A Risk-Based Topology Control Algorithm in Ad Hoc Networks

Hongbiao Li*
School of Information Engineering, Northeast Dianli University, Jilin 132012, Jilin, China
*Corresponding author

***Abstract*—In order to protect social network security from malicious attacks, we study the propagation of malicious software in wireless networks by taking on the attacker's perspective. Some topology control algorithms are designed for the development of effective attack strategies by a malicious mobile node based on the risk function metric, which indicates the network's vulnerability. Our approach, in order to investigate the extent of its attack potentials, could be used for the effective design of network countermeasures. The results of performance evaluation demonstrate that the proposed risk-based topology control algorithms and respective attack strategies effectively balance the tradeoffs between the potential network damage and the attacker's lifetime, and as a result significantly outperform any other flat and threshold-based approaches.**

Keywords-ad hoc; topology; wireless network; attack strategy

I. INTRODUCTION

In this paper we study the propagation of malicious software in wireless ad hoc networks under a probabilistic framework. We design topology control algorithms for the development of effective attack strategies by a malicious mobile node, based on the risk function metric, which indicates the network's vulnerability. Our approach takes on the attacker's perspective, in order to investigate the extent of its attack potentials, which in turn could be used for the effective design of network countermeasures. Our performance evaluation results demonstrate that the proposed risk-based topology control algorithms and respective attack strategies effectively balance the tradeoffs between the potential network damage and the attacker's lifetime, and as a result significantly outperform any other flat and threshold-based approaches.

In this paper, based on a probabilistic framework for the modeling of malicious software (malware), we study the behavior and effectiveness of a malicious node attacking an ad hoc network [1] and the impact of topology in the propagation of malware. We propose a topology-dependent indicator of a node's vulnerability to become infected, the risk, and use a relevant measure, the risk function, to design topology control algorithms for the development of effective attack strategies. The risk function depends solely on the availability of local information, allowing the malicious node to manage efficiently its available resources. Through analysis and simulation we evaluate the algorithms' operation according to an offline metric that indicates the infection efficiency and characterizes the overall performance of an attack strategy [2].

II. SYSTEM MODEL

Therefore, the topology of an ad hoc network can be modeled by a random geometric graph, assuming node locations are determined by their respective coordinates in the geographical deployment region. Consequently, two nodes are connected with probability one if both of them are within transmission radius of each other, and with zero probability in all other cases. The edge set of the induced graph is completely determined by the relative positions of the nodes and their transmission radii.

In our model we assume a single malicious, mobile node with the capability of adjusting its transmission radius, r. The nodes of the underlying network are assumed to be static, having a common transmission radius R. Furthermore, it is possible that these nodes can infect their neighbors, once they become infected themselves, thus propagating potential attacks throughout the network. This behavior models the cases, where the attacker takes control of the exploited machines and uses some of their modules at its own interest [3].

Ad hoc and sensor networks have very limited energy resources. A wireless node needs to consume its energy reserves according to a sophisticated discipline, so as to extend as much as possible its lifetime without sacrificing its operational characteristics. Since in this paper the emphasis is placed on the attacker's behavior and impact, in order to better evaluate such a tradeoff, we assume the mobile attacker to have limited energy resources, while the network nodes have infinite reserves.

The resulting network topology has the form shown in Figure I. There exists a multihop underlying network with nodes having transmission radius *R*, where a mobile attacker moves around the deployment region, varying its transmission radius *r* and thus adapting its attack strategy. The cardinality of the network node set is denoted by *N* and the total area of the network by *A*. Then, assuming the network nodes are uniformly distributed over this area, the density of the network (excluding the attacker) becomes *N/A*.

