate of 0.7. Future studies can focus on modeling work environments having a mix of different learning rates. It will be worthwhile to understand how different learning rates influence the working of various email processing strategies. The simulation model developed for this study assumed that knowledge workers only exchange emails and not primary tasks. Modeling task dependencies in a project management setting will definitely provide important insights into this problem and will potentially raise important questions from a project manager's perspective.

We assumed that all the knowledge workers belonging to the network in this study were located in the same time zones and hence had similar office hours. So emails from within- network knowledge workers did not arrive at night. However, with off-shoring gaining prevalence in modern business organizations, knowledge workers often find themselves working in different time zones and hence processing emails even at night. Future research should investigate how response times and other performance characteristics are impacted by the choice of email processing strategies in the presence of off-shoring.

As pointed out earlier, this study focused on email resolution and not issue resolution. We assumed that an email is resolved if it has gone through one complete cycle of processing, i.e. for an email requiring a response, one complete cycle is going from sender to receiver and then back to the original sender. All the email exchanges within a cycle make up one email thread. Because of this assumption, the maximum thread length at any point of time can not be more than three emails. Future studies should relax this assumption and look at emails with longer threads so that the focus is not just on resolving the emails but also on resolving the issues, which may take longer.

Another restriction that we imposed was that no prioritization or routing strategies were implemented within the current models. All the knowledge workers processed emails on the first come-first served basis and no forwarding of emails within networks occurred. Future studies in this direction could focus on comparing the email processing strategies in the presence of various prioritization and routing schemes.

There are several other directions in which this study can be taken. We assumed that the value of an email drops linearly with time after a certain threshold time has elapsed. It would be interesting to see how value-effectiveness changes when the value of email drops non-linearly with time. Further, we considered only C1, C4 and C policies in this study. We could use other network policies such as C3, C6, C8, etc. to see how a greater heterogeneity within policy-sets can have an influence on the performance. The current study assumes that emails interrupt primary tasks but not visa versa. It would be insightful to study when both can interrupt each other.

The network that was modeled in this study contained only three knowledge workers. It is also important to conduct the same study on a network of comparatively larger size to see whether the findings are robust enough.

In the end, we can say that simulation combined with an analytical approach and statistical analysis can serve as a very useful method to conduct studies such as this, that often become unfeasible to pursue due to the time factor and the requirement that researchers continuously monitor subjects for extended periods of time. Through the use of this approach, we tried to address a very significant problem of email overload and interruptions that several organizations are facing today. Future research in this direction will certainly help in improving the overall productivity of organizations by helping knowledge workers change their poor email processing practices. This study was a small but significant step in this direction.

REFERENCES

-. (2006). "Anti-Phishing Work Group." Retrieved January, 2006, from http://www.antiphishing.org/

Ackoff, R. L. (1967). "Management misinformation systems." Management Science **14**: 147-156.

Adam, R. (2002). "Is e-mail addictive?" Aslib Proceedings **54**(2): 85-94.

Antone, G. (2006). E-Mail Deluge Poorly Handled. InformationWeek**:** 24.

Ash, R. and D. E. Smith-Daniels (1999). "The effects of learning, forgetting, and relearning on decision rule performance in multiproject scheduling." Decision Sciences **30**(1): 47-82.

Axlerod, R. (2003). "Advancing the art of simulation in the social sciences " Japanese Journal for Management Information System Special Issue: Agent-Based Modeling **12**(3): 1-19

Bailey, C. D. (1989). "Forgetting and learning curve: A laboratory study." Management Science **35**(3): 340-352.

Bellotti, V., et al. (2005). "Quality Versus Quantity: E-Mail-Centric Task Management and Its Relation With Overload." Human-Computer Interaction **20**(1/2): 89-138.

Berghal, H. (1997). "Email- The Good, The Bad and The Ugly." Communications of the Acm **40**(4): 11-15.

Cameron, K. S. and R. E. Quinn (1988). Organizational paradox and transformation. Cambridge, MA: Ballinger.

Carley, K. M. and D. Skillicorn (2005). "Special Issue on Analyzing Large Scale Networks: The Enron Corpus." **11**(3): 179.

Carlson, J. G. and A. J. Rowe (1976). "How much does forgetting cost." Industrial Engineering **8**(9): 40-47.

Chervany, N. and G. Dickson (1974). "An experimental evaluation of information overload in a production environment." Management Science **10**: 1335-1344.

Chewning, J., Eugene G. and A. M. Harrell (1990). "The effect of information load on decision makers' cue utilization levels and decision quality in a financial distress decision task." Accounting, Organizations and Society **15**(6): 527-542.

Chung, C. A. (2003). Simulation Modeling Handbook: A practical Approach, CRC Press.

Corragio, L. (1990). Deleterious effects of intermittent interruptions on the task performance of knowledge workers: A laboratory investigation., University of Arizona **Unpublished Ph. D. thesis**.

Cutrell, E. B., et al. (2000). Effects of instant messaging interruptions on computing tasks Proceedings of 2000 CHI Conference on Human Factors in Compu. Sys., ACM press.

