Vulnerability Assessment and Classification based on Influence Metrics in Mobile Social Networks

Keerthiraj Nagaraj
University of Florida
Gainesville, USA
k.nagaraj@ufl.edu

Swapnil Sunilkumar Bhasale
University of Florida
Gainesville, USA
ssb123bhasale@ufl.edu

Janise McNair
University of Florida
Gainesville, USA
mcnair@ece.ufl.edu

Ahmed Helmy
University of Florida
Gainesville, USA
helmy@ufl.edu

ABSTRACT

In emerging 5G wireless systems, Mobile Social Networks (MSN) will play an important role for providing data services and offloading data traffic from cellular networks. MSNs are vulnerable to various security attacks because of the ways users move and collaborate. Since most protocols for MSNs are designed based on social behaviors of users, it is important to understand the impact of user behavior on network vulnerability. This can provide valuable insights into crucial factors, such as how easily a network loses its connectivity, or a network's ability to form strong communities.

We present a novel vulnerability assessment and classification scheme based on structural, social and influence distribution metrics in mobile social networks. We design a vulnerability index metric (VI) to investigate the level of damage inflicted when networks are subjected to both targeted and random attacks. Then, we use influence distribution metrics and various machine learning based classifiers to determine the vulnerability levels for various network states. Finally, we define a Mean Information Diffusion index to determine the information dissemination capability of a network, given the vulnerability state. Our results revealed that campus WLAN traces, represented by the Time Variant Community model, exhibit highly vulnerable states that reduce the network's ability to disseminate information by up to 16%.

CCS CONCEPTS

• **Mathematics of computing** → **Graph theory**; • **Networks** → **Network simulations**; **Network performance analysis**; **Mobile and wireless security**; **Network dynamics**; **Network mobility**; **Mobile networks**; *Denial-of-service attacks*; • **Computing methodologies** → **Supervised learning by classification**; • **Theory of computation** → *Graph algorithms analysis.*

MobiWac '19, November 25–29, 2019, Miami Beach, FL, USA

ACM ISBN 978-1-4503-6905-3/19/11…$15.00
https://doi.org/10.1145/3345770.3356737

KEYWORDS

Mobile Social Networks; Vulnerability assessment; Machine Learning; Influence metrics; Information dissemination.

ACM Reference Format:
Keerthiraj Nagaraj, Swapnil Sunilkumar Bhasale, Janise McNair, and Ahmed Helmy. 2019. Vulnerability Assessment and Classification based on Influence Metrics in Mobile Social Networks. In *17th ACM International Symposium on Mobility Management and Wireless Access (MobiWac '19), November 25–29, 2019, Miami Beach, FL, USA*.ACM, New York, NY, USA, 8 pages. https://doi.org/10.1145/3345770.3356737

1 INTRODUCTION

Currently, billions of mobile devices exist in the world. This number is expected to increase with the introduction of the 5G standard. Most of these devices are equipped with communication and sensor technologies such as Bluetooth, WiFi, and proximity sensors, which allow the users to opportunistically create communication links with other users and form short range ad hoc networks for device-to-device communication, mobile social networks, vehicular communications, and Internet of Things, to name a few applications. One of the features of these networks is the free mobility of devices. Node mobility increases a network's complexity due to the changing dynamics and makes protocol design and a securing network more difficult as the nodes change connections, points of contact and paths for data flow [1].

When mobility is considered in MSNs, the network may reach a state in which small number of nodes, called *Highly Crucial Nodes*, determine most of the network connectivity and information flow. Highly crucial nodes hold connection advantages over other nodes and thus have more impact in *worm propagation*, information dissemination, routing protocols, power flow, security measures, *coordinated attacks* and channel scheduling schemes. The relative importance of these nodes also makes them vulnerable and easy targets for any attacks to degrade network performance [2]. Characterizing the mobility patterns of the network may be the key to predicting a node's future influence and thereby the network's future vulnerability. Other studies have examined centrality metrics and/or social metrics to determining crucial nodes to increase network performance. To our knowledge, previous works have not linked the crucial nodes and mobility patterns to predict network vulnerability. In secure networks, there is a need to identify the relevant structural and social metrics, determine the crucial nodes

of a network and examine the corresponding mobility patterns to better predict and possibly avoid vulnerable states.

In this paper, we present a vulnerability assessment scheme based on structural and social aspects of mobile users in MSNs. We develop a classification scheme to determine how overall network vulnerability, based on influence distribution metrics in the network. The results obtained are quite promising and show the effective use of our scheme and its impact on the information diffusion applications in Mobile Social Networks. Our paper makes the following contributions:

- Presents a novel *vulnerability assessment scheme* for MSNs based on structural and social aspects of users.
- Presents a graph theoretic based vulnerability index metric to assess the degree to which the network's performance is vulnerable to various attacks.
- Develops a set of influence distribution metrics to classify the network's vulnerability states using supervised learning methods such as K-Nearest Neighbor, Random Forest, AdaBoost and Multilayer Perceptron Neural Network classifiers.
- Provides a novel study of the impact of mobility model on network vulnerability. To our knowledge, this is the first study that evaluates how the mobility pattern exhibited by different MSNs impacts their network vulnerability.
- Examines the information diffusion capability in terms of the MSN's network vulnerability state.

