DScentTrail: A New Way of Viewing Deception

S.J. Dixon, M.B. Dixon, J. Elliott, E. Guest and D. J. Mullier[1]

Abstract The DScentTrail System has been created to support and demonstrate research theories in the joint disciplines of computational inference, forensic psychology and expert decision-making in the area of counter-terrorism. DScentTrail is a decision support system, incorporating artificial intelligence, and is intended to be used by investigators. The investigator is presented with a visual representation of a suspect's behaviour over time, allowing them to present multiple challenges from which they may prove the suspect guilty outright or receive cognitive or emotional clues of deception. There are links into a neural network, which attempts to identify deceptive behaviour of individuals; the results are fed back into DScentTrail hence giving further enrichment to the information available to the investigator.

1 Introduction

DScent was a joint project between five UK universities combining research theories in the disciplines of computational inference, forensic psychology and expert decision-making in the area of counter-terrorism. This paper concentrates on phase two of the project and discusses DScentTrail which is an investigator decision support system. A neural network links into DScentTrail and attempts to identify deceptive behavioural patterns. Preliminary work was carried out on a behavioural based AI module that would work separately alongside the neural network with both AI modules identifying deception before updating DScentTrail with their integrated results. For the purpose of data generation along with hypothesis and system testing, the project team devised a closed world game; the Location Based Game.

The Location Based Game was an extension of a board game; the Cutting Corners Board game [1] created within phase one of the project. Phase one consisted of the development of a feed-forward back-propagation neural network used to identify deceptive behavioural patterns within the game data. Due to extra complexities introduced with the location based game a feed-forward back-

[1] All Leeds Metropolitan University, LS6 3QS, UK s.j.dixon@leedsmet.ac.uk, m.dixon@leedsmet.ac.uk, j.elliott@leedsmet.ac.uk, e.guest@leedsmet.ac.uk, d.mullier@leedsmet.ac.uk

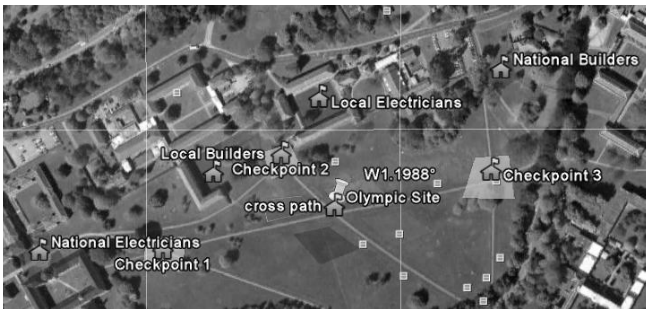
propagation architecture was no longer suitable, therefore research into alternative architectures was required, this is further discussed in the Neural Network section.

2 The Location Based Game

Participants from a variety of different backgrounds were recruited to partake in the game trials. These participants traversed set locations (see figure 1) using GPS enabled devises to communicate, navigate and purchase items. Each participant took either the role of a builder attempting to construct part of an Olympic stadium, or a terrorist masquerading as a builder with the aim of planting explosives. For reasons discussed in the conclusions, only 2 games worth of data was available for testing purposes.

Each game comprised of four teams, and each team comprised of three players, a foreman, and two of the following tradesmen: an electrician; an explosives expert; or a builder. The games were divided into four tasks with the winning team being the first to complete all four. Virtual cash rewards were given to teams upon completion of tasks. Each task involved specific team members being in certain locations at certain times. They involved participants purchasing specific items and unloading these at their site. One team member was given the role of van driver and could purchase items. Vans were virtual and could be transferred between team members via the mobile device.

The GPS locations consisted of four shops; four sections of the Olympic site, one per team; and three fixed checkpoints where players would be checked by investigators and either given a cash reward or penalty. Police investigators performed random checks on players who they suspected to be behaving suspiciously, the same rules applied as with the fixed checkpoints. The four shops consisted of a local and a national electrical store selling dynamite sticks and wiring looms, and a local and a national builder’s yard selling construction blocks, soil and fertiliser. Both local stores only stocked one item type at any given time whilst the national stores carried full stock.



