Using non-verbal social signals and degree centrality to optimise a covert actor’s detection scheme for a healthy networked community

Munene W. Kanampiu
North Carolina A&T State University,
Winston-Salem State University,
Winston-Salem, NC, USA
E-mail: kanampiumw@wssu.edu

Abstract: Passive attacks are of the nature of eavesdropping on, monitoring of, transmissions where the attacker’s goal is to unnoticeably obtain the information in transmission. Passive attacks are difficult to detect and therefore hard to prevent. The main focus for most research in this area has been on preventing the attacks rather than detecting the attacker for example by data encryption. But encryption by itself falls short because passive attack can occur in more ways than just observing exposed data. Encryption is also not always applicable for example in open wireless communication protocols. With these observations, we aim to design a scheme that reduces such attackers’ capability. We introduce the non-verbal social signal deception detection in the investigative module while we optimise the scheme by putting in consideration a participant’s centrality given the topology. The results are an intuitive scalable and optimised scheme that detects, investigates, and expels the guilty suspects.

Keywords: social networks; directed networks; centrality; indegree; outdegree; node popularity; node gregariousness; healthy networked community; active attacks; passive attacks; covert actors; benign covert actor; malicious covert actor; non-verbal social signals; deception detection.

Reference to this paper should be made as follows: Kanampiu, M.W. (2014) ‘Using non-verbal social signals and degree centrality to optimise a covert actor’s detection scheme for a healthy networked community’, Int. J. Information Privacy, Security and Integrity, Vol. 2, No. 1, pp.21–36.

Biographical notes: Munene W. Kanampiu is a Lecturer and IT Support Analyst at Winston Salem State University. He is pursuing his PhD degree at North Carolina A&T State University’s Computational Science and Engineering Department. His research areas are in information security, networking protocols and high powered computing.

1 Introduction

In network security, attacks are generally classified into two main categories, mainly active and passive attacks. As Williams (2006) explains, active network attacks belong to
the class of attacks that involve modification of the data stream or the creation of a false stream and can be subdivided into four categories: masquerading, replay, modification of messages, and denial of service. Although not easy to absolutely prevent because of the wide potential physical, software, and network vulnerabilities, active attacks are not difficult to detect and therefore measures are available to prevent their success. On the other hand, passive attacks are of the nature of eavesdropping on, monitoring of, transmissions where the goal of the attacker is to obtain information that is being transmitted. In their memo, RFC 1704 (Klimt and Yang, 2004), Haller and Atkinson define a passive attack as an attack on an authentication system that inserts no data into the stream, but instead relies on being able to passively monitor information being sent between other parties. This information could be used at a later time in what appears to be a valid session.

Although both a menace to information security, active attackers are mostly detectable and therefore easier to prevent while passive attackers mostly operate without detection and therefore very difficult to deal with since neither sender nor the receiver is aware that a third party has secretly read the message or monitored the traffic pattern. Eavesdropping can result in the release of the learnt sensitive or confidential information by the eavesdropper to some unwanted person or group or in the case of traffic analysis the attacker could determine the location and identity of communicating hosts and could observe the frequency and length of messages being exchanged and thereby guess the nature of the communication that was taking place. In wireless sensor networks (Xi et al., 2010) defines traffic analysis attacks as passive attacks that try to deduce the traffic pattern based on the eavesdropped information. Through analysing the packet traffic it can deduce the location of strategic nodes and then launch an active attack to those locations such as DoS attack. The authors add that defending such a traffic analysis attack would be to prevent the adversary from tracing the location of critical sensor nodes but, they add, would be difficult to use the traditional encryption and authentication to prevent the adversaries from eavesdropping on the wireless communication due to the open wireless communication media exposing the context information to adversaries. In their dynamic data protocol exchange to handle passive attacks (Nasution et al., 2005), the authors explain that the most difficult part of network security is protecting the traffic content from passive attacks. Passive attacks, they explain, cannot be avoided because the data structure (data protocol) of internet transactions is well known and of fixed standard.