This paper is organized as follows. Section 2 describes related work to our analysis. In Section 3, we discuss structural and social metrics that are used to identify critical nodes in the network, graph theoretic concepts that are used to define the *vulnerability index* metric, and influence distribution metrics that are used to classify the network states. In Section 4, we provide a discussion on certain mobility models, we observe the results showing the differences in network vulnerability, performance metrics for classification models and difference in *MID* values for the considered mobility scenarios. Finally, Section 5 concludes the paper.

2 RELATED WORK

In this section, we discuss various research literature from the fields of Mobile Social Networks, Opportunistic networks, Peer-to-Peer networks and Device-to-Device communications, which are related to our proposed vulnerability assessment and classification scheme. We point out the major contributions of these papers and how they form the basis for our study.

In [3], the authors propose a framework for content distribution in content-based MSNs for 5G networks. It is shown that availability of content replicas at Content Centric Nodes greatly affects the caching performance. Also, a caching scheme called Social Content Caching is developed for determining what content to store considering degree centrality of users who request content and mobility of all users in the network. The authors in [4] propose a context-aware information diffusion scheme in 5G Mobile Social Networks. This information diffusion scheme considers networking and socially-related metrics of users to model expected information diffusion time. It develops a metric called social inter-contact time which can predict the interaction frequency between users and a generic social platform to help improve overall information diffusion time in the MSN. The work in [4] shows that social metrics can be used to improve information diffusion applications in 5G MSN. Both [3] and [4] show that MSNs are a viable solution to improve network performance in 5G, as well as use of social metrics and mobility of users to influence network performance.

The authors in [5] propose the SimBetTS delay-tolerant MANET routing algorithm, for routing based on social network analysis techniques. The protocol assigns utility value to all the nodes as a weighted combination of social similarity, betweenness centrality and tie strength to select the message forwarding nodes. The selection of highly central nodes seems to improve the routing performance. Hence, by controlling these nodes we can control the performance of the network. A low energy socially cognizant routing protocol is proposed for delay tolerant networks in [6]. Frequency of collocation and change of degree, which are functions of degree centrality, are used to define utility values for nodes, which are then used as metrics to identify message forwarders in the network. Nodes with higher utility values hold advantage over others as these are often used as intermediate nodes in routing. In [5] and [6], social metrics are used to improve routing are discussed, but they do not consider the impact of mobility or the change in vulnerability levels caused due to modification of influence in the network on information diffusion. These factors will be addressed in our study.

The work in [7] provides a discussion on how to infer implicit social ties in the Mobile Social Networks based on user interaction and activity. The authors explore different basis for developing social ties between users such as location population, co-occurrence diversity and users' mobility behavior. Results show that the developed social ties when tested using real world mobile social network datasets predicted online social interaction between users effectively. This paper shows that the user interactions and mobility behavior could be effectively used to develop meaningful social metrics between users in MSNs. In [8], the authors develop encounter metrics based on the interactions of users extracted from WLAN traces for a duration of 3 months in a campus-based environment. The authors also show that the past encounter values between users are reliable metrics to identify the strength of social relation between them and that they can also be used to improve the performance of routing in the network.

The authors in [9] conduct structural vulnerability assessment of community-based routing in opportunistic networks. The proposed technique aims to find k most critical nodes whose removal results in large separation of network communities. This work does not consider the effect of user movement dynamics, nor enough of the impact of the social aspects of users on the vulnerability assessment scheme, which will be addressed in our analysis. Research on network security due to node influence has focused mainly on centrality metrics and/or social metrics. The authors in [10] analyze the impact of various centrality metrics such as degree centrality, closeness centrality, betweenness centrality and eigenvector centrality as measures to find the important nodes in the network and compare their performance against randomly selected nodes in spreading malicious information in peer-to-peer networks. Results show that nodes with higher eigenvector centrality values needs to be protected in order to avoid misinformation spread in the network

or in other words nodes with higher eigenvector centrality values are highly influential in the network for information dissemination compared to other centrality metrics.

The authors in [11] have studied the vulnerability of networks using centrality metrics by combining the degree of a node and the average degree of its neighbors to measure node importance. This study shows that centrality metrics could be effectively used to understand the network vulnerability. The authors in [12] propose a heuristic algorithm to compute network centrality in real, low-power networking hardware in a 1000-node network, and show that centrality can be effectively used as a building block for security functions in networks.

Authors in [13] show that machine learning techniques such as Support Vector Machines can be effectively applied to very large scale social networks in dealing with prediction and recommendation problems. The authors make use of features such as online factor, content factor, location factor, mobility factor, and social factor to forecast a metric knows as nowcasting that can predict if two users of Device-to-Device communication system might share information in near future.

From the papers that we discussed in this section, we can state that MSN is a viable add-on to traditional 5G networks, mobility and social metrics can impact performance of MSN, connectivity and community forming ability are viable factors to assess network vulnerability, user activity and encounter metrics can help us identify critical nodes in the network for security, routing and information diffusion applications, and machine learning algorithms can be effectively used for prediction problems in MSN. These facts form a basis for our vulnerability assessment and classification scheme using structural, social and influence distribution metrics. To our knowledge, this is the first study that tries to identify the relation between influence distribution metrics and vulnerability in MSNs.

