Tracking and Monitoring E-mail Traffic Activities of Criminal and Terrorist Organisations Using Visualisation Tools[†]

M. J. Lim[1], M. Negnevitsky[1], and J. Hartnett[2]

[1]School of Engineering
University of Tasmania, Australia,
E-mail: mjlim@utas.edu.au

[2]School of Computing
University of Tasmania, Australia,
E-mail: J.Hartnett@utas.edu.au

Abstract

In defensive information operations, knowing about the actions or behaviour of the adversary is important for countering any attacks posed by the adversary. Obtaining information about the activities and behaviour of criminal or terrorist groups from electronic communication sources, such as e-mail, will be useful given that criminal or terrorists may utilise different electronic communication mediums to contact each of their agents or members. In this paper, we describe the development of an e-mail traffic analyser system for analysing the interactions between different e-mail clients in the e-mail system. We discuss how different visualisation tools are used and how the information provided by such tools would be useful to an intelligence analyst. The use of decision trees for locating "interesting" e-mail traffic interactions and the type of information revealed via the technique is also described.

Keywords

Defensive information operations, e-mail, traffic analysis, visualisation, data mining, decision trees, communication behaviour.

INTRODUCTION

Surveillance of communication is an aspect of intelligence gathering that is important for monitoring the communication activities of adversaries and being aware of potential threats from certain organisations or groups. The ability to capture and analyse the communication activities of adversaries is crucial for determining the behaviour of the adversary, developing the appropriate strategies for countering attacks from the adversary, and determining how to prevent the adversary from performing further attacks. However, being able to analyse the large amounts of communication data and making sense of interesting or unusual activities from the data is a difficult task (Coffman, Greenblatt & Marcus 2004). Such a task of analysing the large amounts of information is vital when determining the activities and social structure of criminal and terrorist groups (Mena 2003).

Within the area of analysing criminal and terrorist groups, progress has been made in developing tools and techniques for analysing the social structure or social connections of criminal groups (Xu & Chen 2003; Xu et al. 2004) and terrorist groups (Carley et al. 2003; Krebs 2002) using social network analysis quantitative measures (Scott 2000). The focus of our research is more on analysing the communication aspects of criminal and terrorist groups, rather than their social structure. For our work, we are analysing e-mail traffic data in order to determine suitable techniques that are able to find and recognise e-mail traffic activities from the communication interactions of possible criminal or terrorist groups. The techniques being investigated are artificial intelligence (A.I.) based techniques, such as decision trees or artificial neural networks (Negnevitsky 2004), which will be used to process, analyse, detect, and make sense of interesting or unusual activities in the e-mail traffic data. Such use of A.I. based or machine learning techniques may provide better profiling and prediction of criminal and terrorist behaviour (Mena 2003), in order to prevent attacks or tragic events from occurring. To assist with investigating the A.I. techniques, we propose an e-mail traffic analyser system shown in Figure 1 that extracts basic e-mail traffic behavioural patterns to obtain different types of information on e-mail users. The e-mail traffic analyser system and the information it presents is the focus of this paper.

[†] Proceedings of the 6th Australian Information Warfare & Security Conference, 24th - 25th November 2005, Geelong, Victoria, Australia.

DEFINING E-MAIL TRAFFIC ANALYSIS

Instead of examining the text content of e-mail messages, e-mail traffic analysis focuses on who is sending the e-mail messages and where it is travelling. E-mail traffic analysis involves looking at measures from the e-mail data for information such as whom an e-mail user communicates with, how often does an e-mail user communicate with someone, the size of e-mails sent, how many messages an e-mail user sends over time, and whether the e-mail user has a habit of sending e-mails at a particular time (e.g. in the morning). The traffic analysis of e-mail data is an effective way of profiling the behaviour of e-mail users (Stolfo et al. 2003b, 2003a) since it relies on 'when' and 'where' e-mail messages are being sent, rather than 'what' is being sent by e-mail users. The type of information provided by e-mail traffic analysis may be useful for observing the behaviour of criminals or terrorists, since it can provide the intelligence analyst information on how active certain criminals/terrorists are and their connection with others. The method of e-mail traffic analysis forms the basis of our research for developing the e-mail traffic analyser system and investigating A.I. techniques.

E-MAIL TRAFFIC ANALYSER SYSTEM AND VISUALISATION

The e-mail traffic analyser system shown in Figure 1 extracts and visualises the following e-mail traffic behavioural patterns:

- The social connections between e-mail users ("Social Network Data Processor" component);
- The level or volume of e-mail traffic generated by e-mail users ("Time-Series Data Processor" component);
- The level of interaction between different e-mail users ("Statistical Processor" component).

After these e-mail traffic behavioural patterns have been extracted and processed, they are visualised and presented to the user for analysis.

