On the Machine Learning for Minimizing the Negative Influence in Mobile Cyber Physical Systems

Vijay Chaudhary and Danda B. Rawat

Data Science and Cybersecurity Center (DSC^2), Howard University, Washington DC, USA
vijay.chaudhary@bison.howard.edu, danda.rawat@howard.edu

ABSTRACT

Emerging cyber physical system (CPS) are expected to enhance the overall performance of the networked systems to provide reliable services and applications to their users. However, massive number of connectivities in CPS bring security vulnerabilities and the mobility adds more complexity for securing the mobile CPS. Any mobile CPS can be represented as a graph with connectivity as well as with interactions among a group of mobile CPS nodes that plays a major role as a medium for the propagation of wrong/right information, and influence its members in the mobile CPS. This problem has wide spread applications in viral information disseminating in mobile CPS, where a malicious mobile CPS node may wish to spread the rumor via the most influential individuals in mobile CPS. In this paper, we design, develop and evaluate a machine learning approach that is based on a set theoretic approach for optimizing the influence in mobile CPS. This problem has applications in civilian and military systems.

Keywords: Mobile CPS security, influence minimization, graph theory for mobile CPS

1. INTRODUCTION

Mobile cyber physical systems (CPS) also known as intelligent transportation systems with vehicular ad hoc networks (VANETs) are emerging applications to provide traffic efficiency and road safety.[1] In mobile CPS, information dissemination heavily relies on wireless communications to exchange information among CPS nodes[2,3] via device-to-device and/or device-to-base-station communications. Vehicles as mobile CPS nodes periodically broadcast their periodic status to their neighbors.[1] Furthermore, CPS nodes[*] are also expected to disseminate information related to traffic jams, accidents, possible detours, weather conditions, and facilities such as gas stations and restaurants.[4] Upcoming traffic information could be leveraged to avoid traffic congestion[5] and could help avoid traffic accidents,[1,6] eventually enhancing passenger comfort, traffic efficiency and safety of passengers. Therefore, there is a constant dissemination of information over the network of mobile CPS nodes where CPS nodes are heavily interdependent on the validity of information received to make informed decisions. Any false information related to sensitive information can lead to wrong decisions disrupting the optimal flow of traffic and even causing accidents of serious nature. Thus, it is crucial to check false information in a mobile CPS and maximize the flow of valid information over the network of mobile CPS nodes. In this paper, we investigate a probabilistic model (used in machine learning) for diffusion of information over a network, and its use to contain false information to limited number of devices within the network.

Recent related works include.[3,7–9] However, none of the related works consider minimizing the negative influence in mobile CPS using machine learning based approach, which is the subject of interest of this paper. Note that the machine learning process, in this paper, is considered to be used by mobile CPS nodes while estimating parameters and selecting threshold values on the fly for making informed decision whether the message should be discarded or propagated in the mobile CPS network.

The rest of the paper is organized as follows. Section 2 presents a system model. Section 3 presents a formal analysis for information diffusion. Section 4 presents performance evaluation results. Finally, Section 5 concludes the paper.

[*] CPS node and vehicle represent the same in this paper and will be used interchangeably throughout the paper.

2. SYSTEM MODEL

A typical system model for mobile CPS is depicted in Figure 1. Consider $G = (V, E, w)$ as a network of mobile CPS nodes, where V is the set of mobile CPS nodes, E is the set of edges connecting the CPS nodes, and $w(u, v)$ is the non-negative weights of the directed edges (u, v) between CPS nodes. For any CPS node $v \in V$, $\eta^{in}_{(v)}$ is the set of incoming links from neighbors and $\eta^{out}_{(v)}$ is the set of outgoing neighbor links to v. Let us consider that the A represents the set of seed nodes from where the information start to pervade over the mobile CPS network.

