A vulnerability evaluation framework for online social network profiles: axioms and propositions

Sophia Alim*, Daniel Neagu and Mick Ridley

AI Research Centre,
University of Bradford,
Bradford BD7 1DP, UK
E-mail: S.Alim@Student.Bradford.ac.uk
E-mail: D.Neagu@Bradford.ac.uk
E-mail: M.J.Ridley@Bradford.ac.uk
*Corresponding author

Abstract: Online social network (OSN) usage has led to personal details being presented on online profiles readily. This can cause profile owners to be vulnerable to social engineering attacks. Our approach to quantifying vulnerability consists of a model with three components: individual, relative and absolute vulnerabilities.

The individual vulnerability is calculated by allocating weights to profile attribute values disclosed which may contribute towards the personal vulnerability of the profile owner. The relative vulnerability is the collective vulnerability of the profiles' friends. The absolute vulnerability is the overall vulnerability for the profile which considers the individual and relative vulnerabilities.

This paper extends research done on axioms based on the vulnerability model, by stating propositions to explore the effects of different operators on the profiles relative and absolute vulnerabilities. The case studies show that our approach offers a formal background for estimating how attributes and operator changes influence the individual, relative and absolute vulnerability of OSN profiles.

Keywords: vulnerability; online social networks; OSN; axioms; propositions; information disclosure.

Reference to this paper should be made as follows: Alim, S., Neagu, D. and Ridley, M. (xxxx) 'A vulnerability evaluation framework for online social network profiles: axioms and propositions', *Int. J. Internet Technology and Secured Transactions*, Vol. X, No. Y, pp.000–000.

Biographical notes: Sophia Alim has a PhD from the University of Bradford in the area of online social networks and privacy. Her research focuses on calculating the vulnerability of online social network profiles in terms of information disclosing behaviour of a profile user and its friends. In 2006, she received her BSc (Hons) in Business Information Systems from the University of Salford, UK. In 2007, she received her MSc in Computing from the University of Bradford, UK. Her research interests include web accessibility and social networking.

Daniel Neagu is Professor of Computing at the University of Bradford. His research interests include knowledge discovery, information retrieval, data mining applications in multidisciplinary projects (with a focus in online social networks, healthcare and web profiling) by fusion of human experts knowledge and AI tools. He has published over 100 peer-reviewed papers and is the

principal investigator in projects funded by EU FP7, EPSRC, BBSRC and UK industry. He is IEEE CS, ACM and BCS member and Higher Education Academy Fellow.

Mick Ridley is the Head of the Department of Computing at the University of Bradford. He is the co-author of two books on databases. His research interests include databases, web database systems, bibliographic databases, XML and databases. In bibliographic applications, this included BOPAC and he was a member of the Joint Steering Committee for Revision of Anglo-American Cataloguing Rules: Format Variation Working Group.

1 Introduction

With the rise in usage and popularity of online social networks (OSNs) come the issue of privacy and the consequences of having your privacy breached via uncontrolled online spreading of your personal details.

Facebook, currently the most popular OSN, which has over 750 million active users according to Facebook (2011), has quite an interesting and controversial history when it comes to privacy.

For example, a significant event, which highlighted privacy breaches for Facebook users in November 2007, was the use of Beacon (the advertising system that monitored Facebook users): when users went for online shopping, Facebook shared the data of what they bought with the users' friends and other businesses (BBC News, 2007). In July 2010, the security consultant Rob Bowes highlighted publically available profiles on Facebook by extracting personal details from them and publishing the data online. The extraction occurred from 100 million profiles that were open and publically available. Bowe's motivation for this action was to highlight the privacy issues associated with Facebook (Emery, 2010).

In 2011, two doctoral students at Indiana University had discovered a security vulnerability in Facebook, which allowed malicious website to access the real name and also they could post bogus messages on their wall (Indiana University, 2010). These events have highlighted the issue of privacy in OSNs and how personal details can get without control into the wrong hands. Displaying personal details can make users more vulnerable to social engineering attacks like identity theft and re-identification by linking (Sweeney, 1997). These would enable people to extract personal details from the OSN profiles and use external sources to find out more about that person's identity.

In this paper, we firstly introduce our proposed vulnerability model which aims to measure the vulnerability of an OSN profile based upon measuring vulnerability because of its own personal information disclosure and the vulnerability of OSN friends network. We define the friends network in an OSN as a directed multigraph $G = (V, E)$, where V represents a set of nodes and E is a set of edges which link the nodes together. Each node represents an OSN profile and each edge is the link between two profiles. The OSN graph will be used in the calculation of profile vulnerability. We aim to propose an original set of axioms to formulate a formal background which will facilitate progress into algorithmic approaches towards real-time vulnerability measurement. The motivation behind this research study is to build upon our previous work on the vulnerability model in Abdul-Rahman et al. (2010) and Alim et al. (2011) but to now introduce a formal approach to the model and explore how the model reacts to the dynamic nature of an OSN.

Similar work on the introduction of an axiomatic approach has been proposed by Calvo and Dercon (2005) for the definition of an axiomatic approach for the economic vulnerability of a person. In terms of vulnerability measurement, Gundecha et al. (2011) have carried out research into vulnerability in OSN profiles, but their work and measures focuses on the definition and concept of a vulnerable friend and an investigation into how much a user's security can improve by deleting the vulnerable friend from its network. They define a vulnerable friend as a friend in the user's network of friends that has a lack of privacy and security settings to protect the network of friends. The vulnerability measures that Gundecha et al. (2011) proposed deal more with how public a profile's personal details and friend links are to other users in the network.

In comparison, our work on a vulnerability measure concentrates on the disclosure of personal details by the profile and its friends that can contribute towards the likelihood of privacy and social engineering attacks, as well as the spreading of the personal details through the OSN network. Also OSN graphs are used in the vulnerability calculation so the number of friends and how well connected the friends are to each other are taken into account in measuring vulnerability.

There have been a variety of models produced in various research areas of OSNs ranging from models associated with explaining the small world theory by Kleinberg (2002), through to modelling the relationship strength between nodes in an OSN as illustrated by Xiang et al. (2010). With the move in OSN research towards the analysis of node interaction as shown by Yun et al. (2010), there needs to be more models based on the issue of privacy and the issue of the self disclosure of personal details. Previous work on privacy trust modelling by Dwyer et al. (2007), highlighted how online relationships can exist, where 'perceived trust and privacy safeguards are weak'. This alone highlights the issue of privacy and social engineering attacks occurring and this is where our model would be useful.

The structure of the paper is as follows: Section 2 contains related work associated with the various privacy attacks that can occur by displaying personal details on OSN profiles. Section 3 details the vulnerability model and the model validation. Section 4 presents our axioms on the vulnerability model. Section 5 extends the research done on axioms by introducing the propositions, which explore changes to the vulnerability model components and the effect of different operators. Section 6 details the experimental work that focuses on the application of some propositions to real life cases. Section 7 discuses the axioms and results regarding the experimental work associated with the propositions. Section 8 concludes the paper and suggests ideas for further research.

2 Privacy attacks associated with personal details disclosure

Personal details which are presented and available online can raise social engineering issues as they can be used for identity fraud (Narayanan and Shmatikov, 2009). Some personal details (e.g., first name and date of birth) are considered as 'personal identifiable information' which can be used to identify one's individual data (Krishnamurthy and Willis, 2009). Due to the increase in the number of web system for applications (online shopping and banking), keeping a person's identity values unavailable for general browsing has never been so important. Personal details can be commonly used by web systems to authenticate users. By an individual displaying their personal details publically on OSN profiles, he or she is making themselves vulnerable to privacy and social engineering attacks. There are a variety of privacy breaches that can occur with personal details as illustrated below:

- Phishing via the use of an OSN: personal details are extracted from the user via the use of deception: in 2006, a phishing attack targeted MySpace users, tricking them into submitting personal details to a web page that looked like MySpace; these personal details were then sent to the hacker (Kirk, 2006).
- Online identity loss: hackers can make use personal details publically available on OSN profiles and other public resources in order to gain access to online systems, e.g., e-mail and banking systems. In 2008, David Kernell used Sarah Palin's postcode, date of birth and other details publicly available to reset the password of her Yahoo e-mail account. Mrs. Palin's personal details were found on Wikipedia (BBC News, 2010).
- Re-identification by linking: anonymous data can be used in conjunction with data from external sources in order to derive the identity of an individual. Hospital discharge records which included patient details were used to identify humans by linking their common demographic attributes to a database of public voter details from Massachusetts, USA. This can happen with OSN data due to attributes like date or birth, gender and location (Sweeney, 1997).

There are other privacy risks that can occur especially if the OSN user is careless about keeping some attributes (e.g., e-mail address) private. These risks are illustrated by Balduzzi et al. (2010) and Jagatic et al. (2007). If the e-mail addresses of users were made public, then spammers could crawl the OSNs and collect e-mail addresses and who they belong to from user profiles. Then this information could be used to construct phishing e-mails or targeted spam by using personal details which could include real names and names of friends. This act is known as social phishing.

By allowing users (including outside users) to search through OSNs for profiles by name or e-mail address, this can encourage spammers. Spammers can query the OSN to validate if the e-mails collected through the crawl of an OSN network, belong to real profile owners. The range of privacy risks illustrated previously have showed how there can be many dangers when presenting your personal details online. This fact motivates our work to define a vulnerability model for OSN users.

3 The vulnerability model

The vulnerability model proposed in Abdul-Rahman et al. (2010) was developed from the need to address privacy in OSNs by exploring the interaction and propagation effects of the users' friends. This will be achieved by applying the vulnerability model to OSNs. This will form the basis to validate the relationship between vulnerable nodes and the amount of personal details that can spread through the network.