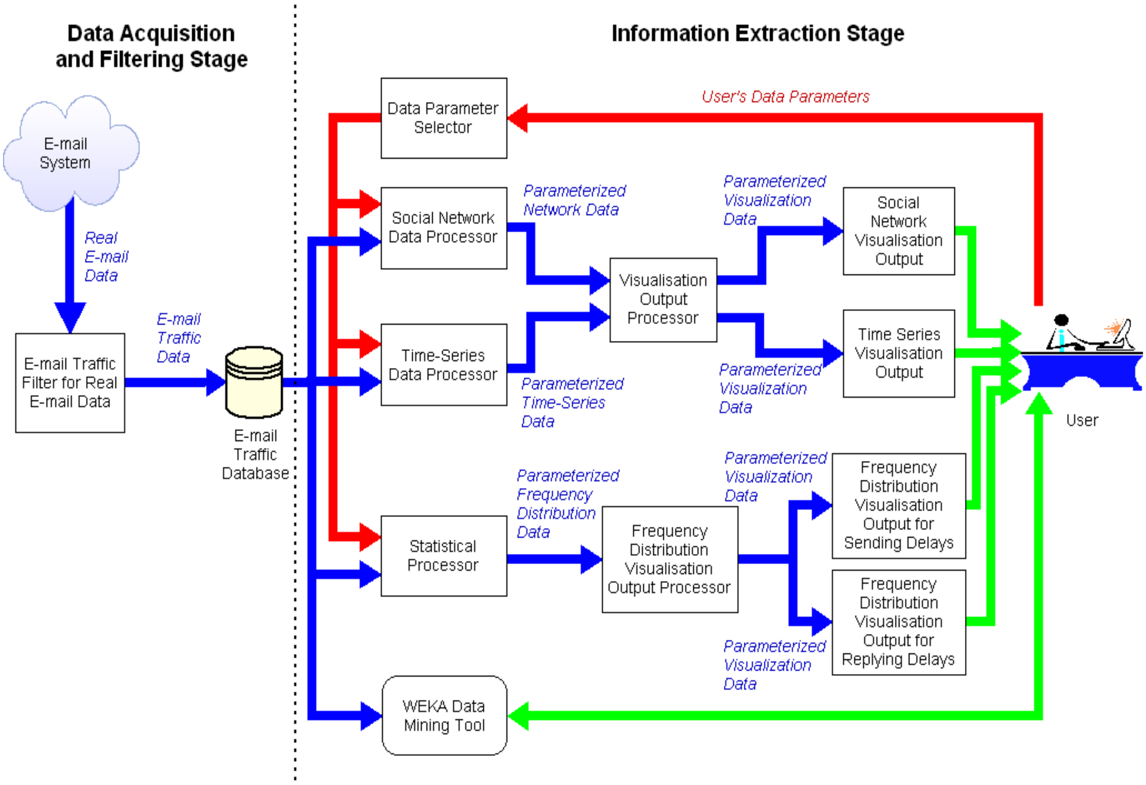

Data Acquisition and Filtering Stage
Information Extraction Stage
E-mail System
Real E-mail Data
E-mail Traffic Filter for Real E-mail Data
E-mail Traffic Data
E-mail Traffic Database
Data Parameter Selector
User's Data Parameters
Social Network Data Processor
Parameterized Network Data
Parameterized Visualization Data
Social Network Visualisation Output
Visualisation Output Processor
Time-Series Data Processor
Parameterized Time-Series Data
Parameterized Visualization Data
Time Series Visualisation Output
User
Parameterized Frequency Distribution Data
Statistical Processor
Frequency Distribution Visualisation Output Processor
Parameterized Visualization Data
Frequency Distribution Visualisation Output for Sending Delays
Frequency Distribution Visualisation Output for Replying Delays
Parameterized Visualization Data
WEKA Data Mining Tool

Figure 1: Overview of e-mail traffic analyser system.

The purpose of extracting different traffic behavioural patterns and visualising them is to allow us to analyse the e-mail traffic data for "interesting" behavioural patterns from each e-mail user. Examples of "interesting" e-mail traffic behaviour 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 significant change in the level of interactions between particular e-mail users. This is an extremely difficult and time-consuming task, given that the user must deal with so much information.

To assist the user with searching for the "interesting" patterns, the WEKA Data Mining Tool program (Witten & Frank 2000) is being used in the "Information Extraction Stage" of the e-mail traffic analyser system. After some interesting patterns are found, the user can then focus his/her attention on the details by using the "Data Parameter Selector" component of the e-mail traffic analyser system. 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.

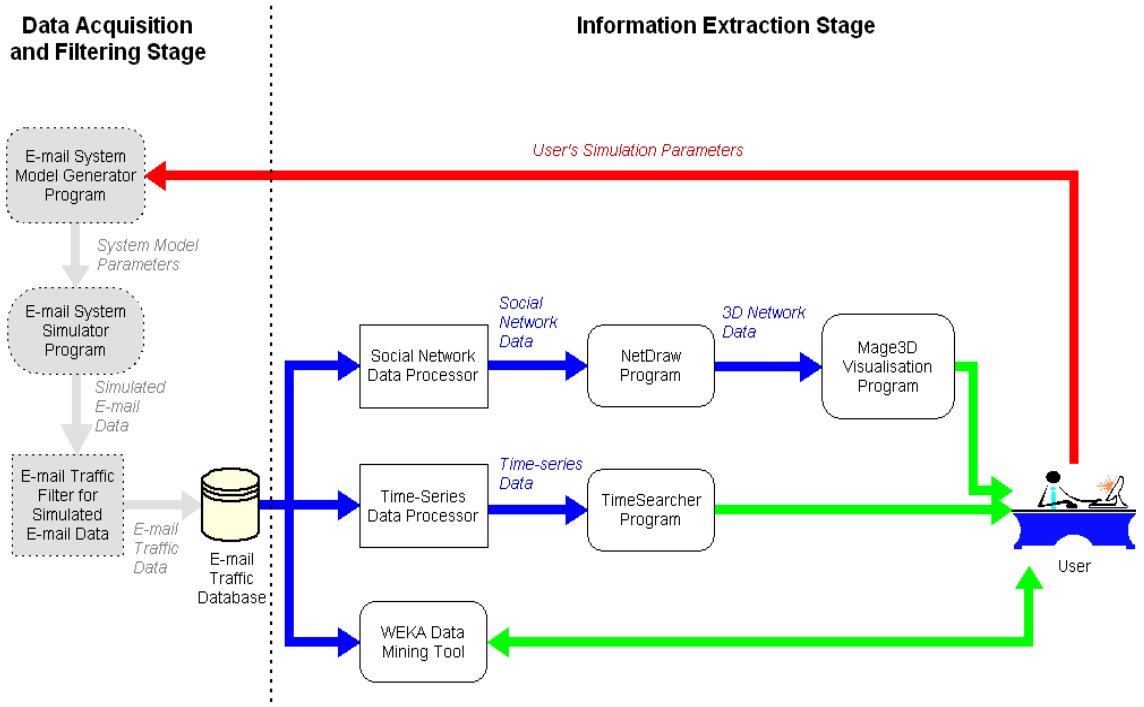

Data Acquisition and Filtering Stage
Information Extraction Stage
User's Simulation Parameters
E-mail System Model Generator Program
System Model Parameters
E-mail System Simulator Program
Simulated E-mail Data
E-mail Traffic Filter for Simulated E-mail Data
E-mail Traffic Data
E-mail Traffic Database
Social Network Data Processor
Social Network Data
NetDraw Program
3D Network Data
Mage3D Visualisation Program
Time-Series Data Processor
Time-series Data
TimeSearcher Program
WEKA Data Mining Tool
User

Figure 2: Overview of programs used in the current implementation of the e-mail traffic analyser system.

In the current implementation of the e-mail traffic analyser system, we have selected some visualisation tools to aid with the visualisation of different e-mail traffic behavioural patterns. An overview of the programs used for different visualisations in the e-mail traffic analyser system is shown in Figure 2. We describe below each method of visualisation and discuss the type of information given to the user or intelligence analyst.

Social Network Visualisation

Social network visualisation provides a visual image of the communication links or social connections between different individuals (Freeman 2000). This type of visualisation is given by the Netdraw program (Borgatti 2002) and the Mage3D program (Richardson 2002) in the implementation of the e-mail traffic analyser. The information provided by social network visualisation 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 through social network visualisation would aid an intelligence analyst gaining an overall view of all the communication links between criminals/terrorists and spot areas of interest in the e-mail social network, such as clustering of different types of e-mail users into distinct social groups or communities (Guimera et al. 2003; Newman & Girvan 2004; Tyler, Wilkinson & Huberman 2005).

Time-series Visualisation

Time-series visualisation presents a one-dimensional view of data by showing how it changes over time. The TimeSearcher 2 program (Aris et al. 2005) is used to provide time-series visualisation for the e-mail traffic analyser system. The use of time-series visualisation 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 would provide an intelligence analyst with a convenient way of analysing the level of e-mail usage generated by different criminal/terrorist suspects, investigate time periods of intense or low e-mail traffic activities, and also pick out interesting temporal patterns by adjusting the time-scale of the time-series visualisation (e.g. the pattern of a person who sends e-mails only on certain days of the of the week is better noticed at the time scale of e-mails per day).

