Diffusion Models for Information Dissemination Dynamics in Wireless Complex Communication Networks

SHIN-MING CHENG
National Taiwan University of Science and Technology, Taipei, 106, Taiwan.

VASILEIOS KARYOTIS∗
School of Electrical and Computer Engineering, National Technical University of Athens (NTUA), Zografou, Greece.

∗Corresponding author: vassilis@netmode.ntua.gr

PIN-YU CHEN
Department of EECS, University of Michigan, Ann Arbor, MI, 48109, USA.

KWANG-CHENG CHEN
Graduate Institute of Communication Engineering, National Taiwan University, Taipei 106, Taiwan.

AND

SYMEON PAPAVASSILIOU
School of Electrical and Computer Engineering, National Technical University of Athens (NTUA), Zografou, Greece.

[Received on 1 May 2013]

Information dissemination has become one of the most important services of communication networks. Modeling the diffusion of information through such networks is crucial for our modern information societies. In this work, novel models, segregating between useful and malicious types of information are introduced, in order to better study Information Dissemination Dynamics (IDD) in wireless complex communication networks, and eventually allow taking into account special network features in IDD. According to the proposed models, and inspired from epidemiology, we investigate the IDD in various complex network types through the use of the Susceptible-Infected (SI) paradigm for useful-information dissemination and the Susceptible-Infected-Susceptible (SIS) paradigm for malicious-information spreading. We provide analysis and simulation results for both types of diffused information, in order to identify performance and robustness potentials for each dissemination process with respect to the characteristics of the underlying complex networking infrastructures. We demonstrate that the proposed approach can generically characterize IDD in wireless complex networks and reveal salient features of dissemination dynamics in each network type, which could eventually aid in the design of more advanced, robust and efficient networks and services.

Keywords: Complex networks, Information diffusion, Epidemic models, Wireless networks, Desired and malicious information

1. Introduction

Information dissemination is a key social process in modern information-centric societies and most of the communication infrastructures have been developed in the last thirty years mainly to allow transfer-
ring diverse types of information. Different information types range in scope (e.g. academic, educational, financial and military), criticality (e.g. confidential, sensitive, public information, etc.) and value (e.g. useful, harmful, indifferent).

Recent advances in networking have been stimulated in order to accommodate emerging trends of increasing volumes and service demands of disseminated information. In general, information may be distinguished in three types, characterized by useful, malicious or indifferent content. The first may consist of news, multimedia, or financial data. People are willing to accept such information, and usually such data is stored for further use, e.g. e-books. Consequently, it is obtained once and a user experiences a single transition from the state of not having it to the state of having received the information. On the other hand, users may get malicious information, such as malware, which however are in principle reluctant to accept and/or use. Unfortunately, sources of malicious-information manage to devise new ways of spreading such data, e.g. through emails, viruses, trapped websites, so that malicious-content is characterized by recurrent behavior, which costs in time and money. In addition, the latter holds for indifferent types of information, such as spam email, etc. The user usually discards such information, but relevant messages are of recurrent nature, e.g. consisting of repeated or newer phishing messages.

In this work, we focus especially on information dissemination in wireless complex communication networks. Modern networks consist of various complex subnetworks [1], which merge as a heterogeneous large-scale network, possibly connecting to a wired infrastructure (Fig. 1). Contemporary users, equipped with devices having various radio interfaces can communicate with each other in more complicated ways than in the past. For example, a user may communicate with another user via mobile phone over cellular networks, while also communicating with a different user in his/her geographic vicinity via WiFi [2] or Bluetooth [3]. To evaluate the Information Dissemination Dynamics (IDD) in such heterogeneous complex communication networks, where communications are affected by both social relations and physical proximity, we establish analytical models with parameters representing IDD for different types of underlying communication networks and exploit them to study the dynamics of IDD, and obtain means of IDD control.

Substantial works [4–9] have employed an analogy of IDD in a single network as infectious spreading diseases by using ordinary differential equation (ODE) in the field of epidemics [10, 11], which could act as a quick reference to efficiently gather approximate knowledge of information dissemination speed and status with various settings of average node degrees and attain further control [12] in those networks. However, IDD are further complicated in wireless complex communication networks due to the presence of dynamic behaviors regarding the variability of topology and user behavior. Our work serves as one of the early attempts to systematically analyze IDD in a generic manner and allow assessing the dissemination and robustness performance in both current and future wireless complex networks.

The main contribution of our work is in realizing that different dynamics are governing the propagation of different types of information and provide appropriate models for each case. Specifically, with regard to useful-data, the objective is to spread such information to as many users as possible, so that useful-information reaches potentially all nodes of a network. On the contrary, in the event of malicious/indifferent-information spreading, the objective is to study the robustness of various network types against harmful or indifferent, but nevertheless network-stressing content. To address the above, in this paper we propose two different paradigms that describe the behavior of systems, a Susceptible-Infected (SI) for the first and Susceptible-Infected-Susceptible (SIS) for the second and third types of information. Both paradigms were inspired by drawing analogies to the field of epidemics [10]. We then propose and analyze two specific approaches for obtaining analytically the behavior of each case and
demonstrate how they can be used in order to identify the parameters and properties of the underlying wireless complex communication networks that govern IDD.

