Personality Trait Based Simulation Model of the E-mail System

Mark Jyn-Huey Lim[1], Michael Negnevitsky[1], and Jacky Hartnett[2]
(Corresponding author: Mark Jyn-Huey Lim)

School of Engineering, University of Tasmania[1]
Hobart, Tasmania 7001, Australia (Email: {mjlim, Michael.Negnevitsky}@utas.edu.au)
School of Computing, University of Tasmania[2]
Hobart, Tasmania 7001, Australia (Email: J.Hartnett@utas.edu.au)

(Received Aug. 26, 2005; revised and accepted Oct. 11, 2005)

Abstract

Within the area of criminal and terrorist social network analysis, there is little research being done on analysing the communication behavior of criminal and terrorist groups. In this paper, we describe the development of a conceptual simulation model of the e-mail system, which is based on the use of personality trait dimensions to model the e-mail traffic behavior of e-mail users. This conceptual simulation model is being used as a first step for further development in simulating criminal and terrorist communication behavior. We also describe the development of an e-mail traffic analyser system, which uses a decision tree to search for interesting e-mail traffic behavioral patterns, and uses social network and time-series visualisation to provide the details of the interesting traffic patterns. We demonstrate that the personality trait based e-mail system model is useful as a tool for generating e-mail traffic behavioral patterns and that decision trees are useful for finding interesting patterns in the e-mail traffic data.

Keywords: e-mail, simulation, personality traits, visualization, traffic analysis

1 Introduction

Intelligence gathering and intelligence analysis have long been an essential part of law enforcement agencies for obtaining information about the nature and activities of criminal and terrorist groups [5]. Much of the problems faced by intelligence analysts today is not caused by the lack of information available, but by the fact that there is too much information to sift through [9]. The difficulty faced by intelligence analysts is in finding suitable tools and techniques to help them sort through a large variety of information sources (e.g. e-mail, Internet web pages, financial records, phone call logs), in order to monitor criminal and terrorist organisations.

In the area of criminal and terrorist social network analysis, research has focused on techniques and tools for analysing the static structure of criminal/terrorist social networks, which provide a snapshot of the social network at a particular point in time. Examples of studies based on the static structure of criminal/terrorist networks are the examination of criminal incident reports to visualise the structure of a criminal network [28], and the examination of news reports to extract the social network structure of the 19 terrorist hijackers involved in the September 11 attacks [16]. The tools and techniques described by [16, 28] can provide intelligence analysts useful ways of examining the structure of a criminal/terrorist organisation, but do not show how the social networks change over time.

Recent work in criminal/terrorist social network analysis have been investigating how the structure of criminal and terrorist social networks evolve over time, by trying to understand the dynamics of the social networks. In [29] for example, they use yearly sequences of criminal incident reports to build up information on the annual changes to the structure of a criminal social network. From their yearly data of the criminal social network, they are able to use social network analysis quantitative measures (degree, betweenness, closeness, link density, group cohesion, and group stability) to analyse how these network properties change as the criminal social network evolves over time. In [7], a different approach is taken to analyse social network dynamics where they use multi-agent technology to simulate the evolution of a terrorist network.

While current research on criminal/terrorist social networks focuses on analysing the stable and changing structure of the social networks, little work has been done on developing tools and techniques for analysing the communication patterns or communication behavior of criminal/terrorist groups. The need for communication analysis tools is important, because without such tools criminal and terrorist organisations can plan attacks or schedule

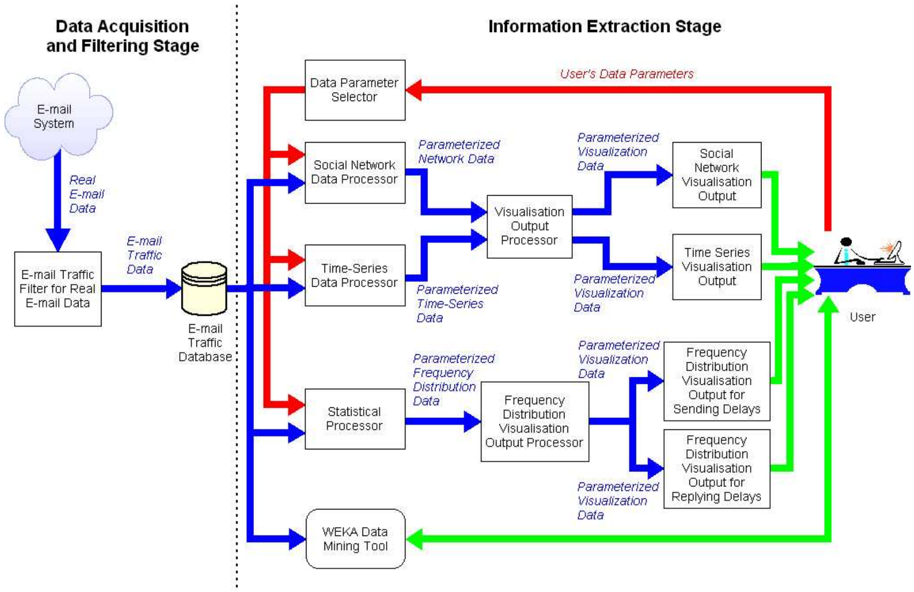

Figure 1: Overview of e-mail traffic analyser system

events without being detected by law enforcement agencies. The availability of a suitable communication analysis tool, will allow law enforcement agencies to analyse the communication patterns and social network structure of criminal/terrorist organisations, in order to better predict the likelihood of an upcoming criminal or terrorist event and identify potential suspects.

Our approach for analysing communication patterns is to analyse e-mail traffic data, given that e-mail is a widely used form of electronic communication and is readily available for data collection. To perform the analysis of e-mail traffic data, we plan to investigate using artificial intelligence techniques such as decision trees and neural networks, to extract information from the e-mail traffic data to identify patterns of behavior for e-mail users. The investigation of these techniques will then enable us to determine which of the artificial intelligence techniques would be most effective in a system that is able to build a profile of e-mail users and learn about their e-mail communication behavior. Such a system using the appropriate type of technique would assist in determining the presence of unusual communication activities (i.e. indications of criminal or terrorist activities).

In this paper, we firstly provide an overview of the e-mail traffic analyser system, which we are initially developing to analyse basic e-mail traffic behavioral patterns. In the second part, we describe the conceptual simulation model of the e-mail system that we have developed, and how it is being used to generate simulated e-mail data for the e-mail traffic analyser system. The third part of the paper describes some visualisation tools that we have used as part of the e-mail traffic analyser system to assist in visually analysing the simulated e-mail data for patterns of behavior. We finally provide a case study describing the simulation of a small network of e-mail users.

2 Overview of E-mail Traffic Analyser System

For our project, we are currently developing an e-mail traffic analyser system that will enable us to extract different types of e-mail traffic behavioral patterns to obtain information on the behavior of e-mail users. The different types of e-mail traffic behavioral patterns we are analysing are: the social connections between e-mail users, the level of e-mail usage by e-mail users (e.g. number of e-mails sent per day, number of e-mails received per day), and the level of interaction between different e-mail users (e.g. how often a user sends e-mail to a particular individual, how quickly a user responds to e-mails received from other users). These e-mail traffic behavioral patterns are being analysed by the e-mail traffic analyser system, to allow us to explore the data and extract "interesting" behavioral patterns from each e-mail user. Examples of "interesting" e-mail traffic behavioral patterns could be where an e-mail user suddenly starts sending more e-mails to a particular individual, where an e-mail user stops communicating with a particular individual, or a period of time where there is a change in the level of interactions between particular e-mail users. The methods for extracting the "interesting" e-mail traffic behavioral patterns will be based on artificial intelligence techniques, which forms the main part of our investigations with the e-mail traffic analyser system.

In the e-mail traffic analyser system we have developed so far, e-mail data is processed through several stages to ensure that the data is clean and is in a suitable format before information is extracted [15]. The diagram in Figure 1 provides an overview of the e-mail traffic analyser system and shows how the e-mail data from the e-mail system is processed to enable for behavioral patterns to be

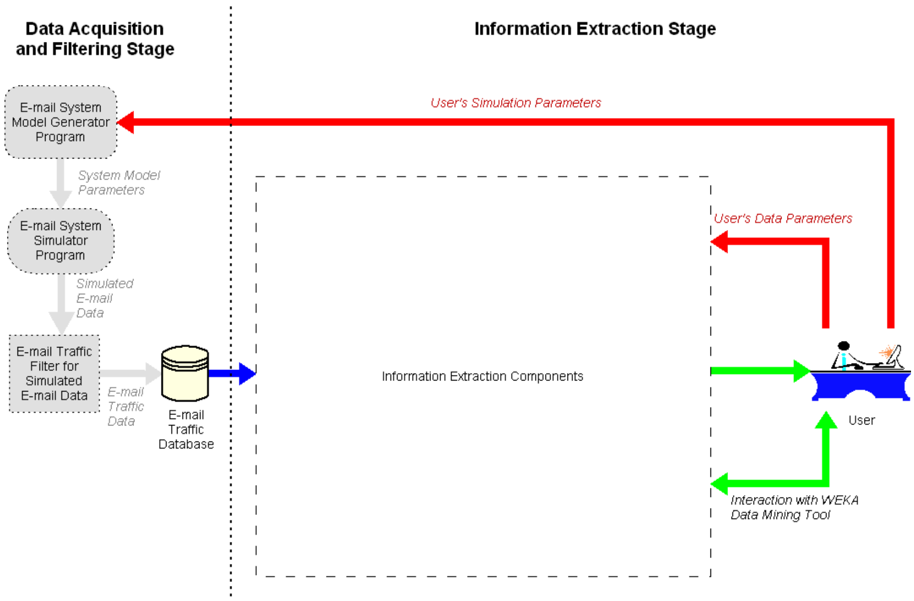

Figure 2: Overview of how the e-mail system simulator is used with the e-mail traffic analyser system

extracted for different e-mail users. There are two main stages for processing the e-mail data in the e-mail traffic analyser system: the "Data Acquisition and Filtering Stage" and the "Information Extraction Stage".

At the first stage, "Data Acquisition and Filtering Stage", e-mail data is collected from the e-mail system and cleaned/filtered to fill in missing values, remove noise, and fix up inconsistent data. Once the e-mail data is filtered and cleaned it is stored into the e-mail traffic database. At the second stage, "Information Extraction Stage", information from the e-mail traffic data is extracted and processed through different components to provide details on the e-mail traffic behavioral patterns of social connections ("Social Network Data Processor" component), level of e-mail usage ("Time-Series Data Processor" component), and level of interaction between e-mail users ("Statistical Processor" component). After these e-mail traffic behavioral patterns have been extracted and processed, they are visualised and presented to the user for analysis (see Section 4 for details on visualisation).