WEKA Decision Tree Visualisation

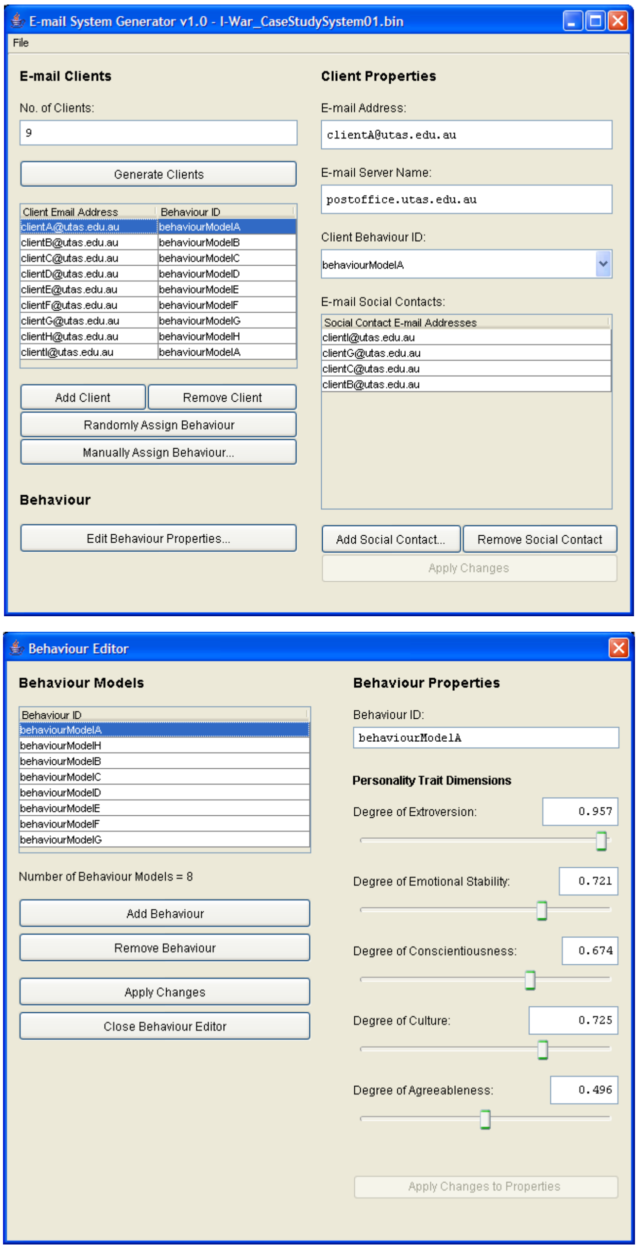
The WEKA Data Mining program (Witten & Frank 2000) provides a library of different machine learning algorithms to assist in the task of discovering knowledge and finding patterns in data. The machine learning algorithm from WEKA being used for our e-mail traffic analyser system are decision trees (Negnevitsky 2004; Witten & Frank 2000), which 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.

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.

GENERATION OF SIMULATED DATA FOR E-MAIL TRAFFIC ANALYSER SYSTEM

To provide data for evaluating our e-mail traffic analyser system, we have developed a simulation tool that allows us to simulate an e-mail system and generate simulated e-mail data for the e-mail traffic analyser system shown in Figure 2. The simulation tool comprises of an E-mail System Model Generator program shown in Figure 3 that allows the user to design the model of the e-mail system, and the E-mail System Simulator program that simulates the e-mail system model. With the E-mail System Model Generator program, the user can specify how many e-mail clients are in the e-mail system and assign different behaviour profiles to the e-mail clients, where each behaviour profile is based on personality trait dimensions (Ajzen 1988).

The reason for using a simulation tool is to allow us to create an e-mail system of any desirable size and enable us to assign different types of behaviour profiles to e-mail clients in the simulated system. By being able to specify what type of behaviour is present in the e-mail system simulation model and knowing of what type of behaviour is present in the e-mail traffic data, we are able to use the e-mail traffic analyser system to analyse the data and verify that the type of behavioural interactions we observe is correct. A demonstration of the use of the simulation tool and the e-mail traffic analyser system is provided in the following case study.



E-mail Clients

No. of Clients: 9

Client Email Address	Behaviour ID
clientA@utas.edu.au	behaviourModelA
clientB@utas.edu.au	behaviourModelB
clientC@utas.edu.au	behaviourModelC
clientD@utas.edu.au	behaviourModelD
clientE@utas.edu.au	behaviourModelE
clientF@utas.edu.au	behaviourModelF
clientG@utas.edu.au	behaviourModelG
clientH@utas.edu.au	behaviourModelH
clientI@utas.edu.au	behaviourModelA

Behaviour

Client Properties

E-mail Address: clientA@utas.edu.au

E-mail Server Name: postoffice.utas.edu.au

Client Behaviour ID: behaviourModelA

E-mail Social Contacts:

Social Contact E-mail Addresses
clientI@utas.edu.au
clientG@utas.edu.au
clientC@utas.edu.au
clientB@utas.edu.au

Behaviour Models

Behaviour ID
behaviourModelA
behaviourModelH
behaviourModelB
behaviourModelC
behaviourModelD
behaviourModelE
behaviourModelF
behaviourModelG

Number of Behaviour Models = 8

Behaviour Properties

Behaviour ID: behaviourModelA

Personality Trait Dimensions

Degree of Extroversion: 0.957

Degree of Emotional Stability: 0.721

Degree of Conscientiousness: 0.674

Degree of Culture: 0.725

Degree of Agreeableness: 0.496

Figure 3: Screenshot of the E-mail System Model Generator program used to generate the e-mail system simulation model.

CASE STUDY

E-mail System Model Set-Up

For this case study, we consider an e-mail system simulation model consisting of 9 e-mail clients and 8 behaviour models. Each of the e-mail clients in the simulation model is given a label of the form "client<letter>@utas.edu.au" and each behaviour model is given a label of the form "behaviour<letter>", where '<letter>' denotes an alphabetical letter identifier. In this case study we will refer to the e-mail clients by the username part of their e-mail address (i.e. the part before the '@' symbol in the e-mail address).

Each e-mail client has been allocated a behavioural profile as shown in Figure 4, which defines each e-mail client's behaviour. Note that each e-mail client has a different behaviour model, except for *clientA* and *clientI* who were both assigned the behaviour model '*behaviourA*'. 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 5.