3 CHARACTERIZING NETWORK VULNERABILITY

3.1 Vulnerability Index *(VI)*

To understand the impact of network attacks in various mobility scenarios, we develop the Vulnerability Index *(VI)*. We formulate *VI* as shown in Eq. 1,

$$VI = W_{AC} * (AC_1 - AC_2) + W_{CI} * (CI) \tag{1}$$

where the weights W_{AC} and W_{CI} are applied to balance the scale of the two metrics. AC_1 is the average clustering coefficient before disrupting the network and AC_2 is the average clustering coefficient after after disrupting the network. The average clustering coefficient (*AC*) shows the ability of nodes to form strong and large communities with neighboring nodes in the network [14]. *AC* is defined as the ratio of closed triplets in the network to number of all triplets (open and closed).

CI is based on the concept of articulation points and is defined as the fraction of number of articulation points to total number of nodes in the network. A given node is considered as an articulation point if its removal splits the network into 2 or more disconnected components. *CI* is defined as the ratio of number of articulation points to total number of nodes in the network. Higher the value of *VI*, more the network loses its ability to form strong communities and connections which is unfavorable in MSN.

3.2 Consolidated Influence Metric *(CIM)*

We now formulate *CIM* as a measure to indicate the relative importance/influence of a given node in the network. *CIM* is calculated based on 2 components namely, Structural influence and Social influence. We propose *CIM* as a consolidated influence metric to represent a node $i's$ collective influence in the network. We define *CIM* as shown in Eq.2.

$$CIM_i = W_{STR} * STR_i + W_{SOC} * SOC_i \tag{2}$$

*3.2.1 Structural influence **(STR)**.* *STR* provides a measure of node's influence based on its position in the current network configuration, and it is formulated based on the concepts of network centrality. Degree Centrality *(DC)* [11] is defined as the number of links incident upon a node. The degree can be interpreted in terms of the immediate risk of a node for catching whatever is flowing through the network, for both useful information or malware. The idea of Betweenness Centrality *(BC)* [5] for a given node *i* is that how many pairs of nodes in the network are connected through the shortest path to each other passing through the node *i*. Closeness Centrality *CC* [10] is based on the idea that nodes with short distance to other nodes can spread information very productively through the network. Eigenvector Centrality *(EC)* [10] is based on the concept that a node's influence increases when it gets connected to other highly critical nodes. [10–12] shows that the aforementioned centrality metrics can be effectively used for security related applications. *STR* is calculated as shown in Eq. 3.

$$STR_i = W_{BC} * (BC)_i + W_{CC} * (CC)_i + W_{DC} * (DC)_i + W_{EC} * (EC)_i \tag{3}$$

*3.2.2 Social influence **(SOC)**.* *SOC* provides a node's influence as a measure of its social aspects and its relation with other mobile nodes. We develop 3 social influence metrics namely *Social Encounter* metric, *Social Willingness* metric, and *Social Popularity* metric.

Social Encounter metric (SE): In [8], we noticed that encounter metrics could be used as stable social metrics to improve routing in the mobile networks. In our analysis, every time 2 nodes come in range of each other, we decide that those nodes encountered with a probability of *'p'*. To calculate *SE*, we sum the total number of encounters a node had with each of its current neighbors in the past iterations and is shown in Eq. 4.

$$SE_i = \sum_{v \in CN_i} en(i, v) \tag{4}$$

where CN_i represents the current neighbors of node i and $en(i, v)$ shows the number of encounters between node i and node v in the past iterations.

Social Willingness (SW): Social network data can reveal willingness of a node to share information, form strong connections or be a part of community with its neighbors. By the principles of homophily, it is safe to assume that people who are friends in a social network have better chances of forming strong connections or being in same community even in mobile networks. To test this

approach we used open source friendship network data from Facebook. We pulled designated social network users as nodes in our experiments. Data from any social network that network nodes are a part of can be used to define this metric. *(SW)* is calculated as shown in Eq. 5:

$$SW_i = |SNF_i \cap CN_i| \tag{5}$$

where SNF_i and CN_i represents set of all social networking friends and current neighbors of node i respectively.

Social Popularity metric (SP): *SP* is the count of how many times a given node was selected as critical node in the previous iterations based on *CIM* values. *SP* is crucial in applications such as information caching. For example, if a node was selected a large number of times for cache storage, then it is more likely to be selected again, or it could have useful information in the network for other nodes. *SP* is calculated as shown in Eq. 6,

$$SP_i = \sum_{v=1}^{k} p_i(v) \tag{6}$$

where k is the current iteration number, $p_i(v) = 1$ if node i was selected as critical node in iteration v, otherwise $p_i(v) = 0$. Higher values of *SW*, *SE* and *SP*, indicate greater social influence of a given node. The Social influence, *SOC*, is calculated as shown in Eq. 7.

$$SOC_i = W_{SW} * (SW)_i + W_{SE} * (SE)_i + W_{SP} * (SP)_i \tag{7}$$

All the individual structural and social metrics are normalized before calculating *STR* and *SOC* to bring all of them to similar scale. The weights in Eq. (2), Eq. (3) and Eq. (7) can be adaptively varied according to the network environment and application. In our analysis, we have given slightly higher weights to *STR*, as it provides influence information of nodes more relevant to the current network configuration and slightly higher weight to *EC* as it was shown to be more successful compared to other structural influence metrics for network vulnerability applications [10].