However, it is a difficult task for the user to pick out the "interesting" patterns from the visualisation outputs, because the user is presented with so much visual information. So, to assist with the task of finding the "interesting" e-mail traffic behavioral patterns, the WEKA Data Mining Tool program [27] is being used in the "Information Extraction Stage" of the e-mail traffic analyser system, to assist the user in finding the "interesting" patterns from the e-mail traffic data. After the user has found some interesting patterns through the use of the WEKA Data Mining Tool, they can then focus their attention on the areas in the e-mail data where the interesting e-mail traffic behavioral patterns were found. This can be done through the use of the "Data Parameter Selector" component, which allows the user to "zoom-in" on the details of the interesting patterns by the specifying a combination of different data selection parameters (e.g. a specific period of time, a specific group of e-mail users, selecting all e-mail messages being received a particular user). The process of extracting information from the e-mail traffic data, presenting the information to the user, and allowing the user to focus on particular details in the e-mail traffic data, provides a more interactive way for the user to analyse the e-mail traffic data for interesting behavioral patterns.

To help develop and test the e-mail traffic analyser system, a conceptual simulation model of the e-mail system has been developed to generate different types of behavioral traffic patterns for the analyser system to examine. The simulated e-mail system is used to substitute for the data provided by the real e-mail system, as shown in Figure 2. The simulation model is described in detail in the next section.

3 Simulation Model

The purpose of the e-mail system simulation model is to provide us with a tool for generating different types of behavioral traffic patterns in the e-mail data and enable us to create different scenarios for analysis by the e-mail traffic analyser system. Our reason for building a simulation model rather than using real e-mail data for our area of work is due to two issues. The first issue is that for the e-mail data we collect, we need to know what kind of behavioral traffic patterns is being generated by e-mail users, in order for us to know whether the artificial intelligence techniques being used are accurately detecting the correct type of "unusual" user behavioral patterns avail-

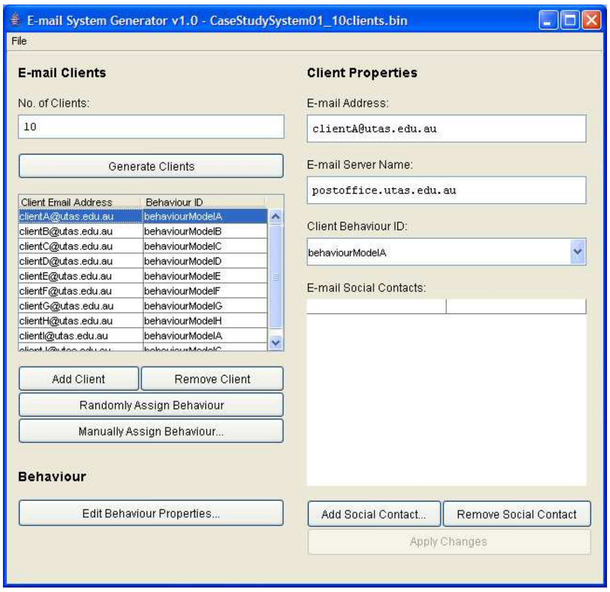

Figure 3: The main graphical user interface for the e-mail system model generation program

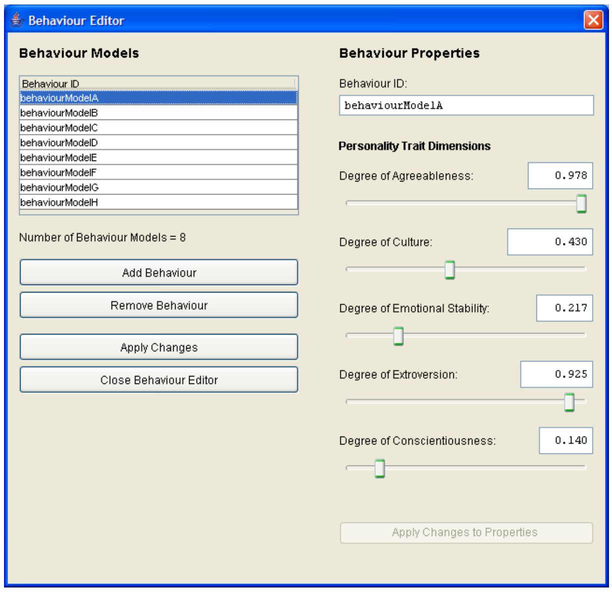

Figure 4: The behavior editor part of the e-mail system model generation program

able in the e-mail data. Without some prior knowledge about the nature of the data collected (e.g. certain e-mail clients are exhibiting criminal or terrorist communication behavior), it will be difficult to test whether the artificial intelligence techniques being used are detecting the correct behavioral traffic patterns. A second issue is that it is difficult for us to readily obtain e-mail data from a reasonably large number of e-mail users (e.g. around 10,000 users). Without enough data for analysis, we would end up with incomplete social network information (i.e. missing nodes/e-mail users from the network) and will miss out on traffic information available at the missing network nodes. Due to these two issues, we have chosen to use a simulation model because it will allow us to generate any number of e-mail users, and the ability to simulate criminal or terrorist type communication behavior using available theories on personality types [1, 8, 20] and the psychology of criminal and terrorist behavior [2, 10, 14].

For the development of our e-mail system simulation model, we are initially developing a conceptual simulation model of the e-mail system, to allow us to focus on modelling the basic personality types of different e-mail users. After the conceptual simulation model has been developed, we will extend the simulation model to include different types of criminal and terrorist behavioral models, in order to simulate the behavior of criminal or terrorist activities. In this section, we will describe the conceptual simulation model of the e-mail system and show how it is being used to generate different types of e-mail traffic behavioral patterns.

3.1 Overview of the E-mail System Simulator

There are several stages required to simulate the e-mail system and generate the simulated e-mail traffic data. These include building the e-mail system model, simulating the e-mail system model, and transferring the simulated e-mail data into the e-mail traffic database (Figure 2). For the building of the e-mail system model, an e-mail system model generator program was developed using the Jython programming language [13], to allow the user to easily design and specify the parameters of the e-mail system model. The graphical user interface for this program is shown in Figures 3 and 4. Once the user has specified the parameters for the e-mail system model and created a scenario that they would like to simulate, the model data is then passed to the e-mail system simulator program (written in Python [21]), which simulates the e-mail system model and outputs the simulated e-mail data. After the simulated e-mail data has been output by the simulator, it is then filtered and stored into the e-mail traffic database.

3.2 Conceptual E-mail System Model

The conceptual simulation model of the e-mail system has been modelled in the e-mail system simulator program as a discrete-event simulation system [4, 18] and comprises of two main types of entities: e-mail clients and behavior models. The e-mail client entities represent the e-mail accounts of e-mail users, through which e-mail messages are sent from and received by each e-mail account. The diagram in Figure 5 shows that the attributes of each e-mail client consist of an e-mail address, an e-mail server

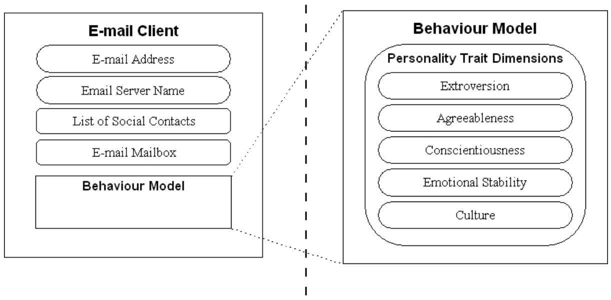

Figure 5: The entities making up the e-mail system model

name, a list of social contact e-mail addresses, an e-mail mailbox, and an assigned behavior model. The behavior model entities, shown in Figure 5, represents the behavioral attributes of different individuals and defines how an individual will interact and respond to others via e-mail. A behavior model is assigned to different e-mail clients, to define how the client's e-mail account will behave. The attributes of each behavior model is represented by a set of five personality traits, which are based upon the five personality trait dimensions described in [1, 8, 20] (personality traits will be described in more detail in Section 3.3).

Our basis for modelling the e-mail system as e-mail client and behavior model entities, is to allow us to focus on modelling the interactions between different e-mail clients, rather than trying to model e-mail client and e-mail server interactions. This is because we are interested in the interaction behavior of individual e-mail clients, in terms of how often they send e-mails to other clients (e.g. several times per day, or once every few days) and how they respond to e-mails that they receive from other e-mail clients (e.g. usually replying back on the same day, or usually replying back within one week of receiving an e-mail). In this way, we are able to set up the e-mail system model as a series of end-to-end communication points between different e-mail clients, as opposed to a more centralised set-up where the e-mail servers are situated in between the lines of communication between e-mail clients.

Another reason for modelling the e-mail user's behavior (behavior model) as a separate entity from the e-mail account details (e-mail client), is to allow us to define behavior characteristics that uniquely represent different individuals. In this way, each individual modelled generates a unique set of e-mail traffic behavioral patterns that can be linked back to attributes that make up the individual's behavioral profile. For example, the behavior for one individual may be that they send out several e-mails per day and respond to e-mails within one or two days after receiving the original e-mail message. This type of e-mail traffic behavior pattern can be linked back to attributes of the individual's behavioral profile, which describe the individual as 'outgoing' and 'conscientious'. Alternatively, there could be another individual that has a different set of e-mail traffic behavioral patterns, in which that individual also sends out several e-mails per day, but always takes at least one week to respond to e-mails they receive or does not respond at all. This type of behavioral traffic pattern can also be linked back to the individual's behavioral profile, which describes this individual as 'outgoing', but 'lazy'. From these examples, we can already see that focusing on the behavior of each e-mail user also provides a useful way for describing the behavioral attributes that cause a person to exhibit certain types of e-mail traffic behavior. As a result of modelling the behavior model and e-mail client as separate entities, a behavior model can be assigned to multiple e-mail accounts to represent how some people use several e-mail accounts, but still exhibit similar types of e-mail traffic behavioral patterns through different e-mail accounts. So through this modelling approach, a model of the e-mail system can be created that consists of individuals with distinct types of behavior profiles, using various e-mail accounts.

3.3 Personality Traits of the Behavior Model

The behavior model in the e-mail system model is made up of five personality trait dimensions, consisting of the dimensions of extroversion, agreeableness, conscientiousness, emotional stability, and culture. Each of these personality trait dimensions is used to describe a particular aspect of person's personality and explain how a person's underlying traits is linked to their observed behavior [1, 8, 20]. For example, if a person is described as being quite extroverted, it implies that they are observed to be very outgoing, social, and talk to a lot of people. Likewise, if a person is described as being introverted (the opposite of being extroverted), it implies that they are observed to be very shy, non social, and don't talk a lot to people. A second example is if a person is described as being very conscientious, it would imply that their behavior is helpful, hard working, and dependable. For a person described as low conscientiousness, it implies their behavior is careless, lazy, and not dependable. In Table 1, adapted from [1, 8], each of the personality trait dimensions is described by trait pair examples, which are used to describe the nature of a person's underlying tendencies and imply about their observed behavior.