 Scale: 45m (Approx)

Figure 1 Map of Location Based Game Playing Area showing the four stores; three fixed checkpoints and the four areas of the Olympics Site.

To complete the game all three types of tradesmen were required, therefore, the foremen were required to sub-contract players between teams to perform specific tasks. The GPS enabled devices recorded all player movements within the game area. The same devices were used to store and transfer money, transfer vans between team members and to purchase, reveal and drop off virtual items at the sites. Before each game all teams were given a short amount of time, approximately 30 minutes, to devise their strategies. The terrorist teams would need to be deceptive in order to cover up their overall objective of causing an explosion, whereas the building teams would devise strategies which were not based on deception, rather maybe slightly bending the rules to gain an advantage. On completion of a game all team members were interviewed separately by police investigators.

3 AI techniques for counter-terrorism

The use of various AI techniques, such as data mining, artificial neural networks, symbolic AI and Case Based Reasoning for counter-terrorism has been advocated by Markman [2] and Marappan [3]. Projects which consider such techniques are discussed below.

Schneier [4] in his article on *Why Data Mining Won't Stop Terror*, writes that data mining works best when you are searching for a well-defined profile, a reasonable number of attacks per year and a low cost of false alarms. Rudmin [5], Professor of Psychology at the University of Tromso, Norway, is also sceptical regarding data mining techniques used for counter-terrorism and disregards them completely. Rudmin states that in order to make a Bayesian computation, he estimates that at best in the USA there would be a base-rate of one terrorist per 300,000 people and that if a surveillance monitoring system had an accuracy rate of 40% positive identification of real terrorists then according to Bayes' Theorem [6] the misidentification rate would be 0.01%, or 30,000 innocent people. Rudmin stresses that these numbers are simply examples based on one particular technology.

Data mining was not used on the DScent project since it is generally used for extracting information from large quantities of data that is collected for reasons other than for the purpose of mining itself, the purpose of the mining being to find extra, useful information. The DScent data was explicitly designed and collated for identifying suspicious behaviour. DScent would not encounter the problems outlined by Professor Rudmin of having to potentially question 30,000 innocent people as the set did not contain the entire population, it is merely a well established sub-set. Ware [7] in his paper on antiterrorism states that neural networks do not lend themselves easily to real-time updated information and has concerns regarding the limited historical data on terrorist attacks, he further comments on how terrorist tactics are not static and change over time. These issues have been carefully considered during the project and are discussed in further detail below. The reason for choosing a neural network as an AI application within DScentTrail, was that a neural network is the most likely type of computer system that will work with a non-polynomial problem such as behavioural patterns of humans. Although Ware's observations may be valid, by identifying the key input factors to the neural network and keeping these to an absolute minimum, then the amount of historical data required for training will be far less. Furthermore, if the neural network can identify deception amongst humans from a small amount of inputs then we are getting closer to that "well-defined profile" of which Schneier speaks.

4 DScentTrail System

A graphically based software product was developed to help visualise game data. Extensive research was carried out to ensure that the interface was designed in such a way that it would benefit investigators in an interview situation and not only serve as a visualisation tool within the project. Various types of information were collated, processed and then presented by means of a 'scent trail'. A scent trail within the project is a collection of ordered, relevant behavioural information over time for a suspect. Viewing this information graphically would allow an investigator to present multiple challenges from which they may prove the suspect guilty outright or receive cognitive or emotional clues of deception [8]. DScentTrail has links into a neural network that attempts to identify deceptive behavioural patterns of individuals, giving further enrichment to the information available to the investigator, not only by supplying them with related information that may not have been possible to find manually but also by reducing their cognitive load, allowing them to concentrate on their interviewing techniques.

4.1 User Interface Design

Various screen designs were created, figure 2 shows the primary suspect screen with two secondary suspects selected. The primary suspect window is located to the left and stays constant throughout the investigation, whereas various secondary suspect windows may be activated as and when required.