The rest of this paper is organized as follows. In section 2, we present the proposed framework for IDD and discuss the analogy to epidemic models for describing such dynamic operation, while in section 3 we describe the complex networks we will consider and the employed assessment metrics. In section 4, an analytical approach for describing useful-information dissemination is presented and evaluated, while in section 5 the case of malicious/indifferent-information spreading is analyzed and evaluated. Finally, section 6 summarizes the contribution of the paper and discusses emerging trends.

2. Epidemic-based IDD Approach for Wireless Complex Communication Networks

Considerably different dynamics govern the propagation of useful and malicious/indifferent types of information. Regarding the spreading process, the dissemination of useful-information resembles that of an epidemic disease, in which population members that get infected by a virus, permanently transit to immunization (after recovery), or termination (if the infection is lethal). In epidemics, such models are readily referred to as Susceptible-Infected (SI) [10] or Susceptible-Removed (SR) [13]. Such behavior is effectively described by a two-state model, where nodes are initially prone to receive epidemics (Susceptible) and then permanently transit to the Infected state, once they receive the epidemic (SI model). The SR model essentially expresses the same behavior in which case nodes are removed from the network in cases of lethal epidemics. The SI (SR) model captures the characteristic behavior, where for each node, only a single and permanent transition takes place from the Susceptible to the Infected (Removed) state.
Recognizing the similarities between IDD and the spread of infectious diseases, ODE models in the field of epidemic [10, 11] are widely adopted as tools to analyze IDD in communication networks [4–9]. Since ODE models provide closed-form formulas for the performance metrics of IDD, a wide range of effects can be encompassed by aggregating individuals in the network into two states (i.e., infected and susceptible), and thus reducing computation time and required resources. In contrast, both agent-based emulation model [3] and simulation experiments [14] try to precisely capture attributes of individuals in the network and the interactions among them via massive experiments, thus the appeals of analytical tractability are neglected. Obviously, the complexity of modeling individual-level details significantly increases with the number of individuals, and thus the ODE model is suitable to act as a quick tool to identify IDD in large complex networks.

The dissemination of useful information resembles the SI process in epidemics, since a rational user waits to receive desired or useful-information and then stores it for further use. If the information is indeed useful and valid, no further transaction for this information will be required/take place. Thus, the user experiences a single state transition when (s)he receives useful-information from the non-informed state (node can potentially receive data, i.e. susceptible) to the informed state (i.e. infected).

This is not the case however in malicious/indifferent-information dissemination. Regarding malicious-information spreading, users are reluctant to retain the malicious-information they receive. However, malicious-information might return (if the user is not properly protected) or adversaries are capable of devising a new malware type or package for the malicious information. With respect to indifferent-information, even tough it is usually of no interest to users, it might further load an already stressed infrastructure, and especially in the case of spam, its repetitive nature might eventually become disturbing, similarly to malware. Consequently, one may consider indifferent-information as a special case of malware with respect to the users’ macroscopic behavior. In our treatment, we will employ this observation, due to the fact that we will focus on modeling the macroscopic behavior of IDD. In the following, we will use the terms malicious, indifferent-information interchangeably.

In principle, a user is able to recover (i.e. dispose malicious-information) and return to the previous state, where it is possible to become infected again. This behavior resembles epidemics, where individuals become infected, then recover from a disease, and become susceptible again to a different or the same disease (the latter if they do not properly “vaccinate”). Such model is denoted by Susceptible-Infected-Susceptible (SIS) in epidemics [10]. Contrary to useful-information, we notice that the SIS model defines two potential transitions between the two possible states of a user, namely switch from susceptible to infected and then switch back to the susceptible state.

In summary, in this paper we introduce two analogies between epidemics and information dissemination. We employ the SI model to describe the macroscopic behavior of useful IDD and the SIS model for malicious (indifferent) IDD. In sections 4 and 5 we analyze in detail the proposed IDD models and obtain both quantitative and qualitative results for IDD in various types of complex communication networks, eventually drawing useful observation for each network type that can be useful for future designs and implementations.

3. Wireless Complex Networks and Assessment Metrics

In Network Science (complex network theory) [15, 16], a network represents a system of interactions and can be modeled as a graph \( G(V, E) \) consisting of a set \( V \) of nodes (i.e. wireless terminals or Internet users) and a set \( E \) of undirected edges (i.e. physical channels). The number of nodes is denoted by \( N \). Without loss of generality, and in order to focus on the IDD rather than on the wireless propagation details, two nodes connected by a link are called neighbors and the number of neighbors for a node is
defined as the degree $k$. We define the degree distribution $P(k)$, as the probability of having $k$ channels for a node and $\bar{k}$ as the mean value of $k$. We assume that there is at most one edge between any node pair and no self loops in $G(V,E)$.

As it will be shown in the sequel, by employing the aforementioned analogy to epidemics, IDD can be effectively described and mathematically analyzed for different types of complex communication networks and their topologies. Each network type is characterized by different topological features and the proposed approach allows in both cases of useful and malicious-information identifying the impact of each topology on the IDD performance and robustness. The evaluation of these models will be based on the following critical parameters involved:

- **Degree distribution**: It provides a suitable representation of the structure of a network, especially for social ones. Based on the degree distribution, a network is of homogeneous mixing if the degree distribution of each node is centered around $\bar{k}$ and the degree variance $\sigma_k ^2 \leq \epsilon$. Otherwise, it is of heterogeneous mixing.