Each of the personality trait dimensions in the behavior model is assigned a degree value between 0 to 1. These degree values provide a scale of how strongly each personality trait dimension affects a person's behavior [1], with values close to 1.0 (high degree values) indicating a strong influence by the personality trait dimension and values close to 0.0 (low degree values) indicating a weak influence by the personality trait dimension. So for example, a degree value of 0.9 for extroversion and 0.7 for conscientiousness trait dimensions, describes a person that is extremely outgoing, hard working, and fairly responsible. Alternatively, a degree value of 0.2 for extroversion and 0.1 for conscientiousness trait dimensions, describes

Table 1: Description of the personality trait dimensions, through the use of trait pair examples

Personality Trait Dimensions	**Trait Pair Examples**
Extroversion-introversion	Talkative-silent; frank-secretive; adventurous-cautious; sociable-reclusive;
Emotional Stability	Calm-anxious; composed-excitable; poised-nervous;
Conscientiousness	Tidy-careless; responsible-undependable; hard working-lazy; persevering-quitting;
Culture	Creative-uncreative; intellectual-nonreflective; well educated-crude; prefers variety-prefers routine;
Agreeableness	Good natured-irritable; gentle-headstrong; cooperative-negativistic; not jealous-jealous;

a person that is rather withdrawn, lazy, and unreliable. By assigning different degree values to each set of personality trait dimensions, unique behavioral profiles can be created that represent different individuals in the e-mail system model, each of whom exhibit varying degrees of behavior for each of the personality trait dimensions.

Since it has been shown that personality trait dimensions and trait degree values provide a way of conceptualising the effects of personality trait dimensions on general behavior, we can now describe how the personality trait dimensions can be modelled to affect e-mail communication behavior. However, given that there has been few empirical studies conducted that examine the effects of personality traits on e-mail communication behavior, some intuitive assumptions were made on how the personality trait dimensions affect e-mail behavior. Table 2 provides a summary of how some of the personality trait dimension attributes from the behavior model will be interpreted to affect e-mail communication behavior, based on our intuitive assumptions.

For the extroversion trait dimension, it has been assumed that this will affect how an individual engages in communication with others by e-mail. This means that it will be interpreted as affecting the frequency of e-mails sent out by an e-mail client and affecting the speed that an e-mail client replies to received e-mails. In the conscientiousness trait dimension, it has been assumed that this will affect the how an individual acknowledges and responds to others via e-mail. This will be interpreted as affecting the probability that an e-mail client will send a reply message when they receive an e-mail message from other e-mail clients. For the emotional stability trait dimension, it has been assumed that this will determine how the emotional element of an individual will affect the individual's ability to be consistent with their communication behavior (i.e. we assume that an emotionally stable person can think more clearly and hence act more consistently, compared to a person who is emotionally less stable and acts more erratically due to their inability to think clearly or rationally). In the case of the emotional stability trait dimension, it will be interpreted as affecting the variability in the time delays between e-mails sent and affecting the variability in the time delay for a reply message.

It has been considered that the personality trait dimensions of the behavior model could also be modelled to affect other types of e-mail communication behavior such as the addition of e-mail attachments, sending e-mails to multiple recipients, and forwarding received e-mails to other e-mail clients. However, it was decided that the conceptual model of the e-mail system should be kept as simple as possible, to allow us to focus on the essential communication behaviors required for e-mail communication interaction. As a result, we have a behavior model that affects the sending and replying behavior of e-mail clients, with each interaction consisting of one sender and one recipient.

3.4 Sending and Replying Delay Distributions

To implement the relationships shown in Table 2 during the simulation of the conceptual e-mail system model, a normal distribution is used to generate the time delays between each e-mail message sent by e-mail clients and also to generate the time delays when e-mail clients reply to e-mail messages received from other e-mail clients. Each of the normal distributions used is controlled by the personality trait dimension degree values taken from the behavior model assigned to each e-mail client. For the sending delay normal distribution, this is controlled by the extroversion and emotional stability trait dimension degree values. Similarly, the replying delay normal distribution is controlled by the extroversion, emotional stability, and conscientiousness trait dimension degree values.

The sending delay and replying delay normal distributions both use a 7-day window interval (see Figures 6 and 7), to limit the possible values for the sending and replying delays to within one week. The use of the 7-day window interval was based on a study from [17], where participants were found to reply to e-mails within one day or within one week. Apart from the study conducted by [17], there have been few other empirical studies that cover either the sending delays or replying delays of e-mail users. So using the results shown by [17], the model implements a delay of up to one week.

For the sending delay normal distribution, this is defined by the normal distribution function $N(m, \sigma)$, rep-

Table 2: The relationship between the effect of personality trait dimensions on e-mail communication behavior, based on intuitive assumptions

Personality Trait Degree Values	**Effect on Sending Out Of New E-mails**	**Effect on Replying To Received E-mails**
High Degree of Extroversion	Higher frequency of sent e-mails	More likely to receive a fast reply
Low Degree of Extroversion	Lower frequency of sent e-mails	More likely to receive a slow reply
High Degree of Conscientiousness	-	Higher probability of returning a reply
Low Degree of Conscientiousness	-	Lower probability of returning a reply
High Degree of Emotional Stability	Lower variability in time delays between e-mails sent	Lower variability in replying delay time
Low Degree of Emotional Stability	Higher variability in time delays between e-mails sent	Higher variability in replying delay time
High Degree of Culture	-	-
Low Degree of Culture	-	-
High Degree of Agreeableness	-	-
Low Degree of Agreeableness	-	-

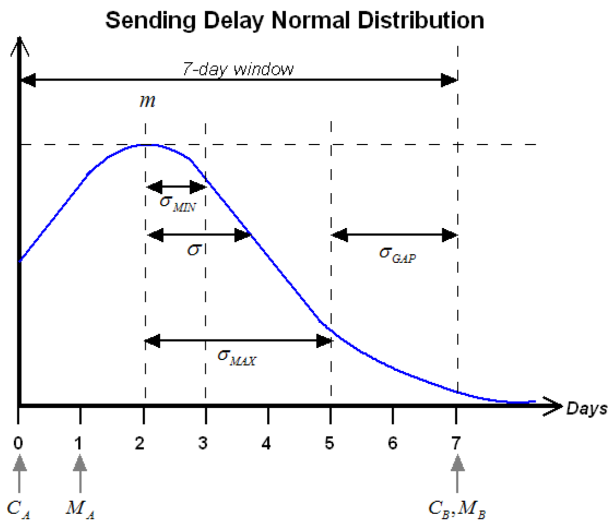

Figure 6: Layout for the sending delay normal distribution

resented in Figure 6, where m is the mean of the normal distribution and σ is the standard deviation of the normal distribution. The mean of the distribution is given by:

$$m = M_B - D_{EX}(M_B - M_A) \tag{1}$$

where M_A and M_B represent the start and end range for mean values $[M_A = 1, M_B = 7]$, and D_{EX} is the degree of extroversion with a value between the interval $[0, 1]$. For the standard deviation of the distribution, this is determined by:

$$\sigma_{MAX} = (C_B - C_A)/2 - \sigma_{GAP} \tag{2}$$

$$\sigma = \sigma_{MAX} - D_{ES}(\sigma_{MAX} - \sigma_{MIN}) \tag{3}$$

where C_A and C_B represent the start and end cut-off points for the 7-day window interval $[C_A = 0, C_B = 7]$, σ_{GAP} is the gap allowance between the standard deviation and the cut-off point for the 7-day window, σ_{MIN} and σ_{MAX} represent the minimum and maximum values for the standard deviation, and D_{ES} is the degree of emotional stability with a value between the interval $[0, 1]$.

To select a sending delay time from the normal distribution, a random number x_S is picked from the distribution $N(m, \sigma)$, such that x_S is within the 7-day window interval $[C_A, C_B]$, where $C_A = 0$ days and $C_B = 7$ days. Once x_S is selected from $N(m, \sigma)$, x_S is used as the sending delay between the previous e-mail message and next e-mail message to be sent by an e-mail client. A new sending delay value x_S is picked from the distribution each time before an e-mail client sends a new e-mail message, to determine when the new e-mail message will be sent.

For the replying delay normal distribution, this is defined by a similar normal distribution function $N(m', \sigma)$,

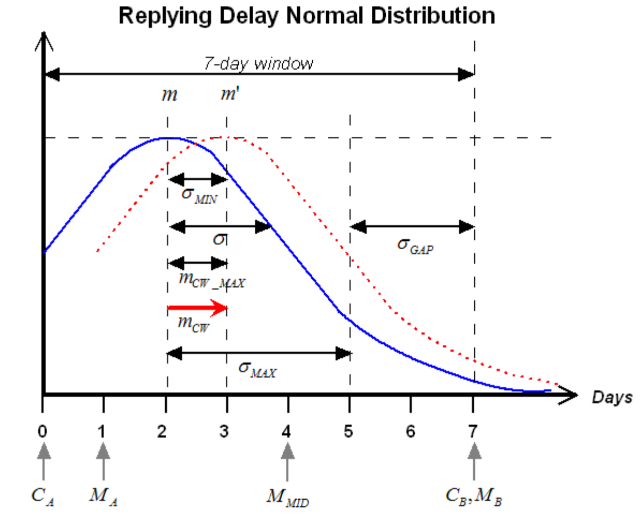

Figure 7: Layout for the replying delay normal distribution

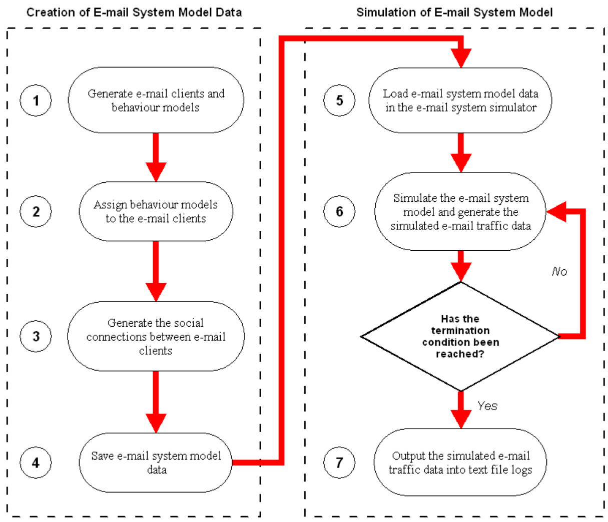

Figure 8: Flow diagram of the simulation model set-up and simulation process

as shown in Figure 7, where m' is the mean and σ is the standard deviation. The mean for the replying delay normal distribution is determined in two steps. The first step calculates the mean in the same way as for the sending delay distribution in Equation (1). The second step makes use of the conscientiousness trait degree value and calculates the resulting mean m' as follows:

$$M_{MID} = \frac{M_B - M_A}{2} + M_A \tag{4}$$

$$m_{CW} = \begin{cases} \frac{D_C - 0.5}{0.5} \times m_{CW\_MAX} \times \frac{M_{MID} - m}{M_{MID} - M_A}, \\ \qquad m < M_{MID} \\ -\frac{D_C - 0.5}{0.5} \times m_{CW\_MAX} \times \frac{m - M_{MID}}{M_B - M_{MID}}, \\ \qquad m \geq M_{MID} \end{cases} \tag{5}$$

$$m' = m + m_{CW} \tag{6}$$

where M_{MID} represents the middle of the interval $[M_A, M_B]$, D_C is the degree of conscientiousness value with a value between the interval $[0, 1]$, m_{CW} is the conscientiousness weight factor for the mean, m_{CW_MAX} is the maximum value for the conscientiousness weight factor, and m' is the mean of the normal distribution after the conscientiousness weight factor m_{CW} has been added to the original mean m. For the standard deviation of the distribution, this is determined in the same way as for the sending delay normal distribution as shown in Equation (3).