3.1 The concept of vulnerability

In different fields, the term vulnerability can imply different concepts. For example, computer networks attack vulnerability presented by Holme et al. (2002) is linked to the reduction of network performance, due to the loss of network nodes and connections. This highlights the similar use of OSN graph structure in vulnerability definitions, but provides limited information about the node contents. This is further justified with some other concepts for vulnerability based on graph theory concepts which include cut-point and vulnerable bridges which is illustrated by Hannerman and Riddle (2005), and the use of outer nodes which are only connected to the main node.

Another example which is utilised in the vulnerability model is the clustering coefficient. The clustering coefficient describes how well connected the neighbours are to each other. The higher the clustering coefficient the more connected the neighbours are. In terms of privacy, a higher clustering coefficient would make a node vulnerable because of the good flow of information between the neighbours and this could imply its further spread throughout the OSN if the node's neighbours display their profiles so publically or at least to friends' friends.

A directed multigraph was used to model the OSN for the aim of calculating the vulnerability of a user profile, because the direction of its relationships would help to analyse the flow of personal details disclosure. Also the multigraph aspect would allow a more detailed analysis of the strength of relationship between two nodes based on OSN interaction.

The vulnerability definition in our opinion should contain information about both, the structure around the node and the node contents (Abdul-Rahman et al., 2010). Vulnerability in an OSN graph is associated with the probability that either the node's owner or one of the neighbours of the node will disclose information about the node. Our initial definition for a vulnerable node in an OSN is:

Definition: a *vulnerable node* is a node that contains attributes and neighbourhood features that breach privacy and provide grounds for a social engineering attack and the opportunity for the attribute values to spread through the network. For such a node, a highly connected neighbourhood in which the neighbours display the attributes readily may increase the risk of vulnerability, as detailed below.

Each node represents an OSN profile which consists of personal details (attributes), a list of friends of the user and interaction elements (e.g., media and the comments wall where the user and their friends can exchange comments with each other). The friends of the profile owner form the node's neighbourhood which can be analysed using an OSN graph. If you have a public profile, which discloses much information about yourself, and highly connected friends, who also have publicly accessible profiles and display many personal details, then your personal details may have a higher probability of spreading by your neighbours. This will increase chances of being vulnerable. The spreading of personal details will also depend on the interaction between a node and its neighbours. Interaction can include writing profile comments or tagging photos.

Also vulnerability is about the loss of control of personal details. The more public you make yourself then the less likely you are to have total control of your personal details. This theory is illustrated by research done by Luo and Lee (2009) into information aggregation attacks and threats to private information. One of the threats is 'Out of context information disclosure' where an OSN user assumes that the OSN is trusted so gives displays their profile information including personal details readily and publically. A very public profile which can be accessed via web searches allows personal details to be gained by social engineering attackers, sexual predators, hackers, etc. The aim is to make your profile and personal details less accessible to unknown users.

3.2 Model components

Our vulnerability model (Abdul-Rahman et al., 2010) consists of three components: the individual vulnerability (which focuses on the vulnerability of a single profile); the relative vulnerability (which explores the collective vulnerability of the neighbours who are friends of the profile owner) and the absolute vulnerability (that takes both the individual and relative vulnerability values into consideration).

3.2.1 Individual vulnerability

The individual vulnerability (V_I) is the result of self disclosure of personal details. It is calculated based on examining each profile for the presence of profile and neighbourhood-based attributes that contribute towards personal information disclosure. Some profile attributes included: *full name*, *gender*, *age*, *profile photo*, *location* and *zodiac*. Some of those attributes were selected, as previous research, Krishnamurthy and Wills (2009), and McCallister et al. (2010) highlighted their significance in breaching privacy and leading to social engineering attacks. From the OSN graph, the *number of friends* and the clustering coefficient which is calculated using equation (1) are both turned into neighbourhood-based attributes. The *clustering coefficient* by Watts and Strogatz (1998) characterises how well connected the neighbours of the node (profile) are in the OSN graph. It is a value between [0, 1]. The higher the *clustering coefficient* then the more connected the neighbours are:

$$C_i = \frac{E_i}{N_i\left(N_i - 1\right)} \tag{1}$$

where E_i is the set of edges between the neighbours of node i and N_i is the number of neighbours of node i.

Further work can be done into how different sets of attributes can make different types of users vulnerable. For a given network, the individual vulnerability of each node $V_I \in \{V_{I1}, V_{I2}, ..., V_{In} \mid V_{I_i} \in [0, 1], i = 1, ..., n\}$ where n is the number of nodes in the network. The V_I value is based upon allocated weights to the attributes mentioned previously. Currently, we assume that the attribute weights are based on the relative frequency of the attributes in a dataset at present but in the future we would also explore the presumption that the weights will be estimated on statistical analysis. If the contents of the node have any of these attributes made available then an attribute weight is allocated to the node. The sum of the weights for the node is the V_I value. The calculation for the V_I value is illustrated using Lam et al. (2008) metric:

$$V_{I_i} = \sum_{j=1}^{m} F_j * W_j \tag{2}$$

For simplicity, V_{I_i} denotes the individual vulnerability of node i where $i = 1, \ldots, n$, n is the number of nodes in the network, m is the number of attributes, F_j is a binary value to show whether an attribute j has been displayed in the profile and W_j is the weight that has been allocated to the attribute if it is vulnerable. The weight W_j can be represented, for example, by the probability that disclosure of attribute j contributes towards vulnerability of the i^{th} profile.

In terms of the attributes, number of friends and clustering coefficient, a weight was allocated if the node had 150 or less friends or had a clustering coefficient greater than 0.5. Having 150 or less friends increases the chances of the personal details of the node spreading across the network though interaction between a node and its neighbours. This value originates from Dunbar's theory presented in Dunbar (1992): 150 is the maximum number of humans a person can have a stable and interactive relationship with.

A node which has a clustering coefficient greater than 0.5 is allocated a weight because more of the neighbours are connected to each other and this increases the spread of personal details. The weights for both the number of friends and the clustering coefficient were based on the relative frequency of the number of nodes in the dataset that had a clustering coefficient greater than 0.5, or 150 or less friends. Overall, the higher the V_I value for a node, the increased chance that the node will become vulnerable to social engineering attacks due to its self disclosure of personal details and displaying features which can help to spread the personal details through the network.

3.2.2 Relative vulnerability

The relative vulnerability (V_{R_i}) of a node i takes into account the individual vulnerabilities of its neighbours (e.g., by using their arithmetical mean as the operator):

$$V_{R_i} = \frac{1}{n} \sum_{j \neq i} V_{I_j} \tag{3}$$

where n is the number of its neighbours and V_{I_i} is the individual vulnerability of the neighbour j. The reason that j is not equal to i is because a node cannot be neighbours with itself but it can be neighbours of other nodes in the OSN. Any $V_R \in \{V_{R1}, V_{R2}, \ldots, V_{Rn} \mid V_R \in [0, 1]\}$ where n is the number of nodes in the network. The higher the V_R value, the more vulnerable the neighbours of a node are collectively because the neighbours readily show the attributes that contribute towards vulnerability. This impacts on the vulnerability of the node because the neighbours' behaviour indicates a lack of concern for privacy.

3.2.3 Absolute vulnerability

The absolute vulnerability (V_{A_i}) for node i is a measure of an individual's information disclosure through the network, based on the individual vulnerability (V_I) and the relative vulnerability (V_R):

$$V_{A_i} = V_{I_i} \bullet V_{R_i} \tag{4}$$

where $i = 1, \ldots, n$, and n is the number of nodes. The operator can be represented by a variety of functions, e.g., product or MAX (the maximum value between V_I and V_R). Each

$V_A \in \{V_{A1}, V_{A2}, \ldots, V_{An} \mid V_A \in [0, 1]\}$ where n is the number of nodes in the network. All vulnerability components have the domain [0, 1].

3.3 Model validation

The aim of the model validation section is to investigate whether a node with a high relative vulnerability will lead to an increased spreading of information. This is because personal details about a node being present on the neighbour's wall can be seen by other users (neighbour's friends and external users if the neighbour's profile is very public in terms of privacy). This facilitates the spreading of the personal details. For an initial investigation, the individual, relative and absolute vulnerability were calculated for 100 random MySpace profiles from Caverlee and Webb (2008) dataset. Each profile represents a MySpace node and the profile's friends represent its neighbours. The dataset represents a top friends' network, so each neighbour represents a profiles' top friend.

The profiles were analysed for the presence of personal information: *name*, *gender*, *profile picture*, *age*, *current location* and *zodiac* as well as the neighbourhood features. The weights in the individual vulnerability, see equation (2), were calculated based on the relative frequency of each attribute presence in the dataset. Also for the 100 MySpace profiles, the extracted profile comments of the profiles' friends were examined, to see if any of the comments leaked information about the profile itself.

The results for the initial investigation indicated that some friends of profiles analysed, having high relative vulnerabilities, had the profile's personal details disclosed in their profile comments. Out of the 100 profiles, 47% of them had a high relative vulnerability (0.9–1.0) and there was interaction between the profile friends and the profile considered. In terms of attribute disclosure:

- 14.8% of the MySpace profiles with high relative vulnerabilities contained profile friends that had the *birthday* of the profile owner displayed in their profile comments. Out of these profiles, the average number of comments disclosing the *birthday* was 1.14.
- 68% of the MySpace profiles with high relative vulnerabilities contained profile friends that that had the *name* of the profile owner displayed in their profile comments. Out of these profiles, the average number of comments counted as above, disclosing the *name*, was 1.53.
- 4.25% of the MySpace profiles with high relative vulnerabilities contained profile friends that that had the *age* of the profile owner displayed in their profile comments. Out of these profiles, the average number of comments disclosing the *age* was 1.
- 25.3% of the MySpace profiles with high relative vulnerabilities contained profile friends that had the *current location* of the profile owner in their profile comments. Out of these profiles, the average number of comments disclosing the *current location* was 1.08.
- 17.0% of the MySpace profiles with high relative vulnerabilities contained profile friends that that had the *education* of the profile owner displayed in their profile comments. Out of these profiles, the average number of comments disclosing the *education* was 1.25.