3.3 Influence Distribution metric

We develop *Peak-to-Average metric (P2A)* as the general influence distribution metric and then calculate *P2A* for each of the individual structural and social influence metrics to understand their impact on *VI* values for different mobility models. *P2A* is defined as shown in Eq. 8,

$$P2A_M = \frac{\frac{1}{x}\sum_{i \in X} M_i}{\frac{1}{n}\sum_{j \in N} M_j} \tag{8}$$

where X is the set of all most critical nodes with a total of x elements, M_i is the influence value of i^{th} critical node, N is the set of all nodes in the network with total of n elements and M_j is the influence value of j^{th} node.

A lower value of P2A indicates that influence has been fairly distributed in the network. For instance, if the value of *P2A* is 1 (lowest possible value), then the influence is distributed uniformly among all nodes. Conversely, *a higher value of P2A indicates that the network is in a state where the selected critical nodes hold most of the influence in the network.*

We define *P2A* values for each of the structural and social influence metrics namely $P2A_{BC}$, $P2A_{CC}$, $P2A_{DC}$, $P2A_{EC}$, $P2A_{SW}$, $P2A_{SE}$ and $P2A_{SP}$ for *BC*, *CC*, *DC*, *EC*, *SW*, *SE* and *SP* respectively according to the Eq. 8.

Finally, we define state of the network for each iteration *'a'* based on the value of *VI* for each mobility model. Three states are defined namely Vulnerable, Semi-Vulnerable and Safe based on the following conditions:

$$state(a) = \begin{cases} Vulnerable & \text{if } VI_a >= (Q_v) \\ Semi-Vulnerable & \text{if } (Q_{sv}) < VI_a < (Q_v) \\ Safe & \text{if } VI_a <= (Q_{sv}) \end{cases}$$

where Q_{sv} and Q_v represent percentile thresholds of *VI* for a given mobility model for all the iterations. The definition of these states are such that when *VI* value of network is high, it falls in Vulnerable state and when *VI* value of network is low, it falls in Safe state. Semi-Vulnerable state acts as a buffer between Safe to Vulnerable states and avoids sharp transition.

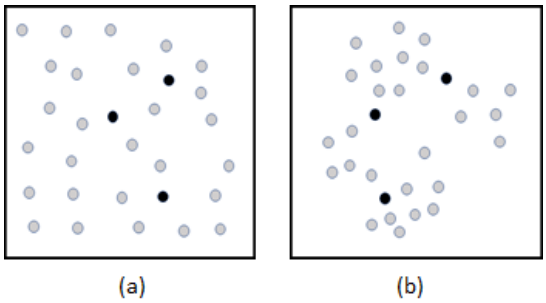

Figure 1: Example of network states. (a) Safe (b) Vulnerable

Fig. 1 (a), (b) shows examples of Safe and Vulnerable network states. In Fig. 1, three nodes are attacked (darkened) in both (a) and (b) cases, but (b) results in more vulnerable network compared to (a) as attacking the darkened nodes results in higher loss of connectivity and community forming capacity among nodes in the network in case (b).

4 RESULTS AND DISCUSSION

The proposed vulnerability assessment and classification scheme based on structural and social metrics was tested by developing MSN simulation test bed using Python in the Anaconda platform. We employ Stanford Network Analysis Platform (*SNAP*) [15], a large scale network data analysis tool and other popular Python library packages such as NumPy, SciPy, Pandas, Scikit-learn and PyMobility [16] to carry out the analysis. We initialize the MSN with 1000 (N) nodes, with the nodes placed randomly in an area of 500x500 (X_{dim}, Y_{dim}) dimensions, which is a reasonable network size for the simulation area for campus/office based environments. All the nodes have same communication range R of 10 units and devices within range of each other are allowed to communicate. The nodes are then allowed to move as per the selected mobility model which changes the node positions accordingly and there by modifying the connections in the network for 10000 (T) iterations.

Nodes are listed in descending order based on the values of *CIM*, 50 (X_{top}) most critical nodes (5% of total users) are selected from this list. The communication links associated with these critical nodes and X_{top} randomly selected nodes are disrupted to imitate the targeted and random attacks in the network respectively. In the adversary model that we consider, the attacker targets X_{top} critical nodes with attacks such as Denial of Service (DoS), Jamming or flooding the communication links of these nodes making them unavailable for any network operations.

Plots of *VI* are presented considering both targeted and random network attacks. A comparison study between the impact of considered mobility models on network vulnerability is presented. Performance metrics obtained for various classification models will be tabulated and the % change in the value of Mean Information Diffusion index will be presented to show the effect of different network states on information diffusion application in MSN.

Algorithm 1 shows the various steps involved in proposed vulnerability assessment scheme. By following the steps of Algorithm 1 for *RWP*, *RPGM* and *TVC* models, we will have *VI* values for targeted and random attacks for each of the considered mobility models.

Algorithm 1 : Vulnerability assessment scheme

Initialization:

(1) Set X_{dim}, Y_{dim}, R, N, and X_{top}.
(2) Select mobility model and initialize node positions

For $it = 1$ **to** T

- Get node positions
- Create a graph G with nodes at these positions
- Create edges between nodes which are in range R of each other in G
- Assign weights W to edges inversely proportional to $'d'$, where d is the distance between the nodes
- Calculate *CIM* for each node
- Sort nodes based on *CIM* values in descending order
- Select X_{top} most critical nodes
- Calculate *CI* and AC_1
- Remove edges incident on selected X_{top} critical nodes from G forming $G_{targeted}$
- Calculate AC_2 in $G_{targeted}$ and then $VI_{targeted}$
- Remove edges incident on X_{top} random nodes from G forming G_{random}
- calculate AC_3 in G_{random} and VI_{random}
- Save $V_{targeted}$ and V_{random} for the current iteration it.