Noticeably all these authors, and many others, are focused on prevention of passive attacks at most. This is done mostly through data encryption. However most authors, for example Xi et al. (2010) contend that encryption cannot always be applicable in preventing a passive adversary. These authors cite their scheme as an example of this indicating that it cannot apply encryption due to the open wireless communication media exposing the context information to the adversary. It is with this premise that we are motivated to introduce a novel protocol, dynamic covert passive actors’ detection protocol for a healthy network community, for networked communities that will focus not only on preventing but also detecting and expelling passive attackers from such a community. The reminder of the paper will be organised as follows: Some selected related works will be presented in Section 2. Section 3 will be composed of our proposed approach that will narrate our aim and the basic investigative operations involved. The implementation and simulation will be explained in Section 4. Section 5 will present the results and analysis there after followed by the conclusion in Section 6.
2 Related work

As seen from the listed works and as far as we can tell, any works related to passive attacks in networking have all been focused on protecting rather than detecting and taking action on the covert actors. The following is a few published examples to support this: The Location Privacy against Traffic Analysis Attacks in Wireless Sensor Networks scheme in Xi et al. (2010) prevents the attacker from tracing the location of critical sensor nodes. In their work on Dynamic data protocol exchange to handle passive attacks on trusted transient simple network (TTSN) transactions (Nasution et al., 2005), the authors have considered an alternative to handle passive attacks through utilising dynamic data protocol exchange but this protocol only increases the time consumed by a passive attacker, i.e., security cryptanalysis or brute force, on encrypted internet traffic. Such a protocol will only allow both parties (requester and sender) to replace or modify the data structure of the traffic package as often as required. In their passive and active attack detection in quantum communication works, Halawa and Elkamchouchi (2008) point out that although their proposed quantum signature scheme can detect passive attacks this is only possible in quantum infrastructure unlike the classical networking infrastructure. These authors point out that although inherently secure against passive attacks, quantum communication is still considered impractical due to the difficulty of converting an electrical signal into a quantum signal and vice versa. In their paper, secure mobile agent and its platform from passive attack of malicious agents (Prem and Swamynathan, 2012), the authors implement a serial encryption technique to ensure the confidentiality of the data that are retrieved from each remote server’s address while they also ensure the remote server’s address is also secured to face the possible eavesdropping threat. In Kong et al. (2003) the authors emphasise the need for countermeasures to passive attacks by demonstrating that existing ad hoc routing protocols are vulnerable to passive attacks. They demonstrate in hostile environments, adversaries can launch passive attacks against interception prone routing information embedded in routing messages and data packets. Stubblefeld et al. (2004) implemented a passive attack against WEP, the link layer security protocol for on IEEE 802.11 networks, using the Fluhrer, Martin, and Shamir passive attack method. They were able to recover the 128 bit secret key used in a production network.

Similarly, Tews and Martin (2009) augmented the ease of a passive attack on both WEP and WPA 802.11 wireless standards in their paper. Their work was aimed at improving the previous such encryption key recovery attacks such as the FSM attack by Fluhrer et al. (2001), the Korek (2004b) attack, the PTW attack (Tews, 2007), and the Chopchop (Korek, 2004a) attack. In Nakayama et al. (2010) the network-based traitor-tracing technique scheme uses traffic patterns analysis to detect passive actors who are watching a streaming content. Generally the scheme uses digital watermarking and encryption keys to observe propagation of the content. However, such a scheme can only be implemented for streaming content and not downloaded content. This is the major limitation in this scheme even before considering other limitations including the probability of watermark attacks such as removal attack, collusion attack, copy attack, closest-point attack, and sensitivity attack (Su et al., 2005; Kutter et al., 2000). All the references in this paper, and to the best of our knowledge many more, only focus on protecting the information from passive attacks through data encryption and very little if any on detection and action taking on the passive attackers. Unlike the existing schemes our approach emphasises on detecting and expelling covert passive actors from the
networked community. Moreover, as indicated earlier, it is not always possible to use encryption schemes on all infrastructures, a situation that would render protection through encryption not feasible and therefore leaving covert actors to act with ease.

3 Our approach

3.1 Aim

Our aim is to maintain a healthy networked community by detecting, investigating, and expelling from the network community malicious passive attackers. The process will involve identifying and categorising all participants in the community as follows:

1. active participants (A) as those that are currently participating in the local community and whose transactions with any external networked community are transparent
2. passive participants (P) as those that have so far not indicated any signs of participation in the local or the external network community
3. covert participants (C) as those that have not participated in their local community but are participating in some external community in the network.