- 8.51% of the MySpace profiles with high relative vulnerabilities contained profile friends that had the *hometown* of the profile owner displayed in their profile comments. Out of these profiles, the average number of comments disclosing the *hometown* was 1.

To investigate if there is an increased information disclosure as the relative vulnerability of profiles increases, we used Spearman rank as used by Ye and Wu (2010) to correlate profiles that have a relative vulnerability greater than 0.8 against the amount of information disclosure for certain attributes displayed in the profiles' friends comments.

$$r = 1 - \frac{6\sum_{i=1}^{n} d_i^2}{n^3 - n} \tag{5}$$

where r is the Spearman rank coefficient, n is the set of observations for variables x and y, and d_i is the difference between the i^{th} rank of x and the i^{th} rank of y. The findings highlighted that the attributes *name*, *age*, *education* and *hometown* had a weak positive relationship with the relative vulnerability where as *location* had a significant positive relationship with relative vulnerability. A positive relationship occurs if as the relative vulnerability increases, the amount of information disclosure displayed in the profiles' friends comments increases whereas a negative relationship signals that as the relative vulnerability increases the amount of information disclosure in the profile's friends comments decreases. The presence of a weak positive relationship for some of the attributes is a good outcome from our initial experiment but currently work is in progress to validate a larger sample size.

4 Axioms for the vulnerability model

We propose in this paper a set of definitions and axioms to address the need of a formal background in identification and processing of the vulnerability concept in OSNs.

Definition 1: Let the individual vulnerability for an OSN profile be defined by the tuple $V = (z, A, P)$, where z can be used to illustrate a vulnerability threshold that indicates the total amount of self disclosure attributes needed for a profile to be labelled as highly vulnerable. The set of attribute values for the i^{th} OSN profile is denoted by a_i. The probabilities set $P_i = (p_{i1}, p_{i2}, \ldots, p_{ij}, \ldots, p_{im})$ where m represents the number of attributes, contains the likelihoods p_{ij}, which measure if the presence of the j^{th} attribute will cause the i^{th} profile to be vulnerable to social engineering or privacy attacks. Consequently, p_{ij} delivers the value for W_j in equation (2). The weighted total of the probabilities p_{ij} is the individual vulnerability of the i^{th} profile: according to equation (2), V_{I_i} is a positive value less or equal to 1, and this meets the closure property of the vulnerability model since our model components (V_I, V_R, V_A) have to be in the domain [0,1].

In this particular research study, we are not concerned with the value of z (vulnerability threshold) in the tuple V or its use. However, this vulnerability threshold will be used further for highlighting the degree of vulnerability a node may have. Definition 1 illustrates that our individual vulnerability measure takes initially into consideration all existing attributes. All profiles have some element of vulnerability but there is an issue about whether a profile which does not disclose any personal details or has any friends, is classed as having any sort of vulnerability.

If a user would like to apply for an OSN profile he or she is encouraged to record personal details in the registration process. The only way to not display the personal details publically is after getting the profile, change the privacy setting to hide the details or give false personal details in the registration process. The latter reason seems the more realistic scenario of the two because currently, for example, for Facebook users, attributes like name and profile picture cannot be removed from an OSN profile; they can only be changed.

We propose three axioms for the probability dependent effect of OSN profile attributes, for the probability change and for the addition of attributes onto profile definition.

4.1 Axiom 1 (probability dependent effect of attributes)

Given two OSN profiles characterised by the vulnerabilities $V = (z, A, P)$ and $V' = (z, A', P')$ respectively, for any change $d > 0$ in one attribute value:

$$\left\{\begin{array}{l} V\left(z,\left(a_1, a_2, ..., a_m\right),\left(p_1, p_2, ..., p_m\right)\right) \\ -V\left(z,\left(a_1+d, a_2, ..., a_m\right),\left(p_1, p_2, ..., p_m\right)\right) \\ =V\left(z,\left(a_1, a_2', ..., a_m'\right),\left(p_1, p_2', ..., p_m'\right)\right) \\ -V\left(z,\left(a_1+d, a_2', ..., a_m'\right),\left(p_1, p_2', ..., p_m'\right)\right) \end{array}\right\} \tag{6}$$

$$\left\{\begin{array}{l} V\left(z,\left(a_1, a_2, ..., a_m\right),\left(p_1, p_2, ..., p_m\right)\right) \\ -V\left(z,\left(a_1+d, a_2, ..., a_m\right),\left(p_1, p_2, ..., p_m\right)\right) \\ \neq V\left(z,\left(a_1, a_2, ..., a_m\right),\left(p_1'', p_2'', ..., p_m''\right)\right) \\ -V\left(z,\left(a_1+d, a_2, ..., a_m\right),\left(p_1'', p_2'', ..., p_m''\right)\right) \end{array}\right\} \tag{7}$$

where V is the profile's individual vulnerability, A is its set of attributes and P is its set of probabilities. The change in attribute value is denoted by d.

A constraint of Axiom 1 is that the attributes have to be independent of one another. This is so the probability of each attribute contributing to vulnerability does not depend on the presence of another attribute. An example of an independent relationship between attributes is *age* and *zodiac* because people of different ages can have different *zodiac* signs. In comparison, the relationship between the *date of birth* and *age* is dependent because the *date of birth* is used to calculate the *age*.

Equation (6) describes the case when two OSN profiles have the same profile attributes and these attributes have the same probability values. This consequently gives the profiles the same individual vulnerability value. For example, both profiles have the *number of friends* as an attribute a1 and the probability that this leads to the profile being vulnerable is p_1.

Over time (e.g., a few days later), the *number of friends* has increased from a_1 to $a_1 + d$ for both profiles. In equation (6), this change in the *number of friends* is represented by d. The consequent effect on individual vulnerability is that the change with the same increment in the same attribute values for both profiles is reflected similarly in the vulnerability of both profiles under the circumstances that both profiles have the same probability $p_1 = p_1'$ associated with the information disclosure because of the attribute a_1.

On the other side, in equation (7) if the two OSN profiles have the same attributes but the probability of the attributes contributing towards the individual vulnerability of each profile differs, as illustrated by the probability notations p' and p'', the changes in the individual vulnerability values for both profiles are not the same. This situation may arise when, for example, one profile belongs to an adolescent and the other one to an adult. The probability of, let us say, the attribute *education information* causing vulnerability is higher for a child or adolescent than an adult. This is backed by Patchin and Hinduja (2010) who claims that having several details of the adolescent, e.g., *name*, *current city*, *profile picture* and *school* is all that is needed to locate the individual and trace their identity.

4.2 Axiom 2 (probability change)

For every $V = (z, A, P)$, the probability change $e \in [0, 1]$ and $p_i + e <= 1$

$$V\left(z,\left(a_1, a_2, ..., a_m\right),\left(p_1,\ p_2, ...,\ p_m\right)\right) \leq V\left(z,\left(a_1, a_2, ..., a_m\right),\left(p_1 + e,\ p_2, ...,\ p_m\right)\right) \quad (8)$$

$$V\left(z,\left(a_1, a_2, ..., a_m\right),\left(p_1,\ p_2, ...,\ p_m\right)\right) \geq V\left(z,\left(a_1, a_2, ..., a_m\right),\left(p_1 - e,\ p_2, ...,\ p_m\right)\right) \quad (9)$$

Equation (8) presents a scenario where there are two users with OSN profiles. Unlike the first profile, the first attribute a_1 in the second profile has a higher probability value of making the profile vulnerable. This means that the individual vulnerability of profile one is the same or smaller than the individual vulnerability value of the second profile. In comparison, equation (9) represents the same scenario with two OSN profiles but this time attribute a_1 in the second profile has a lower probability value. Consequently, the first profile will have the same or a higher individual vulnerability value.

4.3 Axiom 3 (addition of attributes)

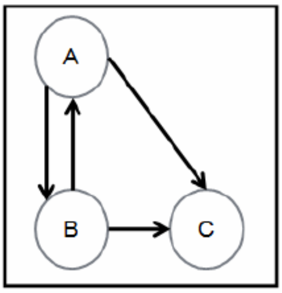
For every $V = (z, A, P)$, $V' = (z', A', P')$, $A' = (a_1, a_2, \ldots, a_m, a_{m+1}, a_{m+2}, \ldots)$ where $A = (a_1, a_2, \ldots, a_m)$, z' and P' are the vulnerability threshold and the probability values of the OSN profile with the addition of new attributes. These changes the individual vulnerability value denoted as V'. The new set including additional independent attributes is denoted as A'. However, if additional independent attributes are added to an OSN profile, then they may or may not contribute towards an increase in the individual vulnerability of a profile. The attributes' effect on the individual vulnerability will depend on the attributes already present in the profile and their probability values. Also the importance of the attribute in social engineering attacks will be a significant factor but more work is required in this area.

4.4 A sample of axioms application

The directed multigraph to model the OSN in Figure 1 can be used alongside the attributes table and neighbourhood features present in the profiles and their respective weights, to demonstrate the vulnerability calculation by applying the axiom notation used above. In Figure 1, there are three profiles represented by nodes *A*, *B* and *C*. The solid lines represent a top friend link between two profiles, e.g., profile *C* is a top friend of profile *A* but profile *A* is not a top friend of profile *C*.