To select a replying delay time from the normal distribution, a random number x_R is picked from the distribution $N(m', \sigma)$, such that x_R falls anywhere along the $N(m', \sigma)$. When $x_R < C_A$ and $x_R > C_B$, it will be considered that the e-mail client will not be replying to a received e-mail message, given that x_R is outside the 7-day window interval. When $C_A \leq x_R \leq C_B$, then x_R will be used as the replying delay time value, since x_R is within the 7-day window interval. A new replying delay value x_R is picked from $N(m', \sigma)$ each time an e-mail client receives an e-mail message from another client.

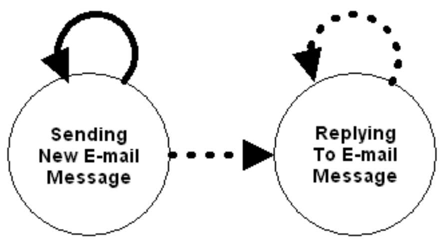

Figure 9: Events diagram for events in the conceptual e-mail system model

3.5 Simulation of E-mail System Model

The process for simulating the conceptual e-mail system model consists of several steps, as shown in Figure 8. In the first stage "Creation of E-mail System Model Data" (Steps 1 to 4), the e-mail system model generator program is used to create the set-up parameters for the conceptual e-mail system model (as seen in Figures 3 and 4). The first step consists of generating the e-mail clients and behavior models, where the number of e-mail clients and behavior models created is such that:

$$N_e \quad \geq \quad N_b \tag{7}$$

where N_e represents the number of e-mail clients and N_b represents the number of behavior models. In Step 2, all behavior models are assigned to e-mail clients, so that

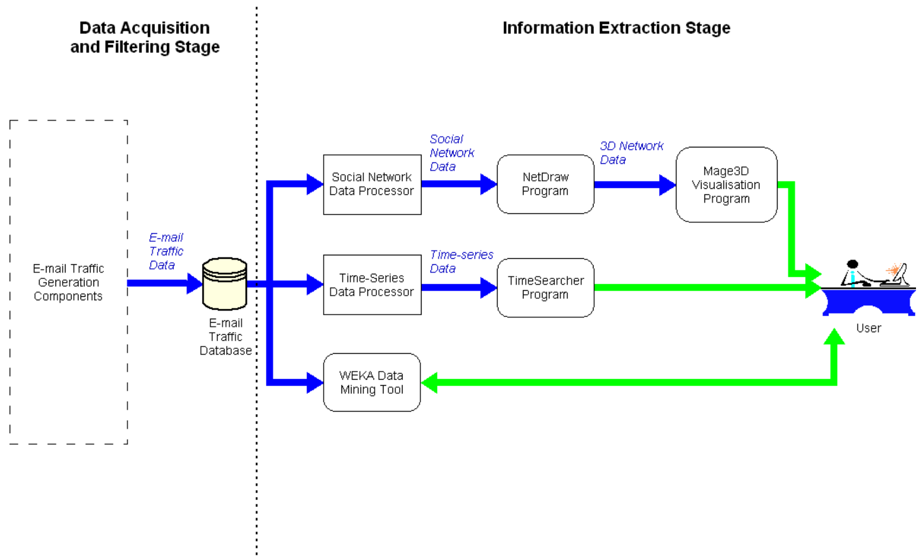

Figure 11: Overview of programs currently used in the implementation of the e-mail traffic analyser system

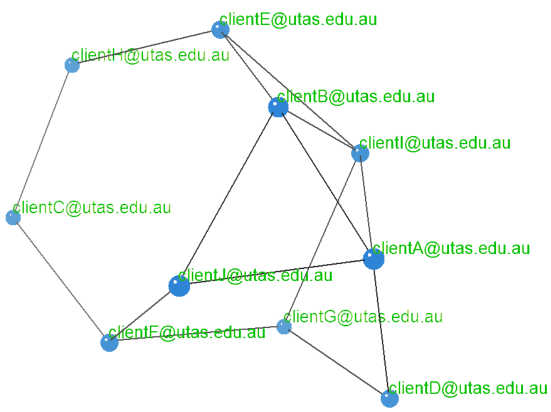

Figure 10: Visualisation of a simulated social network of 10 e-mail clients

each behavior model is at least assigned to one e-mail client each. For the case where $N_e > N_b$, some behavior models may be assigned to more than one e-mail client. Steps 1 and 2 can be repeated as necessary where the e-mail system model generator is used to automatically assign random attribute values for the e-mail clients and behavior models. Alternatively the user can use the e-mail system model generator program to manually adjust the attribute values of the e-mail clients and behavior models. Once the attribute settings and behavior models have been allocated to e-mail clients, the third step consists of generating the social connections between e-mail clients, so that each e-mail client has a list of e-mail addresses of social contacts they will communicate with during the simulation. The final step for the first stage, Step 4, is where the e-mail system model data is saved by the e-mail system model generator program to store the parameter settings and configuration of the e-mail system model.

The next stage of the simulation process, "Simulation of E-mail System Model" (Steps 5 to 7), uses the e-mail system simulator program to simulate the conceptual e-mail system model. In Step 5, the saved e-mail system model data is loaded into the e-mail system simulator program, to read the set-up parameters and configuration settings of the e-mail system model. After the data is loaded into the simulator, the simulation run is started in Step 6.

During the simulation run, two different types of events will be generated by the e-mail clients: 'sending' events (new e-mail messages sent by e-mail clients to randomly selected social contacts) and 'replying' events (e-mail messages sent in reply to a previously received e-mail message). A representation of the events generated by the e-mail clients is shown in Figure 9, which shows how 'sending' events are continuously generated by e-mail clients, whereas 'replying' events can only be triggered by a 'sending' event or another 'replying' event as a result of an e-mail client receiving an e-mail message. Every time a 'sending' or 'replying' event is executed by the simulator, the simulator records a log of the message in the mailbox of each e-mail client involved in the message interaction. The scheduling of delays between e-mail messages sent by each e-mail client and the delay for an e-mail client to reply to a previously received message, is determined by the sending delay and replying delay normal distributions described in Section 3.4.

To terminate the simulation run, the simulator will check whether the termination condition for the simulation has been reached. The termination condition can be specified either as being dependent on M total number of e-mail messages being sent by e-mail clients (e.g.

$M = 10,000$ messages), or as being dependent on when the simulation time reaches the specified value T (e.g. $T = 120$ simulation days). After the simulation run is terminated, in Step 7 the simulated e-mail traffic data from each e-mail client's mailbox is stored into text files. These text files are then filtered and put into the e-mail traffic database as shown in Figure 2, ready to be visualised by the e-mail traffic analyser system.

4 Visualisation Output

The visualisation of data has proven to be a useful way of analysing data for hidden relationships between different variables [23]. There are a variety of different visualisation techniques available for visualising data and each technique has its own advantages for visualising data in a particular way. Here we will describe four different techniques that are being used to visually analyse e-mail traffic data in the e-mail traffic analyser system.

4.1 Social Network Visualisation

Social network visualisation provides a visual image of the communication links or social connections between different individuals [12]. This type of visualisation, shown as the visualisation output from the "Social Network Data Processor" in Figures 1 and 11, is useful for analysing e-mail traffic communication in that it enables the user to observe the relationship in the communication ties between various e-mail clients. The ability to see the relationships between various e-mail clients through social network visualisation, makes it much easier for the user to gain an overall view of all the communication links between e-mail clients and spot areas of interest in the e-mail social network, such as clustering of different types of users into distinct social groups or communities [11, 26]. An example of a visualised social network from our simulated e-mail system is shown in Figure 10, which was visualised by exporting social network data from the Netdraw program [6] into the Mage 3D Visualisation program [22].

4.2 Time-Series Visualisation

Time-series visualisation presents a one-dimensional view of data by showing how it changes over time. This form of visualisation, shown in Figures 1 and 11 as the visualisation output from the "Time Series Data Processor", enables the user to analyse the volume of e-mail traffic (e.g. number of e-mails sent per hour, number of e-mails sent per day) being generated by each e-mail client over a particular period of time. Such use of time-series visualisation of e-mail traffic data provides the user with a convenient way of analysing the level of e-mail usage generated by different e-mail clients and also for investigating time periods of intense or low e-mail traffic activities. The program currently being used for time-series visualisation in our e-mail traffic analyser system is TimeSearcher 2 [3] (Figure 11) and an example of the program output for our simulated e-mail traffic data is displayed in Figure 12.

It should be noted that when dealing with time-series data, it is often quite useful to view time-series data under different levels of granularity [15], in order to sample the data under different time-scales (e.g. e-mails per hour, e-mails per day, e-mails per week, e-mails per month). This provides the user with more options for finding patterns in the time-series data, since some patterns may not be noticed on particular levels of granularity (e.g. the pattern of a person who send e-mails only on certain days of the week is better noticed at the time-scale of e-mails per day). Thus the use of different time granulation can aid the user in visually analysing the e-mail traffic data for different levels of e-mail traffic activities from e-mail clients.