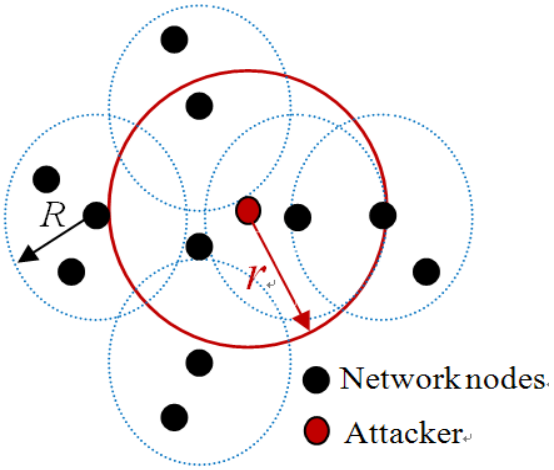

FIGURE I. TOPOLOGY OF ANALYZED AD HOC NETWORK

III. Risk Function and Topological Specialties

Different attacks follow different infection processes and their results vary significantly, depending on their target groups, mechanisms, ultimate objectives and current advances in technology. General means for characterizing and evaluating the propagation of malware are necessary for proper design of efficient countermeasures. In this section, we introduce the concept of the risk of a node, while in the following, based on this concept and properly defined measures, we introduce topology control algorithms that can be used for designing effective attack strategies. Furthermore, in this section an offline metric is described, which can be used for the evaluation of the efficiency of various types of attack strategies.

As the mobile attacker changes locations, its neighborhood varies, depending on the local topological properties of the network graph and the applied mobility model. Specifically, the node degree of the attacker changes non-deterministically, due to the stochastic nature of its movement and the locations of its neighbors. So do the degrees of the network nodes that lie in the 1-hop neighborhood of the malicious node (defined by the disk of radius r centered at the attacker) as this area changes. Thus, the attacker's node degree, Ka, is a random variable, over the sample space of positive integers. Theoretically, this sample space is infinite, since the geometric random graph representation allows a node to have infinite neighbors over a planar region. However, practical limitations on the wireless nodes' dimensions, force Ka to assume finite only values. As the network nodes are uniformly and randomly deployed over the targeted terrain, $Ka \in [0, Ka_{\max}]$, where $Ka_{\max}$ is the maximum degree value, determined by the density of the network nodes.

In graph theoretic terms, the degree of a node is indicative of the topological properties of its neighborhood. In particular, higher node degree means higher node density in the locality (1-hop neighborhood) of the specific node. In terms of malware propagation in an ad hoc network, higher node density allows for greater potentials [4], from an attacker's perspective, to spread successfully malicious software. Furthermore, nodes belonging in network regions of high density have greater risk to become infected, either from the attacker, or from neighbors that have already become infected.

We define the risk of a node in connection with its degree to capture the potential of infection in its region. We note that in this work, we follow the attacker's perspective, in order to realize the potentials for causing network damage, which in turn can be taken into account in designing efficient network countermeasures.

With respect to the attacker, Ka would be the risk to spread malicious software to its neighbors, while the degree, K, of a network node indicates its own risk. According to the previous discussion, high risk is undesirable from the network's viewpoint, whereas it is highly preferable by the attacker.

As explained before, the attacker's risk Ka is a random variable. Thus, Ka is completely defined by its probability density function or the complete set of higher order moments. The stochastic movement of the attacker (or a network node in the case of full network mobility) prohibits the a priori complete knowledge of the probability density function of its risk. Moreover, full knowledge of the higher order moments is computationally inefficient for fast decision making purposes. For these reasons, we define the Risk Function RF, by using only the first two orders of moments, as following equation (1) as a measure of an attacker's risk, where $E[\cdot]$ denotes the expectation of a random variable, $Var(\cdot)$ its variance and c represents a constant. In the following, without loss of generality, we consider $c = 1$.

$$R_F = E[Ka] + cVar(Ka),\ c > 0 \qquad (1)$$

We have added extra sophistication to the attacker, by allowing it to adjust its transmission radius according to the risk function. As the risk function provides additional information about the topological nature of the underlying network, the malicious node can devise topology control algorithms and adapt its strategy accordingly to utilize its available resources in a more effective way. In following section we describe three topology control algorithms (two of them are based on the risk function) that can be used by the malicious node in the process of infecting a network more effectively.

The risk-based topology control algorithms provide the means to properly adjust the attacker's transmission radius according to the measured topological characteristics (expressed by the risk function). As a result these algorithms attempt to dynamically balance the tradeoff between the number of infected nodes and the attacker's lifetime. However, the local operation of the risk-based algorithms does not necessarily indicate how effective these algorithms are in terms of the overall network damage.