Figure 1 OSN graph and table of weights (see online version for colours)

Attributes	A values	B values	C values
Full name	0.2	0.2	0.2
Gender	0.1	0.1	0.0
Age	0.0	0.1	0.0
Profile Photo	0.1	0.1	0.0
Location	0.0	0.1	0.0
Zodiac	0.0	0.1	0.0
Clustering Coefficient	0.1	0.1	0.1
Number of Friends	0.0	0.0	0.2



To calculate the absolute vulnerability of node *A*, the first step is to calculate the individual vulnerability V_I of node *A*. Node *A*'s profile is represented by the tuple *V* (*z*, (*fullname*, *gender*, *age*, *profilephoto*, *location*, *zodiac*, *clustering_coefficent*, *number_of_friends*), (0.2, 0.1, 0.0, 0.1, 0.0, 0.0, 0.1, 0.0). According to equation (2) V_{IA} = 0.5. The attribute/neighbourhood feature weights represent the probability values which state the likelihood that the presence of the attribute or neighbourhood feature will contribute towards the vulnerability of the profile. In this case, the probability values are chosen values between [0, 1] which add up to 1. The attribute *fullname* and neighbourhood feature *number_of_friends* are given higher weights because *fullname* is a common attribute used in identity theft. Having lower than 150 friends may cause information to spread quickly across a network due to the increased chances of having a highly interactive relationship with some of the friends.

The relative vulnerability V_R of node *A* takes the individual vulnerability of the neighbours B and C into consideration. Profile *B* is represented by the tuple *V* (*z*, (*fullname*, *gender*, *age*, *profilephoto*, *location*, *zodiac*, *clustering_coefficent*, *number_of_friends*), (0.2, 0.1, 0.1, 0.1, 0.1, 0.1, 0.1, 0.0) has a V_I value of 0.8.

Node *C*'s profile represented by the tuple *V*(*z*, (*fullname*, *gender*, *age*, *profilephoto*, *location*, *zodiac*, *clustering_coefficent*, *number of friends*), (0.2, 0.0, 0.0, 0.0, 0.0, 0.0, 0.1, 0.2) has a V_I value of 0.5. According to equation (3), V_{RA} value is 0.65. The V_{RA} value is high because of node *B* readily presenting its personal details as well as having a high clustering coefficient and node *C* only presenting one personal detail but having less than 150 friends and so there is an increased chance of having a interaction between friends.

Consequently according to equation (4) the V_{AA} is 0.32 with the product operator. The product operator balances the vulnerability of the node and the collective vulnerability of the neighbours. The choice of operator used in equation (4) can influence the absolute vulnerability value.

5 Propositions for the vulnerability model

The aim of this section is to investigate the effect of different operators on the vulnerability model via new consistent findings which form the propositions below. They provide an insight into how changes to the vulnerability model components can impact on the overall vulnerability of a profile. The propositions depend on the operators used in equations (3) and (4).

Propositions 1 and 2 are based on the axioms which are stated in Section 4, whilst Propositions 3 to 6 aim to explore in more detail other areas of the vulnerability model (e.g., relative and absolute vulnerability).

The list of propositions that we have proposed include: Proposition 1 states that in the context of using the product operator in equation (4), the absolute vulnerability value increases when there is an increase in the individual vulnerability value and the absolute vulnerability value decreases when there is a decrease in the individual vulnerability value.

In the context of using the MAX operator in equation (4), Proposition 2 emphasises that the absolute vulnerability value increases when the individual vulnerability value increases subject to the value being higher than the relative vulnerability value. On the other hand, the absolute vulnerability value decreases when the individual vulnerability value decreases subject to the individual vulnerability value being higher than the relative vulnerability value.

Proposition 3 explores the calculation of the relative vulnerability value by substituting the geometric mean operator into equation (3). The relative vulnerability value increases when the individual vulnerability value of the new neighbour (added to the existing neighbourhood) is greater or equal to the maximum individual vulnerability value of the existing neighbours. Whereas the relative vulnerability value calculated using the geometric mean operator can decrease when the individual vulnerability value of the new neighbour added to the existing neighbourhood, is lower than the minimum individual vulnerability value of the existing neighbours.

In Proposition 4, the arithmetical mean operator is used to calculate using equation (3) the relative vulnerability value. There is an increase in the relative vulnerability value when the individual vulnerability value of the new neighbour added to the existing neighbourhood, is higher than the maximum individual vulnerability value of the existing neighbours. The relative vulnerability value calculated using the arithmetical mean operator can decrease when the individual vulnerability value of the new neighbour added to the existing neighbourhood, is lower than or equal to the minimum individual vulnerability value of the existing neighbours.

Proposition 5 focuses on the change to the absolute vulnerability value. The absolute vulnerability value which is calculated using equation (4) and the product operator, increases when the relative vulnerability which is calculated using geometric mean increases due to the addition of a neighbour (with a higher individual vulnerability value than the maximum value in the existing neighbourhood). The absolute vulnerability value decreases when the relative vulnerability decreases due to the deletion of a neighbour (with a higher individual vulnerability value than the maximum value in the existing neighbourhood).

In Proposition 6 which centres on the absolute vulnerability value, the absolute vulnerability which is calculated using equation (4) and the product operator, increases when the relative vulnerability that is calculated using arithmetical mean, increases due to the addition of a neighbour (with a higher individual vulnerability value than the maximum value in the existing neighbourhood). The absolute vulnerability value decreases when the relative vulnerability decreases due to the deletion of a neighbour (with a higher individual vulnerability value than the maximum value in the existing neighbourhood).

The proposed propositions are presented in more detail below.

Proposition 1: An increase or decrease in the individual vulnerability value determines an increase or decrease in the absolute vulnerability value, absolute vulnerability is calculated using the product operator, •.

If $V_{A_i} = V_{I_i} \bullet V_{R_i}$ for profile i and • is product, then a change in the V_I value will be reflected by a change in the V_A value. Proposition 1 is based on Axiom 2 regarding probability changes.

Proof: $V_{A_i} = V_{I_i} \bullet V_{R_i} \in [0, 1]$ where • indicates the product operator. According to Axiom 2:

$$\begin{aligned} V_{A_e} &= V\left(z, \left(a_1, a_2, ..., a_m\right), \left(p_1 + e, p_2, ..., p_m\right)\right) \bullet V_{R_i} \\ &\geq V\left(z, \left(a_1, a_2, ..., a_m\right), \left(p_1, p_2, ..., p_m\right)\right) \bullet V_{R_i} = V_{A_i} \end{aligned} \tag{10}$$

where the probability change $e \in [0, 1]$, $p_i + e \leq 1$, V_{A_e} is the absolute vulnerability value of profile i as a result of the increase in probability value of the attributes contributing towards vulnerability. V_{R_i} is the relative vulnerability of profile i and V_{A_i} is the absolute vulnerability of profile i without the probability change taken into account. In equation (10), V_{R_i} is a constant calculated using the arithmetical mean. This proposition fits in with the concept that the individual vulnerability value of a profile monotonically increases with the absolute vulnerability value of a profile.

We can identify two cases regarding the change in the V_I value:

a Increase in individual vulnerability value:

Equation (11) highlights how the absolute vulnerability value increases when the individual vulnerability value of a profile increases due to a change in the probability of an attribute or what attributes the profile chooses to present.

$$V'_{I_i} < V_{I_i} \Rightarrow V'_{A_i} = V'_{I_i} \bullet V_{R_i} > V_{I_i} \bullet V_{R_i} = V_{A_i} \tag{11}$$

where V'_{I_i} represents the increased individual vulnerability value and the V'_{A_i} represents the increased absolute vulnerability value.

An example of this case is that profile *A* in Figure 1 decides to present the location attribute on their profile which has an attribute weight of 0.1. This increases the V_I value of profile *A* from 0.5 to 0.6. The change in V_I value (ΔV_I) is an increase of 0.1 due to the addition of the attribute weight for location. This consequently increases the V_A value for profile *A* from 0.325 to 0.390 with a difference of 0.065. This case links in with Axiom 3 which focuses on the addition of attributes and equation (8) in Axiom 2.

b Decrease in individual vulnerability value:

Equation (12) highlights the decrease in absolute vulnerability value when the individual vulnerability value of profile *i* decreases, due to a change in the probability of an attribute or a deletion of a profile attribute that contributes towards vulnerability.

$$V'_{I_i} < V_{I_i} \Rightarrow V'_{A_i} = V'_{I_i} \bullet V_{R_i} < V_{I_i} \bullet V_{R_i} = V_{A_i} \tag{12}$$

where V'_{I_i} represents the decreased individual vulnerability value, the V'_{A_i} represents the decreased absolute vulnerability value and V_{R_i} is constant and calculated using the arithmetical mean.

An example of this case is that profile *B* in Figure 1 decides to delete the *full name* attribute on their profile. This decreases the V_I of profile *B* from 0.8 to 0.6 with a ΔV_I decrease of 0.2. Consequently, this decreases the V_A value for profile *B* from 0.400 to 0.300. This observation illustrates that with a ΔV_I decrease of 0.2 and a V_{R_i} value of 0.5, the V_A value decreases by 0.1. This case links in with equation (9) in Axiom 2 which focuses on probability change.

Proposition 2: An increase or decrease in the individual vulnerability value reflects an increase or decrease in the absolute vulnerability value, when absolute vulnerability is calculated using the MAX operator, • and the individual vulnerability value is higher than the relative vulnerability value.

For profile *i*, the operator • is MAX, then a change in the V_I value will be reflected by a change in the V_A value if and only if the $V_I \geq V_R$.

Proof:

$$\left.\begin{cases} V_A = MAX\left(V_i, V_R\right) \\ V_I \geq V_R \end{cases}\right\} V_A = V_I \tag{13}$$

The proof highlights that in order for the V_I change to lead to a V_A change, the $V_I \geq V_R$. Subject to $V_I \geq V_R$ we have identified two cases of change for individual vulnerability:

a Increase in individual vulnerability:

Equation (14) highlights the increase in absolute vulnerability value when the individual vulnerability value of profile *i* increases, due to a change in the probability of an attribute or what attributes the profile presents.

$$V'_{I_i} > V_{I_i} \Rightarrow V'_{A_i} = MAX\left(V'_{I_i}, V_{R_i}\right) > MAX\left(V_{I_i}, V_{R_i}\right) = V_{A_i} \mid V_{I_i} \geq V_{R_i} \tag{14}$$

where V'_{I_i} represents the increased individual vulnerability value and the V'_{A_i} represents the increased absolute vulnerability value.