4.3 Frequency Distribution Visualisation

Frequency distribution visualisation presents a histogram sample of variables in data that needs to be statistically analysed. The use of this type of visualisation, shown as the visualisation output from the "Statistical Processor" in Figure 1, enables the user to analyse the level of interaction between e-mail clients by presenting the user with histogram samples of the sending delays (time taken between each e-mail sent) and the replying delays (time taken between an e-mail client receiving an e-mail and sending a reply). The ability to analyse the sending delay and replying delay distributions of different e-mail clients allows the user to work out how a particular e-mail client interacts with their social contacts in general, and also how a particular e-mail client interacts with specific social contacts (e.g. client A often communicates more quickly with client B than with client C).

At the time of writing this paper, the statistical processor part of the e-mail traffic analyser was not implemented. However, there are examples of work where frequency histograms have been used to build behavioral profiles of different e-mail clients. An example is the work by researchers at Columbia University [24, 25], where they used histograms of e-mail usage to build profiles of e-mail clients to detect anomalies in e-mail traffic generated by e-mail viruses. The two approaches used by [24, 25] are the creation of histograms of daily 24-hour e-mail usage and histograms of the frequency of e-mails sent to different social contacts. It should be noted that the approach taken by the [24, 25] uses histograms to focus more on the average number of e-mails sent by e-mail clients, whereas we are interested in using histograms to analyse how an e-mail client interacts with other users in terms of the usual delays between each e-mail sent and the usual response time it takes for an e-mail client to respond to another client.

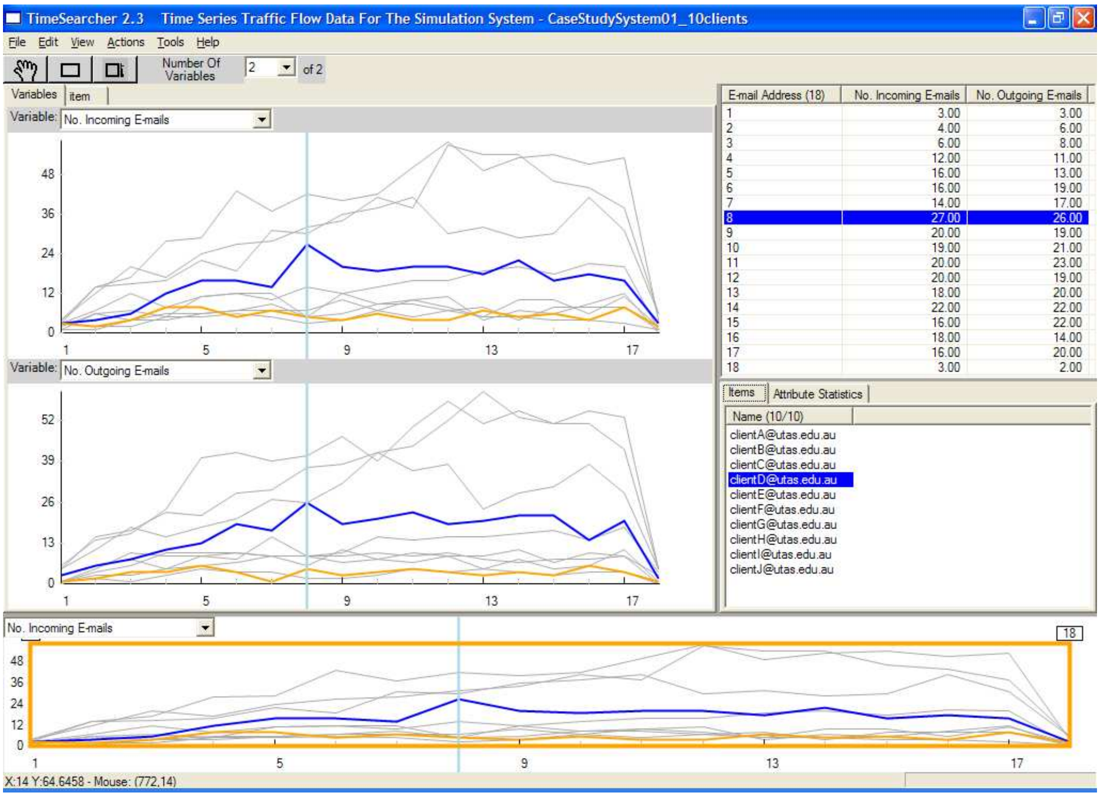

Figure 12: Time-series visualisation of simulated e-mail traffic, using the time-scale of number of e-mails per week

4.4 Decision Tree Visualisation

Decision trees are one of a number of data mining tools used for discovering knowledge and finding patterns in data [19, 27]. When decision trees are used to explore data, they can be used to provide a visual representation of the data set through the use of a tree-like structure. This tree-like structure presents to the user the result of the decision tree algorithm classification, by showing how the data set has been split into segments according to different attributes in the data. The end result from the use of the decision tree is that the user is able to observe the patterns from the data set that were selected by the decision tree algorithm.

For our e-mail traffic analyser system, the program used to perform the decision tree data mining of the e-mail data is WEKA [27], shown as the component "WEKA Data Mining Tool" in Figures 1 and 11. The decision tree algorithm used in WEKA to classify the e-mail traffic data is the J4.8 decision tree algorithm (WEKA's implementation of the C4.5 revision 8 decision tree algorithm) [27], and an example of the output is shown in Figure 13. The decision tree diagram generated by WEKA in Figure 13, shows how the e-mail traffic data from the simulated e-mail client 'clientA@utas.edu.au' has been classified and split according to the following data attributes: direction of the e-mails ('in' or 'out'), the date/time of the messages (units are in simulation time), and the e-mail address of the recipient e-mail client (address taken from

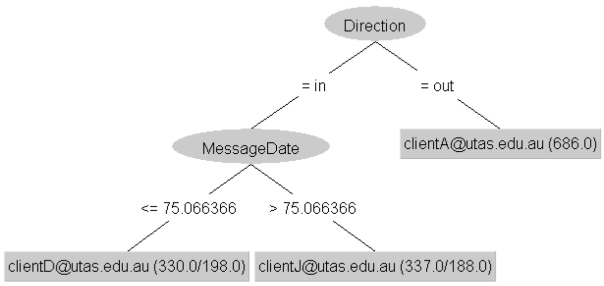

Figure 13: WEKA decision tree visualisation output for the e-mail traffic data of the simulated e-mail client 'clientA@utas.edu.au', using the J4.8 decision tree algorithm

the 'From' field of the generated e-mail messages). The end result from the split data shows interesting periods of time from the simulation run, in which the 'leaves' of the decision tree show the social contacts that the e-mail client 'clientA@utas.edu.au' has been sending the most e-mails to.

In comparison to the other visualisation techniques described, decision trees provide a much quicker way of finding patterns of interest from the e-mail traffic data. In the other visualisation techniques described the user needs to specifically search for items of interest from the visualised e-mail traffic data. This is a difficult process given that the user is presented with a lot of visual information. Decision trees on the other hand, assist the user to quickly focus their attention on areas of interest in the e-mail traffic data, so that the user can concentrate on finding out the details of the interesting part of the data, rather than spending their time searching for such interesting patterns. A demonstration on the use of decision trees to find interesting patterns generated by the conceptual e-mail system simulation model is given in the following case study.

5 Case Study: Small Network of E-mail Clients

5.1 E-mail System Model Setup

For this case study, we consider an e-mail system simulation model comprising of 10 e-mail clients and 9 behavior models. In this e-mail system model we have labelled each of the e-mail clients in the form "client<letter>@utas.edu.au" and labelled each of the behavior models in the form "behavior<letter>", where '<letter>' denotes an alphabetical letter identifier. We will refer to each of the e-mail clients in this case study by the client's e-mail address user name (the part before the '@' symbol).

The allocation of behavior models to the e-mail clients is shown in Figure 15, which represents how the behavioral profiles were assigned to each e-mail client. It should be noted from the stacked column chart in Figure 15 that each of the e-mail clients were assigned a unique behavior model, except for clientA and clientJ who were both assigned the behavior model labelled 'behaviorA'. In Figure 16, the column chart provides an overview of the values of the five personality trait dimensions making up the behavioral profiles of e-mail clients and provides an indication of the relative strength of the personality trait dimensions among the population of e-mail clients. For the social network connections, the social contacts for each of the e-mail clients were randomly assigned and given the configuration shown in Figure 14.

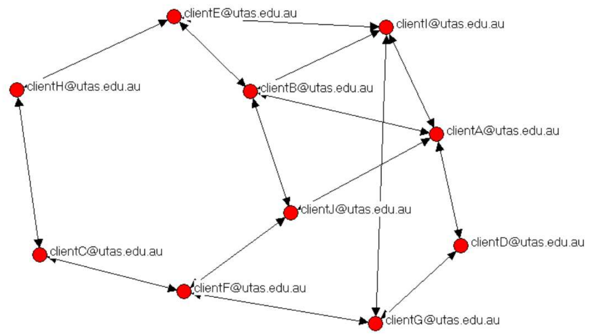

Figure 14: 2D diagram of e-mail system social network set-up configuration

5.2 Simulation Run

The case study e-mail system model was simulated over a period of 120 simulation days with 2748 e-mail messages in total being sent by the 10 e-mail clients. The bar chart in Figure 17 presents a summary of the number of messages sent and received by each of the e-mail clients over the 120 days simulated. A period of 120 simulation days was chosen as the duration of the simulation, to allow for enough time for the e-mail clients to start up their interactions from the initial conditions of the simulation (starting with mostly 'sending' events generated by e-mail clients) to a stable level of interactions ('sending' and 'replying' events generated by e-mail clients).

5.3 Analysing for Interesting Traffic Patterns

Through the use of WEKA, there were some interesting e-mail traffic patterns located by the decision tree technique, which are shown in tabulated form in Table 3. To obtain the results shown in Table 3, the decision tree was applied to analyse all e-mail messages from the mailbox logs of all the e-mail clients. The decision tree was used on e-mail traffic data several times using different data attributes as the class for the decision tree. The data attributes used as the class were 'EmailAccount' (field indicating the owner of the mailbox where the e-mail message was stored), 'Direction' (field indicating the direction of the e-mail message), 'From' (field indicating the sender of the message), and 'To' (field indicating the recipient of the e-mail message). Out of the four fields used as the class, the 'From' and 'To' fields were found to be the most useful, since the decision trees created from the 'From' and 'To' fields presented information on the date/time when interesting interactions between the e-mail clients occurred.