- **Link connectivity**: The finite transmission range of a wireless node determines its neighborhood. Moreover, wireless channel quality affects the success of transmissions. The transmission range and channel quality jointly affect link connectivity. Thus, for a network with variable connectivity, the network is (partially) connectible, if each node has a positive probability to establish an undirected connection to (partial but not all) any other node in the network. We further define a network with (different) same connection probability in each node as (un)equally connectible.

Additional properties may be identified, capturing topological properties of the underlying complex networking infrastructures, and could be exploited for analyzing IDD in such networks. The clustering coefficient (indicative of the clusters building up due to social or other types of interaction) and the average path length between randomly selected node pairs are appropriate quantities [1]. Based on the specific metrics of interest, the complex networks of interest can be classified in five main categories, cumulatively depicted in Table 1.

- **Homogeneous Mixing and Partially Connectible (HoMPC)**: In a (regular) lattice with degree $k$ every node connects to its $k$ nearest neighbors [17]. A lattice is a HoMPC network since the degree distribution is a delta function with magnitude equal to 1 located at $k$. Another example is wireless sensor networks, where sensors are uniformly and randomly distributed on a plane. In this case, for grid based sensor networks the number of neighbors in the communication radius of a sensor is fixed, which acts exactly as HoMPC.

- **Homogeneous Mixing and Equally Connectible (HoMEC)**: Erdős-Rényi (ER) random graphs [18] assume that an edge is present with probability $p_e$ for $N(N-1)/2$ possible edges. The degree distribution of an ER network in relatively large networks is

$$P(k) = \frac{(e^{-\bar{k}_e} \bar{k}_e^k)}{k!}, \quad (3.1)$$

where $\bar{k}_e = Np_e$. It is observed that the degree distribution is centered around $\bar{k}_e$ and a node has equal probability $p_e$ connecting to any other node in the network. Thus, an ER network is a HoMEC network. Some complex wireless networks can be effectively modeled by ER graphs, especially when considering the joint cyber-physical system forming from a social network overlaying a wireless one [19].
Table 1. Complex network classification

<table>
<thead>
<tr>
<th>Connectivity</th>
<th>Mixing type</th>
<th>Partially connectible</th>
<th>Equally connectible</th>
<th>Unequally connectible</th>
</tr>
</thead>
<tbody>
<tr>
<td>Homogeneous</td>
<td>Lattice network</td>
<td>wireles sensor network</td>
<td>ER network</td>
<td>Small-world network</td>
</tr>
<tr>
<td>Heterogeneous</td>
<td>Machine-to-machine network</td>
<td>wireless mesh, Smart Grid</td>
<td></td>
<td>Scale-free network</td>
</tr>
</tbody>
</table>

- **Homogeneous Mixing and Unequally Connectible (HoMUC):** Small-world networks generated by the Watts-Strogatz (WS) model [20] are constructed from a regular ring lattice with $2j$ edges for each node, and randomly rewiring each edge of the lattice with probability $p_w$ such that self-connections and duplicate edges are excluded. The degree distribution of a small-world network is

$$
P(k) = \begin{cases} 
\sum_{n=0}^{\min(k-j, j)} \binom{j}{n} (1-p_w)^n p_w^{k-j-n} \left(\frac{p_w}{k-j-n}\right)! e^{-p_w}, & \text{for } k \geq j, \\
0, & \text{otherwise}.
\end{cases}
$$

Since the degree distribution has a pronounced peak at $k_w = 2j$, it decays exponentially for large $k$ [1], and the connection probability is unequal except $p_w = 1$ (extreme randomness), we regard small-world as HoMUC network. This implies that as $p_w \to 1$, a small-world network becomes HoMEC. Many social network models are extensions of the WS model, since the HoMUC property explains well their clustering features [1]. Small-world topologies emerge very often in wireless complex networks, especially in cyber-physical systems with social networks overlaying the physical ones [19].

- **Heterogeneous Mixing and Partially Connectible (HeMPC):** Machine-to-Machine (M2M) networks, wireless mesh networks, or Smart Grid, have potentially non-structured degree distributions due to finite transmission range, heterogeneity in location, mobility, channel quality variations and time-varying user behavior, eventually classifying the underlying topology as HeMPC.

- **Heterogeneous Mixing and Unequally Connectible (HeMUC):** A power-law distributed network having degree distribution $P(k) \sim k^{-r}$, with $r$ in the range $2 \leq r \leq 3$, is also called a scale-free network. Barabási and Albert observe two essential factors of scale-free network: growth and preferential attachment [21]. In this model, denoted by BA, every new node connects its $m$ edges to existing nodes according to the preferential attachment rule, and $k_h = 2m$. The power-law degree distribution suggests that most nodes have few neighbors, while some “super nodes” (hubs) tend to have a great amount of neighbors. Thus, the power-law distributed network is a HeMUC network.