An example of this case is that profile *A* in Figure 1 decides to present the age and location attributes on their profile. This increases the V_I of profile *A* from 0.5 to 0.7 with a ΔV_I of 0.2. Consequently, the V_A for profile *A* increases because the updated V_I of profile *A* is larger than the VR of profile *A* which is 0.65. The V_A value of profile *A* has increased from 0.65 to 0.70 and this shows that with the MAX operator, as long as the updated V_I value is lower than or equal to the V_R value, then the changes to profile *A* will not be reflected in the overall vulnerability. This case links in with equation (8) in Axiom 2 which centres on probability change.

b Decrease in individual vulnerability:

Equation (15) highlights the decrease in absolute vulnerability value when the individual vulnerability of profile *i* decreases, due to a change in the probability of an attribute or a deletion of an attribute that contributes towards the vulnerability of a profile.

$$V'_{I_i} < V_{I_i} \Rightarrow V'_{A_i} = MAX\left(V'_{I_i}, V_{R_i}\right) < MAX\left(V_{I_i}, V_{R_i}\right) = V_{A_i} \mid V_{I_i} \geq V_{R_i} \qquad (15)$$

An example of this case is that profile *B* in Figure 1 decides to delete the attributes *full name* on their profile. This decreases the individual vulnerability value of profile *B* from 0.8 to 0.6 with a ΔV_I of 0.2.

Consequently, the absolute vulnerability value for profile *B* decreases from 0.8 to 0.6. The change in absolute vulnerability (ΔV_A) = ΔV_I because the individual vulnerability before and after the attribute change for profile *B* was higher than the relative vulnerability value of profile *B* which is 0.5. However, the decrease depends on the comparison between $(V_{I_i} - V_{R_i})$ and $(V'_{I_i} - V_{I_i})$.

In regards to how the absolute vulnerability is affected by the changes to the relative vulnerability of a profile, we selected two operators which can help calculate the relative vulnerability of a profile. One of these operators is the geometric mean, which is illustrated in equation (16).

$$V_{R_i} = \sqrt[n]{\sum_{\substack{j=1 \\ j \neq 1}}^{n} V_{I_j}} \qquad (16)$$

where *n* is the number of neighbours (profile's friends) and V_{I_j} is the individual vulnerability of neighbour *j* where $\forall i - 1, \ldots, n$. Proposition 3 explores how the use of the geometric mean operator can impact on the relative vulnerability value.

Proposition 3: On the change in relative vulnerability value when relative vulnerability is calculated using the geometric mean operator.

Given that $V_{A_i} = V_{I_i} \bullet V_{R_i}$ and $V_{R_i} = \sqrt[n]{\sum_{\substack{j=1 \\ j \neq i}}^{n} V_{I_j}}$.

If $V_{I_{n+1}} > V_{I_i}, \forall i = 1, \ldots, n (=) V_{I_{n+1}} \geq \underset{i=1}{\overset{n}{MAX}}\, V_{I_i}$ where MAX an operator which selects out the maximum individual vulnerability (V_{I_n}) from the neighbours, then $V_{R_{n+1}} > V_{R_n}$, where $V_{R_{n+1}}$ is the relative vulnerability value of neighbourhood with addition of new neighbour and V_{R_n} is the relative vulnerability of the existing neighbourhood.

As long as the V_I of the new neighbour $(V_{I_{n+1}})$ is equal to or higher than the maximum individual vulnerability value from the existing neighbours, then the relative vulnerability value will increase with the addition of a new neighbour and this is shown in the proof below where V'_{R_n} represents the new relative vulnerability value of the profile with the addition of the new neighbour into the existing neighbourhood. The proof which is explained below is based on additional calculations which are presented in Appendix 1.

Proof:

$$V_{I_{n+1}} > V_{I_i}, \forall i = 1, ..., n$$

$$\frac{V_{I_{n+1}}}{V_{I_i}} > 1, \forall i = 1, ..., n$$

$$\text{then } \frac{V_{I_{n+1}}}{V_{I_i}} > 1, \forall i = 1, ..., n$$

$$\text{then } \prod_{i=1}^{n} \frac{V_{I_{n+1}}}{V_{I_i}} > \log 1 = 0$$

$$\log \prod_{i=1}^{n} \frac{V_{I_{n+1}}}{V_{I_i}} > 0 \Rightarrow V'_{R_n} > V_{R_n}$$

An example of this case involves the neighbourhood of profile *A* in Figure 1. The highest individual vulnerability value of the neighbours in *A*'s neighbourhood is 0.8. If a new neighbour with an individual vulnerability of 0.85 joins the neighbourhood, then the relative vulnerability of profile *A* when calculated using the geometric mean operator will increase from 0.632 to 0.697 with a change of 0.065. Additionally, if the new neighbour had an individual vulnerability of 0.8, then the relative vulnerability of profile *A* would still increase from 0.632 to 0.683 with a change of 0.051.

The relative vulnerability value of profile *A* can decrease when the individual vulnerability value of the new neighbour is equal to or lower than the minimum individual vulnerability value in the existing neighbourhood, which in this case is 0.5. A new neighbour with an individual vulnerability value of 0.4 would decrease the relative vulnerability value of profile *A* from 0.632 to 0.542 with a change of 0.090.

Another operator which can be used to calculate the relative vulnerability of a profile is the arithmetical mean, which is illustrated in equation (17).

$$V_{R_i} = \frac{1}{n} \sum_{1}^{n} V_{I_j} \tag{17}$$

where n is the number of neighbours and V_{I_j} is the individual vulnerability of profile j where $\forall i = 1, \ldots, n$.

Proposition 4: On the change in relative vulnerability value when relative vulnerability is calculated using the arithmetical mean operator.

Given that $V_{A_i} = V_{I_i} \bullet V_{R_i}$ and $V_{R_i} = \frac{1}{n} \sum_{1}^{n} V_{I_j}$.

If $V_{I_{n+1}} > V_{I_i}$, $\forall i = 1, ..., n(=) V_{I_{n+1}} \geq \overset{n}{\underset{i=1}{MAX}} V_{I_i}$ where MAX is the operator which selects out the maximum individual vulnerability (V_{I_n}) from the neighbours, then

$V_{R_{n+1}} > V_{R_n}$, where $V_{R_{n+1}}$ is the relative vulnerability value of neighbourhood with addition of new neighbour and V_{R_n} is the relative vulnerability of the existing neighbourhood.

As long as the individual vulnerability value of the new neighbour $(V_{I_{n+1}})$ is higher than or equal to the maximum individual vulnerability value from the existing neighbours, then the relative vulnerability value will increase with the addition of a new neighbour. This is shown in the proof below, where V'_{R_n} signifies the new relative vulnerability value of the profile with the addition of the new neighbour into the existing neighbourhood. The proof which is explained below is based on additional calculations which are presented in Appendix 2.

Proof:

$$V_{I_{n+1}} > V_{I_i}, \forall i = i, ..., n$$

$$\sum_1^n V_{I_{n+1}} > \sum_1^n V_{I_i} \left| \sum_1^n \left(V_{I_{n+1}} - V_{I_i}\right) > 0 \text{ and } \frac{\sum_1^n \left(V_{I_{n+1}} - V_{I_i}\right)}{n(n+1)} > 0 \text{ and } V_{R_{n+1}} - V_{R_n} > 0\right.$$

$$V_{R_{n+1}} > V_{R_n}$$

$$V'_{R_n} = V_{R_{n+1}} \Rightarrow V'_{R_n} > V_{R_n}$$

The proof illustrates that in order for $V_{R_{n+1}} > V_{R_n}$, the difference between $V_{I_{n+1}}$ and V_{I_i} has to be a positive number which is above 0.

An example of this case involves the neighbourhood of profile *A* in Figure 1. The highest individual vulnerability value of the neighbours in *A*'s neighbourhood is 0.8. If a new neighbour joins the neighbourhood with an individual vulnerability value of 0.85, then the relative vulnerability of profile *A* when calculated using the geometric mean operator will increase from 0.650 to 0.716. Even if the individual vulnerability value of the new neighbour was 0.8, then the relative vulnerability of profile *A* would increase from 0.650 to 0.700.

The relative vulnerability value of profile *A* can decrease if the individual vulnerability value of the new neighbour was equal to or lower than the minimum individual vulnerability value in the existing neighbourhood, which in this case is 0.5. A new neighbour with an individual vulnerability value of 0.4 would decrease the relative vulnerability value of profile *A* from 0.650 to 0.566.

Proposition 5: On the change in absolute vulnerability due to addition or deletion of a neighbour when relative vulnerability is calculated using geometric mean operator and absolute vulnerability is calculated using the product operator, •.

Given that $V_{A_i} = V_{I_i} \bullet V_{R_i}$ and $V_{R_i} = \sqrt[n]{\prod_{i=1}^{n} V_{I_j}}$ where *n* is the number of neighbours and V_{I_j} is the individual vulnerability of neighbour *j*. If • is product, then a change (increase or decrease) in the relative vulnerability value (V_{R_i}) will be reflected by a change (increase or decrease) in the absolute vulnerability (V_{A_i}) if and only if $V_{R_{n+1}} > V_R$.

$V_{R_{n+1}}$ is the relative vulnerability of profile *i* with the addition of a new neighbour and V_R is the relative vulnerability of profile *i* without the addition of a new neighbour.