3.2 Basic investigative operations

The scheme’s investigative module will follow as depicted in Figure 1 flow diagram. It begins with the classification of the network members (actors) into either A’s, P’s, or C’s. While it ignores the A’s since transparent hence deemed safe, it will put all P’s on a watch list since these could potentially metamorphose into Cs. All the C’s are put through a rigorous non-verbal social signals deception detection procedure which will refine and distinguish between the good (truth tellers) and the bad (deceiving) C’s. The deceiving actors as determined by this procedure will in turn be placed on a suspicious covert members list, henceforth referred to in this paper as wanted eavesdropping suspects (WEDS) and an investigation is done to identify any members of external communities that these WEDS are in communication with. Once their cohorts are identified a further investigation is done to find whether these identified cohorts contain any classified information leaked from their collaborative WEDS’s network community. As such the next step in our scheme will be to further refine the categorisation of the WEDS either as benign (WEDS that is communicating only non-classified information to the outside), or one that is in fact leaking out domain’s classified information to the outside community. If the latter, such a WEDS will be reclassified as a wanted eavesdropper (WED) and subject to expulsion from the network. The WED investigation process will be carried out by performing an information content (IC) similarity check between the WEDS’s cohort information and that of the WEDS domain’s database. To measure content similarity, our scheme will use an arbitrary preset similarity threshold measure of 0.5. A similarity measure greater than the threshold will be deemed non-trivial and a similarity qualifier between the two. Such a cohort whose IC and that of the WED’s domain passes the similarity threshold will proceed to the second testing, the keyword search procedure. This involves searching the cohort’s data contents for any presence of keywords from the
Using non-verbal social signals and degree centrality

WED’s domain’s classified information list. The presence of such a keyword(s) in the cohort’s data will indicate that the information that was leaked is in fact sensitive and therefore an illegal transaction. The last step will be to identify the WED that was involved in leaking the information (in the case of multiple WEDs communicating with the same external community member). Such a WED when identified, together with the cohorts, will be subject to expulsion from the networked community.

Figure 1 Diagram illustrating the steps in the investigative protocol (see online version for colours)

Degree centrality is among the basic measures of network centrality in social networks analysis (SNA). It is defined as the number of direct links that a particular node has in a network (Scott, 2000; Wasserman and Faust, 2003). It highlights the node with the most links to other nodes in the network which is a reflection of having more direct contrast and adjacency than all the other nodes in the network [new (Wasserman and Faust, 2003; Carrington et al., 2005)]. In a directed network, degree centrality can be assessed for in-degree and out-degree; where ‘in’ represents other actors visit to a particular actor and ‘out’ represents that particular actor visits to other actors in the network (Ahmed et al., 2006). Both these concepts are defined by formulas (1.1) and (1.2) below where x refers to the adjacency matrix for network data.

\[ d_i(\text{out-degree}) = \sum_j x_{ij} \]  

(1.1)
This concept can therefore be used to further define a node’s popularity, and likewise a node’s gregariousness. As noted by Uddin and Hassain in 2011, comparison of degree of centrality of different nodes in different networks with different sizes can be made by dividing the right hand side of the equation 1.1 and 1.2 by N-1, where N is the number of nodes in the network representing the size of the subject network. As such this makes a network-level measure for degree centrality, $D_i$:

$$D_i(\text{out-degree}) = \sum_{j} \frac{jx_{ji}}{N-1}$$

(1.3)

$$D_i(\text{in-degree}) = \sum_{j} \frac{jx_{ij}}{N-1}$$

(1.4)

where $X$ refers to the adjacency matrix for network data.

Our idea of a perfectly healthy networked community is a network with a completely connected directed network topology for the participating nodes. For our healthy network scheme any deviation from this must be scrutinised and if need be investigated for possible eavesdropping. For a perfect healthy networked community our scheme expects the following realisation: For every local network member node $x$, in a $N$ nodes network:

$$\sum_{j} x_{ij} = N - 1$$

(1.5)

$$\sum_{j} x_{ji} = N - 1$$

(1.6)

Edges per node:

$$\sum_{j} x_{ij} + \sum_{j} x_{ji} = 2(N - 1)$$

(1.7)

where by definition $2(N-1)$ is the total network edges

3.3 Our scheme’s nodes anomaly detection and classification stage:

Nodes anomaly detection will be carried out for any network community that does not pass the above network health test. It will involve identifying any node in the network that does not obey or causes a distortion of any of the health metrics as previously set.