Table 1 summarizes the aforementioned complex network categories. Note that the size of the largest connected component (“giant component”) is also an important measure of the effectiveness of the network at information epidemics or cooperations. Regarding this, a network is saturated if the giant component size approaches the total number of nodes in the network. Otherwise, nodes may be isolated from the giant component (non-saturated network). For instance, a giant component emerges almost surely in random graphs, if $\sum k(k-2)P(k) > 0$ [22].
4. Useful-Information Dissemination Epidemic Modeling

As we explained before, the macroscopic behavior of useful-information dissemination can be described by the SI epidemic model, in which, nodes receive the designated data once-in-lifetime. In this section, we provide an analytical approach for quantifying the process of useful-information dissemination in various types of complex communication networks.

4.1 SI Epidemic Spreading Model

We adopt the SI model \([10]\) in order to describe the IDD of useful-information. In the analogy we draw between IDD and epidemics, an uninfected node in epidemics corresponds to a non-informed node in IDD, i.e. a node not having received yet useful-information. On the other hand, an infected node corresponds to a node that has received and stored useful-information. Such model is appropriate for describing a specific type of useful-information spreading, e.g. handbook disseminated to a population, similarly to the SI model describing the dissemination of a single viral threat. In the sequel we will use the terms non-infected and non-informed, as well as the infected and informed terms interchangeably. Assume that \(N\) nodes are initially all susceptible non-informed except for a small number that are infected and contagious (denoted as infectious). These ‘contagious’ nodes represent the nodes that generate the useful content to be disseminated. Thus, by the previous analogy, in this scenario, all nodes would like to eventually become infected (i.e. informed).

We adopt infection parameter \(\lambda\) to characterize the rate of spreading between S-I pairs. Considering the volatile nature of wireless channels and MAC, information transmission over a communication link may not be successful. The availability (or the successful transmission rate) of a link in wireless communication channels is typically modeled as a two-state Markov chain with on and off states \([23]\), which is sufficient for the purposes of our study that does not focus on the impact of the channel propagation effects. For a dynamic topology the set \(V\) of nodes is time-invariant, the set \(E\) of edges varies with time, while however, the network maintains its basic features. The informed nodes adopt a “consistent broadcast” behavior, so that they tend to transmit the useful-information to susceptible nodes in contact consistently, just as in the spread of biological viruses. Hence, the spreading rate (similar to infection rate) \(\lambda\) is equivalent to the probability of an available channel (‘on’ channel state), which is independently determined for all channels such that the long time behavior of channel availability is equal to \(\lambda\).

Expressed mathematically, if \(I(t)\) and \(S(t)\) are respectively the fraction of infected (informed) and susceptible (non-informed) nodes at time \(t\), and we have

\[
\begin{align*}
I(t) + S(t) &= 1; \\
\frac{dI(t)}{dt} &= \lambda \bar{k} I(t) S(t).
\end{align*}
\]

Then the simple analytic solution obtained is:

\[
I(t) = \frac{I(0)}{I(0) + (1 - I(0))e^{-\bar{k}t}}
\]

(4.2)

From (4.2) it is clear that the informed density \(I(t)\) approaches 1 as time evolves, as desired. In this study, we assume initially there is only one informed node, i.e. \(I(0) = 1/N\). Note that the saturation of any complex network should serve as one of the important sufficient conditions when adopting the SI model for IDD, since the first order differential equation (4.2) fails to distinguish the saturation of
Fig. 2. IDD in regular lattice (HoMPC; \( p_w = 0 \)), ER (HoMEC; \( p_w = 1 \)), scale-free (HeMUC), and small-world (HoMUC) networks for different \( p_w \) with mean degree equal to 10, \( \lambda = 0.01 \) and \( N = 2500 \).

a network. In the sequel, we demonstrate the proposed SI framework for static and dynamic saturated complex networks.

4.2 Static Complex Networks

The static network is regarded as a time-invariant graph \( G(V,E) \) where the topology is unchanged in time. Given that, in Fig. 2, numerical results in regular lattice (HoMPC), ER (HoMEC), small-world/WS (HoMUC), and scale-free/BA (HeMUC) networks are presented, based on ensemble averages, obtained by 100 simulations in saturated complex networks. Since an ER network can be totally characterized by parameter \( N \) and \( p_e \), IDD in ER is described through the SI model by adopting \( \bar{k}_e = Np_e \). However, compared to IDD in ER networks, the SI model amplifies the cumulative informed node fraction with time as it implicitly assumes the HoMEC property, resulting in inaccurate estimation. In subsection 4.2.1 we will address this discrepancy, in order to obtain a more accurate IDD SI-based model.

We also observe an interesting phenomenon for the IDD in scale-free networks generated by the BA model. The corresponding IDD curve exceeds that of the SI model in the beginning, but at some instant \( t_0 \), the curve of the SI model transcends. This is in accordance with the fact that the information has higher possibility to be transmitted to super nodes than the nodes with low degrees at early stages.
Once the information spreading begins, the number of informed nodes highly increases as super nodes eventually become informed. Then the spreading rate decreases as information dissemination is now mainly propagated to nodes with lower degrees. Thus, although the SI model is not suitable for describing the IDD in scale-free networks, it still provides useful insights for better understanding the IDD in networks with the HeMUC property. Note that as we adopt the fact that homogeneity holds for nodes of the same degree [13], an enhanced model can be derived to accurately match the IDD curves of HeMUC networks.