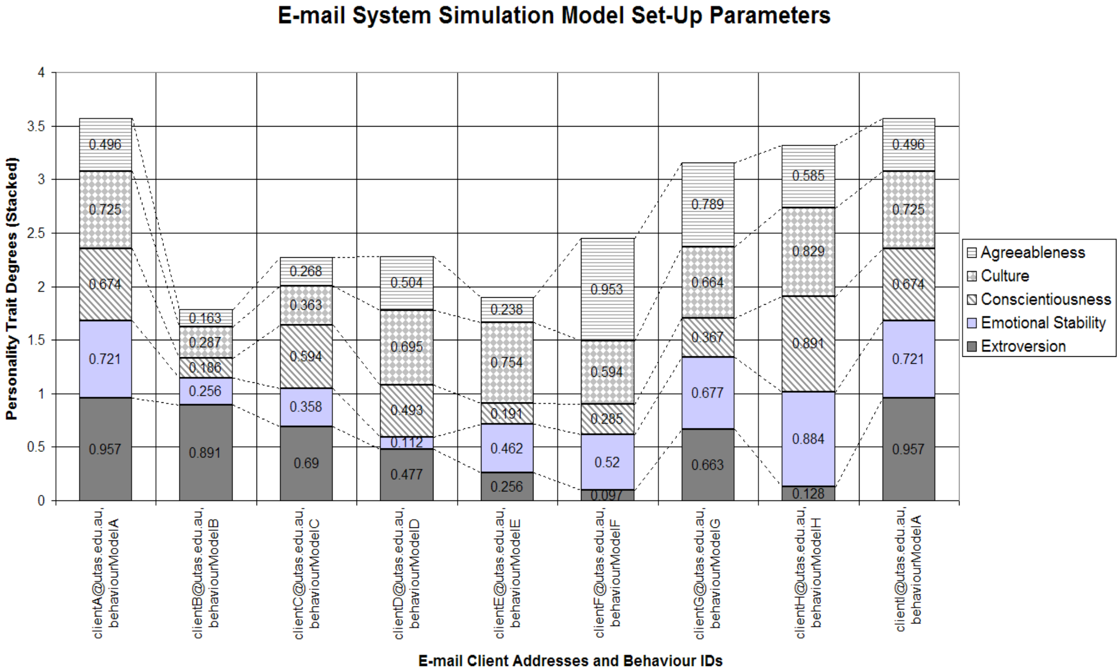


Figure 4: The behavioural profiles of each e-mail client in the simulation model.

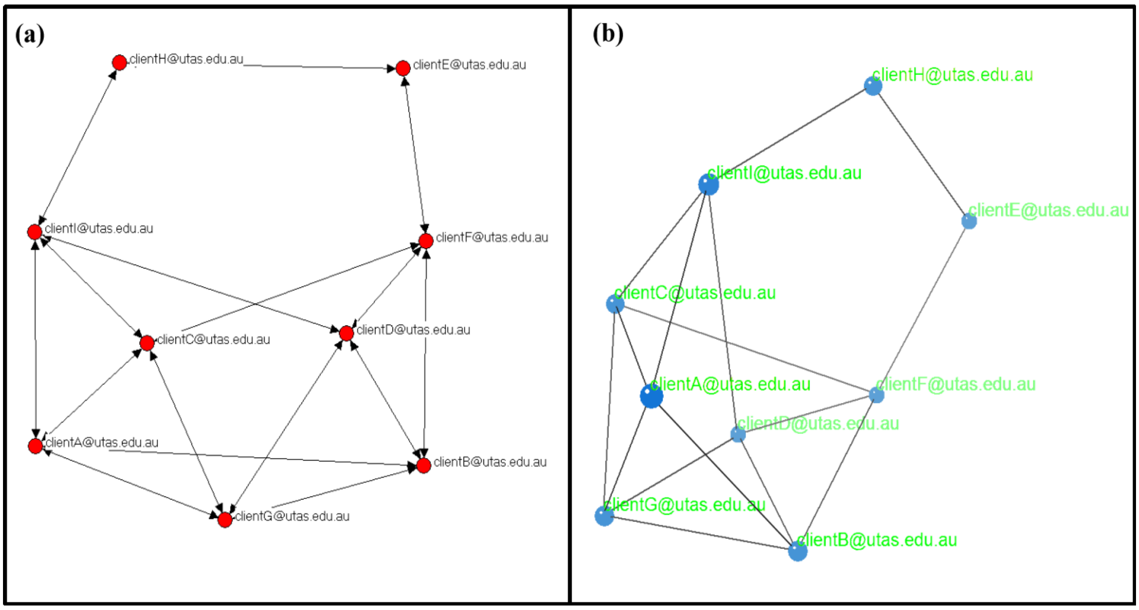


Figure 5: (a) 9 e-mail client social network diagram produced by Netdraw. (b) 3-D version of the same e-mail client social network produced by Mage3D.

Simulation Run

The case study e-mail system model was simulated over a period of 182 simulation days or 26 simulation weeks, with a total of 3257 e-mail messages being sent by all e-mail clients. The bar chart in Figure 6 presents a summary of the number of messages sent and received by each e-mail client over the duration of the 182 simulation days.

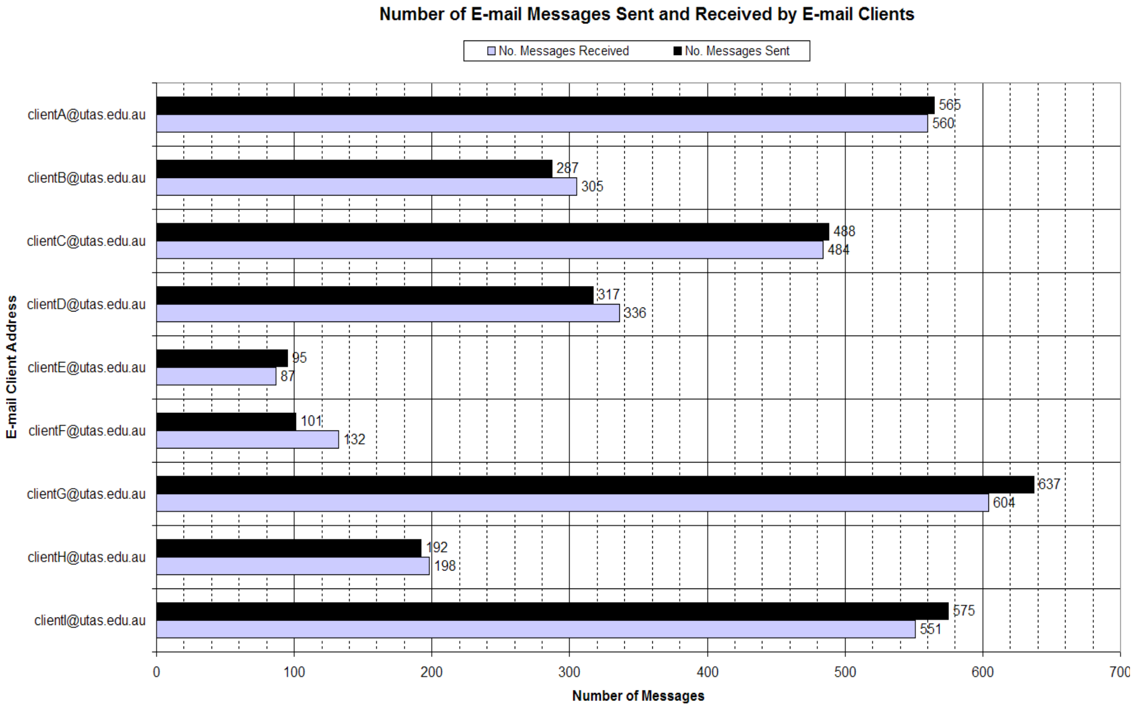


Figure 6: Number of e-mail messages sent and received by e-mail clients over 182 simulation days.

Traffic Pattern Analysis

Using the WEKA Data Mining program, the decision tree was used to analyse all incoming and outgoing the e-mail traffic data to produce two decision tree outputs showing periods of time where there were some "interesting" incoming and outgoing interactions between e-mail clients. Information on "interesting" incoming and outgoing e-mail traffic interactions from the decision tree outputs were tabulated and compiled together in Table 1. The data presented in, are based on the 'To:' address field involving the e-mail account owner's address (incoming traffic) and based on the 'From:' address field involving the e-mail account owner's address (outgoing traffic).