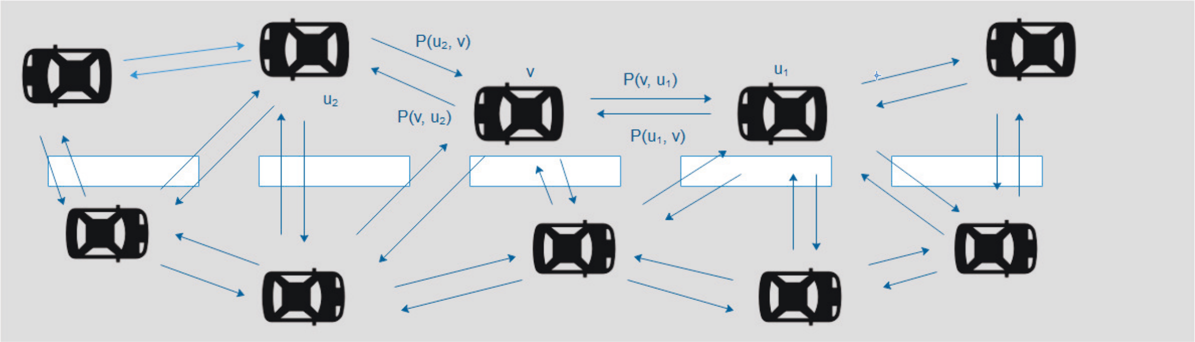

Figure 1. A typical mobile CPS model where u_ks are the neighboring nodes to the CPS node v with bidirectional communication links/edges and weights to the edges are represented as $P(u_k, v)$ or $P(v, u_k)$ values.

In this CPS network represented by the graph, we populate the weights, $w(u, v)$, of the edges as the probability of a CPS node u "trusting" the another CPS node v. In order to evaluate the probabilities, we consider an observation period over which the CPS nodes communicate with each other. While exchanging messages, certain CPS nodes might pass false information to their neighbors. By observing and keeping track of such events, we can evaluate the "trust" level of each CPS node relative to its neighboring reachable CPS nodes. Here $w(u, v)$ represents the probability of CPS node u passing true information to another CPS node v while $w(v, u)$ represents the other way round.

3. ANALYSIS

3.1 Observation Period and Edge Weight Calculation

Over a T period of time, a given CPS node observes the information exchange with other CPS nodes during which nodes might pass false information. Let $\mathcal{M}$ be set of messages where $\mathcal{M}_v \in \mathcal{M}$ which is a set of messages associated with CPS node v. $\mathcal{M}_{v,t}, \mathcal{M}_{v,r} \subseteq \mathcal{M}_v$ are the set of messages transmitted and received by the CPS node v, respectively. $m^v_{i,t} \in \mathcal{M}_{v,t}$ and $m^v_{i,r} \in \mathcal{M}_{v,r}$ are the messages of CPS node v such that $m^v_i \in \{0, 1\}$ where m^v_i is 0 for false information and 1 otherwise.

Then, the probability of a CPS node v passing true information to node u, $P(v, u)$, can be calculated as

$$P(v, u) = \frac{\sum\limits_{m \in (\mathcal{M}_{v,t} \cap \mathcal{M}_{u,r})} m^v_{i,t}}{|\mathcal{M}_{v,t} \cap \mathcal{M}_{u,r}|}. \tag{1}$$

where $|\mathcal{M}_{v,t} \cap \mathcal{M}_{u,r}|$ represents the cardinality of the set. Next, we calculate the probability of a CPS node v passing true information to any other CPS node $(P(v))$ by using the average true information the CPS node v passes over the given observation period as

$$P(v) = \frac{\sum\limits_{m \in \mathcal{M}_{v,t}} m^v_{i,t}}{|\mathcal{M}_{v,t}|} \tag{2}$$

Note that any CPS node $u \in V$ will use $P(v, u)$ given in (1) to make decision to accept messages from node v, as these two nodes have already established connections over the observation period. For new connections where the nodes are communicating for the first time with each other, they can use $P(v)$ given in (2) as an indicator to make decisions.

Formal algorithmic steps, based on above analysis for calculating weights of the edges in the graph/network of the CPS nodes, are presented as *Algorithm 1*.