Czerwinski, M., et al. (2000). Instant messaging and interruption: Influence of task type on performance OZCHI 2000.

Davenport, T. H. and J. C. Beck (2000). "Getting the attention you need." Harvard Business Review **78**(5): 118.

Davenport, T. H. and J. C. Beck (2001). "The Economics of Attention." Harvard Business Review **79**(7): 12.

Denning, P. (1982). "Electronic Junk." Communications of the ACM **25**(3): 163-165.

Denning, P. (2002). "Internet time out." Communications of the ACM **45**(3): 15-17.

Di Paolo, E. A., Noble, J., Bullock, S. (2000). Simulation models as opaque thought experiments. . Artificial Life VII: Proc. of 7th Internat. Conf. on Artificial Life Cambridge, MA., MIT Press.

Ducheneaut, N. and V. Bellotti (2001). "E-mail as habitat: an exploration of embedded personal information management." Interactions **8**(5): 30-38.

Ducheneaut, N. and L. A. Watts (2005). "In Search of Coherence: A Review of E-Mail Research." Human-Computer Interaction **20**(1/2): 11-48.

Edmunds, A. and A. Morris (2000). "The problem of information overload in business organisations: a review of the literature." International Journal of Information Management **20**(1): 17-28.

Eppler, M. and J. Mengis (2005). "The Concept of Information Overload: A Review of Literature from Organization Science, Accounting, Marketing, MIS, and Related Disciplines." Journal of Academic Librarianship **31**(2): 174.

Farhoomand, A. F. and D. H. Drury (2002). "Managerial information overload." Communications of the ACM **45**(10): 127.

Ford, J. D. and R. W. Backoff (1988). "Logics of identity, contradiction, and attraction in change." Academy of Management Review **19**: 756-795.

Galbraith, J. R. (1974). "Organizational design: An information processing view." Interfaces **3**: 28-36.

Gersick, C. J. G. and J. R. Hackman (1990). "Habitual routines in task-performing groups." Organizational Behavior and Human Decision Processes **47**(1): 65-97.

Greve, R. A. (2005). Modeling and simulation of knowledge worker attention for evaluation of email processing strategies. Department of MSIS. Stillwater, Oklahoma State University. **PhD:** 122.

Grise, M. L. and R. B. Gallupe (1999). "Information overload: Addressing the productivity paradox in face-to-face electronic meetings." Journal of Management Information Systems **16**(3): 157-185.

Gupta, A., et al. (2006). "Email management: A techno-managerial research perspective." Under review.

Gupta, A., et al. (2004). An exploratory analysis of email processing strategies. Proceedings of Decision Sciences Annual meeting, Boston.

Gupta, A., et al. (2006). How often should we check our email? Balancing interruptions and quick response times, Oklahoma State University.

Hall, A. and G. Walton (2004). "Information overload within the health care system: a literature review." Health Information & Libraries Journal **21**(2): 102-108.

Hallowell, E. (2005). "Overloaded Circuits: Why Smart People Underperform." Harvard Business Review **83**(1): 54-62.

Hancock, P. A. and N. Meshkati, Eds. (1988). Human mental Workload. Advances in Psychology, Elsevier Sciences.

Hans-Joachim, M., S. Karsten, A. Florin., H. Heinz. (2001). "Computer simulation as a method of further developing a theory: simulating the elaboration likelihood model." Personality & Social Psychology Review, 5(3) 201-215. **5**(3): 201-215.

Hart, S. G. and L. E. Staveland, Eds. (1988). Development of a multi-dimensional workload rating scale: results of empirical and theoretical research. Human Mental Workload. Amsterdam, Elsevier.

Her, C. and S. Hwang (1989). "Application of queuing theory to quantify information workload in supervisory control systems." International Journal of Industrial Ergonomics **4**: 51-60.

Huberman, B. A. and L. A. Adamic (2004). "Information dynamics in the networked world." Lecture Notes in Physics **650**: 371-398.

Huberman, B. A. and T. Hogg (1995). "Communities of practice." Computational & Mathematical Organization Theory **1**(1): 73-92.

Jackson, T., et al. (2001). " The cost of email interruption." Journal of Systems and Information Technology **5**(1): 81-92.

Jackson, T., et al. (2003). "Understanding email interaction increases organizational productivity." Communications of the ACM **46**(8): 80-84.

Jennings, R. (2006, 20th April 2006). "Unreasonable Expectations Contribute to Email Overload." Ferris Research, from http://blog.ferris.com/2006/04/unreasonable_ex.html.

Jett, Q. R. and J. M. George (2003). "Work interrupted: A closer look at the role of interruptions in organizational life." Academy of Management. The Academy of Management Review **28**(3): 494-507.

Johnson, S. L. and S. Faraj (2005). Preferential attachment and mutuality in electronic knowledge networks. 26th International Conference on Information Systems, Las Vegas.

Jones, Q., et al. (2004). "Information Overload and the Message Dynamics of Online Interaction Spaces: A Theoretical Model and Empirical Exploration." Information Systems Research **15**(2): 194.

Kerr, B. and E. Wilcox (2004). Designing remail: Reinventing the email client through innovation and integration CHI 2004, Vienna, Austria, ACM.