Proof:

$V_{A_i} = V_{I_i} \bullet V_{R_i} \in [0,\ 1]$ for profile *i* where • indicates a product operator.

$$V_{I_{n+1}} > V_{I_i},\ \forall i = 1, ..., n$$

$$\frac{V_{I_{n+1}}}{V_{I_i}} > 1,\ \forall i = 1, ..., n$$

$$\text{then } \frac{V_{I_{n+1}}}{V_{I_i}} > 1,\ \forall i = 1, ..., n$$

$$\text{then } \prod_{i=1}^{n} \frac{V_{I_{n+1}}}{V_{I_i}} > \log 1 = 0$$

$$\frac{1}{n(n+1)} * \log \prod_{i=1}^{n} \frac{V_{I_{n+1}}}{V_{I_i}} > 0 \Rightarrow V_{R_{n+1}} > V_{R_n}$$

$$V_{R_{n+1}} > V_{R_n} \Rightarrow V_{R_{n-1}} < V_{R_n}$$

The proof shows that when the relative vulnerability is calculated using the geometric mean operator, the addition of a new neighbour can increase the profile's relative vulnerability value. We have identified two cases regarding the increase in relative vulnerability value due to addition and deletion of a neighbour in the profile's neighbourhood:

a Increase in relative vulnerability through addition of a neighbour

Equation (18) highlights an increase in the V_A when the V_R of profile *i* increases, due to the addition of a neighbour into the profile's neighbourhood.

$$V_{R_{n+1}} > V_{R_n} \Rightarrow V'_{A_i} = V_{I_i} \bullet V_{R_{n+1}} > V_{I_i} \bullet V_{R_n} = V_{A_i} \tag{18}$$

where $V_{R_{n+1}}$ is the relative vulnerability of profile *i* with the additional neighbour, V_{R_n} is the relative vulnerability of profile *i* without the neighbour, V'_{A_i} is the absolute vulnerability of profile *i* as a result of the addition of the neighbour and V_{A_i} is the absolute vulnerability of profile *i* without the addition of the neighbour.

An example of this case involves the neighbourhood of profile *A* in Figure 1. With the addition of a new neighbour with an individual vulnerability value of 0.9, the relative vulnerability value of profile *A* increases from 0.632 to 0.711 with an increase change of 0.079. Consequently, the absolute vulnerability value increases from 0.316 to 0.355 with an increase of 0.039. An important observation between these results, is that the difference between V'_{A_i} and V_{A_i} is half the difference between $V_{R_{n+1}}$ and V_{R_n}.

b Decrease in relative vulnerability through the deletion of a neighbour

Equation (19) states a decrease in the absolute vulnerability value of profile *i* when the relative vulnerability of profile *i* decreases, due to the deletion of a neighbour in the profile's neighbourhood.

$$V_{R_{n-1}} < V_{R_n} \Rightarrow V'_{A_i} = V_{I_i} \bullet V_{R_{n-1}} < V_{I_i} \bullet V_{R_n} = V_{A_i} \qquad (19)$$

where $V_{R_{n-1}}$ is the relative vulnerability of profile *i* with a neighbour deleted, V_{R_n} is the relative vulnerability of profile *i* before the neighbour was deleted, V'_{A_i} is the absolute vulnerability of profile *i* as a result of the deletion of the neighbour and V_{A_i} is the absolute vulnerability of profile *i* before the neighbour was deleted.

Equation (19) highlights that deleting a neighbour in the neighbourhood with an individual vulnerability value that is higher than the maximum individual vulnerability value of the neighbourhood, decreases the relative vulnerability value of the profile and this results in a decrease in the absolute vulnerability value of the profile.

An example of this case involves the neighbourhood of profile *A* in Figure 1. If profile *B* which has the highest individual vulnerability value is removed from profile *A*'s neighbourhood, then the relative vulnerability value of profile *A* decreases from 0.632 to 0.500 with a difference of 0.132. Consequently, the absolute vulnerability value decreases from 0.316 to 0.250, with a difference of 0.066. The difference between V'_{A_i} and V_{A_i} is half the difference between $V_{R_{n+1}}$ and V_{R_n}.

Proposition 6: On the change in absolute vulnerability due to addition or deletion of a neighbour when relative vulnerability is calculated using arithmetical mean operator and absolute vulnerability is calculated using the product operator, •.

Given that $V_{A_i} = V_{I_i} \bullet V_{R_i}$ and $V_{R_i} = \frac{1}{n}\sum_{1}^{n} V_{I_j}$ where *n* is the number of neighbours and V_{I_j} is the individual vulnerability of neighbour *j*. If • is product, then a change in the relative vulnerability (V_{R_i}) will be reflected by a change in the absolute vulnerability (V_{A_i}) if and only if the $V_{R_{n+1}} > V_R$ where $V_{R_{n+1}}$ is the relative vulnerability of profile *i* with the addition of a new neighbour and V*R* is the relative vulnerability of profile *i* without the addition of a new neighbour.

Proof:

$V_{A_i} = V_{I_i} \bullet V_{R_i} \in [0,\ 1]$ for profile *i* where • indicates a change.

$$V_{I_{n+1}} > V_{I_i},\ \forall i = i, ..., n$$

$$\sum_{1}^{n} V_{I_{n+1}} > \sum_{1}^{n} V_{I_i} \left| \sum_{1}^{n} \left(V_{I_{n+1}} - V_{I_i}\right) > 0 \text{ and } \frac{\sum_{1}^{n}\left(V_{I_{n+1}} - V_{I_i}\right)}{n(n+1)} > 0 \text{ and } V_{R_{n+1}} - V_{R_n} > 0\right.$$

$$V_{R_{n+1}} > V_{R_n} \Rightarrow V_{R_{n-1}} < V_{R_n}$$

The proof show that when the relative vulnerability is calculated using the arithmetical mean operator, the addition of a new neighbour in to the existing neighbourhood can increase the profile *i* relative vulnerability value if and only if the individual vulnerability value of the new neighbour is higher than the maximum individual vulnerability value of the existing neighbourhood.

This implies that the deletion of a neighbour will decrease the relative vulnerability value of the profile. In order for the relative vulnerability value to increase, the individual vulnerability value of the new neighbour has to be higher than the maximum individual vulnerability value of the existing neighbours in the neighbourhood.

There are two cases of change for relative vulnerability:

a Increase in relative vulnerability through addition of a neighbour

Equation (20) highlights the increase in absolute vulnerability when the relative vulnerability value of a profile increases due to the addition of a neighbour into the profile's neighbourhood.

$$V_{R_{n+1}} > V_{R_n} \Rightarrow V'_{A_i} = V_{I_i} \bullet V_{R_{n+1}} > V_{I_i} \bullet V_{R_n} = V_{A_i} \tag{20}$$

where $V_{R_{n+1}}$ is the relative vulnerability of the profile with the additional neighbour, V_{R_n} is the relative vulnerability of the profile without the neighbour, V'_{A_i} is the absolute vulnerability of the profile as a result of the addition of the neighbour and V_{A_i} is the absolute vulnerability of the profile without the addition of the neighbour. Equation (20), illustrates that adding a neighbour with a higher individual vulnerability value to a neighbourhood, increases the relative vulnerability value of the profile and this results in an increase in the absolute vulnerability value of the profile.

An example of this case involves the neighbourhood of profile *A* in Figure 1. With the addition of a new neighbour with an individual vulnerability value of 0.9, the relative vulnerability value increases from 0.650 to 0.733 which is a difference of 0.083. Consequently, the absolute vulnerability value of profile *A* increases from 0.320 to 0.366 with a difference of 0.046.

b Decrease in relative vulnerability through deletion of a neighbour

Equation (21) highlights the decrease in absolute vulnerability when the relative vulnerability value of a profile decreases due to the deletion of a neighbour with the highest individual vulnerability value, in a profile's neighbourhood.

$$V_{R_{n-1}} < V_{R_n} \Rightarrow V'_{A_i} = V_{I_i} \bullet V_{R_{n-1}} < V_{I_i} \bullet V_{R_n} = V_{A_i} \tag{21}$$

where $V_{R_{n-1}}$ is the relative vulnerability of the profile with a neighbour deleted, V_{R_n} is the relative vulnerability of the profile before the neighbour was deleted, V'_{A_i} is the absolute vulnerability of the profile as a result of the deletion of the neighbour and V_{A_i} is the absolute vulnerability of the profile before the neighbour was deleted. Equation (21) illustrates that deleting a neighbour with the highest individual vulnerability value in the neighbourhood, decreases the relative vulnerability value of the profile and this results in a decrease in the absolute vulnerability of the profile.

> An example of this case involves the neighbourhood of profile *A* in Figure 1. If profile *B* is removed from profile *A*’s neighbourhood, then the relative vulnerability value of profile *A*, decreases from 0.650 to 0.500 which is a difference of 0.150. Consequently, the absolute vulnerability value of profile *A* decreases from 0.325 to 0.250 which is a difference of 0.075.

Overall, this section has detailed several propositions, which focus on changes in the vulnerability model. What the propositions have demonstrated is that the changes in what a user present on their profile and the adding and deletion of friends into a neighbourhood, impacts on the overall profile vulnerability.

6 Experimental work and findings regarding application of propositions

In order to explore how the operators from the propositions affect the vulnerability of real life cases, different operators for the relative and absolute vulnerabilities calculations were used. These calculations were tested on 76,263 profiles from the Caverlee and Webb (2008) dataset. For the 76,263, the average age of the profile owners is 25.6. 50.1% of the profile owners are male, 48.6% of the profile owners are female and 1.3% profile owners do not state their gender. 16.3% of the profiles are private profiles. Private profiles may present basic information (e.g., name, gender, profile picture and age) on their profiles but not the list of friends or interactions.