Algorithm 1 Calculating the weights of the edges in the graph/network of the CPS nodes.

```
procedure EvaluatingWeightsOfEdges($G, \mathcal{M}$)
    for all $v \in V(G)$ do
        for $x \in \eta_{out}^{v}$ do
            Evaluate $P(v, x)$ using (1)
            $w(u, x) \leftarrow P(v, x)$
            $v_{trust} \leftarrow P(v)$ using (2)
        end for
    end for
    return $G$
end procedure
```

3.2 Information Diffusion

Information spread over the network of mobile CPS nodes can be represented in discrete steps starting from initial adopters based on a threshold value. In this paper, we present model that relies on cascades to observe and analyze the spread of any information over a mobile CPS network. As a CPS node (vehicle) communicates with its neighboring CPS nodes (vehicles), either the message will be accepted or dropped depending on the "trust" level of the transmitting CPS node (vehicle). We also assume that a vehicle might receive a message from multiple neighboring vehicles where the receiving vehicle have to make the decision (to accept or drop) based on the votes of trustworthy neighbors. After evaluating the weights to the edges of the graph, $G = (V, E, w)$, we can use the cascade approach to either minimize the influence of false information/message or maximize the true information/message within the network based on the weights of the edges.

Let $A_v : 2^{\eta^{in}(v)} \longrightarrow [0, 1]$ be the activation function, such that a CPS node v is influenced (i.e. accepts a particular message received by one or more neighboring CPS nodes) only if $A_v(S) \geq \theta$, where $S \subseteq \eta^{in}_{(v)}$ is the set of mobile CPS nodes transmitting a particular message to node v, and θ is the threshold score to trust a message. First, we write the activation function to influence node v by its neighbors as follow

$$A_v(S) = \frac{\displaystyle\sum_{u \in S \subseteq \eta^{in}_{(v)}} \Gamma(P(u, v))}{|S|} \geq \theta \tag{3}$$

where the $\Gamma(x)$ is given as

$$\Gamma(x) = \begin{cases} 1 & \text{if } x \geq \alpha \\ 0 & \text{if } x < \alpha. \end{cases}$$

In this approach, we can observe the diffusion of information over the mobile CPS network. Once the weights are estimated, we do not categorize any message as a lie or truth, rather let the CPS nodes/vehicles make their decision based on who is communicating the message. Thus, the spread of any message should depend on tolerance value (α) whether to "trust" any vehicle. Here, α is a threshold value of reputation/trust level of a communicating node, depending on which a receiving node consider the communicating node for voting. In the equation (3), $x, \alpha \in [0, 1]$, and the choice of α depends on the sensitivity of a message. In a mobile CPS (or vehicular) network, there can be different priorities to messages according to their (urgency in terms of time or) significance.[10] While α checks the sensitivity of a message, θ decides the number of communicating vehicles in the neighbors to trust in order to accept their messages. The spread of a message depends on these parameters within the network.

3.3 Greedy Approach for Positive (Negative) Influence Maximization (Minimization)

In order to observe the maximum spread, $\sigma(A)$, of any particular information for a given initial adopters (A), we use a greedy approach to approximate the spread. In the network, $G = (V, E, w)$, we are looking for a seed set $A \subset V$ where $|A| \leq k$ and $k \geq 0$ such that $\sigma(A)$ is maximized. $\sigma(A)$ is the estimated number of nodes that accept a particular message spread by the initial seed nodes. $\sigma(.)$ is the global influence function that estimates the number nodes influenced by initial adopters after the influences is saturated within the network. It is unclear how to exactly estimate $\sigma(.)$ in polynomial time but it has been shown that $\sigma(A)$ is generally NP-complete as shown in[11] in a different context. However, through Monte Carlo simulation, we could obtain arbitrarily good approximation of $\sigma(.)$ in some polynomial time by using Monte Carlo simulations with random choices and diffusion process.

Therefore, we use a greedy approach to find the initial adopters by approximating the maximum marginal gain $\sigma(A \cup \{u\}) - \sigma(A)$ and updating set of seeds, A, starting with an empty set of seeds.