Example:

For any node $d_i$,

1. If $\left( \sum_{j} x_{ij} = N - 1 \right)$ for local network, $d_i$ is regarded as completely transparent hence safe.

2. If $\left( \sum_{j} x_{ij} < 1 \right)$ for local network and $\left( \sum_{j} x_{ij} < 1 \right)$ for external networks, $d_i$ is regarded as dormant and is classified as passive but harmless. Because of the general concept of this paper a node’s gregariousness of zero is not a concern since it signals no current potential for eavesdropping. Our scheme therefore will only put such a
node on the eavesdroppers watch list instead of sequestering it for current investigation.

3. If \( \sum_{j} x_{ij} = 0 \) for local network but \( \sum_{j} x_{ij} > 0 \) for external networks then \( d_i \) is deemed suspicious and is classified as wanted eavesdropper suspect (WEDS). This group comprises of those nodes that are not dormant but yet not participating in their local network (i.e., their out-degree is greater than zero but none of their communication is with the local networked nodes). It is this behaviour that creates the suspicion that they might have joined the local network for the purpose of eavesdropping and sharing local network information to their conspirators in some other networks (denoted in red in our diagram). It is important to note here that not all acts of communicating information to the outside should be automatically deemed malicious by the scheme hence the suspect status coined to these nodes until proved in fact malicious. This is the case only when the information passed belongs to the domain classified group. As such our next step in the scheme is to investigate these WEDs to see if their communication with their external cohorts has in it any classified information from the local network thereby reclassifying the node from WEDS to WED. It is this reason that all these WEDS will undergo a social signal deceit detection procedure where they will be subjected to a simple yes/no questionnaire as to whether they are sending out any domain classified information from the local network to the outside networked cohorts.

Figure 2  An example of a networked community with active (green), passive (purple), and covert (red) members (see online version for colours)

A complete example of such a network’s actors categorisation is illustrated in the example Figure 2, where the A’s, P’s, and C’s are represented in green, purple, and red colours, respectively in a three communities (A, B, and C) networked group. The local
domain network communication links are labelled in red colour while the inter-network communication links are labelled in blue colour.

3.4 Deception detection on WEDS

To check whether a suspect actor is communicating classified information our investigation continues by subjecting such actors through a social signal deceit detection procedure. Here they will undergo a simple yes/no questionnaire as to whether they are sending out any domain classified information. As explained earlier, the A’s are transparent and therefore their activities can be easily seen and corrected therefore our protocol takes no action on them. As for the P’s our protocol will put them on a watch list suggesting that such users are benign since their activities are presently unknown and therefore have an equal probability of leaning either towards the active or the covert actors. They could be waiting to gather enough information to share but they could just as well be gathering information to leak out. Our target then remains the C’s (marked in the diagram with red colour and labelled ‘R’) category whose local out-degree coupled with some degree of gregariousness with external networked communities is suggestive of non-transparency. For clarification their links are denoted by a thick and non-continuous red line in our diagram. Our scheme deems such behaviour as suspicious and therefore warranting further investigation.

Figure 3 An example of a horizontally partitioned data topology layout representing non-verbal (deceit/truth) social signal data elements

<table>
<thead>
<tr>
<th>Subject-1</th>
<th>Subject-2</th>
<th>Subject-S</th>
</tr>
</thead>
<tbody>
<tr>
<td>P1 (HP, HV), (R, HP, R, HV), (LHP, LHV)</td>
<td>(HP, HV), (RHP, RH, HV), (LHP, LHV)</td>
<td>(HP, HV), (RHP, RH, HV), (LHP, LHV)</td>
</tr>
<tr>
<td>P2 (HP, HV), (R, HP, R, HV), (LHP, LHV)</td>
<td>(HP, HV), (RHP, RH, HV), (LHP, LHV)</td>
<td>(HP, HV), (RHP, RH, HV), (LHP, LHV)</td>
</tr>
<tr>
<td>.</td>
<td>.</td>
<td>.</td>
</tr>
<tr>
<td>Pn (HP, HV), (R, HP, R, HV), (LHP, LHV)</td>
<td>.</td>
<td>(HP, HV), (RHP, RH, HV), (LHP, LHV)</td>
</tr>
</tbody>
</table>