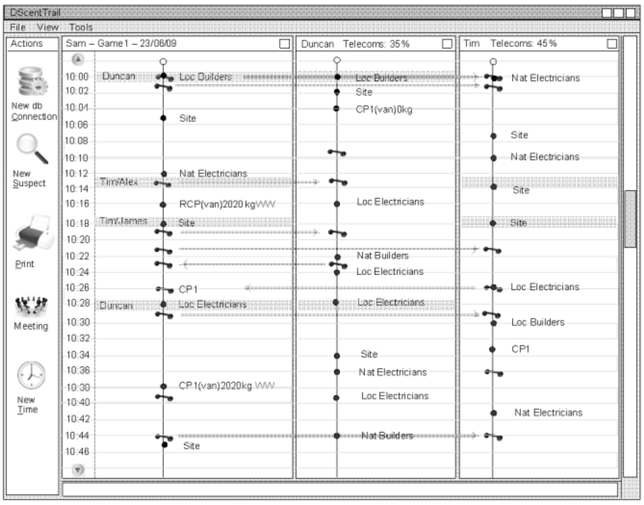


Figure 2 DScentTrail screen design showing both primary and secondary suspects movements and actions through time with their associated communication and location links.

For all windows within the DScentTrail system, time is displayed down the y axis and suspect information along the x axis, both of which are scrollable. All suspect windows display a time-line. A time-line represents the 'scent trail' and shows a series of events for a participant, the name of the suspect is displayed at the top of the window. A time-ordered list of locations and police checks is displayed down the right side of the time-line, these locations are listed in the 'The Location Based Game' section above. If a participant driving a van enters a fixed or has an investigator initiated checkpoint additional information is displayed, consisting of the weight of the van and up to two items which must be revealed. In figure 2 above, Sam at 10.16am had an investigator initiated *random* checkpoint, had a van weight of 2020kg and revealed two wiring looms. At 10.26am Sam entered checkpoint1 but as she was not driving a van, no additional information was displayed. Table 1 shows the codes used for the various stock items.

Stock Item	**Code**
Dynamite	D
Wiring Loom	W
Construct ion Block	B
Fertiliser	F
Soil	S

Table 1 The Stock Item Code Cross Reference Table showing the codes which would be displayed against the checkpoint events whenever the primary or secondary suspects revealed one or two stock items.

The information down the left side of the primary suspect's time-line shows potential meetings. A meeting is defined by the investigator; it is where two players are within *x* meters for greater than *y* seconds. Certain locations may be excluded, for example shops, checkpoints and sites, as these are areas where participants may naturally gather. To display a secondary suspect's time-line the investigator would right click the mouse over a name down the left side of the primary suspect's time-line, alternatively they may select 'New Suspect' from either the top menu bar under 'File' or from the side menu bar. Multiple secondary suspect time-lines may be displayed at one time.

The horizontal arrows in figure 2 show telecommunication activity between primary and secondary suspects with the arrow head indicating the direction of the call. The horizontal bars indicate potential meetings, again between the primary and secondary suspects. Hovering the mouse over either type of highlighter bar provides additional information, for example call or meeting duration and detailed meeting location information. Nodes on the time-line are either shown in black or red, with black indicating normal behaviour and red indicating potentially deceptive behaviour; the red nodes varying in hue depending on the combined certainty factor generated from the AI modules, drawing the investigators attention to a potential terrorist.

The investigator has the option to highlight alerts for all movements into locations which have taken greater than the calculated maximum travel time; figure 3 displays a detailed trajectory view. Here the dotted arrow shows the player leaving the local electrical store at 10.28am, stopping for two minutes, continuing, stopping for a further three minutes before arriving at checkpoint1 at 10.38am. The investigator may choose to hover their pointer over the rest events to view all other participants within a close proximity from the primary suspect during that rest period, which may indicate a reason for the rest period.