Algorithm 2 Finding the seed set with maximum (or minimum) influence in the mobile CPS network.

```
procedure FindMinimumSeedSet($G, k$)
    Initialize $A = \emptyset$
    while $|A| \leq k$ do
        For each $v \in V$, approximate $\sigma(A \cup v)$ by using repeated sampling
        Pick $v$ where $\sigma(A \cup \{u\}) - \sigma(A)$ is maximum
        $A \leftarrow A \cup \{v\}$
    end while
    Get the set $A$ of CPS nodes for positive influence maximization or
    Get the set $V - A$ of CPS nodes for negative influence minimization.
end procedure
```

4. PERFORMANCE EVALUATION

In order to corroborate our analysis in previous sections, we simulated different scenarios using Monte Carlo simulations. We created a graph of CPS nodes of size between 0 and 500 with random edge weights to represent the randomness of trustworthiness of nodes. The graphs were generated using the NetworkX Python Library.[12]

We plotted the variation of active nodes vs the initial seeds for different values of trust threshold for the message α and normalized number of CPS nodes threshold $\theta = 0.2$ as shown in Figure 2 and different values of α and $\theta = 0.9$ as shown in Figure 3. The maximum number of active nodes influenced by the initial adopters/seeds depended on the choice of θ and α. As mentioned, θ is the measure of trusting the number of neighboring CPS nodes/vehicles communicating to a particular CPS node/vehicle receiving a message from them. With stricter (higher) value of θ, a vehicle needs more votes on a message transmitted by the neighbors with at least α to make a decision about it. α measures the sensitivity of a message; with higher value of α, the receiving node v can only trust a neighbor u with edge weight, $w(u, v)$, of at least specified value of α. Both of these parameters constrain the information diffusion with their higher values, which we found with the help of results using Monte Carlo simulations. Furthermore, the results suggest that information can be contained to a limited number of nodes with higher values of θ and α within the network, as shown in Figures 2 and 3. We consider these parameters for the performance evaluation because they represent important conditions under which the initial adopters or seed CPS nodes are trying to have a maximum influence over the CPS network. The choice of these parameters can be crucial to propagate a particular message over the network as they create restrictions. These parameters can be used as a check mechanism to constrain the influence of seeds while spreading a message. With lower value of α, any node with lower trustworthy connection can be trusted for any message, while with higher value of α, a node is considering only highly trusted connections to receive messages. Similarly, with a lower value of θ, a node is considering only a small fraction of trustworthy connection to confirm a message while with higher value, it requires a larger fraction to confirm a message. Thus, when a node is required to make a decision on the message, it can use an appropriate value of α and θ; the value of α depending on the priority of the message and the value of θ depending on how many votes the receiving node want to consider to confirm the message (which might depend on the nature of the mobile CPS). With these results, we observed that having a model of weighing the trust of neighboring vehicles in communication help create a secure network to propagate information in mobile CPS. In other words, this helps to minimize the negative influence caused by false information propagation in mobile CPS. The higher values of θ ensures that vehicles can accept a message only if multiple nodes are confirming the same message considering the nodes are trustworthy enough which is ensured by the value of α. Therefore, a misinformation originating from a node with a reputation of a liar is more likely to dampen to a limited number of nodes in the network. But, the reputation is a relative notion in the network as a node might be liar to one of its neighbors and a trustworthy communicator to the other, therefore the voting system is there to mitigate such situation where a lie is trying to get propagated through a trustworthy connection.