4.1 Background on mobility models

The network architecture for this study is a collection of similarly able mobile nodes that are members of the same network. We incorporate selected mobility models that provide various perspectives to understand the movement patterns of the network nodes, and to capture the collective mobility behavior of nodes in the network. The resulting network dynamics because of mobility is then evaluated. The models considered for this study are: Random Way Point *(RWP)*[17], Reference Point Group mobility *(RPGM)* [18], and Time Variant Community *(TVC)* [19]. Further, we briefly describe these models, as well as the motivation for including them in our analysis.

4.1.1 Random Way Point. RWP is a model that includes random changes in location, velocity and acceleration of networks nodes over time. It is the most popular mobility model to evaluate mobile network protocols, because of its simplicity and wide availability. The destination, speed and direction are all chosen randomly and independently of other nodes in the successive iterations. The parameters associated with this model are as follows:

- Speed = $[V_{min}, V_{max}]$
- Destination = [Random X, Random Y]
- Pause Time ≥ 0

Each node begins by pausing for a fixed number of seconds. The node then selects a random destination in the simulation area and a random speed between minimum Speed (V_{min}) and maximum Speed (V_{max}. The node moves to this destination and again pauses for a fixed period before selecting another random location and speed. This behavior is repeated for the entire length of the simulation.

RWP has been shown to not match many realistic scenarios well. However, we used it since it is the most commonly used mobility model, and also it is useful as a reference to compare the other models. RWP has little to no spatial and temporal dependence, which means the node positions for successive iterations doesn't depend too much on the previous location or time of the simulation.

4.1.2 Reference Point Group Mobility. RPGM is an example for highly spatial dependent mobility model. In this model, nodes tend to form groups and each group has a logical center. The logical centers' mobility follows *RWP* mobility model. The nodes of each group follow their logical centers' mobility closely, with some deviation. The following steps shows a one simple way how *RPG* mobility works:

- $\tau -> \tau$+1
- $\vec{RP}(\tau) + \vec{GM} -> RP(\vec{\tau} + 1)$
- New destinations for nodes: $\vec{RM} + RP(\vec{\tau} + 1)$.

where τ is the time step, *RP* is the reference point vector, *GM* is the group motion vector, and *RM* is the random motion vector whose magnitude is chosen randomly from an uniform distribution $U_m[0, R]$ (R - Range of logical center) and direction is chosen randomly from an uniform distribution $U_m[0, 360^o]$.

The reason for selecting RPGM is to understand the dependence of network vulnerability on a highly spatial dependent mobility model. This model is also applicable in many scenarios such as campuses, museums, theme parks etc.

4.1.3 Time Variant Community. TVC is a synthetic mobility model developed using the trends observed in campus WLAN traces and hence using TVC is similar to working with campus based real world traces. TVC has both strong spatial and temporal dependence. This mobility model mainly captures two important factors in mobility namely, Skewed location visiting preferences of nodes and Periodic re-appearance of nodes at same location. In this model, communities are defined that are periodically revisited by nodes which also have skewed location visiting preferences for nodes. It results in

formation of groups over time at same locations with repetitive behavior.

Two types of time periods are defined for nodes namely, Normal Movement Period (*NMP*) and Concentration Movement period (*CMP*). In *CMP*, the certain locations are given high priority to be chosen as destination for nodes which results in the skewed location preferences. In each time period, communities are assigned to nodes and community locations are chosen at random. In each time period, a node has two different movement modes namely, Local epoch mode (nodes movement is confined to its community) and Roaming epoch modes (nodes are free to explore the entire simulation area). At beginning of each epoch, nodes randomly choose velocity from $U_v[Vmin, Vmax]$ and randomly choose direction from $U_d[0, 360^o]$, Movement length is chosen from an exponential distribution with the parameter L (L-average epoch length). If a node hits the boundary in local epoch, then it is re-inserted from the other end of the community. If a node hits the boundary in roaming epoch, then it is re-inserted from the other end of the simulation area. At the end of epoch, pause time is randomly chosen from $U_p[0, Tmax]$ and movement mode for the next epoch is picked according to a two-state Markov model with predefined probabilities. The selection of epoch along with combination of effect of *CMP* results in periodical re-appearance of nodes.

TVC model is selected to understand the impact of highly spatially and temporally dependent mobility model on network vulnerability.

4.2 Vulnerability Index for Targeted Versus Random attacks

Fig. 2, 3, and 4 shows the *VI* plots for *RWP*, *RPGM* and *TVC* models respectively when critical nodes *('Targeted')* and randomly selected nodes *('Random')* are attacked.

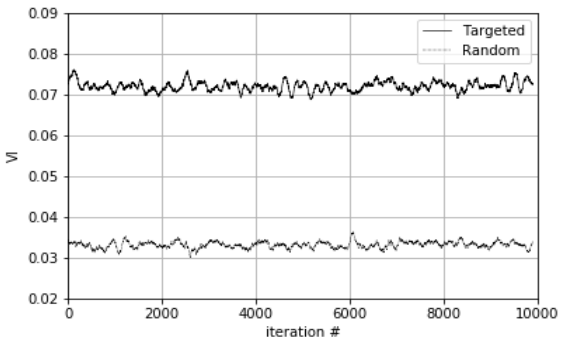

Figure 2: VI for Random Way Point mobility model

In each plot, we can observe that the *VI* is usually higher when critical nodes *('Targeted')* are selected than the case with randomly selected nodes *('Random')*. As the time progresses and *RPGM* starts affecting the node configuration to form groups, the difference between the *('Targeted')* and *('Random')* increases, which can be observed in Fig. 3. This increase in difference indicates that over the time *RPGM* results in a more vulnerable network. Similar behavior is also observed in Fig. 4 for the *TVC* model because of skewed location preferences of nodes which results in formation of communities and hence the difference between *('Targeted')* and *('Random')* increases over time.