From the decision trees created from the 'From' and 'To' fields of the e-mail traffic data, the branching information containing date/time information was taken and put into Table 3. The "Incoming Interactions" column presents results from using the 'From' field of the e-mail

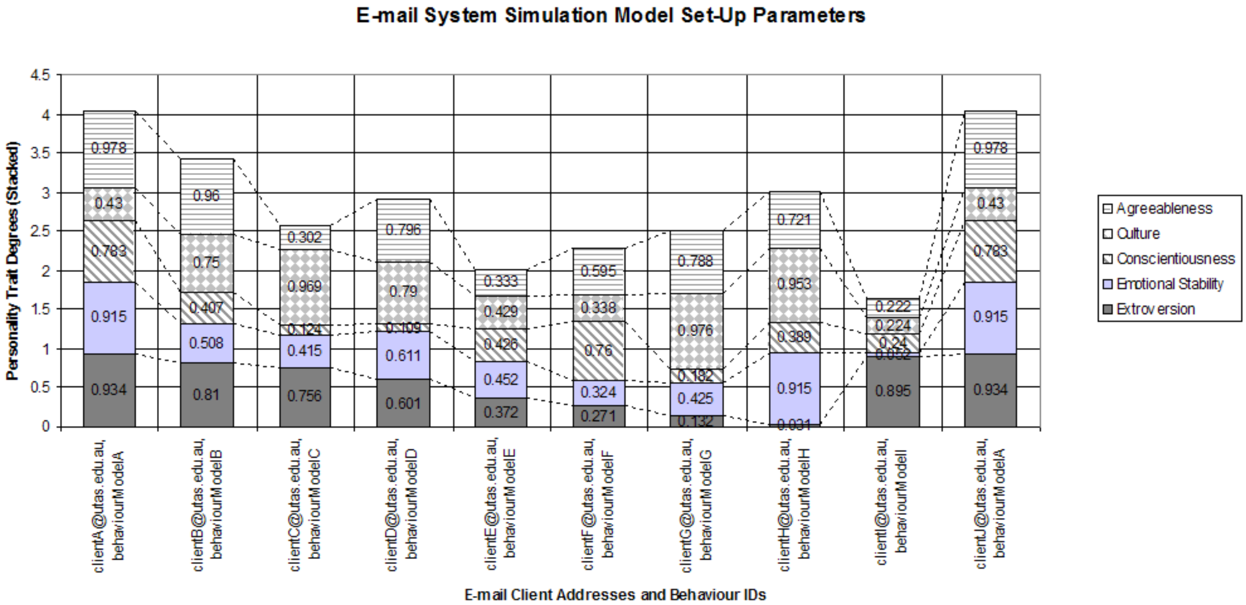

Figure 15: Stacked column chart showing the behavioral profiles of each e-mail client and their behavior model ID

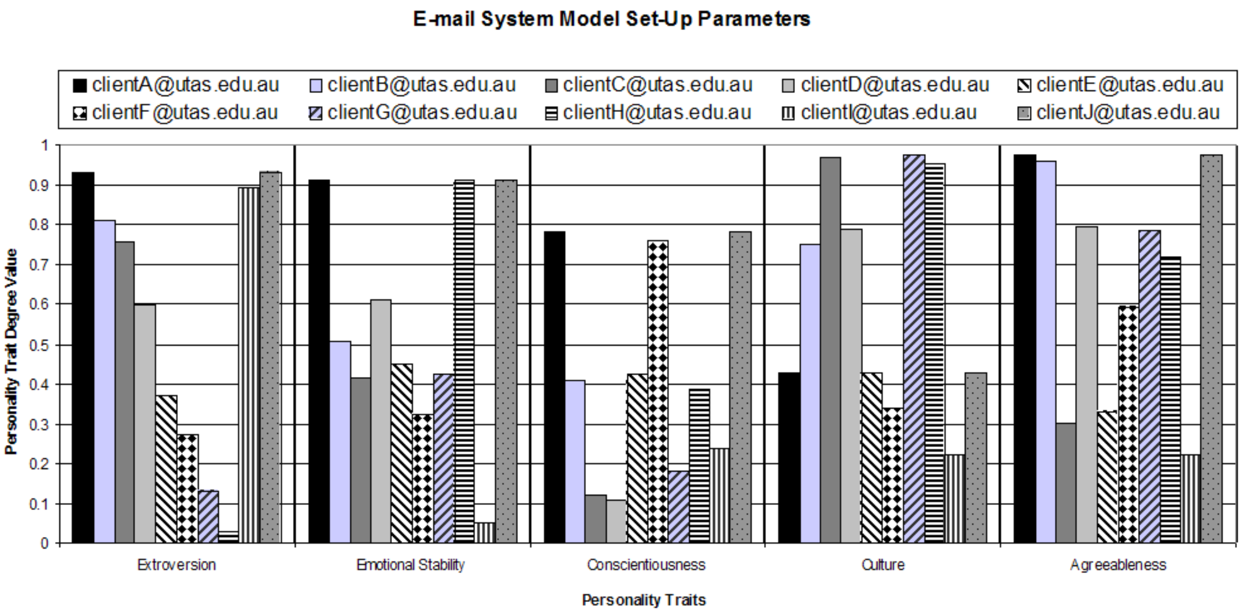

Figure 16: Column chart showing the personality trait values of all the e-mail clients

Table 3: Interesting e-mail traffic behavioral patterns found by decision tree for the case study

E-mail Account of Interest	**Incoming Interactions**	**Outgoing Interactions**
clientA@utas.edu.au	clientD to clientA, where date <= day 75, 132 messages	-
	clientJ to clientA, where date > day 75, 149 messages	-
clientF@utas.edu.au	clientC to clientF, where date <= day 41.15, 28 messages	clientF to clientC, where date <= day 66.7, 40 messages
	clientJ to clientF, where date > day 41.15, 119 messages	clientF to clientJ, where date > day 66.7, 84 messages
clientG@utas.edu.au	clientI to clientG, where date <= day 37.34, 13 messages	clientG to clientI, where date <= day 48.3, 9 messages
	clientD to clientG, where date > day 46.86, 43 messages	clientG to clientD, where date > day 48.3, 30 messages
	clientF to clientG, where date > day 37.34 and date <= day 46.86, 5 messages	-
clientI@utas.edu.au	-	clientI to clientG, where date <= day 33.7, 12 messages
	-	clientI to clientA, where date > day 33.7, 44 messages
clientJ@utas.edu.au	clientB to clientJ, where date <= day 82.5, 136 messages	clientJ to clientB, where date <= day 78.6, 139 messages
	clientA to clientJ, where date > day 82.5, 121 messages	clientJ to clientA, where date > day 78.6, 135 messages

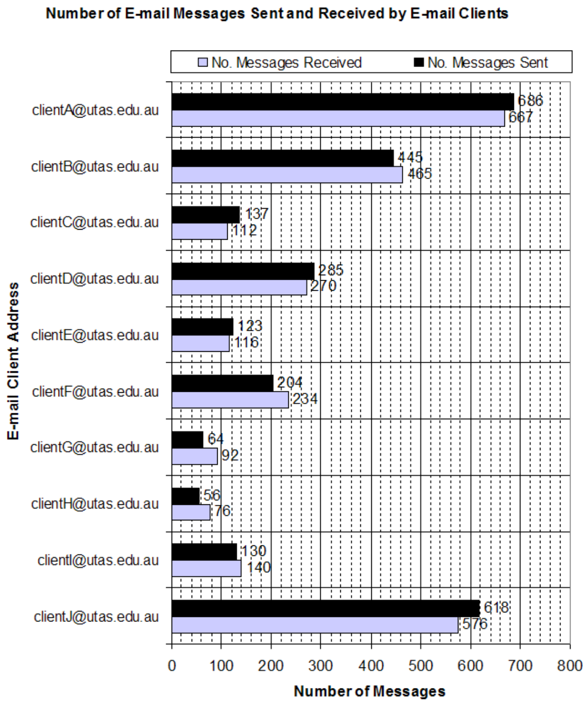

Figure 17: Number of e-mail messages sent and received by the e-mail clients over 120 simulation days

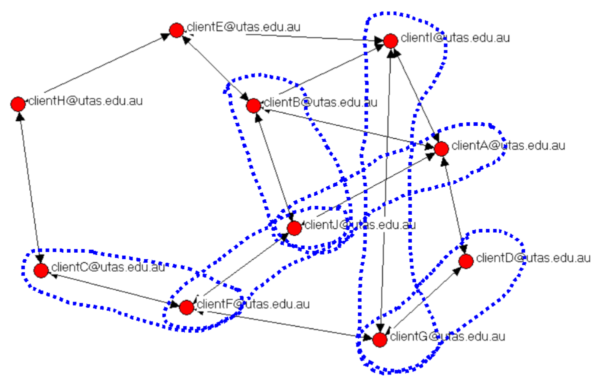

Figure 18: Interesting interactions found from both the inboxes and outboxes of e-mail clients (i.e. interesting patterns from two directions)

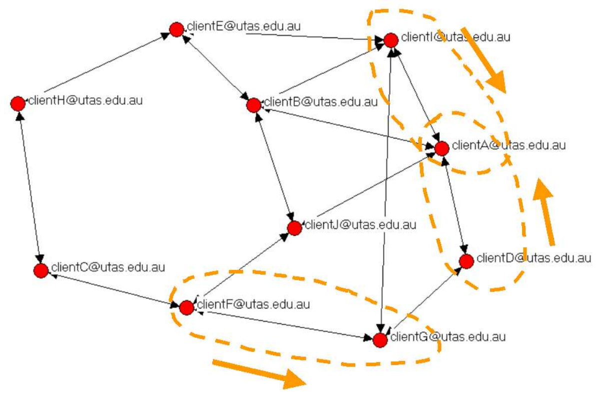

Figure 19: Interesting interactions found from either the inbox or outbox of e-mail clients (i.e. interesting patterns from one direction)

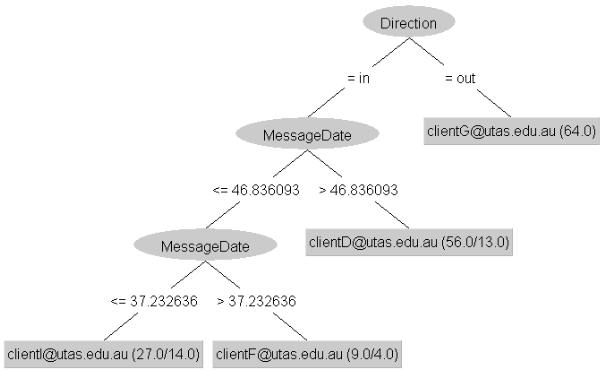

Figure 20: Decision tree for clientG using the “From” message header field as the class

messages as the decision tree class, while the “Outgoing Interactions” column presents results from using the ‘To’ field of the e-mail messages as the decision tree class. By compiling the interesting results into Table 3, it can be observed where the interesting interactions are occurring in the e-mail social network. Figures 18 and 19 show the areas of the e-mail social network where the interesting interactions are occurring. It is noted that some of the interesting interactions were found from both the inboxes and outboxes of e-mail clients (Figure 18), indicating that the interactions were of particular significance for incoming and outgoing e-mail traffic. Other interesting interactions were found only from either the inbox or outbox of some e-mail clients (Figure 19), indicating that the interactions were only of particular significance for certain incoming or outgoing e-mail traffic. We now analyse in detail the the interesting interactions occurring for clientG.