Regarding small-world networks, effects of different rewiring probabilities \( p_w \) are investigated (Fig. 2). Ranging from extreme regularity \( (p_w = 0) \) to extreme randomness \( (p_w = 1) \), IDD accelerates due to the enhancement of connectivity and the small-world network transforms from HoMPC network \( (p_w = 0) \) to HoMUC network \( (p_w \in (0,1)) \) and finally HoMEC network \( (p_w = 1) \).

4.2.1 Corrected SI Model

To overcome the above discussed discrepancies between the SI model and the proposed epidemic IDD model in complex networks, the emerging “degree correlation” problems need to be addressed. The traditional SI model implicitly assumes that each node is uncorrelated. However, when a node is informed in a static network, this suggests that at least one of its neighbors has been informed, and hence the mean degree has to be corrected accordingly. Specifically, the average number of susceptible neighbors of an infected node is less than \( \bar{k} \) and thus an effective mean degree \( \hat{k} \) is proposed to accommodate this phenomenon. This phenomenon should be carefully modeled otherwise the overestimation problem [6] leads to significant deviation. We take HoMEC network as an example, since it can be characterized by its mean degree \( \bar{k} \). The rate of informed nodes will be given by:

\[
\frac{dI(t)}{dt} = \lambda \hat{k} I(t) S(t) = \lambda (\bar{k} - f(t)) I(t) S(t),
\]

where \( f(t) \) is a time varying function accounting for the average number of informed neighbors of an informed node. By setting \( f(t) \) as a constant, a tight upper bound on IDD can be obtained when \( f(t) = 1 \) for HoMEC networks since intuitively, a node that is infected implies that at least one of its neighbors is infected. Thus, we have the following observations.

- **Observation 1**: For a saturated and HoMEC network given the informed rate \( \lambda \) and mean degree \( \bar{k} \) parameters, there exists a tight upper bound on IDD at any time instance.

By setting appropriate values of \( f(t) \) according to the features of different networks, the adjusted SI models successfully capture the corresponding IDD. This implies that upper bounds on IDD of HoMUC and HoMPC networks also exist. Take IDD in sensor network (HoMPC) as an example, where sensors are uniformly and randomly distributed on a plane with node density \( \sigma \), the communication radius of a sensor \( R \) and the radius of the circle containing the informed sensors \( r(t) \). Obviously, \( \sigma \pi r(t)^2 = N \cdot I(t) \). This is approximated into our model by having only the infected nodes that lie on the periphery of an informed circle can communicate with the susceptible nodes located at a distance of at most \( R \) outside the infection circle, and thus have the potential to inform (infect) them. In other words, the spatial broadcasting of the information is only contributed from the wavefronts of informed circles, while the infected nodes located in the interior of the informed circles are not engaged in further spatial dissemination. Thus, we could calculate \( f(t) \) as

\[
f(t) = \bar{k} \cdot \frac{\sigma \pi r(t) - R^2}{N I(t)} \approx \bar{k} \cdot (1 - c \sqrt{NI(t)})
\]
where \( c = 2\sqrt{\sigma \pi R} \) and \( \sigma \pi R^2 \) is usually negligible compared with \( NI(t) \). Applying \( f(t) \) to the corrected SI model, we thus obtain the same result derived in [8]. From this example, we have the following observation.

- **Observation 2**: For a saturated and homogeneous mixing network, the time \( T \) needed to inform a fraction \( s \) of nodes can be directly obtained from (4.3) as \( T = I^{-1}(s) \), if \( I^{-1}(\cdot) \) exists.

As the cumulative informed fraction approaches 1 in a saturated network, Observation 2 serves as a more accurate benchmark to any broadcasting mechanisms in wireless complex networks for evaluating the performance of information dissemination. This can be justified by the fact that Observation 2 greatly mitigates the biased estimation due to degree correlations.

### 4.3 Dynamic Complex Networks

This section discusses the dynamic case where network topology changes with time, while maintaining the basic structure and properties. As it will be shown, a time-varying topology provides great chances for information dissemination to the entire network and therefore is suitable for describing complicated interactions within large-scale networks with mobility support (e.g., routing protocols in MANET). Two nodes originally disconnected, might eventually establish a virtual link between them due to mobility, thus yielding a virtual giant component. Given the condition that a network is originally non-saturated (e.g., MANET in sparsely populated area), mobility may make the network virtually saturated, since all nodes eventually receive the information due to change of connectivity in the dynamic sense.

- **Observation 3**: A dynamic network is virtually saturated in the sense that the virtual giant component size approaches the number of nodes.

The mobility of nodes facilitates connectivity in homogeneous mixing networks originally not saturated and thus dynamic HoMEC, HoMPC, and HoMUC networks are virtually saturated. In the following, we focus on and analyze IDD in such networks where benefits from mobility are more obvious.

- **Observation 4**: The IDD of dynamic homogeneous mixing networks can be characterized by the corrected SI model.

Due to the virtual saturation property in Observation 3, the SI model failing to describe the IDD of non-saturated networks is suitable to characterize the IDD of dynamic homogeneous mixing networks. As we define \( f(t) \) according to the properties of the networks, the IDD curves can be accurately matched by using equation (4.3).