Langfitt, F. (2005). "Blackberry or Crackberry? A PDA Culture War." Retrieved March, 2006, from http://www.npr.org/templates/story/story.php?storyId=4279486.

Law, A. M. and W. D. Kelton (2000). Simulation modeling and analysis.

Lawrence, T. B., et al. (2001). "The temporal dynamics of institutionalization." Academy of Management Review **26**(4): 624-644.

Lewis, M. W. (2000). "Exploring paradox: Toward a more comprehensive guide." Academy of Management. The Academy of Management Review **25**(4): 760-776.

Malhotra, N. K. (1984). "Reflections on the Information Overload Paradigm in Consumer Decision-Making." Journal of Consumer Research **10**(4): 436-440.

Malhotra, N. K., et al. (1982). "The Information Overload Controversy - an Alternative Viewpoint." Journal of Marketing **46**(2): 27-37.

Markus, M. L. (1994). "Electronic Mail as the Medium of Managerial Choice." Organization Science **5**(4): 502-527.

Markus, M. L. (1994). "Finding a happy medium: explaining the negative effects of electronic communication on social life at work." ACM Trans. Inf. Syst. **12**(2): 119-149.

McFarlane, D. C. (2002). "Comparison of four primary methods for coordinating the interruption of people in human-computer interaction." Human-Computer Interaction **17**(1): 63-139.

McManus, D. J., et al. (2002). "Intraorganizational versus interorganizational uses and benefit of electronic mail." Information Resources Management Journal **15**(3): 1.

Nasrallah, W., et al. (1998). "Diversity and Popularity in Organizations and Communities." Computational & Mathematical Organization Theory **4**(4): 347-372.

Nasrallah, W., et al. (2003). "Interaction Value Analysis: When Structured Communication Benefits Organizations." Organization Science **14**(5): 541-557.

Nasrallah, W. F. (2006). "When does management matter in a dog-eat-dog world:An “Interaction Value Analysis” model of Organizational Climate." Computational & Mathematical Organization Theory(Forthcoming).

Nasrallah, W. F. and R. E. Levitt (2001). "An Interaction Value Perspective on Firms of Differing Size." Computational and Mathematical Organization Theory **7**(2): 113.

Neath, I. and A. M. Surprenant (2002). Human Memor: An introduction to research, data, and theory, Thompson-Wadworth.

Ostroff, C. and N. Schmitt (1993). "Configurations of organizational effectiveness and efficiency." Academy of Management Journal **36**(6): 1345.

Paul, M. (2006). E-Mail Overload Or Avoidance? InformationWeek**:** 24.

Perlow, A. L. (1999). "The time famine: Towards a sociology of work time." Administrative Science Quarterly **44**(1999): 57-81.

Peschl, M., Scheutz, M. (2001). Explicating the epistemological role of simulation in the development of theories of cognition. Proc. of Internat. Colloq. on Cognitive Science

Rohall, S. L. (2002). Redesigning email for the 21st Century. CSCW 2002, New Orleans.

Rosen, S. M. (1994). Science, paradox, and the moebius principle, Albany: State University of New York Press.

Rubio, S., et al. (2004). "Evaluation of Subjective Mental Workload: A comparasion of SWAT, NASA-TLX and Workload Profile methods." Applied Psychology :An International Review **53**(1): 61-86.

Saunders, C. S. (2002). "Accounting for time." Information Resources Management Journal **15**(3): V.

Schick, A. G., et al. (1990). "Information overload: A temporal approach." Accounting, Organizations and Society **15**(3): 199-220.

Sharda, R., et al. (1999). "Group knowledge networks: A framework and an implementation." Information Systems Frontiers **1**(3): 221-239.

Simpson, C. W. and L. Prusak (1995). "Troubles with information overload--Moving from quantity to quality in information provision." International Journal of Information Management **15**(6): 413-425.

Sipior, J. C. and B. T. Ward (1995). "The ethical and legal quandary of email privacy." **38**(12): 48.

Smith, S. (2006). "Email's Mid-Life Crisis." EContent **29**(2): 20.

Speier, C., I. Vessey, et al (2003). "The Effects of Interruptions, Task Complexity, and Information Presentation on Computer-Supported Decision-Making Performance." Decision Sciences . **342**(4): 623-81.

Speier, C., et al. (1999). "The influence of task interruption on individual decision making: An information overload perspective." Decision Sciences **30**(2): 337-360.

Spira, J. (2005). The high cost of interruptions (Content, Document and Knowledge Management). KMWorld. **14**.

Spira, J. and J. Feintuch (2005). The cost of not paying attention: How interruptions impact knowledge worker productivity, Bassex Research Report. .

Steedman, I. (1970). "Some improvement curve theory." International Journal of Production Research **8**(3): 189-205.

Steven, M. B. (2004). "Spamming, Phishing, Authentication, and Privacy." **47**(12): 144.

Swartz, N. (2006). "Survey: E-mail Hampers Productivity." Information Management Journal **40**(2): 22.

Tassabehji, R. and M. Vakola (2005). "Business Email: the Killer Impact." Communications of the ACM **48**(11): 64-70.