The different operators were used in the following ways: the relative vulnerability value of the cases in the dataset was calculated using the geometric and arithmetical mean. These operators are used in Propositions 3 to 6. The absolute vulnerability value for the cases was calculated using the product and MAX operators used in Propositions 1, 2, 5 and 6. The operators were applied to 76,263 cases with varying individual vulnerability values.

In terms of the relative vulnerability calculation, unlike some operators (e.g., MAX and MIN) which select just the maximum or minimum individual vulnerability value of the neighbourhood, the geometric and arithmetical mean takes into account all the individual vulnerability values of the neighbours in the calculation. With the arithmetical mean operator, 78.0% of the profiles had a relative vulnerability of 0.9 or above and with the geometric mean, 75.9% of the profiles had a relative vulnerability of 0.9 and above. The results highlight that most neighbours in profiles’ neighbourhoods self disclose the attributes and display structural features, that contribute towards vulnerability readily.

The effect of the calculation of the relative vulnerability on the absolute vulnerability of profiles is covered in Propositions 5 and 6. Table 1 presents statistics for the whole dataset, regarding different combinations of operators for the relative and absolute vulnerability calculations. The MAX operator is mainly used in Proposition 2. In Table 1, the absolute vulnerability of profiles is denoted as V_A.

What Table 1 demonstrates, is the big difference between the effect of the product and MAX operator in regards to the absolute vulnerability of profiles. The product operator can act as a reducing effect. An example is that if a profile has friends which self disclose personal details readily (V_R value of 0.8), but the profile itself is very private (V_I value of 0.2), the overall vulnerability of the profile would be 0.16. This emphasises that there is a low likelihood of the profile’s personal details spreading through the OSN.

Table 1 Statistics regarding application of different operators

Absolute vulnerability operators	*Product*		*MAX*	
Relative vulnerability operators	*Geometric*	*Arithmetical*	*Geometric*	*Arithmetical*
Features corresponding to whole dataset				
% of profiles that have a V_A of 0.9 or above	47.4%	50.2%	99.1%	99.2%
Average V_A value	0.871	0.872	0.964	0.969
Standard distribution of V_A values	0.134	0.132	0.034	0.034
Skewness	–3.386	–3.454	–2.455	–2.105

On the other hand, the MAX operator selects the higher value between the V_I and V_R. The high percentage of profiles that have an absolute vulnerability of 0.9 or above highlights that there are cases in which the profile’s neighbours many not collectively disclose many attributes that contribute towards vulnerability but the profile itself will and this causes the overall vulnerability of the profile to increase.

The average absolute vulnerability values of the dataset for the product and MAX operators demonstrate that the MAX operator may not be an effective operator for the calculation of absolute vulnerability. This is because it does not take into effect the meaning of both the individual and relative vulnerability values together and it would be hard to select the profiles which are truly vulnerable.

With the product operator, 15.5% of the profiles in which the relative vulnerability is calculated using arithmetical mean have an absolute vulnerability value which is less or equal to 0.8. Therefore, the product operator is more realistic than the MAX operator in selecting vulnerable nodes. This is validated by the standard deviation for the V_A values calculated using the product operator. It illustrates that in this dataset there are a variety of cases with varying V_I and V_R values.

The negative skew for the product and MAX operators indicates that there is a long distribution tail to the left and consequently more profiles have a higher absolute vulnerability value. Overall, what the statistics have shown is that the choice of operator especially for the calculation of absolute vulnerability can influence the number of profiles which are classed as having a high overall vulnerability (V_A of 0.9 or above).

7 Discussion

Based on the vulnerability model, the axioms and propositions presented have helped to investigate the impact of attribute and probability value change on the individual vulnerability of a profile. Axiom 1 highlights the issue of attributes and the probability that the attribute change can contribute towards the profile’s vulnerability. In our vulnerability model, we make the assumption that an attribute’s contribution towards the vulnerability of a profile is independent of the type of user, as illustrated in equations (6) and (7). This assumption helps when modelling an OSN because different users have different attitudes towards privacy.

On the other hand, Axiom 2 illustrates the effect that a probability change can have on the individual vulnerability value of the profile. An increase in an attribute weight will increase the individual vulnerability of the profile. Since the weights of the attributes that contribute towards vulnerability have to add up to one, an increase in one attribute weight will lower the weights of the other attributes. The generation of the weights is an issue which will require further investigation. Even though the relative frequency approach is stated in Section 3.2.1, there are other approaches to generate weights. One approach is based on human perception and involves distributing a questionnaire which asks subjects about their willingness to share certain profile attributes. Another way is to use is an information theory-based approach to investigate how distinctive a profile attribute is.

The propositions focus on how changes in the individual and relative vulnerabilities of a profile, can impact on the overall (absolute) vulnerability of the profiles. The mathematical operators (geometric mean and arithmetical mean) used for the calculation of the relative vulnerability, highlighted the issue of a profile adding friends that have a high individual vulnerability. Propositions 3 and 4 show that if a new friend added to the neighbourhood has an individual vulnerability that is higher than or equal to the maximum individual vulnerability of the existing neighbourhood, then the relative vulnerability of the profile will increase. This may increases the likelihood of the personal details of the profile spreading through the OSN.

In order for OSN users to lower their relative vulnerability, the user has to unfriend the friends who self disclose readily and have a high individual vulnerability. This finding is validated by Gundecha et al. (2011) who carried out research into defining vulnerable friends and states that an "individual is vulnerable if any friends in the network of friends has insufficient security and privacy settings to protect the entire network of friends".

Propositions 5 and 6 demonstrated that with the operators presented in the proposition, if the addition of a new neighbour causes the relative vulnerability to rise, then this will lead to an increase in the absolute vulnerability of the profile. The experimental work regarding the propositions highlight how important it is to use the right operators for the calculation of absolute vulnerability, in order to model the vulnerability of profiles in a realistic way. The choice of using simple weighted average functions (geometric and arithmetical means) for calculating relative vulnerability was used by Gundecha et al. (2011) in their calculations which expanded their definition of vulnerable friend.

In terms of the MAX and product operators, even though product operator has a reduction effect, it is suitable in this case because the concept of vulnerability centres on the spreading of personal details of the profile through comments written by friends or friends of friends of the profile.

Concentrating on just an online relationship between a profile and its friends, if a profile is very private in terms of self disclosure and the friends are very public in terms of their disclosure, then the overall vulnerability will be quite low because there would not be many personal details to spread through the OSN network. The MAX operator would just focus on the actions of one or more users regardless of what the other users are doing.

8 Conclusions and future research

With the increase in the amount of personal details displayed on an OSN profile comes the privacy issue and the threat of social engineering attack which can make users vulnerable. Our proposed vulnerability model considers that the vulnerability of a profile can be quantified in such a way to also take into account how the behaviour of the profile's friends can impact on the vulnerability of a profile. The axioms presented have highlighted various areas that can be developed and implemented into the model in the future, to accurately reflect the vulnerability of a profile. One of the issues is incorporating the profile interaction into the model. Analysis of interactions has helped to validate that the personal details of a profile can be spread through the OSN network. This is shown by the profile's personal details being displayed in the neighbours' profile comments. Another open issue is the attribute weight and the reality of using OSNs.

The propositions which are based on the axioms, highlighted the changes in what a user present on their profile and the adding and deletion of friends into a neighbourhood which impacts on the overall profile vulnerability. The experimental work which involved investigating the effect of different operators on the overall vulnerability of a profile, demonstrated that the product operator in comparison to the MAX operator was the most effective in the absolute vulnerability calculation. This was because the product operator realistically modelled the concept of vulnerability.

The work presented in this paper has provided grounds for future research. The research includes developing more axioms associated with the vulnerability model in order to reflect the changes in the relative vulnerability and the operators used to calculate the relative and absolute vulnerability. Also incorporating the interaction between two profiles into the vulnerability model will help to model vulnerability realistically.

Investigating which attributes contribute towards the vulnerability of a profile for different types of users can help to incorporate the psychology of information retrieval because different users can self disclose in different ways. Other ideas include analysing different approaches to derive the attribute weights for the individual vulnerability calculation and using probabilistic approaches, e.g., Bayesian theory and Bayesian nets to explore privacy risks.

References

Abdul-Rahman, R., Alim, S., Neagu, D. and Ridley, M. (2010) 'Algorithms for data retrieval from online social network graphs', *IEEE. International IEEE Conference on Computer and Information Technology. In: Proceedings of the 10th International IEEE Conference on Computer and Information Technology (CIT 2010)*, 29 June–1 July, Bradford, UK, pp.1660–1666, IEEE CS.

Alim, S., Neagu, D. and Ridley, M. (2011) 'Axioms for vulnerability measurement of online social network profiles', *IEEE, International Conference in Information Society, in Proceedings of the International Conference in Information Society (i-Society) 2011*, 27–29 June, London, UK, pp.241–247.

Balduzzi, M., Platzer, C., Holz, T., Kirda, E., Balzarotti, D. and Kruegel, C. (2010) 'Abusing social networks for automated user profiling', *13th International Symposium on Recent Advances in Intrusion Detection (RAID), in Proceedings of Springer LNCS*, Ottawa, Canada, 15–17 September, Vol. 6307, pp.422–441.

BBC News (2007) 'Protests force Facebook to change', available at http://news.bbc.co.uk/1/hi/7120916.stm (accessed on 10 February 2011).

BBC News (2010) 'Palin e-mail hack details emerge', available at http://news.bbc.co.uk/1/hi/technology/7624809.stm (accessed on 3 July 2010).

Calvo, C. and Dercon, S. (2005) 'Measuring individual vulnerability', available at http://website1.wider.unu.edu/conference/conference-2005-3/conference-2005-3 papers/Calvo%20&%20 Dercon.pdf (accessed on 8 April 2011).

Caverlee, J. and Webb, S. (2008) 'A large-scale study of MySpace: observations and implications for online social networks', in *Proceedings of the 2nd AAAI International Conference on Weblogs and Social Media (ICWSM 2008)*, Seattle, USA, 30 March–2 April, pp.36–44, Association for the Advancement of Artificial Intelligence, California, CA.