However, the decision tree information compiled in Table 1 does not really provide a sense of the nature of the interactions. This can be aided by the use of social network visualisation to show where the "interesting" interactions are occurring as can be seen in Figure 7. From Figure 7, it is observed that many of the "interesting" interactions involve *clientA*, *clientI*, and *clientG*, whom are highly extroverted individuals (see Figure 4). This type of result was what was expected from the assigning those behaviour profiles.

Table 1: Interesting e-mail traffic interactions derived from two decision tree outputs produced from the e-mail traffic data. Highlighting has been used for interactions found in two accounts.

E-mail Account of Interest	Incoming Interactions	Outgoing Interactions	Type of Interaction
clientA@utas.edu.au	clientI to clientA; where date <= day 24.42; 18 messages	clientA to clientI; where date <= day 22.65; 18 messages	*Two-way -> found in one account
	clientG to clientA; where date > day 24.42 and date <= day 109.01; 95 messages	clientA to clientG; where date > day 22.65 and date <= day 108.52; 93 messages	**Two-way -> found in two accounts
	clientI to clientA; where date > day 109.01; 92 messages	clientA to clientI; where date > day 108.52; 92 messages	*Two-way -> found in one account
clientB@utas.edu.au	clientG to clientB; where date <= day 110.6; 68 messages	clientB to clientG; where date <= day 107.58; 62 messages	**Two-way -> found in two accounts
	clientA to clientB; where date > day 110.6; 72 messages	clientB to clientA; where date > day 107.58; 67 messages	*Two-way -> found in one account
clientC@utas.edu.au	clientI to clientC; where date <= day 107.1; 93 messages	clientC to clientI; where date <= day 109.09; 93 messages	*Two-way -> found in one account
	clientG to clientC; where date > day 107.1; 161 messages	clientC to clientG; where date > day 109.09; 156 messages	**Two-way -> found in two accounts
clientD@utas.edu.au	-	clientD to clientB; where date <= day 23.23; 7 messages	One-way interaction
	-	clientD to clientI; where date > day 23.23 and date <= day 26.31; 4 messages	*Two-way -> found in one account
	-	clientD to clientF; where date > day 26.31 and date <= day 28.97; 2 messages	One-way interaction
	-	clientD to clientB; where date > day 28.97 and date <= day 31.59; 2 messages	One-way interaction
	clientI to clientD; where date <= day 48.92; 31 messages	clientD to clientI; where date > day 31.59 and date <= day 49.71; 19 messages	*Two-way -> found in one account
	clientG to clientD; where date > day 48.92; 128 messages	clientD to clientG; where date > day 49.71; 119 messages	*Two-way -> found in one account
clientF@utas.edu.au	-	clientF to clientC; where date <= day 104.08; 15 messages	One-way interaction
	-	clientF to clientD; where date > day 104.08; 19 messages	One-way interaction
clientG@utas.edu.au	-	clientG to clientC; where date <= day 7.7; 3 messages	One-way interaction
	clientB to clientG; where date <= day 32.12; 19 messages	clientG to clientB; where date > day 7.7 and date <= day 35.64; 21 messages	**Two-way -> found in two accounts
	clientA to clientG; where date > day 32.12 and date <= day 105.18; 79 messages	clientG to clientA; where date > day 35.64 and date <= day 108.97; 84 messages	**Two-way -> found in two accounts
	clientC to clientG; where date > day 105.18; 161 messages	clientG to clientC; where date > day 108.97; 159 messages	**Two-way -> found in two accounts
clientH@utas.edu.au	-	clientH to clientE; where date <= day 66.72; 27 messages	One-way interaction
	-	clientH to clientI; where date > day 66.72; 104 messages	One-way interaction

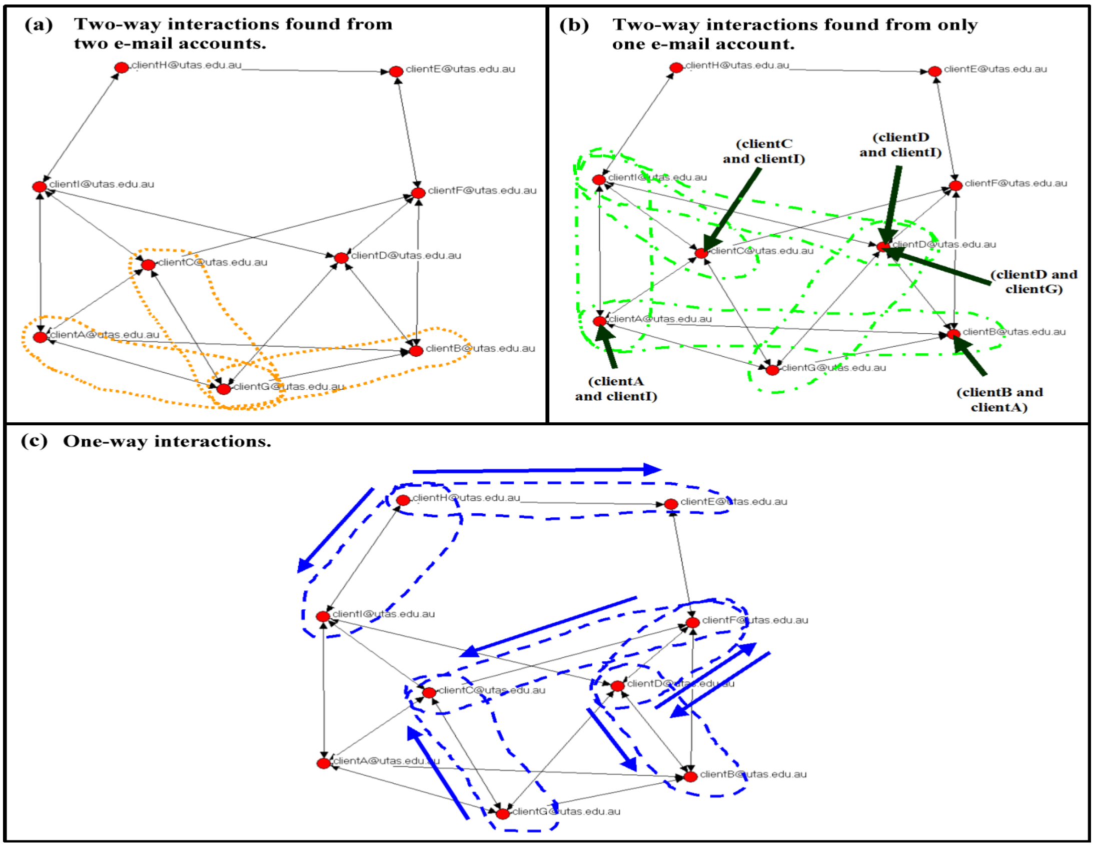
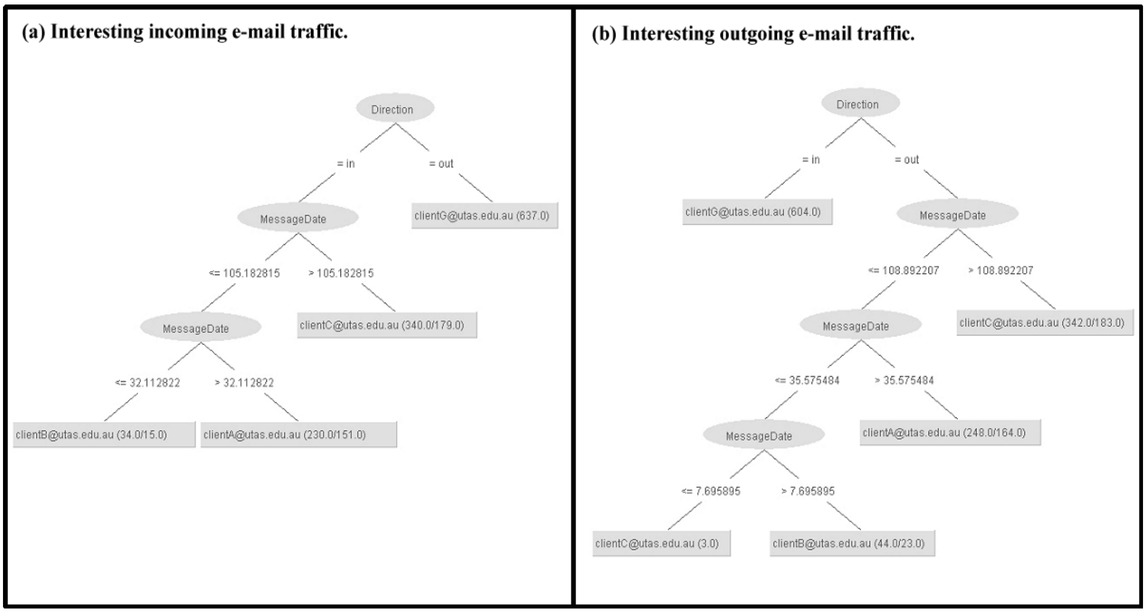