Trafton, J. G., et al. (2003). "Preparing to resume an interrupted task: effects of prospective goal encoding and retrospective rehearsal." International Journal of Human-Computer Studies **58**(5): 583-603.

Tuttle, B. and F. G. Burton (1999). "The effects of a modest incentive on information overload in an investment analysis task." Accounting, Organizations and Society **24**(8): 673-687.

Tyler, J. R. and J. C. Tang. (2003). When can i expect an email response? a study of rhythms in email usage. . Proceedings of ECSCW 2003, Kluwer, B.V.

Varian, H. R. (1992). Microeconomic Analysis, W.W. Norton & Company.

Venolia, G., L. Dabbish., J. J. Cadiz, A. Gupta (2001). Supporting email workflow, Microsoft Research Tech Report

Weber, R. (2004). "The grim reaper: the curse of e-mail." MIS Quarterly **28**(3): III-XIII.

Weisband, S. P. and B. A. Reinig (1995). "Managing user perceptions of email privacy." Communications of the ACM **38**(12): 40-47.

Welch, P. D. (1983). The statistical analysis of simulation results. The computer performance modeling handbook. S. S. Lavenberg. NY, Academic press.

Whittaker, S. (2005). "Supporting Collaborative Task Management in E-mail." Human-Computer Interaction **20**(1/2): 49-88.

Whittaker, S., et al. (2006). "Email in Personal Information Management." Communications of the ACM **49**(1): 68.

Whittaker, S., et al. (2005). "Introduction to This Special Issue on Revisiting and Reinventing E-Mail." Human-Computer Interaction **20**(1&2): 1-9.

Whittaker, S., et al. (2005). "Introduction to This Special Issue on Revisiting and Reinventing E-Mail." Human-Computer Interaction **20**(1/2): 1-9.

Whittaker, S. and C. Sidner (1996). Email overload: exploring personal information management of email. CHI 96, Vancouver, BC Canada, ACM.

Wilson, P. (1996). "Interdisciplinary research and information overload." Library Trends **45**(2): 192-203.

Wilson, R. (1934). "A scientific routine for stock control " Harvard Business Review **13** (1): 16-34.

Winsberg, E. (2003). "Simulated experiments: methodology for a virtual world." Philosophy of Science **70**(1): 105-121.

Wright, T. P. (1936). "Factors affecting the cost of airplanes." Journal of Aeronautical Science **3**(4): 112-128.

Wu, F., et al. (2004). "Information flow in social groups." Physica A: Statistical and Theoretical Physics **337**(1-2): 327-335.

Xie, B. and G. Salvendy (2000). "Review and reappraisal of modelling and predictive mental workload in a single and multi- task environments." Work & Stress **14**(1): 74-99.

Yadati, C. N. (2006). Switching Between Email Processing and Other Knowledge Work Tasks: A Semi Markov Decision Process Department of Industrial Engineering and Management. Stillwater, Oklahoma State University. **Master of Science**.

Zellmer-Bruhn, M. E. (2003). "Interruptive Events and Team Knowledge Acquisition." Management Science **49**(4): 514.

APPENDIX

*** 1. If you wish to participate in this study, please complete the following steps:**
1) Click on “I have read the above conditions and agree to participate” and print a copy of consent form for you.
2) Complete the survey that starts from the next screen.

○ I have read the above conditions and agree to participate in the study

○ No Thanks!

1. Information About Your Work Environment

1. In what country are you located?

2. Which best describes your personal involvement with your organization? I...

○ Am a member of a messaging support team

○ Work for a vendor of messaging and collaboration products or services

○ Am an IT professional,

○ Am not in IT,

○ Have a different involvement (if so, please describe)

3. Which best describes what your industry is or what your organization does? If several apply to you, choose the one that's most accurate.

4. You can be best described as:

○ Manager

○ CIO

○ CEO

○ Staff

○ Other (please specify)

5. What is the total number of employees in your entire world-wide organization (i.e., not just in your own department)?

6. What is the total number of employees in your entire department?

7. How many people typically work with you in the following different roles:

As subordinates

As supervisors (direct and indirect)

As immediate peers

Others

8. Please check the type of projects you are involved with (check all that apply).

☐ Centralized team(s) (e.g., teams located at one physical location)

☐ Decentralized team(s) (e.g., "virtual groups")

☐ I am not involved in any collaborative projects

☐ Other (please specify)

Please skip question 9 and 10 if you are not involved in any collaborative projects

9. How many teams of following type are you typically involved with (Check all that apply)?

Centralized team(s)

Decentralized team(s)

Other

10. How many employees are there in each team you are typically involved with (Please use the text box in Q.11, if you are involved with more than 5 different teams) ?

	Team 1- No. of Employees	Team 2- No. of Employees	Team 3- No. of Employees	Team 4- No. of Employees
Centralized team(s)				
Decentralized team(s)				
Other				

11. Please use the space below if you work in more than five teams as reported in previous question or for providing any additional information.