4 Implementation

Node anomaly detection will be implemented as stated earlier with a mathematical analysis to distinguish between the active and transparent, passive and dormant, passive and benign, and passive and malicious actors of the community. We will assume a deceit detection procedure carried out on the passive and malicious group as introduced and described in the non-verbal deceit deception signal data extraction and analysis by Meservy et al. in Williams (2006). It subjects the suspect through a high quality digital video to extract the hand and face regions from an image sequence then applies colour segmentation process for the classification process. Without delving into the details of the
Using non-verbal social signals and degree centrality

Using non-verbal social signals and degree centrality, we will only briefly explain the general concept of the procedure. The system collects raw video data obtained from the subjects and segments it into discrete units in readiness for further analysis. For our scheme this data breakdown will be as depicted in Figure 3 where the row heading (Pi) represents participant i and the field (Subject–i) values represent the three extracted data segments from the i\textsuperscript{th} individual.

Each field makes up a data segment (data unit) comprised of the head position (HP) and its velocity (HV), the left hand position (LHP) and its velocity (LHV), and the right hand position (RHP) and its velocity (RHV). Some general metrics will be extracted from each of these units and in turn these metrics will be used to compute each unit’s features. The features will then be used to classify the raw data as described by the authors. It is this classification of the raw data extracted from each participant that will be used to determine the participant’s deceit/truth categorisation. Figure 4 shows an example of such raw data features classification. It shows some typical differences observed from a few extracted features from two short videos of a deceiver and a truth teller.

Figure 4  An example of typical differences between a few features extracted from two short videos (a deceiver and truth teller) (see online version for colours)

Our scheme will then sequester suspected actors (WEDS) identified from this procedure for the next phase which will involve identifying all their cohorts that they are in communication with from other networked communities. Data from all the cohorts will be investigated for similarity with the domain dataset in question, in our case e-mails selected from the Enron e-mails. The protocol will set an arbitrary similarity threshold number of 0.5. If two e-mail strings (cohort e-mail versus Enron e-mail) under similarity comparison produce a similarity measure below the threshold, they will be deemed dissimilar. Two strings with a similarity measure above the threshold will be deemed similar and will be subject to a keyword search that will be done by searching the cohort’s e-mail string for pre-established keywords from the domain e-mail dataset. This procedure will be carried out for every cohort e-mail. If a cohort’s e-mail data is deemed similar to the domain’s dataset e-mail and has in it at least one domain’s dataset keyword, then all the WEDS in collaboration with this cohort are examined to see which one actually communicated the secret (keyword). When found, this WEDS will be reclassified from a WEDS to a WED. Both the WED and its cohort will be expelled from their respective networked communities.

Our simulation data are selected from the domain of the Enron’s personal e-mail search that was made public by the U.S government. Our running examples include the
following list of keywords that we established from observing most of the Enron’s e-mails: confidential, ‘intended’, protected, privilege, privileges, privileged, prohibited, sensitivity, and destroy. Here we wish to reiterate that sensitive data is domain dependent. We only selected these keywords from what we deemed sensible and more so for the purpose of our simulation. For example in the example shown in Figure 5, our protocol will detect the words ‘confidential’ and ‘protected’ as domain sensitive keywords mainly because the sentence insinuates so. The figure is an example of a partial e-mail thread from the Enron e-mail collection (Klimt and Yang, 2004). The line containing the deemed sensitive keywords are marked in red font. Any WEDS containing these keywords will be subject to expulsion if its contents similarity ratio to the dataset e-mail reaches or passes the preset threshold.