To show the significance of these observations and the flexibility of the proposed model, Fig. 3 illustrates analytic and simulation plots depicting the epidemic routing via broadcasting in large-scale MANETs. Epidemic routing aiming at exploring the advantages of path diversity with concrete analysis [7] has been regarded as one practical way to achieve routing in dynamic HoMPC [24, 25]. In the simulation experiments, we assume that each node has the same transmission range \( R \) with uniform random deployment in a \( 100 \times 100m^2 \) plane with wrap-around condition. According to a stationary and ergodic mobility model, such as the Truncated Lévy Walk model [26], we set the step length exponent \( s = 1.5 \) and pause time exponent \( \varphi = 1.38 \), which fit the trace-based data of human mobility pattern collected in UCSD and Dartmouth [27]. We incorporate the successful packet transmission rate to the spreading rate \( \lambda \). Fig. 3 shows that the MANET is non-saturated (however, virtually saturated), and our model captures the complicated interactions among numerous mobile nodes precisely.
4.4 Hybrid Complex Networks

When nodes are capable of communicating with each other using multiple heterogeneous connections, a hybrid complex network consisting of multiple complex sub-networks is built via heterogeneous links. As in the example we mentioned in Section 1, people could exploit mobile smart phones to communication with individuals in the address books via traditional phone-call and short message, as well as the individuals in geographic proximity via WiFi [2] or Bluetooth [3]. As shown in Fig. 4, the IDD in such complicated networks can be investigated by separately considering IDD in social network constructed by contacts and IDD via broadcasting and then aggregating the results for the combined cyber-physical system.

According to the proposed categories, we exploit ER network (HoMEC) and sensor network (HoMPC) to respectively model the delocalized and broadcasting dissemination patterns. Thus, the subpopulation function \( I(t) = I_e(t) + I_s(t) \), where \( I_e(t) \) and \( I_s(t) \) are those that have been disseminated via ER and sensor networks at time \( t \), respectively. The average degree \( \bar{k}_e \) describing the social relationships between handsets means the average number of contacts in the address book. According to (4.3), the basic differential equation that describes the dynamics of informed subpopulation is

\[
\frac{dI_e(t)}{dt} = \lambda \frac{S(t)(\bar{k}_e - 1)}{N} I(t).
\]  

(4.5)
When an informed node intends to disseminate via broadcasting, it first scans to search the nearby nodes within its transmission range $R$ and connects to the neighbor so as to determine the susceptible neighbors for propagation. In this case, the average number of neighbors $\bar{k}_s$ equals $\rho \pi R^2$. The behavior of such spontaneous spreading can be regarded as a ripple centered at the infected source node which grows with time. As shown in Fig. 4, the spatial spreading of the information here is only contributed from the wavefronts of informed circles, while the infected nodes located in the interior of the informed circles are not engaged in further spatial infections.

Without loss of generality, we assume that a single informed circle is generated at time $t_1$ by a point source infected through and kept stretching for $t_2$ time units. Then its incremental spatial infection at time $t_1 + t_2$ is

$$G'(t_1, t_2) \triangleq \frac{dG(t_1, t_2)}{dt_2} = \lambda \frac{S(t_1 + t_2) \cdot \bar{k}_s}{N} c \sqrt{G(t_1, t_2)}.$$

(4.6)
where $\frac{1}{2}k_e$ accounts for the fact that for an infected node on a periphery, roughly half of neighbors outside the infection circle are susceptible. The incremental spatial infection at time $t$ of all infection circles is given by

$$\frac{dI_s(t)}{dt} = \int_0^t I'_e(\tau)G'(\tau, t-\tau)d\tau.$$  \quad (4.7)

It means that there are $I'_e(\tau)d\tau$ point sources originated at time $\tau$ and each contributes $G'(\tau, t-\tau)$ incremental spatial infection at time $t$.

To validate the analytical model, we develop experiments to simulate IDD in a hybrid network among 2000 individuals uniformly deployed in a $50 \times 50$ plane. The constructions of social contact networks and setup of parameters (e.g., $\bar{k}_e = 6$) follow the data sheet in [14]. Fig. 5 illustrates analytic and simulation plots depicting the IDD via both delocalized communication and broadcasting in hybrid (HoMEC and HoMPC) complex networks. We observe that the curves of propagation dynamics closely match our analytical model, where limited discrepancy exists, mainly due to the fact that information may propagate to individuals who have already been informed and uncertain boundary conditions could not be considered in the analysis. Comparing with the traditional SI in Fig. 2, the corrected SI could capture the IDD of ER network (HoMEC) more precisely.
5. Malicious-Information Propagation Modeling

In this section, we adopt and extend an analytical model which is able to capture the behavior of malicious IDD (modeled as SIS epidemics) in wireless multi-hop networks. Contrary to the case of useful-information, regarding malicious/indifferent-information, one is interested in the robustness capabilities of the network to sustain such traffic, which in both cases, it is of no use (even harmful for malware). The proposed analytical model is based on queuing theory and we apply it on various types of complex wireless networks, shown cumulatively in Table 1.