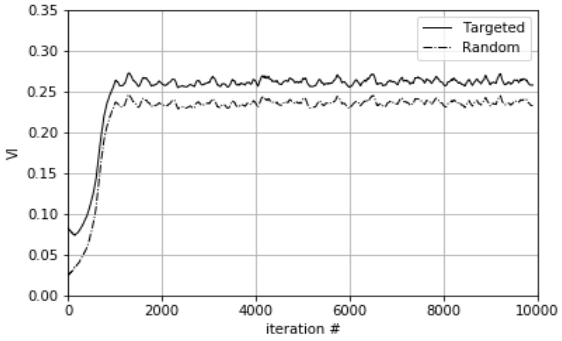

Figure 3: VI for Reference Point Group mobility model

For *TVC*, we can also observe that *VI* increases and decreases repetitively and this behavior can be explained by the periodical re-appearance property of users in the campus based environments.

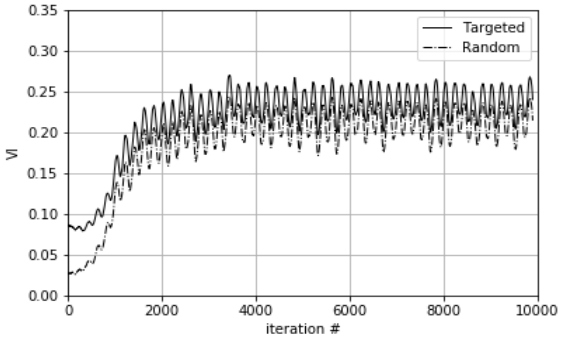

Figure 4: VI for Time Variant Community mobility model

4.3 Vulnerability Index for different mobility scenarios

Here, we present the plot which helps to identify the differences in network vulnerability caused by considered mobility models for targeted network attacks. We smoothened the *VI* values over time using running mean function with a window size of 100. Although minor local information loss occurs when running mean is used, the global trend remains unaffected. Fig. 5 shows the global trend of *VI* for all the mobility models considered in our analysis for 10000 iterations.

From Fig. 5, we can observe that *initially the vulnerability index for all the considered mobility models are at similar level, but as the time progresses they divert from each other showing each mobility model has different level of impact on network vulnerability.*

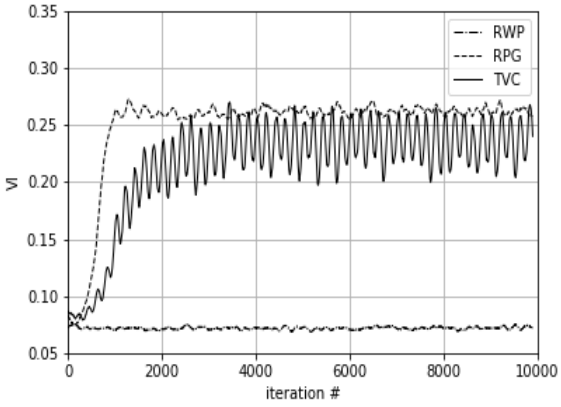

Figure 5: Comparison of VI for different mobility scenarios for targeted attacks

4.4 Vulnerability Classification scheme

We use $P2A_{BC}$, $P2A_{CC}$, $P2A_{DC}$, $P2A_{EC}$,$P2A_{SW}$, $P2A_{SE}$ and $P2A_{SP}$ as the feature set and the network state in each iteration as the target for developing classification models for each mobility scenario. We collected 10000 samples of which we used 67% to train and 33% to test the classification models. We randomized the samples before splitting them into training and testing to remove the bias from the dataset.

Algorithm 2 : Vulnerability classification scheme

(1) Select 1 mobility model among *RWP*, *RPGM* and *TVC*
(2) Calculate $P2A_{BC}$, $P2A_{CC}$, $P2A_{DC}$, $P2A_{EC}$,$P2A_{SW}$,$P2A_{SE}$ and $P2A_{SP}$ for T iterations
(3) Define network vulnerability states based on the value of VI in each of the T iterations
(4) Prepare dataset for classification using influence distribution metrics as feature set and vulnerability states as target variable
(5) Randomize the dataset to avoid bias and conduct correlation analysis
(6) Use $\frac{2}{3}^{rd}$ of the dataset to form training data and the remaining $\frac{1}{3}^{rd}$ to form test data
(7) Set hyper-parameters for KNN, RF, ADA and MLPNN classification models
(8) Train KNN, RF, ADA and MLPNN classification models with training data using scikit-learn library
(9) Record the performance metrics for test data
(10) Repeat steps [2-9] for remaining 2 mobility models

Four popular classification techniques namely K-nearest neighbors *(KNN)* (with *'K'* = 10) , Random Forest *(RF)* (with 100 estimators), AdaBoost classifier *(ADA)* (with 200 Decision Tree classifier estimators) and Multilayer Perceptron Neural Networks *(MLPNN)* (with 2 hidden layers, *ReLU* activation function, adaptive learning rate, and *ADAM* optimizer) are used for developing classification models for network states. We use *F1-score* as the performance metric to compare classification capability of different models. *F1-score* is the harmonic mean of Precision and Recall. *F1-score* is more useful than accuracy, precision or recall since we have uneven class distribution for different states in our analysis. High F1-score indicates better performance accuracy of the model.