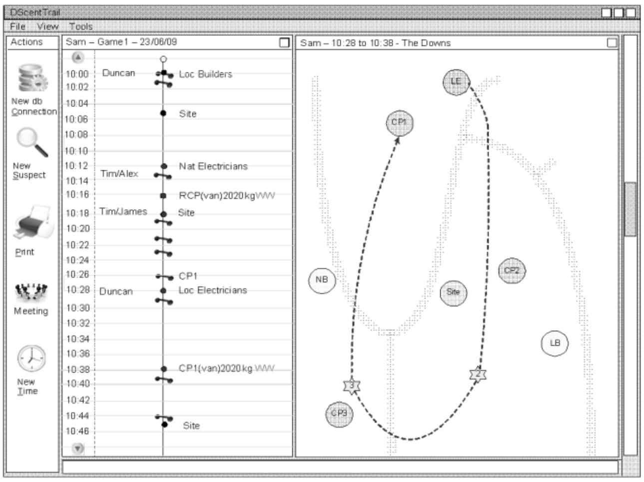


Figure 3 DScentTrail screen design detailing the trajectory of a slowly walked route between two points; the *local electrical store* and *checkpoint1* including a rest event of two minutes followed by a second rest event of three.

Various reporting screens are available to the investigator. Figure 4 shows a telecommunications bar chart for a primary suspect within a game. Other participants who have either made or received calls from the primary suspect during a game are represented along the x axis and the number of calls is displayed along the y axis. A similar chart is available for meetings behaviour during a game. These reporting summary screens may be accessed via the View menu. In addition, the telecommunications chart may be accessed by right clicking the mouse on any of the handset icons down the primary suspect's time-line then selecting 'view summary report'. The meetings chart may be selected by right clicking the mouse on any of the names down the left side of the primary suspect's time-line and selecting 'view summary report'.

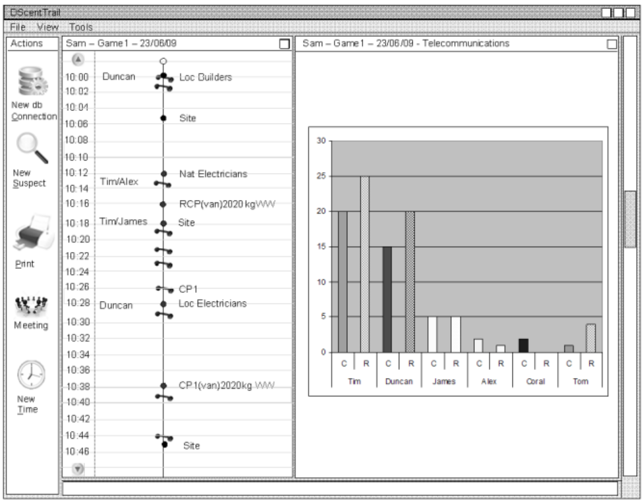


Figure 4 DScentTrail screen design showing the summary of all telecommunication activities for a participant. This chart divides the communications into calls made and calls received.

4.2 *Technical System Design*

DScentTrail is an Object Oriented [9] (OO) system, designed and specified using various techniques from the Unified Modelling Language [10] (UML), such as class and object modelling within the QSEE Superlite Development Environment [11]. All user interface design was created using Microsoft Visio and was written in the Java programming language [12] within the Eclipse Integrated Development Environment [13] (IDE). The game data was captured and stored in an Oracle Spatial database [14]. The DScentTrail system connects to this Oracle database using the Java Database Connectivity (JDBC) application programmer's interface (API) [15]. It was never the intention to implement the entire design for

DScentTrail as this would not have been achievable within the timeframe and with the allocated resources. However, all major areas were implemented, this allowed for the data to be imported from an external non-tailored database and a dynamic class model [2] built, from this meaningful information could be drawn.

5 Neural Network

A neural network was built which attempted to identify deceptive behaviour for a suspect. This neural network was integrated with the DScentTrail system, potentially deceptive scent trails were then highlighted for the investigator's attention. A regression network architecture [16] was adopted for the neural network which takes the output from the previous row of input data and uses this as input with the next. This allows the network to see time series data as opposed to discrete chunks of data. The problem with chunking the data is that there is no distinct point at which to do this, other than to present a complete trail. Presenting a complete trail was not practical in this case as the network would become far too complex to successfully train. By contrast a regression neural network takes a subset of inputs and the complete trail is then passed through the inputs, rather like a person viewing the trail an element at a time whilst retaining a memory of what they have previously seen and forming an opinion based on potentially increasing evidence. A trail is passed to the neural network an element at a time and for each presentation it outputs a certainty that the trail contains suspicious behaviour. For a genuinely suspicious trail the neural network would output a steadily strengthening certainty as it is presented with more data. This process is transparent to the DScentTrail system, all the user would see would be a final output from the neural network.