In order to evaluate the efficiency of the various topology control algorithms and corresponding attack strategies, we use a general attack-evaluation metric that captures the overall impact caused by the attacker to the network. Since unavailability of a percentage of the network for a time period causes reduced network operation, a combined metric for the number of infected nodes over a time interval is more suitable for the overall performance characterization. In this paper we use the *Infection Efficiency*, I_E, of an attacker, which is defined as the integral of the time function of the number of infected network nodes [5], to characterize the overall performance of an attack

strategy. Intuitively, I_E is the product of the number of infected network nodes in a time interval by the duration of the corresponding interval, and therefore provides a combined measure of the instantaneous damage (i.e. absolute value of the number of infected nodes) and the corresponding interval that the damage takes place.

IV. Attack Modeling

Without harming the validity of the analysis, we focus on the infection of a node itself and not on the process or type of infection. According to this, a recovering node might become infected again throughout the attacker's lifetime. We are not interested in the particular infection mechanism (absence of immunization, new threats), but in the event that a node might become re-infected at some point of time. Within this context, we consider the malicious node to have the means to infect again already recovered nodes. This also, justifies the fact that a malicious node usually has a higher probability to infect other nodes compared to the infection capability of already infected network nodes.

The state of a node is considered to be binary, namely 'infected' or 'non-infected' [6]. Then the system state is given by a binary vector whose components correspond to the network nodes and the value of each row indicates the current node state. For each network node, there exists a link infection probability to become infected from either the attacker or an already infected neighbor, if the node's state is 'non-infected'. As the node has more than one neighbor, the probability of infection increases with the number of links. Thus, the number of infections that a node receives from a single link is a random variable. We assume this variable to be distributed according to a Poisson process. As various nodes follow different recovery procedures, we model the recovery process so that successive recoveries take place in exponential intervals.

The designated node model can be mapped to that of an *M/M/1* queue, where an infection corresponds to an arrival that requires service, and the recovery to the service itself. Both arrivals and departures are exponentially distributed. Since the attacker is allowed to move throughout the network, infecting the rest of the nodes and depleting its energy, the network will recover completely in finite time after the mobile attack-node exhausts its energy.

There are two types of events taking place: infection of a node and recovery of an infected node. If m denotes the total number of non-infected nodes, then in a network of N nodes and one attacker, N- m nodes of the network are infected (excluding the attacker). Furthermore, let n denote the number of nodes that are non-infected at a given instant and have the malicious node as a neighbor. Assuming an event has just taken place, with the above assumptions for the infection and recovery processes, the time interval for the next event is exponentially distributed with rate described as equation (2), where k_i is the number of infected neighbors of node i (including the attacker if applicable), μ_i denotes its recovery rate, λ_i denotes the link-infection rate between node i and any other neighboring network node and λ_{mal} denotes the link-infection rate between a network node and the attacker. The summation index spans the sequence of the network nodes, where without loss of generality, we assumed that the non-infected nodes that have the attacker as a neighbor are the first n, the non-infected not having the attacker as neighbor follow in the sequence up to index m, and the rest are the infected nodes for the time instant under consideration.

$$\sum_{i=m+1}^{N} \mu_i + \sum_{i=n+1}^{m} k_i * \lambda_i + \sum_{i=1}^{n} [(k_i - 1) * \lambda_i + \lambda_{mal}] \quad (2)$$

V. Topology Control Algorithms

Based on our previous discussion, in this section we introduce and describe in detail three different topology control algorithms. The first one is a single-threshold based topology control algorithm that uses the available energy resources to adapt the attacker's transmission radius, while the other two are based on the risk function.

Specifically, the *Threshold-based Topology Control (TTC)* algorithm starts with a large transmission radius, denoted by R_m and when the available energy resources drop below the threshold of 50% of the initial reserves, the algorithm switches to a smaller radius, denoted by R_s. We used two radii values (i.e. single-threshold) as a base case, even though the algorithm can be easily extended to a multi-threshold approach. The intuition behind TTC is that as long as the attacker has enough energy, it can use a large radius to increase its node degree as much as possible and therefore increase the number of nodes that can be directly infected. However, when its available energy reduces below a certain threshold, it is more efficient to reduce its radius to conserve energy and extend its lifetime at the cost of smaller node degree and consequently smaller instantaneous network damage.

The operation of the risk-based topology control algorithms is driven by the risk function, R_F. Ideally an attacker could utilize its maximum possible transmission radius to increase R_F as much as possible. However, this would lead to the quick exhaustion of its energy resources and the reduction of its lifetime. Thus, an efficient risk-based topology control algorithm should attempt to balance the demand for high risk with the need for conservative energy consumption.