2. Time Spent on Emails And The Frequency of Email Processing

1. On a typical workday, how many hours do you work at your office?

2. On a typical workday, how many hours do you work at home?

*** 3. On a typical workday, how much time do you spend on work-related emails?**

*** 4. On a typical workday, how much time do you spend on non-work-related emails?**

▾

5. On a typical workday, do you process emails at places other than your desktop computer at work?

○ Yes

○ No

Please skip question 6 and 7 if you selected "No" in the previous question.

6. Please indicate which methods you use to check emails while you are away from your desktop computer (check all that apply).

☐ PDA

☐ Blackberry, etc.

☐ Checking at home

☐ While traveling

☐ Other (please specify)

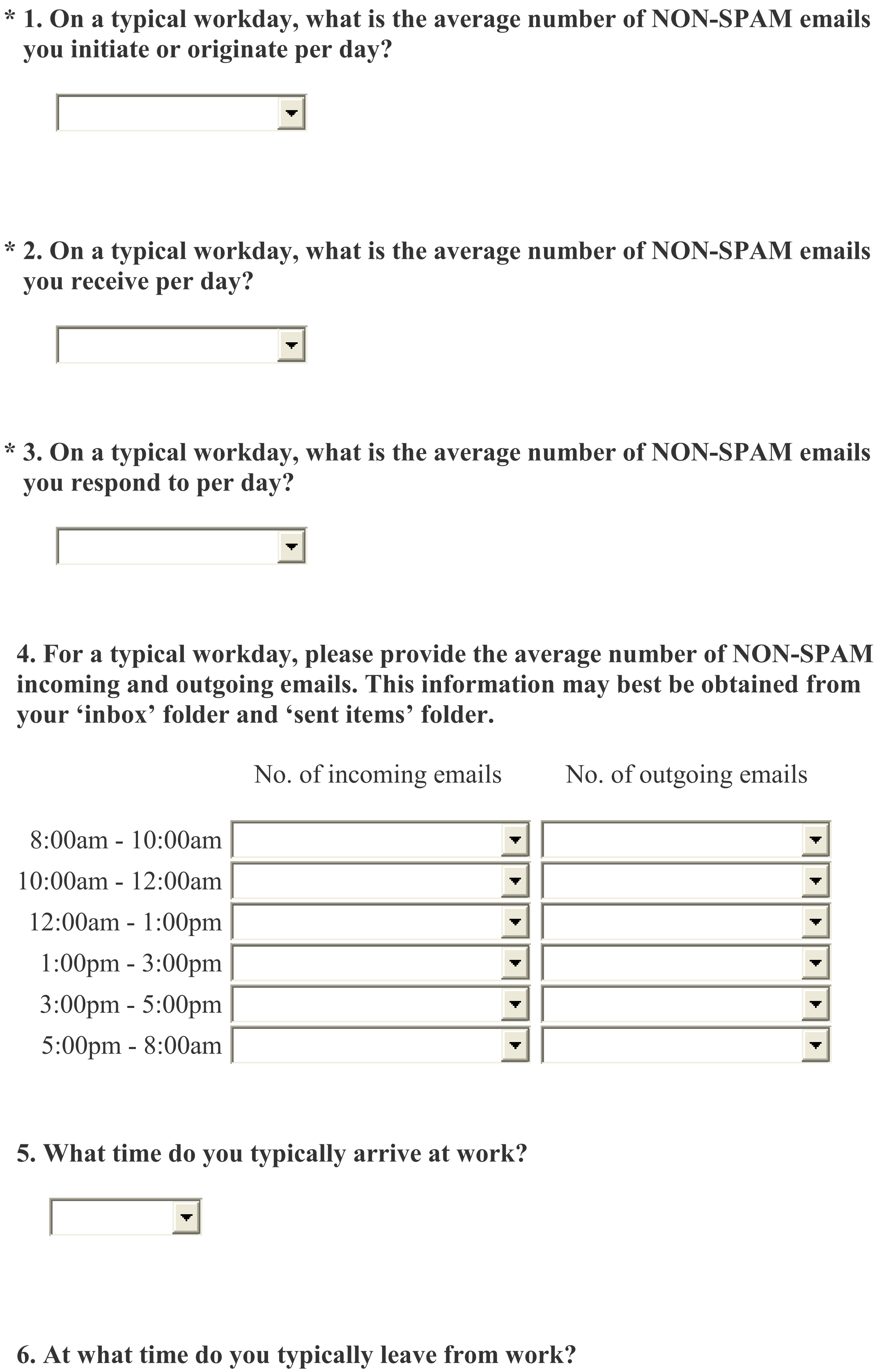
7. On a typical workday, what percentage of your overall email processing time is spent using wireless devices like a PDA, Blackberry, etc. as mentioned in previous question.

*** 8. On a typical workday, what is the average number of times you process [or check] work-related emails?**

- ○ Once a day
- ○ One to two times a day
- ○ Two to four times a day
- ○ Four to eight times a day
- ○ Greater than eight times per day
- ○ A few times per hour
- ○ More or less continually.
- ○ Any other comments(please specify)

9. Why do you think you process [or check] emails multiple times a day (for example- is it because of your habit or the nature of your work or your need or the work culture within your organization or impression management, etc.)?