The neural network was developed using Encog [17], which is a powerful neural network and artificial intelligence application programmer's interface. The neural network architecture used was an Elman Recurrent Network [18], see figure 5 for a graphical representation of the network architecture.

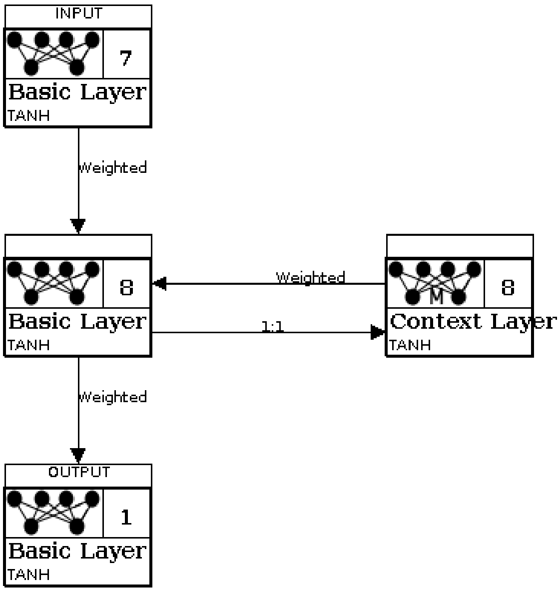


Figure 5 Neural Network Architecture for DScentTrail.

Due to the severe lack of data available, it proved impossible to train the network. Though a regression neural network was implemented and integrated into the DScentTrail system. Future work[2] is underway to develop a method for automatically generating behavioural data, building on the rules of the location based game [19]. This will be done by combining intelligent agents [20] with gene expression programming [21] and the use of an Emdros database [22]. The intention is to train and fully test the neural network on the receipt of this game data. To allow the network a chance of generalising, the number of columns in the input file were reduced to a bare minimum. The resulting columns were 'event'; 'time'; 'items'; 'award'; 'call duration'; 'call location'; 'call count' and 'close players'. The 'game Id', 'player Id' and 'time' were present for all rows, but not

---

2 An EPSRC Standard Proposal is currently being reviewed to continue the work of DScent. This proposal has the following main objectives: 1) Incrementally develop techniques for generating deceptive behavioural data starting with the DScent board game, extending to the location based game, and finally making the data more closely model everyday life. This will include developing methods for generating new deceptive behavioural patterns, as we cannot assume that all past and future patterns are present in the data. 2) Incrementally extend DScentTrail so that it can handle more types of data and more complexity bringing it away from the constraints of the game towards real life. Develop a hybrid AI module consisting of the existing neural network and a (yet to be decided on) behavioural based AI module. 3) Validate the automatically generated data against existing data obtained from the DScent project and against real-life criminal datasets. The DScent system including the hybrid AI module will be fully tested at the end of each development phase with the automatically generated data.

presented. A player's worth of game data was shown to the network in time order; see [2] for more details.

5.1 Behavioural Based Artificial Intelligence

Phase two incorporated the preliminary stages of design for a symbolic AI [23] system, here the decision making process behind the output would be visible to the investigators; this was in direct contrast to the 'black box' nature of the neural network. By analysing the relationships of the variables within the game, patterns and behavioural rules could be extracted and type classifications derived. These models would then be embedded, attached with probabilistic information to identify emergent deceptive behaviour. This would become a refining, iterative process for future research and development.