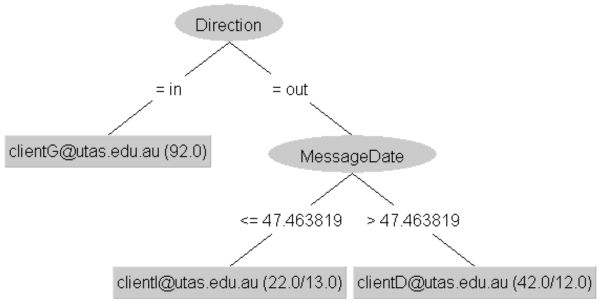

Figure 21: Decision tree for clientG using the “To” message header field as the class

5.4 Details of Interesting Patterns

The decision tree visualisations provided by WEKA for clientG are shown in Figures 20 and 21. It is seen from the decision trees that there is a lot of communication between clientG and clientI prior to day 47.46. However, after day 47.46, it appears that clientG communicates more with clientD. A close inspection of the weekly time-series data for clientG shows a very interesting change in outgoing e-mail traffic activity, as seen in Figure 22. In the weekly time-series data, it is shown that there is a drop in outgoing e-mail traffic for clientG on week 7, which corresponds with the same week where clientG changes from communicating more with clientI to communicating more with clientD. Changing the time resolution from week resolution to day resolution in Figure 23, shows that shows that clientG stops sending e-mails for 5 days from day 42 to 46, which explains about the drop in outgoing communication in week 7 of the weekly time-series data.

In Figure 20, it shows that there is some communication between clientG and clientF during the time period day 37.23 to day 46.84, where 5 e-mail messages are sent from clientF to clientG. An examination of the daily time-series data in Figure 23, shows that from day 37 to day 47 there are incoming e-mail messages, but clientG provides very little response until day 47 when it starts communicating again after a 5-day break from communication. This confirms from what was observed from Table 3 and Figure 19, where it was found that the communication interaction between clientG and clientF was mostly from one direction from clientF.

6 Discussion

6.1 E-mail Traffic Analyser

From the case study, it was shown that the decision tree visualisation technique proved to be very useful in finding “interesting” traffic behavioral patterns from the e-mail traffic data generated by the e-mail system conceptual simulation model. The decision tree output from WEKA assisted the user in being able to “pinpoint” the exact locations of interesting e-mail traffic interactions between

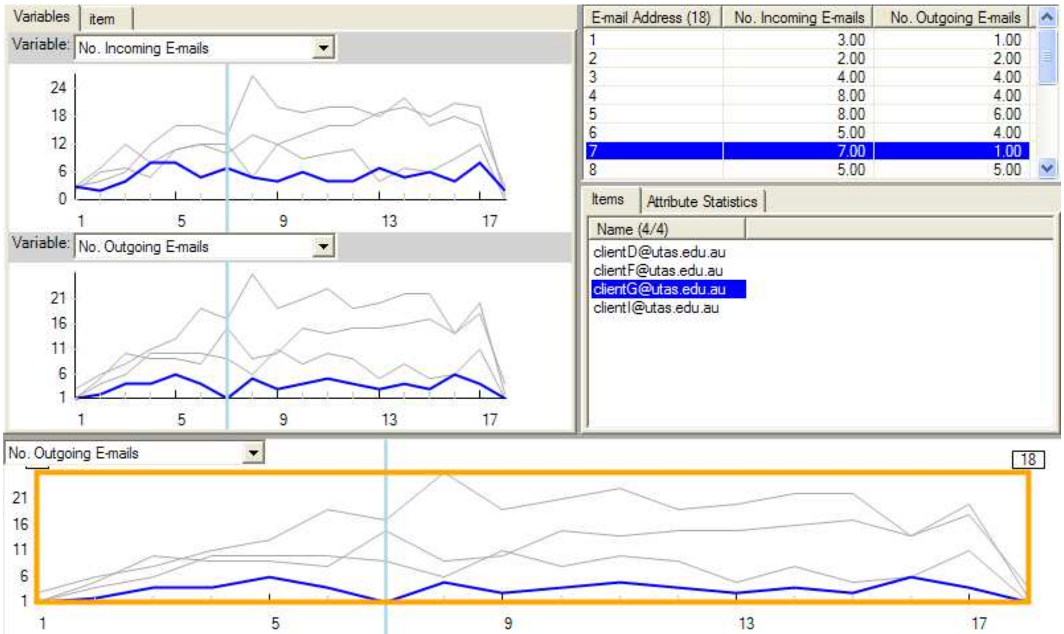

Figure 22: Weekly time-series data for clientG, showing that there is a drop in outgoing e-mail traffic on week 7 or around day 49

e-mail clients. With the assistance of additional visualisation techniques such as social network visualisation (using Netdraw [6]) and time-series visualisation (using TimeSearcher 2 [3]), we were able to obtain detailed information about the hidden relationships in the e-mail traffic data. The use of different time scales for analysing the time-series data of clientG in the case study proved to be very useful for observing both long-term trends and short-term changes in the e-mail traffic activity levels, in order to correlate those time-series patterns with the interesting patterns selected by the decision trees.

Although the current implementation of the e-mail traffic analyser system (Figure 11) was useful for finding interesting behavioral traffic patterns in the e-mail data, the process of analysing the e-mail data for detailed information on the interesting patterns was limited by the fact that not all the features of the e-mail traffic analyser system (i.e. "Statistical Processor" and "Data Parameter Selector" components from Figure 1) were implemented at the time of writing this paper. If the "Statistical Processor" component of the e-mail traffic analyser system was implemented, then the information on sending delays and replying delays would have provided more details about the level of interactions and response between e-mail clients (e.g. between clientG and clientF in the case study). Also, if the "Data Parameter Selector" component was implemented, it would have allowed the user more choices in how information is visualised for social network visualisation and time-series visualisation. For example, the "Data Parameter Selector" component would have been used in the case study to restrict the data to visualise all incoming and outgoing e-mail traffic between day 35 to day 56 from clientG, rather than providing visualisation of information that is not of interest to the user. Having these features implemented for the e-mail traffic analyser would provide the user a much better understanding about the interactions between e-mail clients and about the underlying nature of the behavior of e-mail clients.

6.2 Conceptual Simulation Model

In the case study, the creation of simulated e-mail clients with different personality trait degree values allowed us to observe the linkage between personality traits and e-mailing behavior. By observing the degree values for the personality trait dimensions assigned to different e-mail clients (Figures 15 and 16) and the total number of e-mail messages sent by clients during the case study simulation (Figure 17), it was easy to examine the effects that certain personality traits, such as extroversion, had on overall e-mail behavior (e.g. extroverted individuals generally sent more e-mails, introverted individuals generally sent less e-mails). However, the number of e-mails sent by e-mail clients is not just dependent on how extroverted an e-mail client is, but also dependent on to whom the e-mail client is socially connected and how they respond to e-mail messages sent by the more extroverted individuals. This type of behavior can be observed by simply examining the number of e-mails sent and received by e-mail clients (Figure 17) socially connected to clientA and clientJ from the case study (Figure 14). In Figure 17, it can be seen that clientB, clienF, and clientG are influenced in some way by their social connections to clientA and clientJ. An exception to this assumption is clientI, who is supposed to

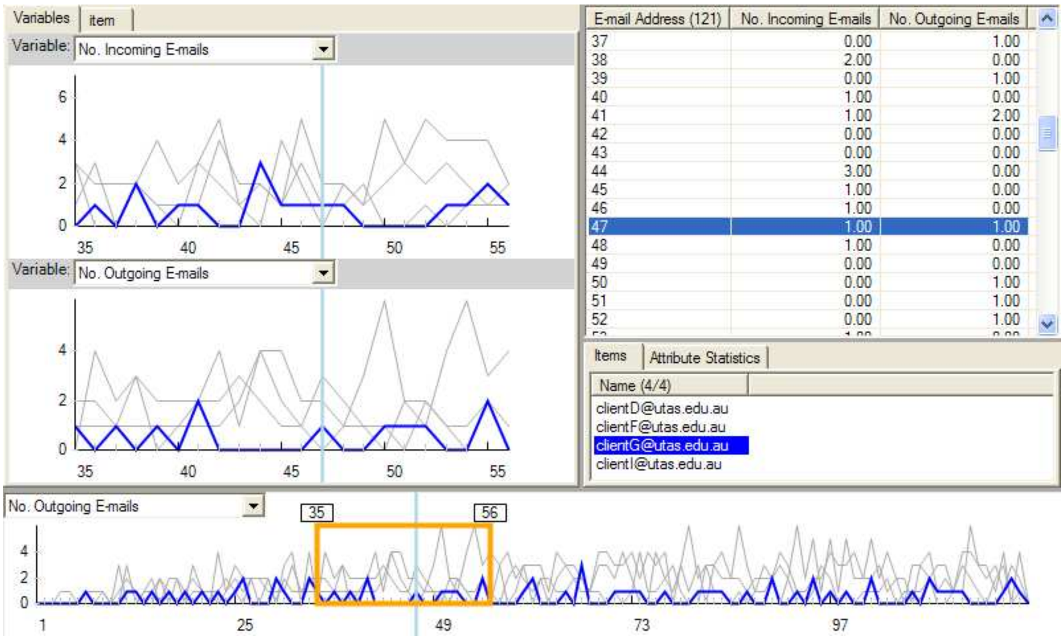

Figure 23: Daily time-series data for clientG for days 35 to 56 (i.e. weeks 6 to 8), showing that clientG stops sending e-mails for 5 days from day 42 to 46

be highly extroverted but does not send or receive as many e-mails as other clients connected to clientA and clientJ. A suggestion for this exception in behavior is that clientI has an extremely low emotional stability trait ($D_{ES} = 0.052$), indicating that clientI would be considered to be highly temperamental and easily upset, hence not always communicating with other clients.

What was not as easy to predict and understand was the individual interactions between different e-mail clients. With the use of the decision tree, social network, and time-series visualisation tools as shown in Section 5.4, these tools provided a way for better observation of the e-mail traffic behavior of e-mail clients and linking the behavior back to the e-mail clients' behavior profile. For example, clientG is described as having a fairly introverted, reclusive, and lazy behavior profile ($D_{EX} = 0.132, D_{ES} = 0.425, D_C = 0.182$), explaining why clientG generated a low volume of e-mail traffic and had a drop in e-mail traffic activity on week 7 (Figure 22). The case study therefore demonstrates that our e-mail system simulation model generates some very interesting e-mail traffic behavioral patterns, each of which can be linked back to the behavioral profiles of each e-mail client.