3. Number of Incoming and Outgoing Email



* 1. On a typical workday, what is the average number of NON-SPAM emails you initiate or originate per day?

* 2. On a typical workday, what is the average number of NON-SPAM emails you receive per day?

* 3. On a typical workday, what is the average number of NON-SPAM emails you respond to per day?

4. For a typical workday, please provide the average number of NON-SPAM incoming and outgoing emails. This information may best be obtained from your 'inbox' folder and 'sent items' folder.

	No. of incoming emails	No. of outgoing emails
8:00am - 10:00am		
10:00am - 12:00am		
12:00am - 1:00pm		
1:00pm - 3:00pm		
3:00pm - 5:00pm		
5:00pm - 8:00am		

5. What time do you typically arrive at work?

6. At what time do you typically leave from work?

4. Types and Priority of Email

1. For a typical workday, please provide information about arriving emails, categorized as 'Sender' -
(1- highest priority, 8- least priority)

	% of Emails Received per Day	% of Time Spent on Email Processing	Priority given to Emails while Processing
1. Within organization			
1(a) Within department			
(i) Supervisor			
(ii) Peers			
(iii) Subordinate			
1(b) Outside department			
2. Outside organization			
3. Other			

2. For a typical workday, please provide information about arriving emails, categorized as 'content of emails'-

	% of Emails Received per Day	% of Time Spent on Email Processing	Priority given to Emails While Processing
1. Complex email (take longer time to process)			
2. Short email (take shorter time to process)			
3. FYI			
4. Listserv email			

5. CC email			
6. Personal email			
7. Spam			
8. Other			

3. For a typical workday, please provide information about arriving emails, categorized as 'collaborative emails'-

	% of Emails Received per Day	% of Time Spent on Email Processing	Priority given to Emails While Processing
1. Team 1			
2. Team 2			
3. Team 3			
4. Team 4			
5. Team 5			
6. Team 6			
7. Team 7			

4. For a typical workday, please provide information about arriving emails, categorized as 'expected response time'-

	% of Emails Received per Day	% of Time Spent on Email Processing	Priority given to Emails While Processing
1. Urgent (e.g. response needed ASAP)			
2. Within 0 to 2 hours			
3. Within 0 to 4 hours			
4. Within 1 workday			
5. Within 2 workdays			

6. Within 1 week			
7. Other (e.g. 2 week, 1 month, etc.)			

5. What email processing strategy or criteria do you use to determine the order in which a particular email should be processed (Please check all that apply)?

☐ First come first serve

☐ First come last serve

☐ Randomly

☐ Depends on who the sender is

☐ Depends upon how long the email is sitting un-responded in the inbox

☐ Depends upon the content of email

☐ Depends upon which group it is coming from

☐ Depends upon the urgency of email (expected response time)

☐ Any other comments (please specify)

6. Please rank the above mentioned criteria in order of importance to you (Please check all that apply).
(1- most important, 8 - least important)

	Rank
First come first serve	
First come last serve	
Randomly	
Depends on who the sender is	
Depends upon how long the email is sitting un-responded in the inbox	

Depends upon the content of email
Depends upon which group it is coming from
Depends upon the urgency of email (expected response time)

7. Please include any other comments you may have here.

Oklahoma State University Institutional Review Board

Date: Thursday, November 03, 2005
IRB Application No BU0611
Proposal Title: Understanding Email Management Within Organizations

Reviewed and Processed as: Exempt

Status Recommended by Reviewer(s): Approved Protocol Expires: 11/2/2006

Principal Investigator(s

Ashish Gupta
42 S. Univ. Place #4
Stillwater, OK 74075

Ramesh Sharda
320 CBA
Stillwater, OK 74078

The IRB application referenced above has been approved. It is the judgment of the reviewers that the rights and welfare of individuals who may be asked to participate in this study will be respected, and that the research will be conducted in a manner consistent with the IRB requirements as outlined in section 45 CFR 46.

☑ The final versions of any printed recruitment, consent and assent documents bearing the IRB approval stamp are attached to this letter. These are the versions that must be used during the study.

As Principal Investigator, it is your responsibility to do the following:

1. Conduct this study exactly as it has been approved. Any modifications to the research protocol must be submitted with the appropriate signatures for IRB approval.
2. Submit a request for continuation if the study extends beyond the approval period of one calendar year. This continuation must receive IRB review and approval before the research can continue.
3. Report any adverse events to the IRB Chair promptly. Adverse events are those which are unanticipated and impact the subjects during the course of this research; and
4. Notify the IRB office in writing when your research project is complete.

Please note that approved protocols are subject to monitoring by the IRB and that the IRB office has the authority to inspect research records associated with this protocol at any time. If you have questions about the IRB procedures or need any assistance from the Board, please contact Beth McTernan in 415 Whitehurst (phone: 405-744-5700, beth.mcternan@okstate.edu).

Sincerely,

Sue C Jacobs

Sue C. Jacobs, Chair
Institutional Review Board

Informed Consent Form

You are being asked to participate in a study that I am completing as part of my doctoral program requirements. The study is being supervised by Professor Ramesh Sharda of the Management Science and Information Systems Department, Spears School of Business, Oklahoma State University. The following information summarizes the purpose of the study, how long it will take to participate, what you are asked to do in the study, the risks/discomforts and benefits for being in the study, your right to not be in the study or stop at any time, and who to call if you have any questions.