Theme 5.0 software [24] was used for detecting and analysing hidden patterns of behaviour within the game data. Theme detects statistically significant time patterns in sequences of behaviour and provides basic analysis tools. This behavioural based AI module would contain probabilistic information and would be centred on pattern matching and relationship modelling of entities within their environment. Two files are required to analyse data using Theme software; a category table and a data file [2]. The category table contains coded metadata used to record the subject (participants); the behaviours (events) and the modifiers (variables). Mutual exclusivity between the three is enforced within this file. The data file contains behavioural data, scored according to the codes defined in the category table. Separate data files were required for each participant resulting in 21 for analysis purposes.

Analysis was performed on full game and individual team game data. When entire game data was analysed Theme detected over 170 patterns. Due to the nature of the game; each player being part of a three person team there was no disadvantage in splitting the analysis down into teams, this resulted in more manageable data sets and subsequently a simplified analysis process. The results were analysed using both temporal and event based analysis.

5.1.1 Temporal Analysis of Results

Theme detected a total of 329 patterns over the two games by setting the following search parameters: "Minimum Occurrences = 3"; "Significance Level = 0.0005" and "Exclude Frequent Event Types = Yes". Due to the simplicity of the game rules, Theme did not identify any patterns which the development team were not already aware, though it did serve as verification. It was apparent that by identifying these patterns Theme recognised the individual teams, which would be extremely significant in a more complex system outside the constraints of the game. Theme provides various ways of viewing pattern information; figure 6 shows a Pattern Length Distribution graph giving an overview of patterns grouped by their number of internal elements.

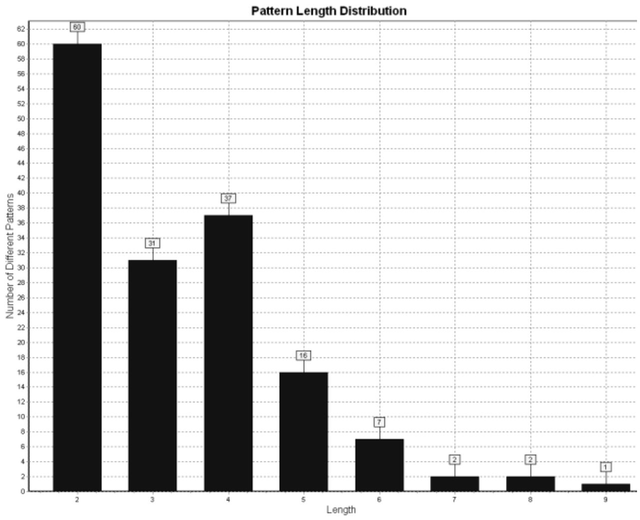


Figure 6 Theme pattern length distribution graph showing the number of different patterns containing various lengths.

From here analysis was carried out within the individual length categories by viewing the separate pattern breakdown diagrams, an example of which can be seen in figure 7 below:

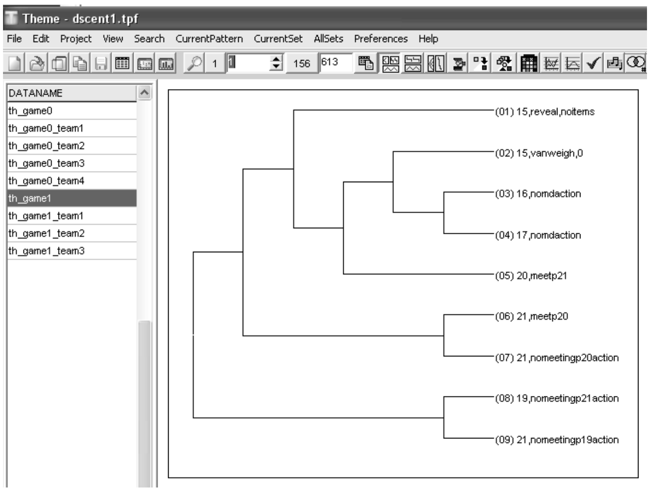


Figure 7 Theme pattern breakdown diagram showing the structure of patterns found within the data.

Once relevant patterns had been identified, it was then important to know how many of each specific relevant pattern had been identified, this was done by viewing the data using the Current Pattern Statistics view.