Figure 7: Social network diagrams showing where the decision tree identified "interesting" e-mail traffic interactions.

Details of Interesting Patterns From ‘clientG@utas.edu.au’ Account

The diagrams and screenshots from Figures 8 to 12 provide an account of the e-mail traffic interactions that *clientG* has been involved in over the 182 simulation days. In the first week, *clientG* received 3 e-mails, as shown in Figure 9, and sent out 3 e-mails to *clientC* prior to day 7.7 (Figure 8b). Moving along to week 15 (around day 108), *clientG* has a slight dip in outgoing e-mails (Figure 10) before having a significant increase in outgoing e-mails. Around the same time, *clientA* has a significant dip in incoming e-mails on week 15 (Figure 11), before having a sharp spike in incoming e-mails. A close look at the daily e-mail traffic for *clientC* (Figure 12) reveals that leading up to the 108th day *clientC* had a decrease in incoming e-mail traffic, before receiving more e-mails after the 108th day. These details show there is an interesting change in interaction from *clientG->clientA* to *clientG->clientC* on week 15, suggesting that *clientG* must have had a significant influence on the nature of these interactions.



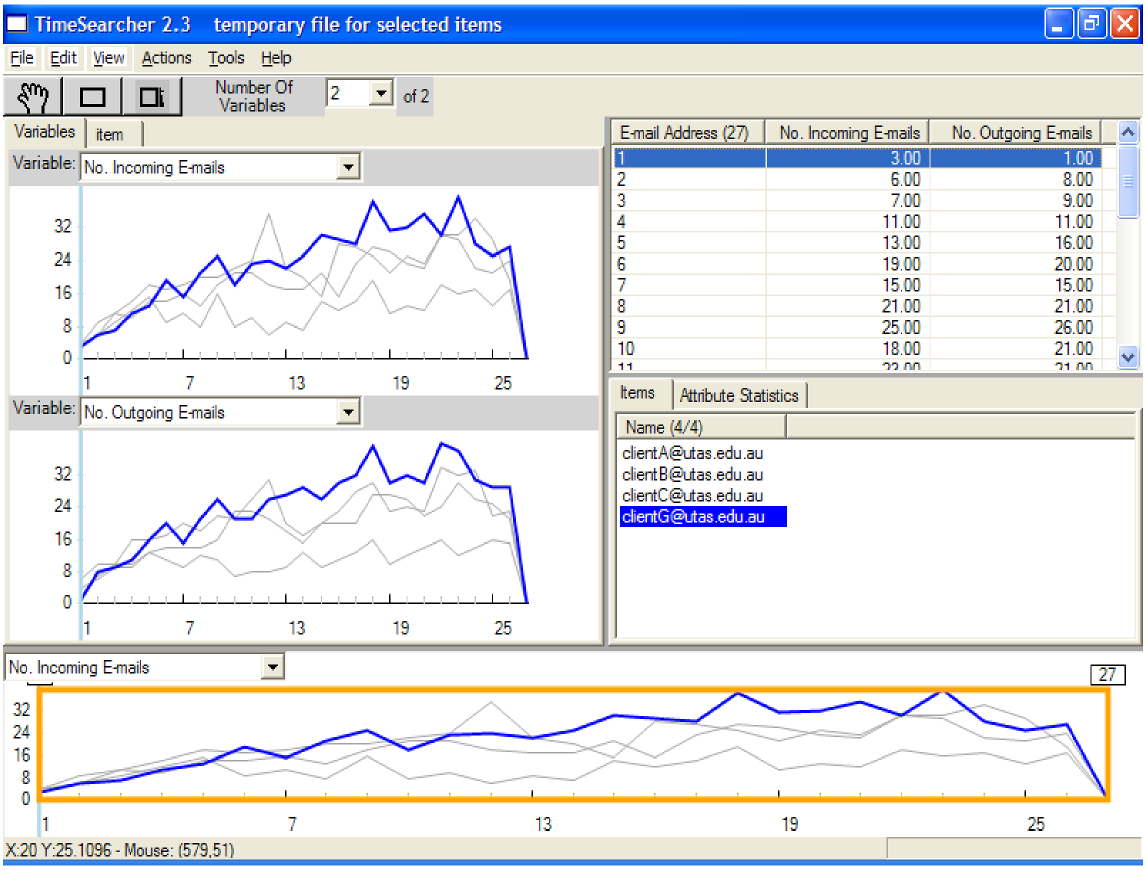
(a) Interesting incoming e-mail traffic.

Direction
= in
= out
clientG@utas.edu.au (637.0)
MessageDate
<= 105.182815
> 105.182815
clientC@utas.edu.au (340.0/179.0)
MessageDate
<= 32.112822
> 32.112822
clientB@utas.edu.au (34.0/15.0)
clientA@utas.edu.au (230.0/151.0)

(b) Interesting outgoing e-mail traffic.

Direction
= in
= out
clientG@utas.edu.au (604.0)
MessageDate
<= 108.892207
> 108.892207
clientC@utas.edu.au (342.0/183.0)
MessageDate
<= 35.575484
> 35.575484
clientA@utas.edu.au (248.0/164.0)
MessageDate
<= 7.695895
> 7.695895
clientC@utas.edu.au (3.0)
clientB@utas.edu.au (44.0/23.0)

Figure 8: WEKA decision tree output for the e-mail traffic belonging to the ‘*clientG@utas.edu.au*’ account.