**Figure 5** Enron mail thread example with sensitive keywords (see online version for colours)

<table>
<thead>
<tr>
<th>Michelle Cash</th>
</tr>
</thead>
<tbody>
<tr>
<td>Enron North America Corp.</td>
</tr>
<tr>
<td>1400 Smith Street, EB 3823</td>
</tr>
<tr>
<td>Houston, Texas  77002</td>
</tr>
<tr>
<td>(713) 853-6401</td>
</tr>
<tr>
<td><a href="mailto:michelle.cash@enron.com">michelle.cash@enron.com</a></td>
</tr>
<tr>
<td>This message may contain confidential information that is protected by the attorney-client and/or work product privileges.</td>
</tr>
<tr>
<td>----- Forwarded by Michelle Cash/HOU/ECT on 03/19/2001 01:08 PM -----</td>
</tr>
<tr>
<td>=09=09 Subject: FW: NYTimes.com Article: Companies Turn to Grades, and Employees Go to Court</td>
</tr>
<tr>
<td>To: Oxley, David</td>
</tr>
<tr>
<td>Subject: RE: NYTimes.com Article: Companies Turn to Grades, and Employees Go to Court</td>
</tr>
</tbody>
</table>

Variable listing with definitions:

- **WEDS**: wanted eavesdropping suspect
- **WED**: wanted eavesdropper
- **P**: network community participant
- **WEDSCohort**: a WEDS cohort in an external network community
- **Keyword**: a word deemed secret or classified by a networked community
- **LC**: local networked community
Using non-verbal social signals and degree centrality

- **XC**: external networked community
- **WEDSArray**: an array to store all WEDS
- **totalWEDS = 0**: total number of WEDS initialised to 0
- **n**: number of participants
- **similarityThreshold**: a literal measure of similarity between two file strings (e-mails)
- **keywordCount**: count of all keywords
- **checkSimilarity()**: returns similarity measure (similarity)
- **categoriseWEDS(WED, WEDS)**: re-categorises a WEDS to a WED
- **expel(expulsion, x, y)**: expels x and y from their respective network communities.

The premise of our protocol:

For a passive attack to be deemed eminent the following must be true:

a. **Similarity >= similarityThreshold**

b. **keywordsCount > 0**

Explanation:

1. high similarity without any keywords is benign at most
2. even with a high keyword count a low similarity measure could result in a different context of the encompassed message and therefore would require a context analysis to qualify as eavesdropped information.

As seen in Algorithm following, the Algorithm will comprise of two main phases:

a. identifying, sequestering, and keeping count of WEDS

b. identifying all cohorts of the WEDS in external networked community, doing a content similarity check with the chosen dataset to establish the similarity threshold and if necessary a keyword search.

**Algorithm**

**Algorithm 1** Identify, sequester, and keep count of WEDS

```
1 Start algorithm
2 for every Ps {1,2,........n}
3 If Pi is passive to LC and Pi is active to XC
4 WEDSArray[i] = Pi
5 totalWEDS += totalWEDS
```

**Algorithm 2** Identify and investigate each cohort of the WEDS in external networked community

```
6 For every WEDS/
7 ID WEDSCohorts
8 For every WEDSCohort
```
While e-mail string dataset != null
  checkSimilarity(WEDSCohort’s data, e-mail set dataset)
  Search(WEDSCohort’s data, keywords)
  If (similarity < similarityThreshold) and (keywordCount = 0))
    abort investigation
  Else If((similarity > similarityThreshold) and (keywordCount > 0))
    Investigate which WEDS sent the secret information
    CategoriseWEDS to WED(WED, WEDS)
    expel(expulsion, WED, WEDCohort
  Stop algorithm

5 Results and analysis

On running the code the results displayed the data similarity value of each of our e-mails compared to the Enron data set e-mails. It also checked for the existence of any domain sensitive words in the suspect’s data to reinforce the similarity findings and to link the contextual meaning of the two data as it is possible to have similar data with contextually different meanings. Actors found to have a similarity value greater than the preset similarity threshold and containing at least 1 domain classified or sensitive word are deemed malicious passive attackers and therefore subject to expulsion from the network.

To explain our results more explicitly we migrated some of our results obtained from running the java simulation code to a simple excel spread sheet as in Table 1. For reasons of information redundancy and space optimisation, the table only depicts a prototypical situation with just two WEDS looping through 10 Enron e-mails. Our actual excel spread sheet results was composed of a 10*10 dense matrix data producing 100 such results.