Although the conceptual e-mail system model provides a way of simulating the basic e-mail behaviors of sending e-mails and replying to e-mails, there are some limitations to the conceptual model. Firstly, there are other e-mail behaviors that could be modelled to provide other useful metrics for observing e-mailing behavior. These e-mail behaviors are sending attachments, forwarding e-mails, and sending to multiple recipients, which were described in section 3.3. Secondly, the conceptual model does not have a specific way of modelling social groups, which are an important factor in communications in real e-mail social networks. An individual's e-mailing behavior may be affected by the social groups they belong to, or may be affected by their involvement in the certain social groups. For example, if an individual is involved in organising a conference they would spend a lot of time sending e-mails to other conference committee members while the conference is being organised. Once the conference has finished, the individual's e-mailing behavior will change since they no longer need to send e-mails to committee members. Finally, the conceptual model does not model any specific type of e-mailing habits, such as sending e-mails when arriving to work in the morning, and not sending e-mails on weekends. These factors do have an affect on certain individuals' e-mailing behavior as described by [24, 25], but are not always consistent among all individuals.

6.3 Other Considerations

With our current conceptual simulation model of the e-mail system and the e-mail traffic analyser system, we have demonstrated how different e-mail traffic behavioral patterns can be generated and demonstrated how the interesting traffic behavioral patterns can be detected by the use of a decision tree. However, the simulated system we have demonstrated is an extremely small e-mail system consisting of 10 clients. Obviously the 10-client e-mail system we have used for the case study provides a useful way for explaining how our e-mail traffic analyser system operates. But once we start analysing a simulated system consisting of hundreds or thousands (e.g. 10,000) of e-mail clients, then we must really consider whether using the decision tree is the most effective way of finding the "interesting" patterns. Is there a better and more efficient way for designing a system that can find the "interesting" patterns that are of most interest to the analyst user, rather than every single "interesting" pattern it comes across? Is there a way of just focusing the analysis on a specific group of e-mail clients? Is there a way of focusing the analysis to search for a specific type of e-mail traffic behavior, such as criminal or terrorist behavior? This is something that our future work must address, in order to more efficiently analyse very large size e-mail social networks and yet quickly point the user to areas of most interest.

7 Conclusion and Future Work

In this paper we have described the development of a conceptual simulation model of the e-mail system and have shown how it can be used to generate different e-mail traffic patterns, based on the use of personality traits for the behavior model of each e-mail client in the simulated system. Through the use of visualisation tools such as Netdraw, Mage 3D, TimeSearcher, and WEKA for our e-mail traffic analyser system, we have also shown how visualisation tools can be useful for observing e-mail traffic behavioral patterns of social connections between e-mail clients, the volume of e-mail traffic generated by e-mail clients in the simulation model, and identifying patterns from the simulated e-mail traffic output.

Future work will involve continued development of the e-mail traffic analyser system, further investigations using decision trees, and investigations using other artificial intelligence techniques such as neural networks and neuro-fuzzy systems to analyse e-mail traffic data. Other future work to be considered are adding additional attributes to the behavior model of each e-mail client to enable for more complex behavior (e.g. social groupings of e-mail clients based on common interests), and adding the ability to simulate criminal or terrorist behavior in the simulated e-mail system.

References

[1] I. Ajzen, *Attitudes, Personality, and Behavior*, Mapping Social Psychology, Stony Stratford, Milton Keynes: Open University Press, 1988.

[2] R. L. Akers, *Criminological Theories: Introduction, Evaluation and Application*, 3rd ed., Los Angeles: Roxbury, 2000.

[3] A. Aris, A. Khella, P. Buono, B. Shneiderman, and C. Plaisant, *Timesearcher 2*, Human-Computer Interaction Laboratory, Computer Science Department, University of Maryland, 2005, http://www.cs.umd.edu/hcil/timesearcher/.

[4] J. Banks, J. Carson, B. L. Nelson, and D. Nicol, *Discrete-event System Simulation*, 4th (international) ed., Upper Saddle River, NJ: Prentice Hall, 2005.

[5] B. D. Berkowitz and A. E. Goodman, *Best Truth: Intelligence in the Information Age*, New Haven: Yale University Press, 2000.

[6] S. P. Borgatti, *Netdraw: Graph Visualization Software*, Harvard: Analytic Technologies, 2002, http://www.analytictech.com/netdraw.htm.

[7] K. M. Carley, M. Dombroski, M. Tsvetovat, J. Reminga, and N. Kamneva, "Destabilizing dynamic covert networks," in *8th International Command and Control Research and Technology Symposium*, National Defense War College, Washington DC, 2003, http://www.casos.cs.cmu.edu/publications/resources_others/a2c2_carley_2003_destabilizing.pdf.

[8] S. C. Cloninger, *Theories of Personality: Understanding Persons*, 2nd ed., Upper Saddle River, N.J.: Prentice Hall, 1996.

[9] T. Coffman, S. Greenblatt, and S. Marcus, " Graph-based technologies for intelligence analysis," *Communications of the ACM*, vol. 47, no. 3, pp. 45-47, 2004.

[10] M. Crenshaw, "The psychology of terrorism: An agenda for the 21st century," *Political Psychology*, vol. 21, no. 2, pp. 405-420, 2000.

[11] J. P. Eckmann, E. Moses, and D. Sergi, "Entropy of dialogues creates coherent structures in e-mail traffic," in *Proceedings of the National Academy of Sciences of the United States of America*, vol. 101, no. 40, pp. 14 333 -14 337, 2004.

[12] L. C. Freeman, "Visualizing social networks," *Journal of Social Structure*, vol. 1, no. 1, 2000, http://www.cmu.edu/joss/content/articles/volume1/Freeman.html.

[13] *Jython Programming Language*, viewed 24 Feb. 2005, http://www.jython.org/.

[14] T. K. Ghosh, M. A. Prelas, D. Viswanath, and S. K. Loyalka, *Science and Technology of Terrorism and Counterterrorism*, Marcel Dekker, 2002.

[15] J. Han and M. Kamber, *Data Mining: Concepts and Techniques*, Morgan Kaufmann Series in Data Management Systems, San Francisco, CA: Morgan Kaufmann, 2001.

[16] V. E. Krebs, "Mapping Networks of Terrorist Cells," *Connections*, vol. 24, no. 3, pp. 43-52, 2002.

[17] A. Lantz, "Does the use of e-mail change over time?," *International Journal of Human-Computer Interaction*, vol. 15, no. 3, pp. 419-431, 2003.

[18] A. M. Law and W. D. Kelton, *Simulation Modeling and Analysis*, 3rd ed., McGraw-Hill, 2000.

[19] M. Negnevitsky, *Artificial Intelligence: A Guide to Intelligent Systems*, 2nd ed., Essex: Addison Wesley, 2004.

[20] W. T. Norman, "Toward an adequate taxonomy of personality attributes: Replicated factor structure in peer nomination personality ratings," *Journal of Abnormal and Social Psychology*, vol. 66, no. 6, pp. 574–583, 1963.

[21] *Python Programming Language*, viewed 24 Feb. 2005, http://www.python.org/.

[22] D. C. Richardson, *Mage: 3-D Visualization Program*, Durham: Biochemistry Department, Duke University, 2002, http://kinemage.biochem.duke.edu/software /mage.php.

[23] W. Schroeder, K. Martin, and B. Lorensen, *The Visualization Toolkit: An Object-Oriented Approach to 3D Graphics*, 3rd ed., Kitware, 2004.

[24] S. J. Stolfo, S. Hershkop, K. Wang, O. Nimeskern, and C. W. Hu, "A behavior-based approach to securing email systems," in *Computer Network Security*, LNCS 2776, pp. 57–81, Springer-Verlag, 2003.

[25] S. J. Stolfo, S. Hershkop, K. Wang, O. Nimeskern, and C. W. Hu, "Behavior profiling of email," in *Intelligence and Security Informatics, Proceedings*, LNCS 2665, pp. 74–90, Springer-Verlag, 2003.

[26] J. R. Tyler, D. M. Wilkinson, and B. A. Huberman, "E-mail as spectroscopy: Automated discovery of community structure within organizations," *Information Society*, vol. 21, no. 2, pp. 133–141, 2005.

[27] I. H. Witten and E. Frank, *Data Mining: Practical Machine Learning Tools and Techniques with Java Implementations*, San Francisco, Calif.: Morgan Kaufmann, 2000.

[28] J. Xu and H. C. Chen, "Untangling criminal networks: A case study," in *Proceedings of Intelligence and Security Informatics*, LNCS 2665, pp. 232–248, Springer-Verlag, 2003.

[29] J. Xu, B. Marshall, S. Kaza, and H. C. Chen, "Analyzing and visualizing criminal network dynamics: A case study," in *Proceedings of Intelligence and Security Informatics*, LNCS 3073, pp. 359–377, Springer-Verlag, 2004.

Mark Lim received his B.E. (Hons) from the University of Tasmania, Australia, in 2003. The author is presently a Ph.D. candidate and his major research interests are computer security and artificial intelligence. Mr Lim is a student member of the Institution of Engineers Australia and also a student member of the IEEE.

Michael Negnevitsky received his B.S.E.E. (Hons), 1978, and Ph.D., 1983, from Byelorussian University of Technology, Belarus. Currently, he is an Associate Professor in the School of Engineering at the University of Tasmania, Australia. From 1984 to 1991, he was a Senior Research Fellow and Senior Lecturer in the Department of Electrical Engineering, Byelorussian University of Technology, Belarus. After his arrival in Australia, Dr. Negnevitsky was with Monash University, Australia. His major interests are computational intelligence, fuzzy logic, neural networks and genetic algorithms. Dr. Negnevitsky is a Chartered Professional Engineer, a Senior Member of the Institution of Engineers Australia, Member of the IEEE, and a Member of CIGRE AP36 (Electromagnetic Compatibility), Australian Technical Committee.

Jacqueline Hartnett first degree is a B.A. (Hons) in Geography from Exeter University in the UK and a MComp from the University of Tasmania. After graduating, she worked in the then new computing department of the Royal Dutch Shell group and then for IBM both in Australia and the UK. She has taught computer security courses in the University of Tasmania since 1993. Her current research interests are the use of authentication and access control as a means of maintaining privacy and confidentiality of personally identified data and the development of intrusion detection techniques for groups of collaborating network gateways. She is a member of Australian Standard Review Committees; IT-014 04 System & Data Security, Integrity and Privacy subcommittee, and IT-014-06-08 Electronic Communications in Health working groups. She is also currently chair of the Tasmanian branch of the Australian Computer Society.