Dunbar, R.I.M. (1992) 'Neocortex size as a constraint on group size in primates', *Journal of Human Evolution*, Vol. 22, No. 6, pp.469–493.

Dwyer, C., Hiltz, S.R. and Passerini, K. (2007) 'Trust and privacy concern within social networking sites: a comparison of Facebook and MySpace', *Americas Conference on Information Systems, in Proceedings of the Thirteenth Americas Conference on Information Systems (AMICS 2007)*, 9–12 August, Colorado, USA.

Emery, D. (2010) 'Details of 100m Facebook users collected and published', available at http://www.bbc.co.uk/news/technology-10796584 (accessed on 10 February 2011).

Facebook (2011) 'Press room', available at http://www.facebook.com/press/info.php?statistics (accessed on 12 September 2011).

Gundecha, P., Barbier, G. and Liu, H. (2011) 'Exploiting vulnerability to secure user privacy on social networking site', ACM, available at http://engineering.asu.edu/sites/default/files/shared/ASUCIDSE-2011-001.pdf (accessed on 1 July 2011).

Hannerman, R.A. and Riddle, M. (2005) 'Introduction to social network methods', available at http://www.faculty.ucr.edu/~hanneman/nettext/C7_Connection.html#geodesic (accessed on 29 December 2010).

Holme, P., Kim, B.J., Yoon, C.N. and Han, S.K. (2002) 'Attack vulnerability of complex networks', *Physical Review*, Vol. E, No. 65, pp.1–14.

Indiana University (2011) 'Informatics students discover, alert Facebook to threat allowing access to private data, bogus messaging', available at http://newsinfo.iu.edu/news/page/normal/17192.html (accessed on 10 February 2011).

Jagatic, T.N., Johnson, N.A., Jakobsson, M. and Menczer, F. (2007) 'Social phishing', *Communications of the ACM*, Vol. 50, No. 10, pp.94–100.

Kirk, J. (2006) 'Phishing scam takes aim at MySpace.com', available at http://www.pcworld.com/article/125956/phishing_scam_takes_aim_at_myspacecom.html (accessed on 25 June 2010).

Kleinberg, J. (2002) 'The small-world phenomenon: an algorithm perspective', *Procs. of the Thirty-second Annual ACM Symposium on Theory of Computing (STOC 2002)*, pp.163–170, Portland, USA.

Krishnamurthy, B. and Wills, C.E. (2009) 'On the leakage of personally identifiable information via online social networks', *ACM, The 2nd ACM SIGCOMM Workshop on Online Social Networks (WOSN 2009), in Proceedings of the 2nd ACM Workshop on Online Social Networks (WOSN 2009)*, Barcelona, Spain, 17 August, pp.7–12, ACM Press, New York, NY.

Lam, I.F., Chen, K.T. and Chen, L.J. (2008) 'Involuntary information leakage in social network services', *Third International Workshop on Security, in Proceedings of the 3rd International Workshop on Security: Advances in Information and Computer Security (IWSEC 2008)*, Springer, Kagawa, Japan, 25–27 November, pp.167–183.

Luo, B. and Lee, D. (2009) 'On protecting private information in social networks: a proposal', *IEEE, IEEE Conference on Data Engineering, in Proceedings of the 25th IEEE Conference on Data Engineering (ICDE 2009)*, Shanghi, China, pp.1603–1606.

McCallister, E., Grance, T. and Scanfore, K. (2009) 'Guide to protecting the confidentiality of personally identifiable information (PII)', available at http://www.scribd.com/doc/10968241/NIST-Guide-to-Protecting-the-Confidentiality-of-PII (accessed on 14 November 2010).

Narayanan, A. and Shmatikov, V. (2009) 'De-anonymizing social networks', *IEEE Symposium on Security and Privacy, in Proceedings of the 30th IEEE Symposium on Security and Privacy*, 17–21 May, pp.173–187, IEEE CS.

Patchin, J.W. and Hinduja, S. (2010) 'Changes in adolescent online social networking behaviours from 2006 to 2009', *Computers in Human Behaviour*, Vol. 26, pp.1818–1821.

Sweeney, L. (1997) 'Weaving technology and policy together to maintain confidentiality', *J. of Law, Medicine and Ethics*, Vol. 25, pp.98–110.

Watts, D.J. and Strogatz, S. (1998) 'Collective dynamics of small networks', *Nature*, Vol. 393, No. 1998, pp.440–442.

Xiang, R., Neville, J. and Rogati, M. (2009) 'Modelling relationship strength in online social networks', ACM, *World Wide Web Conference, in Proceedings of 19th International Conference on World Wide Web (WWW'10)*, New York, USA, pp.981–990, ACM Press, New York, NY.

Ye, S. and Wu, F. (2010) 'Measuring message propagation and social influence on Twitter.com', *Procs. of the Second International Conference on Social Informatics (SocInfo'10)*, Laxenburg, Austria, pp.216–231.

Yun, S., Do, H. and Kim, H.G. (2010) 'Analysis of user interactions in online social networks', available at http://channy.creation.net/blog/data/channy/2010/sns-social-interaction.pdf (accessed on 17 September 2010).

Appendix 1

Additional calculations for geometric mean operator used in Proposition 3

$$\left.\begin{array}{l} V_{R_i} = \left(\coprod\limits_{\substack{i=1 \\ i\neq j}}^{n} V_{I_i} \right)^{\frac{1}{n}} \\[2ex] V'_{R_i} = \left(\prod\limits_{\substack{i=1 \\ i\neq j}}^{n+1} V_{I_j} \right)^{\frac{1}{n+1}} \end{array} \right| \frac{V_{R_i}}{V'_{R_i}} = \frac{\left(\prod\limits_{1}^{n} V_{I_i} \right)^{\frac{1}{n}}}{\left(\prod\limits_{1}^{n+1} V_{I_i} \right)^{\frac{1}{n+1}}}$$

$$\log \frac{V_{R_i}}{V'_{R_i}} = \frac{1}{n} \log \prod_{1}^{n} V_{I_i} - \frac{1}{n+1} \log \prod_{1}^{n+1} V_{I_i}$$

$$= \frac{(n+1) * \log \prod\limits_{1}^{n} V_{I_i} - n * \log \prod\limits_{1}^{n+1} V_{I_i}}{n(n+1)}$$

$$= \frac{n * \log \prod\limits_{1}^{n} V_{I_i} + \log \prod\limits_{1}^{n} V_{I_i} - n * \log \left(\left(\prod\limits_{1}^{n} V_{I_i} \right) * V_{I_{n+1}} \right)}{n(n*1)}$$

$$= \frac{n * \log \prod\limits_{1}^{n} V_{I_i} + \log \prod\limits_{1}^{n} V_{I_i} - n * \log \prod\limits_{1}^{n} V_{I_i} - n * \log V_{I_{n+1}}}{n(n*1)}$$

$$= \frac{\log \prod\limits_{1}^{n} V_{I_i} - \log V_{I_{n+1}}^{n}}{n(n+1)} = \frac{\log \prod\limits_{1}^{n} \left(\frac{V_{I_i}}{V_{I_{n+1}}} \right)}{n(n+1)}$$

$$V_{I_i} < V_{I_{n+1}} \Rightarrow \frac{V_{I_i}}{V_{I_{n+1}}} < 1$$

$$\prod_{1}^{n} \frac{V_{I_i}}{V_{I_{n+1}}} < 1$$

$$\log \prod_{1}^{n} \frac{V_{I_i}}{V_{I_{n+1}}} < 0$$

$$\frac{V_{R_i}}{V'_{R_i}} < 1 \Rightarrow V_{R_i} < V'_{R_i}$$

Appendix 2

Additional calculations for arithmetical mean operator used in Proposition 4

$$V_{R_{n+1}} = \frac{1}{n+1}\sum_{1}^{n+1} V_{I_i}$$

$$V_R = \frac{1}{n}\sum_{1}^{n} V_{I_i}$$

$$V_{R_{n+1}} - V_{R_n} = \frac{1}{n+1}\sum_{1}^{n+1} V_{I_i} - \frac{1}{n}\sum_{1}^{n} V_{I_i} = \frac{n*\sum_{1}^{n+1} V_{I_i} - (n+1)*\sum_{1}^{n} V_{I_i}}{(n+1)*n}$$

$$= \frac{n*\left(\sum_{1}^{n} V_{I_i} + V_{I_{n+1}}\right) - n*\sum_{1}^{n} V_{I_i} - \sum_{1}^{n} V_{I_i}}{n(n+1)}$$

$$= \frac{n*\sum_{1}^{n} V_{I_i} + n*V_{I_{n+1}} - n*\sum_{1}^{n} V_{I_i} - \sum_{i}^{n} V_{I_i}}{n(n+1)}$$

$$= \frac{n*V_{I_{n+1}} - \sum_{1}^{n} V_{I_i}}{n*(n+1)} = \frac{\sum_{1}^{n} V_{I_{n+1}} - \sum_{1}^{n} V_{I_i}}{n*(n+1)}$$

$$= \frac{\sum_{1}^{n} V_{I_{n+1}} - V_{I_i}}{n*(n+1)}$$

$$V_{I_{n+1}} > V_{I_i}$$

$$\sum V_{I_{n+1}} > \sum_{1}^{n} V_{I_i} \Rightarrow \sum_{1}^{n}\left(V_{I_{n+1}} - V_{I_i}\right) > 0$$

$$\frac{\sum_{1}^{n}\left(V_{I_{n+1}} - V_{I_i}\right)}{n(n+1)} > 0$$

$$V_{R_{n+1}} - V_{R_n} > 0$$

$$V_{R_{n+1}} > V_{R_n}$$