Table 1: F1-scores of classification models

Classifier	RWP	RPGM	TVC
KNN	0.52	0.88	0.81
RF	0.51	0.92	0.84
ADA	0.51	0.91	0.85
MLPNN	0.51	0.84	0.75

The *F1-score* values for test dataset are shown in the Table. 1. As the behavior of nodes vary randomly in RWP without any spatial and temporal dependency, *VI* values have a lot of random variation and influence distribution metrics fail to capture the network vulnerability correctly, hence the classification models result in such a low F1-scores. In Table. 1, we can observe that for both RPGM and TVC mobility scenarios, classification models result in excellent *F1-scores*, proving that influence distribution metrics could be used as reliable features to classify the network states in MSN.

As the behavior of nodes vary randomly in *RWP* without any spatial and temporal dependency, *VI* values have a lot of variation and influence distribution metrics fail to capture the network vulnerability correctly, hence the classification models result in such low values for performance metric as shown in Table. 1. In Table. 1, we can observe that for RPGM and TVC mobility, classification models result in excellent values for performance metrics, proving that influence distribution metrics could be used as reliable features to classify the vulnerability states in MSN. *It is important to notice that even though influence distribution metrics were not directly used to calculate VI, these metrics proved to be reliable features to predict the network vulnerability states.*

4.5 Mean Information Diffusion index *(MID)*

It is crucial to understand the importance of classifying the network states and being able to predict the future network states based on influence distribution metrics. We use Mean First Passage Time *(MFPT)* [22] to define Global Information Diffusion (*GID*) index which is used to differentiate the information dissemination ability of network in different vulnerability states. *MFPT* from node i to node j is the expected number of steps it takes for a random message starting at node i to arrive for the first time at node j. *GID* for a network is the average of all the inverses of pairwise *MFPT* values between nodes and calculated as shown in Eq. 9.

$$GID = \sum_{(i,j) \in Nodes} \frac{1}{MFPT(i,j)} \tag{9}$$

where $MFPT(i, j)$ represents the Mean First Passage time between nodes i and j in the current network configuration.

Higher the value of *GID*, better is the information diffusion capabilities in the network as the information takes relatively lesser

Table 2: MID % change for different mobility scenario

Mobility model	MID % change
RWP	4.48 ↓
RPGM	12.41 ↓
TVC	15.94 ↓

number of steps to reach the targets. We define *MID* for each network state as the mean of global information diffusion index of all the iterations that were classified as belonging to that particular state.

$$MID\%change = \frac{MID_S - (MID_{SV} + MID_V)}{MID_S} \quad (10)$$

We define MID_s, MID_{sv} and MID_v for Safe, Semi-Vulnerable and Vulnerable network states. MID % change is percentage change in the value of *MID* when the network leaves Safe state and stays in Semi-vulnerable and Vulnerable states and is defined as shown in Eq. 10. Table. 2 shows the values of MID % change for three considered mobility models and the ↓ symbolizes the decrease in *MID* value. In Fig. 5, we can notice that *RPGM* and *TVC* models result in more vulnerable network compared to *RWP* because of their spatial and/or temporal dependency properties. From Table. 2, we can see similar performance deterioration among *RPGM* and *TVC* models for information diffusion applications in MSN.

5 CONCLUSIONS AND FUTURE WORK

In this paper, we presented a novel vulnerability assessment and classification scheme based on structural, social and influence distribution metrics for Mobile Social Networks *(MSN)*. We formulated Consolidated Influence Metric *(CIM)*, a measure which can provide the relative importance for any given node in the network based on its structural and social significance. The network was subjected to targeted and random attacks. Mobility models in which node movement has high spatial and temporal dependence (which is usually the case in MSN) results in more vulnerable networks compared to network in which nodes move randomly.

Influence distribution metrics were developed and were successfully used to classify network states using K-Nearest Neighbor, Random Forest, AdaBoost and Multilayer Perceptron Neural Network classification models. Influence distribution metrics provide us a way to detect the vulnerability state of the network and how much damage a possible attack might cause. Mean Information Diffusion (*MID*) index values decreases approximately by 5%, 12% and 16% when network leaves Safe state and stays in Vulnerable or Semi-Vulnerable states in Random Way-Point, Reference Point Group Mobility and Time Variant Community models respectively. Lower value of *MID* indicates lower quality of information diffusion. The *CIM* metric developed in our analysis can also be used to identify message forwarders for routing and cache storage nodes in information caching applications as it is helpful to identify influential nodes in the network. In future work, we would like to investigate how to improve the quality of information diffusion during various network attacks even for the mobility scenarios with high spatial, temporal and social dependencies.

REFERENCES

[1] F. Bai, Narayanan Sadagopan and A. Helmy, "IMPORTANT: a framework to systematically analyze the Impact of Mobility on Performance of Routing Protocols for Adhoc Networks," IEEE INFOCOM 2003. Twenty-second Annual Joint Conference of the IEEE Computer and Communications Societies (IEEE Cat. No.03CH37428), San Francisco, CA, 2003, pp. 825-835 vol.2.