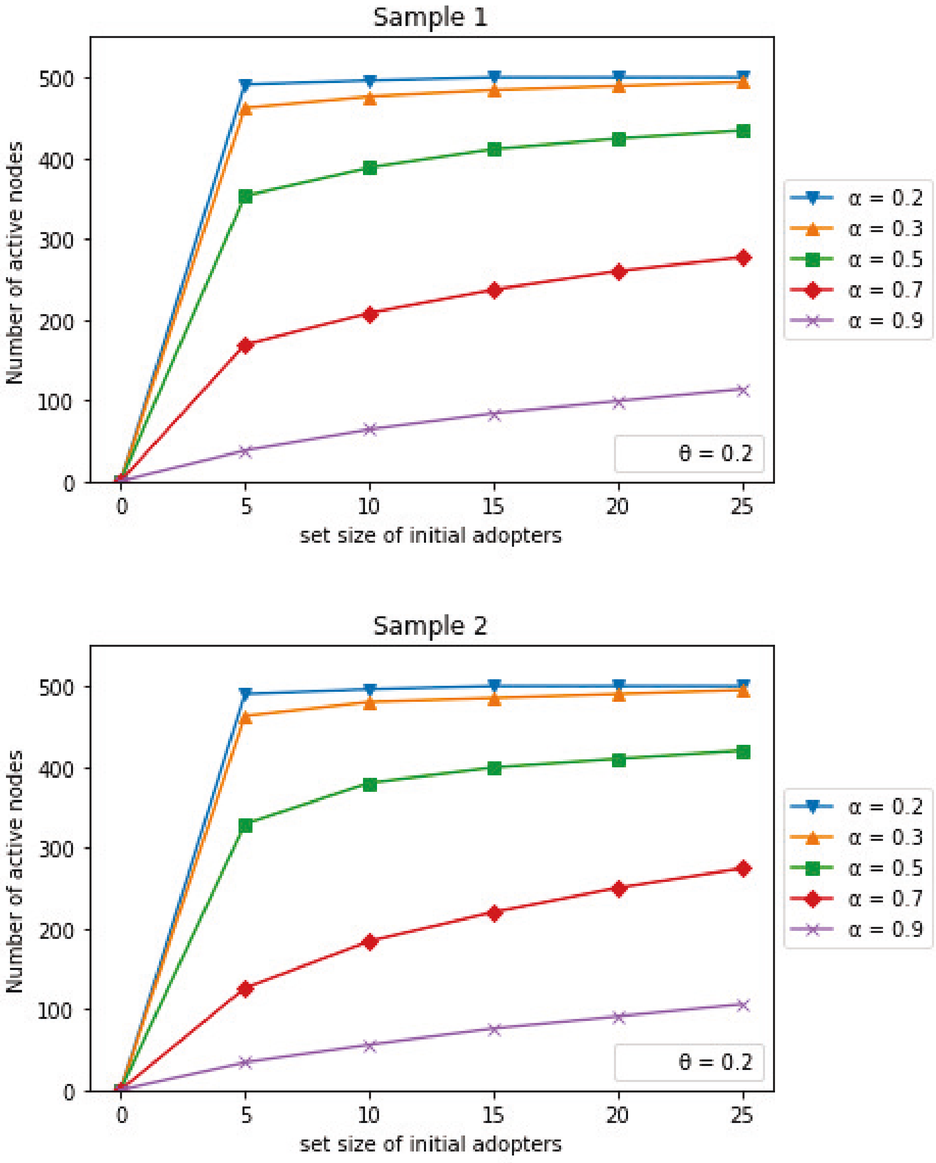

Figure 2. Variation of the number of active nodes for $\theta = 0.2$. vs the initial seed nodes with $\theta = 0.2$ and different values of α. The number of active nodes is saturated to the total number of nodes even with the lowest number of seeds whereas the spread is contained to a fraction of total nodes with higher value of α for given lower value of $\theta = 0.2$.