5.1.2 Event Based Analysis of Results

Key events with their average number of occurrences were identified and calculated for each of the three players within a team, before calculating team averages. The totals were then considered to see whether any teams were behaving differently to others, i.e. differences between the building and the terrorist teams. Numbers were highlighted to indicate significant variants which required further analysis [2]. Alternative categorisation scenarios were analysed but with the sparse amount of data available, no significant patterns were detectable for identifying deceptive behaviour. The following results were drawn from this analysis:

1. Visits to Checkpoint1 and the National Electricians were higher with the terrorist team, this was because the building teams were spreading their visits between the electrical and building stores to purchase what they needed for their building tasks, whereas the terrorist teams had the option of buying all their items from just the electrical stores. Checkpoint1 was on route to the National Electricians; see figure 1 for details.
2. The National Builders and Checkpoint3 were visited much less with the terrorist team, this again was because the terrorist teams did not need to visit the building stores. Checkpoint3 was on route to the National Builders and therefore was visited far less by the terrorist team.

By performing the above analysis process; particularly with more complex data, key events and combinations could be identified and from these combinations, the rules and intelligence of the system could start to be derived.

6 Conclusion

DScentTrail presents a new way of viewing deceptive behaviour both by individuals and by groups. The system proved to be extremely effective when studied by psychologists and experts in the field of interrogation and serious crime investigation[3]. The AI modules working to identify deception would provide DScentTrail with intelligent information attracting the investigator’s attention to a subset of potential terrorist suspects. Future work would include separating the

---

[3] Presentations were given to the stakeholders, including personnel from CPNI and the MoD, showing screen designs and discussing the functionality of the DScent system along with its AI modules. Discussions were had with criminologists to arrive at the optimal visualisation of scent trail information, taking into account different interviewing techniques.

scent trail information into chunks and training the neural network to identify deceptive patterns within a scent trail; which would be necessary when used in the real world.

Meeting the dual requirement of making the location based game playable while enabling it to generate suitable data for all the various analysis required on the project proved not possible within the timeframe. The cognitive load placed on the participants for the location based game was much higher than for the original board game. In addition the participants had less time to think because with the location based game participants played continuously rather than waiting for their throw of a dice. This resulted in the data from the board game being much richer than the location based game, by which is meant containing greater variations in strategies of play. The software development effort required for programming the mobile devices was greatly underestimated resulting in incomplete and unreliable data for system testing purposes.

The observations above suggest that it is very difficult, if not impossible to generate suitably complex data via game playing. Future plans are underway to complete and extend the work of the DScent project. These plans include continuing the work started with Theme and developing a behavioural based AI module to work alongside the neural network in identifying deception; creating a method for automatically generating behavioural data, building on the rules of the location based game and incrementally bringing it in line with reality. This automatically generated data would then be used to fully test the DScent system.

A tentative conclusion drawn from the analysis is that the deceptive behaviour of terrorists is camouflaged by the dishonest behaviour of much of the general population. Artificial intelligence is a powerful tool and can be extremely useful in today's mass of information, though the results generated by AI techniques are difficult for regular users to interpret without an effective method of visualisation such as DScentTrail.

7 Acknowledgements

The DScent project was funded by the EPSRC, grant number: EP/F014112/1
Project partners included Lancaster University, University of Nottingham, University of St. Andrews and University of Leicester.

References

1. Dixon, S., Dixon, M., Elliott, J., Guest, E., Mullier, D.: DScent Final Report (2011). Available via: http://www.leedsmet.ac.uk/aet/computing/Computing_DScentFinalReportv2_0_2011.pdf Accessed 8 March 2011.