[2] T. N. Dinh and M. T. Thai, "Network Under Joint Node and Link Attacks: Vulnerability Assessment Methods and Analysis," in IEEE/ACM Transactions on Networking, vol. 23, no. 3, pp. 1001-1011, June 2015.

[3] Z. Su and Q. Xu, "Content distribution over content centric mobile social networks in 5G," in IEEE Communications Magazine, vol. 53, no. 6, pp. 66-72, June 2015.

[4] G. Araniti, A. Orsino, L. Militano, L. Wang and A. Iera, "Context-Aware Information Diffusion for Alerting Messages in 5G Mobile Social Networks," in IEEE Internet of Things Journal, vol. 4, no. 2, pp. 427-436, April 2017

[5] E. M. Daly and M. Haahr, "Social Network Analysis for Information Flow in Disconnected Delay-Tolerant MANETs," in IEEE Transactions on Mobile Computing, vol. 8, no. 5, pp. 606-621, May 2009.

[6] C. Baker, J. Almodovar-Faria, P. S. Juste and J. McNair, "Low Energy Socially Cognizant Routing for Delay Tolerant Mobile Networks," MILCOM 2013 - 2013 IEEE Military Communications Conference, San Diego, CA, 2013, pp. 299-304.

[7] T. Pi, L. Cao, P. Lv, Z. Ye and H. Wang, "Inferring implicit social ties in mobile social networks," 2018 IEEE Wireless Communications and Networking Conference (WCNC), Barcelona, 2018, pp. 1-6.

[8] U. Kumar, G. Thakur, A. Helmy."PROTECT: proximity-based trust-advisor using encounters for mobile societies,"ACM IWCMC,2010.

[9] M. A. Alim, X. Li, N. P. Nguyen, M. T. Thai and A. Helal, "Structural Vulnerability Assessment of Community-Based Routing in Opportunistic Networks," in IEEE Transactions on Mobile Computing, vol. 15, no. 12, pp. 3156-3170, Dec. 1 2016.

[10] M. Kas, L. R. Carley and K. M. Carley, "Monitoring social centrality for peer-to-peer network protection," in IEEE Communications Magazine, vol. 51, no. 12, pp. 155-161, December 2013.

[11] A. V. Sathanur and D. J. Haglin, "A novel centrality measure for network-wide cyber vulnerability assessment," 2016 IEEE Symposium on Technologies for Homeland Security (HST), Waltham, MA, 2016, pp. 1-5.

[12] L. Maccari, Q. Nguyen and R. Lo Cigno, "On the Computation of Centrality Metrics for Network Security in Mesh Networks," 2016 IEEE Global Communications Conference (GLOBECOM), Washington, DC, 2016, pp. 1-6.

[13] X. Wang, H. Wang, K. Li, S. Yang and T. Jiang, "Serendipity of Sharing: Large-Scale Measurement and Analytics for Device-to-Device (D2D) Content Sharing in Mobile Social Networks," 2017 14th Annual IEEE International Conference on Sensing, Communication, and Networking (SECON), San Diego, CA, 2017, pp. 1-9

[14] K. Y. K. Hui, J. C. S. Lui and D. K. Y. Yau, "Small world overlay P2P networks," Twelfth IEEE International Workshop on Quality of Service, IWQOS 2004, pp. 201-210, 2004.

[15] J. Leskovec, and R. Sosic, "SNAP: A General Purpose Network Analysis and Graph Mining Library," ACM Transactions on Intelligent Systems and Technology. 8. 10.1145/2898361, 2016.

[16] A. Panisson, "panisson/pymobility", GitHub, 2012. [Online]. Available: https://github.com/panisson/pymobility. [Accessed: 29- June- 2019].

[17] C. Bettstetter, G. Resta and P. Santi, "The node distribution of the random waypoint mobility model for wireless ad hoc networks," in IEEE Transactions on Mobile Computing, vol. 2, no. 3, pp. 257-269, July-Sept. 2003

[18] X. Hong, M. Gerla, G. Pei, and C. C.Chiang, "A group mobility model for ad hoc wireless networks," In Proceedings of the 2nd ACM international workshop on Modeling, analysis and simulation of wireless and mobile systems. ACM, New York, NY, USA, pp53-60, 1999.

[19] W. J. Hsu, T. Spyropoulos, K. Psounis and A. Helmy, "Modeling Spatial and Temporal Dependencies of User Mobility in Wireless Mobile Networks," in IEEE/ACM Transactions on Networking, vol. 17, no. 5, pp. 1564-1577, Oct. 2009.

[20] R. A. Nugrahaeni and K. Mutijarsa, "Comparative analysis of machine learning KNN, SVM, and random forests algorithm for facial expression classification," 2016 International Seminar on Application for Technology of Information and Communication (ISemantic), Semarang, 2016, pp. 163-168.

[21] P. Natesan and P. Rajesh, "Cascaded classifier approach based on Adaboost to increase detection rate of rare network attack categories," 2012 International Conference on Recent Trends in Information Technology, Chennai, Tamil Nadu, 2012, pp. 417-422.

[22] A. Avena-Koenigsberger, J. Goni, R. Sole, and O. Sporns, "Network morphospace," Journal of the Royal Society Interface, 12(103), 20140881, 2015.