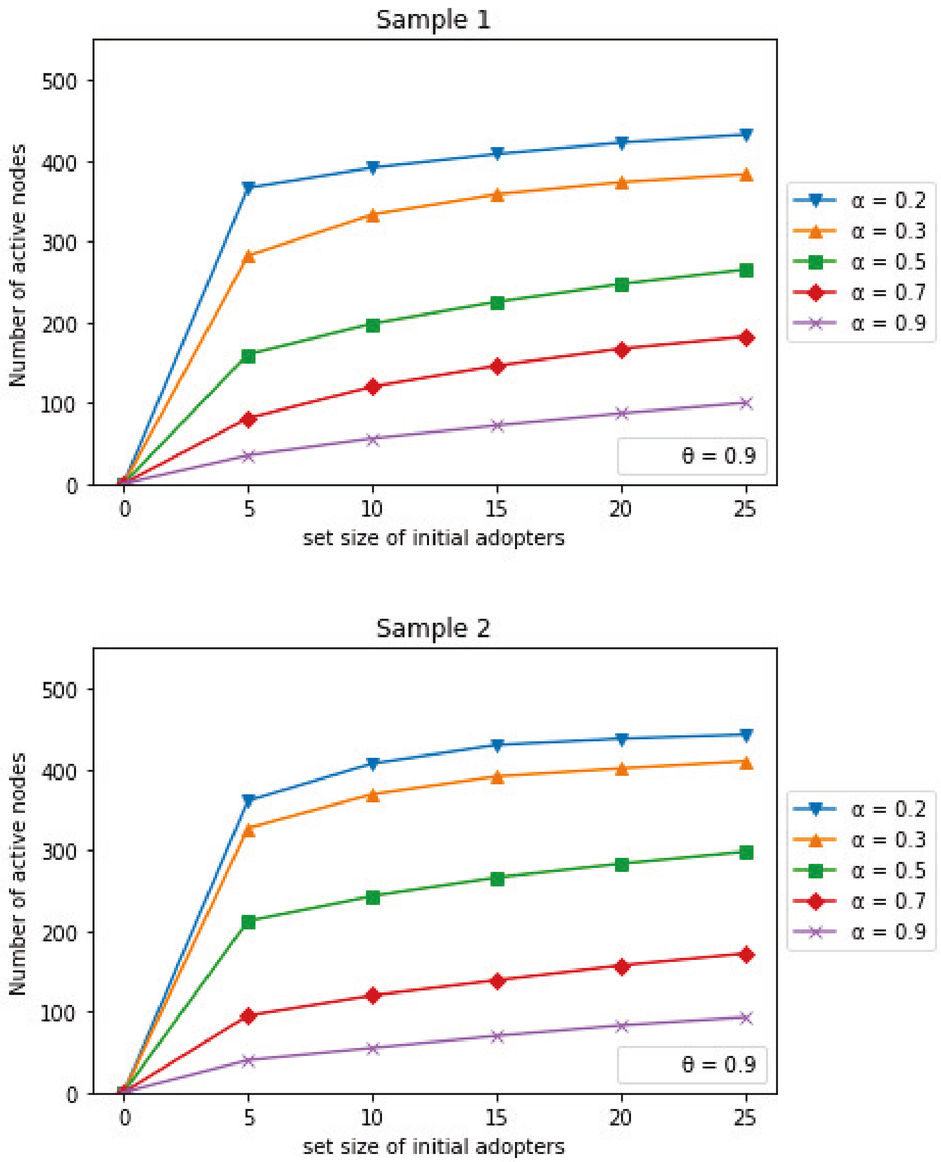

Figure 3. Variation of the number of active nodes for $\theta = 0.9$ vs the initial seed nodes with the maximum spread saturates to a lower value than the total number of nodes with stricter θ, i.e., $\theta = 0.9$. Even with the maximum number of seeds, while the trend of decreasing spread of false information is similar to previous experiment with increasing value of α.

Next, we plotted the variation of expected reputation value of the seed CPS nodes for different initial samples, as shown in Figure 4. The results are from 6 randomly generated graphs of size 500 CPS nodes. For each sample (graph of 500 CPS nodes), we chose 10 seed nodes that had influence over the mobile CPS network as initial adopters for given α and θ values. Figure 4 shows that, for higher α, say $\alpha = 0.9$ and lower θ, say $\theta = 0.2$, the expected reputation of initial adopter or seed CPS nodes is higher and the message was spread to a fraction of total CPS nodes due to the higher restriction. Furthermore, the lower values of α, say $\alpha = 0.2$ and lower value of θ, say $\theta = 0.2$, the reputation of initial adopter or seed CPS nodes is lower and the message was spread to a large number of CPS nodes due to the lower restriction but their reputation is lower compared to that of other combination of θ and α values, as shown in Figure 4.

Thus, the proposed approach can be an effective way to check misinformation or false information by leverag-

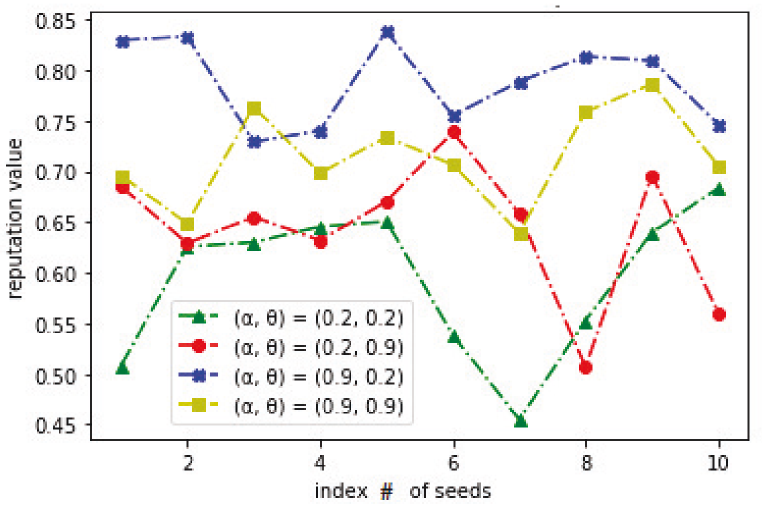

Figure 4. Variation of expected reputation value vs the seed CPS node # for different combination of θ and α values.