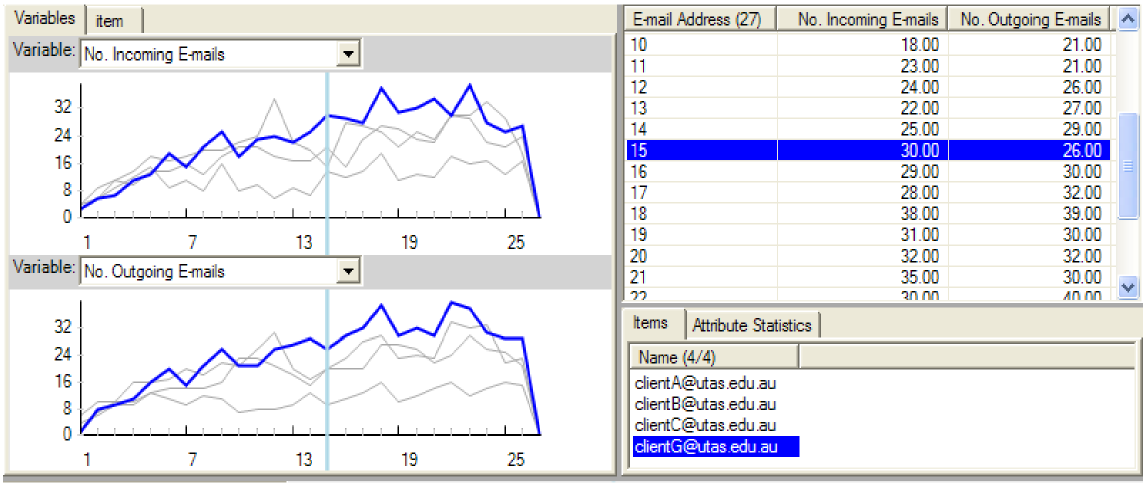
E-mail Address (27)	No. Incoming E-mails	No. Outgoing E-mails
1	3.00	1.00
2	6.00	8.00
3	7.00	9.00
4	11.00	11.00
5	13.00	16.00
6	19.00	20.00
7	15.00	15.00
8	21.00	21.00
9	25.00	26.00
10	18.00	21.00

Name (4/4)

clientA@utas.edu.au
clientB@utas.edu.au
clientC@utas.edu.au
clientG@utas.edu.au

X:20 Y:25.1096 - Mouse: (579,51)

Figure 9: Weekly time-series visualisation for *clientG*, highlighting the 1st week.

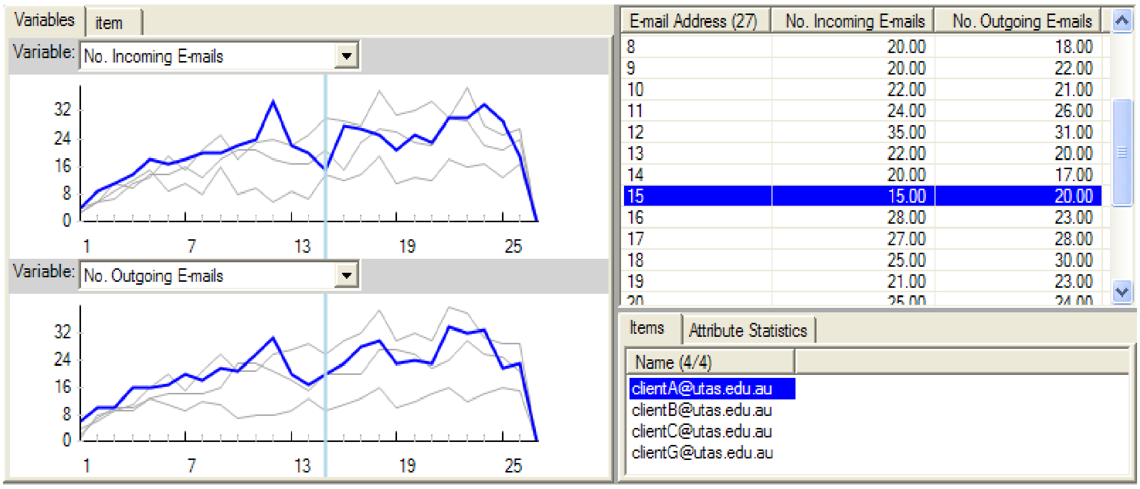


E-mail Address (27)	No. Incoming E-mails	No. Outgoing E-mails
10	18.00	21.00
11	23.00	21.00
12	24.00	26.00
13	22.00	27.00
14	25.00	29.00
15	30.00	26.00
16	29.00	30.00
17	28.00	32.00
18	38.00	39.00
19	31.00	30.00
20	32.00	32.00
21	35.00	30.00

Name (4/4)

clientA@utas.edu.au
clientB@utas.edu.au
clientC@utas.edu.au
clientG@utas.edu.au

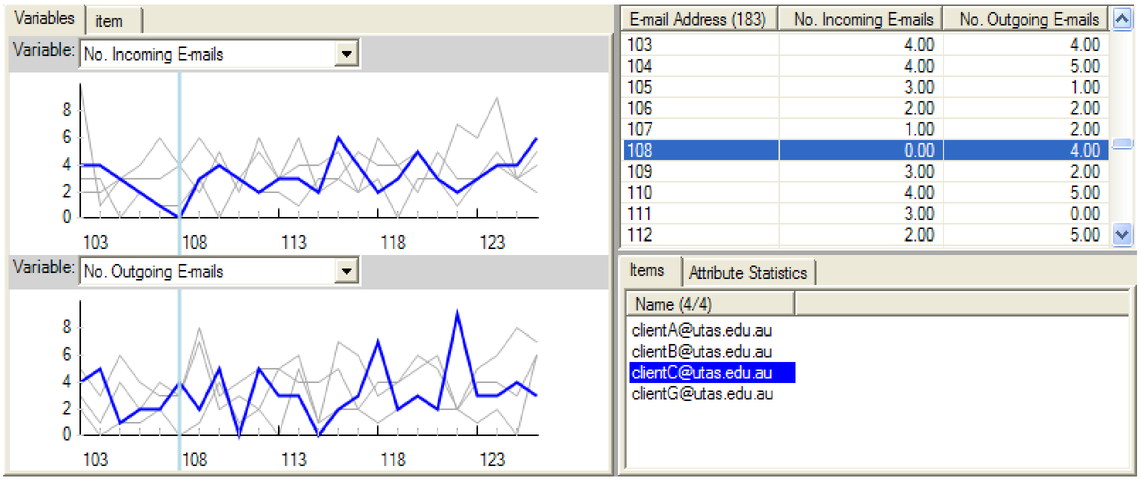
Figure 10: Weekly time-series visualisation for *clientG*, highlighting the 15th week (~day 108) with a slight dip in number of outgoing e-mails.