We are currently conducting research on understanding email communication within organizations. As a participant, you will be asked to complete a survey which will take approximately 10-12 minutes.

I assure you that your name will not be shared in any way; all data are completely confidential and will be used only for the purposes of data analysis. Your survey is identified only by an ID number assigned by the researcher, making it impossible for anyone other than the researcher to establish your identity. If you supply your name and email address, it would be used only to share the results of this survey and any related reports. This information will not be linked to your responses in any way. Furthermore, you will return the completed survey through a secure website, accessible only to the researcher. The information from the surveys will be used for the purposes described here and will not be shared in any way with others. Your participation is voluntary and you may refuse to participate, or may discontinue participation at any time, without penalty or loss.

There are no foreseeable discomforts or risks involved in participating in this study. You will receive no direct benefit from your participation in this study, but your participation will help researchers and managers better understand how email use within organizations may impact employee performance.

If you would like additional information about the study or your rights as a participant before or after the study is completed, please feel free to contact Ashish Gupta (the primary investigator), or Dr. Ramesh Sharda (advisor), or the Director of University Research Compliance, Dr. Sue Jacobs. The contact information for these individuals is listed below:

Ashish Gupta
Visiting Assistant Professor
Department of Management
Science and Information
Systems
Oklahoma State University-Tulsa
Tulsa, OK 74106-0702
Ph: (918) 594-8583
Email: ashish.gupta@okstate.edu

Dr. Ramesh Sharda
Regents Professor
Department of Management
Science and Information
Systems
Oklahoma State University
Stillwater, OK 74078
Ph: (405) 744-8850
Email: sharda@okstate.edu

Dr. Sue Jacobs
Director of University
Research Compliance
415 Whitehurst

Oklahoma State University
Stillwater, OK 74078
Ph: (405) 744-1676
Email:sue.c.jacobs@okstate.edu

If you wish to participate in the study, please complete the following steps:

1) Click on "I have read the above conditions and agree to participate" and print a copy of consent form for you.
2) Complete the survey that starts from the next screen.

We thank you for your participation in this research study!

OSU
Institutional Review Board
Approved 11/3/05
Expires 11/2/06
Initials oj
BU0611

Script of Email to be sent out to all other subjects

To: Educators, Students, Managers, and other Working Professionals (All departments)

From: Ashish Gupta, Visiting Assistant Professor at Oklahoma State University-Tulsa.

Subject: Research study on emails being conducted by Ashish Gupta

This study is being conducted by Ashish Gupta at the Spears School of Business at the Oklahoma State University, which examines how people use email at work. If you use emails at work, we invite you to share your opinions on characteristics and volumes of email that you receive. Participation will involve completing a survey at http://fp.okstate.edu/guptaa/students.htm. This survey should take about 10-12 minutes to complete. If you wish to participate in this study, please complete the following steps: 1) Click on "I have read the above conditions and agree to participate" and print a copy of consent form for you. 2) Complete the survey that starts from the next screen.

Your help in this effort will be greatly appreciated. However, your participation is completely voluntary and all information you provide, including your email address, will be kept completely confidential. Only reports of study results based on aggregated data will be shared, so that no individual respondent can ever be identified.

Thank you in advance for your cooperation with this project. If you have any comments about the survey or project, feel free to contact principal investigator, Ashish Gupta, Department of Management Science and Information Systems, Oklahoma State University-Tulsa, 372A North Hall, 700 North Greenwood Avenue, Tulsa, Ok 74106-0702, ph. 918-594-8583, Email: ashish.gupta@okstate.edu or Dr. Ramesh Sharda at sharda@okstate.edu, or Dr. Sue Jacobs at sue.c.jacobs@okstate.edu.

OSU
Institutional Review Board
Approved 11/3/05
Expires 11/2/06
Initials oj
BU0611

OKLAHOMA STATE UNIVERSITY

OSU
TULSA

Academic Affairs
700 North Greenwood Avenue
Tulsa, Oklahoma 74106-0700
918-594-8013; Fax: 918-594-8023

November 9, 2005

Dr. Raj Basu, Vice-President
Academic Affairs
OSU-Tulsa
Tulsa, Ok 74106

Dear Dr. Basu,

I am conducting a study titled "Understanding email usage within organizations", which is a part of my dissertation work under the supervision of Dr. Ramesh Sharda (Department of MSIS). I am interested in using OSU-Tulsa faculty and staff members as subject for this study. An IRB approval has already been granted for this purpose. I want to request you to please grant me permission to circulate an email and a possible follow-up reminder to OSU-Tulsa faculty and staff members. The script of the email is presented below:

Email Subject: Need help in filling survey on Email Usage at work

Dear OSU-Tulsa faculty and staff members,

I am conducting a study, which is a part of my dissertation work under the supervision of Dr. Ramesh Sharda (Department of MSIS) to examine the nature of email work environment of individuals and to understand how they process their emails. I would like to request you to please share your opinions on characteristics and volumes of email that you receive by completing a survey at http://fp.okstate.edu/guptaa/other.htm . This survey should take about 10-12 minutes to complete. Your help in this effort will be greatly appreciated. However, your participation is completely voluntary and all information you provide, including your email address, will be kept completely confidential. Only reports of study results based on aggregated data will be shared, so that no individual respondent can ever be identified.