5.1 SIS Closed Queuing Network Model

In the SIS paradigm, susceptible (non-informed) nodes essentially wait until the arrival of malicious-information, in which case they transition to the infected (informed) state. We consider a propagative network, where nodes spread further the malicious-information they receive. Consequently, a node might become infected from malicious software either from an attacker or an already infected legitimate node. This holds for several types of viruses and worms that have appeared [28, 29]. Infections are assumed to arrive in a non-deterministic fashion. The recovery process (disposing malicious software) is of similar nature, but not necessarily of the same waiting behavior. Throughout this work, following current literature [30], we assume that the infection arrival process is Poisson, while the recovery process of each network node is exponentially distributed. We denote by \( \lambda_i \) the infection arrival rate in link \( i \) and by \( \mu_j \) the recovery rate of node \( j \).

Legitimate nodes can be separated at any instant in two subsets, i.e. infected and susceptible. Following the aforementioned modeling approach, the operation of the system can be mapped to a closed two-queue packet network, as shown in Fig. 6 [31], where \( N \) packets, i.e. network nodes, circulate. This closed feedback system of queues does not model the actual packets exchanged in a network, but rather the transition of nodes themselves from one state to the other and back (SIS behavior), as desired. For simplification purposes and in order to visually consider the behavior of legitimate nodes as they become infected and recover, one may map each node to an arbitrary packet/customer (not the packets of the specific IDD network, but the packets of a hypothetical one with two queues in a closed loop fashion) circulating in this queuing system and thus properly map queuing terminology to IDD concepts.

At any instant, if \( i \) nodes are infected, then \( N - i \) are susceptible. Both service rates are state-dependent according to the number of packets (user nodes) that exist in the corresponding queue at each time instance. Explicit definition of each queue’s equivalent service rate, i.e. \( \lambda(N - i + 1) \) and \( \mu(i) \), depends on the underlying complex network and employed infection paradigm. Without loss of generality, we assume that the lower queue represents the group of infected legitimate nodes and denote it as “infected”, while the upper queue represents the susceptible nodes and we denote the queue as “susceptible”. The state-dependent service rate of the susceptible queue can be extended to the case of multiple malicious-information sources (i.e. attackers), in which case the service rate will have the form \( \lambda(N - i + M) \), \( M \) being the number of attackers. Our analysis here is focused on the case of a single attacker (i.e. \( M = 1 \)).

Standard approaches from queuing theory may be employed to analyze the two queue closed network. The focus is on the infected queue. Its steady state distribution, denoted by \( \pi(i) \), represents the probability that there are \( i \) packets (nodes) in this queue. Using balance equations for the respective
Fig. 6. Closed queuing model for SIS malicious-information propagative networks.

Markov chain, the explicit expression for the steady state distribution can be obtained as:

$$\pi(i) = \pi(0) \cdot \prod_{j=1}^{i} \frac{\lambda(N+1-i+j)}{\mu(j)}$$

(5.1)

where $\pi(0)$ is the probability of no infected nodes in the network. Applying the normalization condition $\sum_{i=0}^{N} \pi(i) = 1$, and appropriately specifying the total infection and recovery rates (where $\lambda_i = \lambda$ and $\mu_i = \mu \forall i \in \{1, 2, \ldots, N\}$), and by setting $\alpha^{-1} = \frac{\lambda \pi}{\mu} \left(\frac{R}{L}\right)^2$ and considering a large number of legitimate nodes, the probability of no infected nodes can be approximated as:

$$\pi(0) \approx \frac{\alpha^N}{N!} \cdot e^{-\alpha}$$

(5.2)

and the steady state distribution can be obtained as:

$$\pi(i) = \frac{\alpha^{N-i}}{(N-i)!} \cdot e^{-\alpha} = \pi'(N-i)$$

(5.3)

where $\pi'(N-i)$ is the steady state distribution for the non-infected queue. Using relation (5.3), the probability of a completely infected network equals $\pi(N) = e^{-\alpha}$. It is noted that the error introduced by the above approximation is negligible for values of $\alpha$ and $N$ commonly used in practice (in the order of less than $10^{-3}$).

Expression (5.3) clearly indicates the critical parameters that affect the behavior of the system. Assuming a fixed area of the network deployment region, the number of legitimate nodes (i.e. the density of the network) along with the common transmission radius and the ratio of the link infection rate to the node recovery rate are decisive factors regarding the overall behavior and stability of the system.

Based on such model, the expected number of infected (informed) nodes (corresponding to the expected number of packets in the lower queue) for different types of networks may be obtained. The general expression yielded is:

$$E[i] = \sum_{i=1}^{N} i \cdot \pi(i) = N - c$$

(5.4)

where $\pi(i)$ the steady state distribution of the underlying Markov chain, and $c = \frac{\mu \pi}{\lambda}$ for a HeMPC network (wireless multi-hop) over a square deployment region of side $L$ and each node having a transmission radius $R$, and $c = \frac{\mu N}{\lambda \bar{K}}$ for all other types of networks, where $\bar{K}$ is the average node degree for
each network under discussion ($\lambda_i = \lambda$ and $\mu_j = \mu$ for all $i, j$ without loss of generality). The average throughput of the susceptible queue $E'[\gamma]$, corresponding to the average rate that nodes receive malicious-information can be computed as:

$$E'[\gamma] = \sum_{i=1}^{N} \lambda(i) \cdot \pi'(i) = \frac{\lambda \bar{k}}{N} \left[ c(N+1) - (1-e^{-c})(N+2) - c - c^2 \right]$$