The *Risk Function-based Topology Control (RFTC)* is a two stage algorithm that adjusts the measured value of the risk function. In a realistic operational environment, an attacker does not have prior knowledge of the R_F value interval. Thus, predefined thresholds on the measured R_F cannot be applied here for correct decision making. As a result, the first stage of RFTC is to adjust the lower and upper value bounds of R_F, so that the attacker has up-to-date knowledge of the interval that R_F belongs to. The next step is to determine, based on the current value of R_F , whether this value is closer to the upper bound and needs to be reduced or closer to the lower bound and needs to be increased. The adjustment of R_F can be done implicitly. That is, by increasing the transmission radius r the local neighborhood of the node increases, leading to higher values of R_F. Conversely, decreasing r, decreases the attacker's node degree, leading eventually to lower values of R_F. The operation of RFTC is shown in Table I, where RF_{min}, RF_{max}, and RF_{curr} is the minimum, maximum and current measured values of the risk function respectively. Parameter dr denotes the step value (constant for RFTC) by which the attacker's transmission radius is adjusted at each step as long as its minimum or maximum values are not reached.

TABLE I. RFTC ALGORITHM

Algorithm: RFTC

```
if (R_Fcurr > R_Fmax)
   R_Fmax := R_Fcurr;
elseif (R_Fcurr < R_Fmin)
   R_Fmin := R_Fcurr;
if (|R_Fcurr − R_Fmin| > |R_Fcurr − R_Fmax|)
   /*closer to RFmax → decrease RF*/
   r:=r-dr;
elseif (|R_Fcurr − R_Fmin| < |R_Fcurr − R_Fmax|)
   /*closer to R_Fmin → increase R_F*/
   r:=r+dr;
```

An immediate enhancement to the operation of RFTC refers to the discipline used to increase or decrease the step value of the attacker's transmission radius. *Risk Function-based Topology Control-Enhanced (RFTC-E)* algorithm takes into account the remaining energy resources when determining the level (i.e. step value) of radius adjustment (increase/decrease) required. When the available energy is more than 50% of the initial reserves, RFTC-E increases *r* by *dr* and decreases it slowly by *dr*/2 . When the remaining energy reserves are less than 50% of the initial energy, the attacker increases slowly the radius by *dr*/2 and decreases it quickly by *dr*.

Our performance evaluation results demonstrate that the proposed risk-based topology control algorithms and respective attack strategies effectively balance the tradeoffs between the potential network damage and the attacker's lifetime, and as a result significantly outperform any other flat and threshold-based approaches.

VI. SUMMARY

In this work, we assumed a static underlying network, in order to better reveal how the mobile malicious node's operational characteristics affect the attack propagation process. We first introduced the concept of the risk of a node that can be used for realizing the potentials of an attack in a network region. Based on this concept we defined the risk function as a relative measure and used it to design topology controlled attack strategies.

REFERENCES

[1] Michelle Effros, Ralf Koetter and Muriel Médard, Breaking Network Logjams. Scientific American, 296:6 (2007) 78-85.

[2] C. C. Zou, W. Gong, D. Towsley, Code Red Worm Propagation Modeling and Analysis. Proc. of the 9th ACM Conference on Computer and Communications (ACM CCS), Washington, DC, USA, 2002, pp.138-147.

[3] V. Karyotis, S. Papavassiliou, M. Grammatikou, B. Maglaris, On the Characterization and Evaluation of Mobile Attack Strategies in Wireless Ad-Hoc Networks. Proc. of the 11th IEEE Symposium on Computers and Communications (ISCC), Pula-Cagliari, Sardinia, Italy, June 2006.

[4] A. Ganesh, L. Massoulie, D. Towsley, The effect of network topology on the spread of epidemics. Proc. of 24th IEEE Conference on Computer Communications (INFOCOM), 2 (2005) 1455-1466.

[5] C.C. Zou, D. Towsley, W. Gong, Email virus propagation modeling and analysis. Technical Report TR-CSE-03-04, University of Massachusetts, Amherst, MA, USA, 2004.

[6] J.-Y. Le Boudec, M. Vojnovic, Perfect simulation and stationarity of a class of mobility models. Proc. of 24th IEEE Conference on Computer Communications (INFOCOM), 4 (2005) 2743-2754.