In return for your time, you will receive reports that let you compare your experience with that of organizations of similar size. Alternatively, if you're responding as a private individual, then the reports will compare your experience with those of other users. The reports will cover:

- Email volumes handled by people
- The time spent on emails
- The types of email people receive
- How people prioritize and categorize their email?

Thank you in advance for your cooperation with this project. If you have any comments about the survey or project, feel free to contact principal investigator, Ashish Gupta, Department of Management Science and Information Systems, Oklahoma State University-Tulsa, 372A North

Hall, 700 North Greenwood Avenue, Tulsa, Ok 74106-0702, ph. 918-594-8583, Email: ashish.gupta@okstate.edu or Dr. Ramesh Sharda at sharda@okstate.edu, or Dr. Sue Jacobs at sue.c.jacobs@okstate.edu.

Regards,

Ashish Gupta
Visiting Assistant Professor
Department of MSIS
Oklahoma State University-Tulsa
Phone: (918)-594-8583
Fax: 918-594-8281
URL: http://fp.okstate.edu/guptaa/

Thank you so much in helping me in this research effort and look forward to hearing from you!

Sincerely,

Ashish Gupta
Visiting Assistant Professor
Department of MSIS
Oklahoma State University-Tulsa
Phone: (918)-594-8583
Fax: 918-594-8281
URL: http://fp.okstate.edu/guptaa/

Approved

VITA

Ashish Gupta

Candidate or the Degree of

Doctor of Philosophy

Thesis: UNDERSTANDING THE RHYTHMS OF EMAIL PROCESSING STRATEGIES IN A NETWORK OF KNOWLEDGE WORKERS

Major Field: Business Administration

Biographical:

Education:

Bachelor of Engineering (Mechanical Engineering) from Marathwada University in June, 1998
Master of Science (Industrial Engineering & Management) from Oklahoma State University in May 2001
Completed the requirements of a Doctor of Philosophy at Oklahoma State University in May, 2007

Experience:

Graduate Research Assistant, Oklahoma State University, Stillwater, (Jan 2001 – May 2001)
Graduate Teaching Associate, Oklahoma State University, Stillwater (Aug 2002- Dec 2004)
IRIS Research Associate, Oklahoma State University, Stillwater (Jan 2005- Aug 2005)
Visiting Assistant Professor of MSIS, Oklahoma State University-Tulsa Campus (2005- 2006)
Assistant Professor of Operations Management, Minnesota State University Moorhead (Fall 06- Present)

Professional Affiliation:

Association of Information Systems, Decision Science Institute, INFORMS, INSNA

Name: Ashish Gupta Date of Degree: May, 2007

Institution: Oklahoma State University Location: Stillwater, Oklahoma

Title of Study: UNDERSTANDING THE RHYTHMS OF EMAIL PROCESSING STRATEGIES IN A NETWORK OF KNOWLEDGE WORKERS

Pages in Study: 178 Candidate for the Degree of Doctor of Philosophy

Major Field: Management Science and Information Systems

Scope and Method of Study: While emails have improved the communication effectiveness of knowledge workers, they have also started to negatively impact their productivity. Emails have long been known to provide value to the organization, but the influence of the overwhelming amount of information shared through emails and the inefficiencies surrounding the everyday use of emails at work has remained almost completely unanalyzed so far. Frequent announcements of new emails and then a user's checking her email leads to an escalation in the interruption issues, the resulting overall effectiveness derived from email communication needs to be re-explored. This study uses a computational modeling approach to understand how various combinations of timing-based and frequency-based email processing strategies adopted within different types of knowledge networks can influence average email response time, average primary task completion time, and the overall effectiveness, comprising value-effectiveness and time-effectiveness, in the presence of interruptions. Earlier research on the topic has focused on individual knowledge workers. This study performs a network-level analysis to compare different sender-receiver relationships to assess the impact of different overall email policies on the entire network. Computational models of three different email exchange networks were developed, namely, homogeneous networks with higher users of email, homogeneous networks with low users of email and heterogeneous networks utilizing various combinations of email strategies. A new method, referred to as forward and reverse method, to evaluate and validate model parameters is also developed.

Findings and Conclusions: Findings suggest the choice of email checking policy can impact time and value effectiveness. For example, rhythmic email processing strategies lead to lower value-effectiveness but higher time-effectiveness for all types of networks. Email response times are generally higher with rhythmic policies than with arrhythmic policies. On the other hand, primary task completion times are usually lower with rhythmic policies. On an average, organizations could potentially save 3 to 6 percent of overall time spent per day by using email strategies that are more time effective but could lose 2.5 to 3.5 percent in the communication-value. These values cumulate into significant time saving or value loss for large organizations.

ADVISER'S APPROVAL: Dr. Ramesh Sharda