The 1st column has the Enron e-mail id numbers as that is how we identified the Enron e-mail excerpts we used in our experiment. The Data Similarity Coefficient column has the data coefficient values obtained from our simulation while the Data Similarity Value column displays a value depending on if the value of the coefficient met or failed to meet the threshold criteria (1,0) respectively. The Number of Sensitive Keywords column displays the number of sensitive keywords found while the Sensitive Keywords Value column displays a 1(one) if the value is greater than 0 or a 0(zero) if otherwise. The row headings are for the WEDS (e-mail suspects) in question. Notice that each WEDS traverses through the 10 selected Enron e-mails for comparison. The Expel WEDS column has the verdict (true/false) computed by comparing the Data Similarity Value column and the Sensitive Keywords Value column. For both columns, if each column has 1 as its value then the associated Expel WEDS returns true, else false. The final column indicates the true/false numerical value of the previous cell for computational purposes. The true/False value column indicates whether a WEDS was found guilty and hence subject to expulsion or not. Although a single instance of ‘true’ in the column is enough to declare the WEDS expulsion bound, our scheme displays all such results for the WEDS.
Using non-verbal social signals and degree centrality

<table>
<thead>
<tr>
<th>Simulation results tabulation</th>
</tr>
</thead>
<tbody>
<tr>
<td>Enron: Email</td>
</tr>
<tr>
<td>----------------</td>
</tr>
<tr>
<td>WEDS1 Enron1</td>
</tr>
<tr>
<td>WEDS1 Enron2</td>
</tr>
<tr>
<td>WEDS1 Enron3</td>
</tr>
<tr>
<td>WEDS1 Enron4</td>
</tr>
<tr>
<td>WEDS1 Enron5</td>
</tr>
<tr>
<td>WEDS1 Enron6</td>
</tr>
<tr>
<td>WEDS1 Enron7</td>
</tr>
<tr>
<td>WEDS1 Enron8</td>
</tr>
<tr>
<td>WEDS1 Enron9</td>
</tr>
<tr>
<td>WEDS1 Enron10</td>
</tr>
</tbody>
</table>

Total expulsion punishable offenses for WEDS1...
We carried out the simulation by using a single Dell Latitude E6410 equipped with Intel Core2Duo 6700@3.64GHZ, 4GB RAM, and 64-bit Windows7 operating system.

We used 10 Enron e-mails from the Enron e-mails dataset against our 10 self-created suspect e-mails, a total of 100 observations. We need to mention here that although we could have used large numbers of data both from the Enron dataset and our own this would only increase the execution time without any enhancement to the results of the simulation. The reason being that for the experimental purposes we created our model based on the pre-selected Enron data set, created the purported covert actors using parts of the same data, and used our own pre-selected sensitive keywords from the same data set. It therefore follows that on using the same data to assess the performance of the model then the procedure will always yield an overly optimistic result, 100% of the expected value. For real practical purposes this case would be more realistic since the model would be evaluated with data it has not been trained on or seen before. In this case then the amount of data would probably have an impact on the results.

6 Conclusions

We have developed and simulated a dynamic covert passive-actors detection scheme for a healthy networked community. By employing degree centrality the scheme categorises and sequesters all suspect actors. By use of non-verbal social signal deceit detection the scheme narrows down to determine potentially malicious covert actors in the networked communities. By carrying out a content similarity check coupled with a keyword search on the selected group, our scheme determines benign or malicious covert actors. Benign actors are left in a watch list while malicious actors are expelled from their networked communities.

The following were the observed limitations of our proposed scheme:

1 our scheme suffers the limitation that like all other schemes and protocols before it is a post-test scheme in that to discover a passive attacker, the attack must first happen at least once
2 there will be some cost incurred on some false-positives when a WEDS is cleared or declared benign following the WEDS investigation
3 due to privacy and other legal limitations on personal e-mails data our scheme was only evaluated with data it has been trained on or seen before therefore producing overly optimistic results of 100% performance.

For our future work we would like to extend this scheme to consider inter-nodal communication cost optimisation. We would also like to improve the scheme by enforcing a one strike rule where the scheme expels a WEDS as soon as it has been classified as a WED and not to continue with any further investigation on it on any other networked nodes.

Acknowledgements

This research was partially supported by NSF grants (No. 1247663, 1238767, 1137443, 1137516).
References