E-mail Address (27)	No. Incoming E-mails	No. Outgoing E-mails
8	20.00	18.00
9	20.00	22.00
10	22.00	21.00
11	24.00	26.00
12	35.00	31.00
13	22.00	20.00
14	20.00	17.00
15	15.00	20.00
16	28.00	23.00
17	27.00	28.00
18	25.00	30.00
19	21.00	23.00
20	25.00	24.00

Name (4/4)
clientA@utas.edu.au
clientB@utas.edu.au
clientC@utas.edu.au
clientG@utas.edu.au

Figure 11: Weekly time-series visualisation for *clientA*, highlighting the 15th week (~day 108) with a significant dip in number of incoming e-mails, then a spike afterwards.



E-mail Address (183)	No. Incoming E-mails	No. Outgoing E-mails
103	4.00	4.00
104	4.00	5.00
105	3.00	1.00
106	2.00	2.00
107	1.00	2.00
108	0.00	4.00
109	3.00	2.00
110	4.00	5.00
111	3.00	0.00
112	2.00	5.00

Name (4/4)
clientA@utas.edu.au
clientB@utas.edu.au
clientC@utas.edu.au
clientG@utas.edu.au

Figure 12: Daily time-series data for *clientC*, with a downward trend in incoming e-mails prior to the 108th day, then a rise in incoming e-mails after the 108th day.

DISCUSSION

From the case study, it was shown that the combination of different visualisation methods proved to be very useful in analysing the details of the interesting e-mail traffic interactions occurring in the simulated e-mail data. The decision tree technique was useful in assisting the user in being able to "pinpoint" the location of interesting e-mail traffic behavioural patterns between e-mail clients. The additional information provided by the social network visualisation and time-series visualisation methods also helped make more sense of what was actually occurring inside the e-mail traffic data.

However, there were some problems associated with using and analysing the decision tree output. The decision tree outputs produced by WEKA created a very large decision tree of size 88 leaves and 105 nodes for incoming e-mail traffic going into e-mail clients' accounts, and a decision tree of size 95 leaves and 119 nodes for outgoing e-mail traffic coming from e-mail clients' accounts (both too large to sufficiently display in this paper). Not all the leaves of the decision tree (i.e. where the "interesting" information is usually located) contained useful information, so this required a lot of manual labour and time (at least 5 – 6 hours) to compile together the useful information from the decision trees for incoming and outgoing e-mail traffic data, and to produce the data shown in Table 1. Although the decision tree technique in the case study of 9 e-mail clients shows that it does provide useful information, the application of this data mining technique for much larger data sets of hundreds or even thousands of e-mail clients clearly brings up the following questions: Are decision trees feasible for analysing large amounts of e-mail traffic data? Would there be a much more efficient and faster way of analysing much larger data sets? Is there a much better approach or technique for mining and presenting the "interesting" e-mail traffic interactions to the user?

CONCLUSION:

Through the use of the e-mail traffic analyser system, we have demonstrated the usefulness of visualisation tools such as Netdraw, Mage 3D, TimeSearcher, and WEKA, and discussed how the information provided by such tools may be useful for analysing the communication behaviour of criminals/terrorists. We have also shown how the use of decision trees aids the user in being able to "pinpoint" the location of interesting e-mail traffic interactions and demonstrated the type of information revealed by this technique. However, there are some considerations from the use of decision trees that we must take into account when analysing large data sets. The issues described in the Discussion section are important if the tools developed with A.I. based or machine learning based techniques are to be practical and aid an intelligence analyst in quickly tracking the communication behaviour of criminals or terrorists. These are the considerations we will take into account in our future work when continuing developing the e-mail traffic analyser, investigating other A.I. techniques (e.g. artificial neural networks or neuro-fuzzy systems), and analysing much larger data sets of e-mail clients.

REFERENCES:

Ajzen, I 1988, *Attitudes, personality, and behavior*, Mapping social psychology, Open University Press, Stony Stratford, Milton Keynes.

Aris, A, Khella, A, Buono, P, Shneiderman, B & Plaisant, C 2005, *TimeSearcher 2*, Human-Computer Interaction Laboratory, Computer Science Department, University of Maryland, <http://www.cs.umd.edu/hcil/timesearcher/>.

Borgatti, SP 2002, *NetDraw: Graph Visualization Software*, Analytic Technologies, Harvard, <http://www.analytictech.com/netdraw.htm>.

Carley, KM, Dombroski, M, Tsvetovat, M, Reminga, J & Kamneva, N 2003, 'Destabilizing dynamic covert networks', paper presented to 8th International Command and Control Research and Technology Symposium, National Defense War College, Washington DC.

Coffman, T, Greenblatt, S & Marcus, S 2004, 'Graph-based technologies for intelligence analysis', *Communications of the ACM*, vol. 47, no. 3, pp. 45-7.

Freeman, LC 2000, 'Visualizing Social Networks', Journal of Social Structure, vol. 1, no. 1, <http://www.cmu.edu/joss/content/articles/volume1/Freeman.html>.

Guimera, R, Danon, L, Diaz-Guilera, A, Giralt, F & Arenas, A 2003, 'Self-similar community structure in a network of human interactions', *Physical Review E*, vol. 68, no. 6.

Krebs, VE 2002, 'Mapping Networks of Terrorist Cells', *Connections*, vol. 24, no. 3, pp. 43-52.

Mena, J 2003, *Investigative Data Mining for Security and Criminal Detection*, 1st edn, Butterworth-Heinemann.

Negnevitsky, M 2004, *Artificial Intelligence: A Guide to Intelligent Systems*, 2nd edn, Addison Wesley, Essex.

Newman, MEJ & Girvan, M 2004, 'Finding and evaluating community structure in networks', *Physical Review E*, vol. 69, no. 2.

Richardson, DC 2002, *MAGE: 3-D Visualization Program*, Biochemistry Department, Duke University, Durham, <http://kinemage.biochem.duke.edu/software/mage.php>.

Scott, J 2000, *Social Network Analysis: A Handbook*, 2nd edn, SAGE Publications, London.

Stolfo, SJ, Hershkop, S, Wang, K, Nimeskern, O & Hu, CW 2003a, 'A behavior-based approach to securing email systems', in *Computer Network Security*, SPRINGER-VERLAG BERLIN, Berlin, vol. 2776, pp. 57-81.

---- 2003b, 'Behavior profiling of email', in *Intelligence and Security Informatics, Proceedings*, SPRINGER-VERLAG BERLIN, Berlin, vol. 2665, pp. 74-90.

Tyler, JR, Wilkinson, DM & Huberman, BA 2005, 'E-mail as spectroscopy: Automated discovery of community structure within organizations', *Information Society*, vol. 21, no. 2, pp. 133-41.

Witten, IH & Frank, E 2000, *Data Mining: Practical Machine Learning Tools and Techniques with Java Implementations*, Morgan Kaufmann, San Francisco, Calif.

Xu, J & Chen, HC 2003, 'Untangling criminal networks: A case study', in *Intelligence and Security Informatics, Proceedings*, SPRINGER-VERLAG BERLIN, Berlin, vol. 2665, pp. 232-48.

Xu, J, Marshall, B, Kaza, S & Chen, HC 2004, 'Analyzing and visualizing criminal network dynamics: A case study', in *Intelligence and Security Informatics, Proceedings*, SPRINGER-VERLAG BERLIN, Berlin, vol. 3073, pp. 359-77.

COPYRIGHT