(5.5)

Observing the analytic form of the average number of infected nodes for each network type, according to the specific expression of $c$, a major difference between HeMPC networks and the rest should be noted. More specifically, in a HeMPC the spatial dependence among nodes (due to their multi-hop nature) is reflected by the fraction $\frac{\pi R^2}{L^2}$ representing the coverage percentage of each node with respect to the whole network area. On the contrary, in the expression of $c$ for the rest of the networks for which the topology is mainly based on connectivity relations and not spatial dependence as in the multi-hop case, the corresponding quantity expressing the local neighbor impact, is expressed by $\frac{N}{k}$. It is evident, that for the rest of network types, quantity $\frac{N}{k}$ expresses solely the special connectivity properties of the employed network through the value of $\bar{k}$, since for these networks, no spatial dependence is expressed in their connectivity graph and thus, no such spatial feature has impact on IDD.

5.2 Demonstration

Contrary to the SI model, in the SIS paradigm employed for the malicious information spreading, average quantities are of interest, while in the SI model instantaneous quantities were considered, as in the long run the network converges to a pandemic (all nodes are informed) state. Especially the average number of infected legitimate nodes is the most important quantity, since malicious-information is of recurrent nature and if one observes the system macroscopically, nodes oscillate between the $\{S\}$, $\{I\}$ states.

Fig. 7, shows the average number of infected nodes with respect to the infection/recovery ratio for various types of complex networks. It is evident that as the ratio $\lambda/\mu$ increases, $E[i]$ increases for all types of networks, denoting greater probability of users to get malicious-information from a communication link. Parameters for each type of network are in accordance with the notation employed in the classification of section 2.

With respect to ad hoc networks (HeMPC), the greatest the transmission radius of nodes, the denser the network, and thus the easier the malicious-data to be spread. Especially for ad hoc networks, the dependence of the average number of infected nodes on the number of legitimate nodes is linear as shown in Fig. 8. The combination of larger values of $R$ and $\lambda/\mu$ yields the higher number of $E[i]$ for all values of legitimate nodes, while the combination with smaller values of $R$ and $\lambda/\mu$ yields the lower values of $E[i]$. It should be noted that the behavior of the spreading dynamics of the network is more sensitive to changes in the value of $\lambda/\mu$ rather than variations in $R$.

Similarly to HeMPC networks, in ER networks, the greatest the probability two nodes are connected (thus the denser the network is), the easier for malicious-information to propagate on average. Such property may be also identified with the rest of the network types, leading to the following:

• Observation 5: The denser a network becomes, irrespective of its type, the easier for malicious-information to spread.

For regular, ER, WS and BA networks with the same $\bar{k}$, the same result would be obtained, due to the expression of $c$ given before. In order to better demonstrate the different robustness properties of
these network types, in Fig. 7 a different $\bar{k}$ value has been employed. Regarding the different types of networks, HeMPC (ad hoc) are close to ER (HoMEC), exhibiting that in general, randomness aids the spreading of malicious-information. This is a very useful outcome with significant practical value for designing efficient countermeasures for malign IDD. On the contrary, a regular lattice (HoMPC) makes the spread of malicious-information more difficult, since each node is only connected to a typically small number of other nodes and it would take significant effort to quickly spread malicious-information throughout the whole network. Similarly, WS (HoMUC) and BA (HeMUC) exhibit robustness closer to lattice networks, as their topologies are derived from such regular arrangements [1]. Among the latter three categories, a scale-free (BA) network may be more prone to spreading than a lattice (as shown in Fig. 7), because for the specific network instances, the given BA network is more dense than the lattice (the BA has mean degree $\bar{k} = 16$, while the lattice $\bar{k} = 4$ for the same number of network users). The following observations may be derived from the above analysis:

- **Observation 6**: Topological randomness favors the spreading of malicious-software.

- **Observation 7**: Among similar network types, the relative (local) density of each topology determines the robustness of the system against malicious information spreading.

### 6. Conclusion

In this work we introduced novel epidemic-based models for modeling and understanding information dissemination dynamics (IDD) in wireless complex networks. Our approach was inspired by epidemic
approaches, developed for the study of viruses in social communities. Useful-information dissemination was modeled according to the SI epidemic model, while malicious and indifferent types of propagated information were modeled according to the SIS infection paradigm. We provided analytical approaches for obtaining the behavior of spreading dynamics in both paradigms and in order to characterize the spreading of information in specific and diverse types of wireless complex networks. Numerical results depicted the effectiveness of the proposed approaches for analyzing and utilizing the developed processes in the described environment, yielding the most important characteristics of complex networks that affect and control the propagation process. Useful directions were identified for similar studies and developing practical countermeasures/infrastructures.

The proposed methodology and the respective analytical results could be further exploited for defining more complex and application-oriented problems that arise in information diffusion processes in wireless complex networks. For instance, these could be used in order to optimize the spreading of useful-information and designing more robust networks capable of sustaining large-scale attacks of malicious spreading information. Depending on the context of each application more effective dissemination campaigns can be designed and more robust infrastructures against malicious intentions.

REFERENCES