2. Markman, A., Rachkovskij, D., Misuno, I., Revunova, E.: Analogical Reasoning Techniques in Intelligent Counterterrorism APNN. International Journal ITA Vol. 10, (2) (2003). Available via: http://www.foibg.com/ijita/vol10/ijita10-2.pdf#page=19 Accessed 27 November 2010.
3. Marappan, K., Nallaperumal, K., Kannan, S., Bensujin, B.: A Soft Computing Model to Counter Terrorism. International Journal of Computer Science and Network Security Vol. 8, May 2008, pp.141, (2008).
4. Schneier, B.: Why Data Mining Won't Stop Terror. Wired (2006). Available via: http://www.wired.com/politics/security/commentary/securitymatters/2006/03/70357 Accessed 23 August 2011.
5. Rudmin, F.: Why Does the NSA Engage in Mass Surveillance of Americans When It's Statistically Impossible for Such Spying to Detect Terrorists? Counterpunch (2006). Available via: http://www.counterpunch.org/rudmin05242006.html Accessed 30 October 2010.
6. Joyce, J.: Bayes' Theorem. In: The Stanford Encyclopedia of Philosophy, Fall 2008 Edition. Edward (2008). Available via: http://plato.stanford.edu/archives/fall2008/entries/bayes-theorem/ Accessed 09 June 2011.
7. Ware, B.S., Beverina, A., Gong, L., Colder, B.: A Risk-Based Decision Support System for Antiterrorism (2002). Available via: http://www.dsbox.com/Documents/MSS_A_Risk-Based_Decision_Support_System_for_Antiterrorism.pdf Accessed 30 January 2009.
8. Ekman, P.: Telling lies: Clues to deceit in the marketplace, marriage, and politics. New York, Norton (2001).
9. Ambler, S.W.: Introduction to Object-Orientation and the UML (2009). Available via: http://www.agiledata.org/essays/objectOrientation101.html Accessed 26 January 2011.
10. Booch, G., Jacobson, I., Rumbaugh, J.: The Unified Modeling Language User Guide. Object Technology Series. 2nd Ed., Addison Wesley (2005).
11. Dixon, M.: QSEE Superlite home. QSEE Technologies (2004). Available via: http://www.leedsmet.ac.uk/qsee Accessed 26 October 2010.
12. Flanagan, D.: Java In a Nutshell, 4th Ed., O'Reilly Media (2002).
13. The Eclipse Foundation: Eclipse.org home (2004). Available via: http://www.eclipse.org/ Accessed 14 October 2010.
14. Oracle Corporation: Oracle Spatial and Oracle Locator (2010). Available via: http://www.oracle.com/us/products/database/options/spatial Accessed 26 October 2010.
15. Reece, G.: Database Programming with JDBC and Java, 2nd Ed., Sebastopol, CA, USA, O'Reilly & Associates, Inc. (2000).
16. Ripley, B.D.: Pattern Recognition and Neural Networks. New York, Cambridge University Press (2008).
17. Heaton, J.: Programming Neural Networks with Encog2 in Java. Heaton Research (2010).
18. Elman, J.L.: Distributed representations, simple recurrent networks, and grammatical structure. Machine Learning Vol. 7, no. 2-3 pp.195-224, doi: 10.1007/BF00114844 (1991).
19. Guest, E., Bolotov, A., Dixon, S., Sandham, A.: Realistic Deceptive Data Generation for Counter Terrorism (2011). EPSRC Standard Proposal. Wiltshire, UK. Ref: M1504908 July 2011.
20. Evertsz, R., Pedrotti, M., Busetta, P., Acar, H.: Populating VBS2 with Realistic Virtual Actors. Proceedings of the 18th Conference on Behavior Representation in Modeling and Simulation pp.1–8, (2009).
21. Ferreira, C.: Gene Expression Programming: Mathematical Modeling by an Artificial Intelligence (Studies in Computational Intelligence), 2nd Ed., Springer (2006).
22. Petersen, U.: Emdros: a text database engine for analyzed or annotated text. Proceedings of the 20th international conference on Computational Linguistics (2004). Available via: http://emdros.org/petersen-emdros-COLING-2004.pdf Accessed 24 January 2011.
23. Haugeland, J.: Artificial Intelligence: The Very Idea. Cambridge, MA, MIT Press (1985).
24. Noldus Information Technology: Theme. Noldus Information Technology (2004). Available via: http://www.noldus.com/content/theme-0 Accessed 27 October 2010.