ing the estimated expected edge weights and appropriate choice of different parameters to limit the propagation of false information or misinformation in mobile CPS networks.

5. CONCLUSION

In this paper, we have presented an approach that helps to minimize the influence of misinformation (false information). The proposed approach limits the false message propagation in mobile CPS through a robust estimation of edge weights, and appropriate choice of parameters that measure not only the trustworthiness of CPS nodes on the fly but also the sensitivity of messages to be propagated over the CPS network. Simulation results show that the proper choice of the parameters and dynamic adaptation of edge weights of the CPS nodes on the fly help minimize the influence of the negative messages in mobile CPS.

Acknowledgment

This work is partly supported by the U.S. National Science Foundation (NSF) under grants CNS 1650831 and HRD 1828811. However, any opinion, finding, and conclusions or recommendations expressed in this document are those of the authors and should not be interpreted as necessarily representing the official policies, either expressed or implied, of the NSF.

REFERENCES

[1] Rawat, D. B. and Bajracharya, C., [*Vehicular Cyber Physical Systems: Adaptive Connectivity and Security*], Springer (2016).

[2] Tseng, Y.-C., Ni, S.-Y., Chen, Y.-S., and Sheu, J.-P., [*The Broadcast Storm Problem in a Mobile Ad Hoc Network*], Wirel. Netw. 8(2/3):153-167 (2002).

[3] Rawat, D. B., Yan, G., Bista, B. B., and Weigle, M. C., "Trust On the Security of Wireless Vehicular Ad-hoc Networking," *Ad Hoc & Sensor Wireless Networks Journal* **24**(3-4), 283–305 (2015).

[4] Chowdhury, N., Mackenzie, L., and Perkins, C., [*Requirement Analysis for Building Practical Accident Warning Systems Based on Vehicular Ad-Hoc Networks*], Wireless On-Demand Network Systems and Services (WONS), 2014 11th Annual Conference on, pages 81 (2014).

[5] Chen, P.-Y., Liu, J.-W., and Chen, W.-T., [*A fuel-saving and pollution reducing dynamic taxi-sharing protocol in vanets*], Vehicular Technology Conference Fall (VTC 2010-Fall), 2010 IEEE 72nd, pages 1 (5, Sept 2010).

[6] Pierce, A., “crash survival systems + car-to-car communication = a next gen crash avoidance system,” 8–9, Technology Today (2011).

[7] Yao, Q., Shi, R., Zhou, C., Wang, P., and Guo, L., “Topic-aware social influence minimization,” in [*Proceedings of the 24th ACM International Conference on World Wide Web*], 139–140 (2015).

[8] Wang, B., Chen, G., Fu, L., Song, L., and Wang, X., “Drimux: Dynamic rumor influence minimization with user experience in social networks,” *IEEE Transactions on Knowledge and Data Engineering* **29**(10), 2168–2181 (2017).

[9] Kempe, D., Kleinberg, J., and Tardos, E., “Maximizing the spread of influence through a social network. in proc. of the ninth acm sigkdd international conference on knowledge discovery and data mining,” ACM.

[10] Rawat, D. B., Popescu, D. C., Yan, G., and Olariu, S., “Enhancing VANET performance by joint adaptation of transmission power and contention window size,” *IEEE Transactions on Parallel and Distributed Systems* **22**(9), 1528–1535 (2011).

[11] Wang, C., Chen, W., and Wang, Y., [*Scalable influence maximization for independent cascade model in large-scale social networks*], Data Mining and Knowledge Discovery Journal (2012).

[12] [*NetworkX Python Library*], https://networkx.github.io/.