Modeling Node Compromise Spread in Wireless Sensor Networks Using Epidemic Theory

Pradip De, Yonghe Liu, and Sajal K. Das Department of Computer Science and Engineering The University of Texas at Arlington 416 Yates Street, Arlington, TX 76019 {pradipde, yonghe, das}@cse.uta.edu

Abstract

Motivated by recent surfacing viruses that can spread over the air interfaces, in this paper, we investigate the potential disastrous threat of node compromise spreading in wireless sensor networks. Originating from a single infected node, we assume such a compromise can propagate to other sensor nodes via communication and preestablished mutual trust. We focus on the possible epidemic breakout of such propagations where the whole network may fall victim to the attack. Based on epidemic theory, we model and analyze this spreading process and identify key factors determining potential outbreaks. In particular, we perform our study on random graphs precisely constructed according to the parameters of the network, such as distance, key sharing constrained communication and node recovery, thereby reflecting the true characteristics therein. The analytical results provide deep insights in designing potential defense strategies against this threat. Furthermore, through extensive simulations, we validate our model and perform investigations on the system dynamics.

Index Terms— Sensor Networks, Epidemiology, Random Key Predistribution, Random Graph.

I. Introduction

As wireless sensor networks are unfolding their vast potential in a plethora of application environments [1], security still remains one of the most critical challenges yet to be fully addressed. In particular, a vital problem in the highly distributed and resource constrained environment is node compromise, where a sensor node can be completely captured and manipulated by the adversary. While extensive work has focused on designing schemes that can either defend and delay node capture or timely identify and revoke compromised nodes themselves [5], little attention has been paid to the node compromise process itself. Inspired by recently emerged viruses that can spread over air interfaces, we identify in this paper the threat of epidemic spreading of node compromises in large scale wireless sensor networks and present a model that captures the unique characteristic of wireless sensor

networks in conjunction with pairwise key schemes. In particular, we identify the key factors determining the potential epidemic outbreaks that in turn can be employed to devise corresponding defense strategies.

A. Motivation

Due to its scarce resources and hence low defense capabilities, node compromises can be expected to be common phenomena for wireless sensor networks in unattended and hostile environments. While extensive research efforts, including those from ourselves [15], have been engineered toward designing resilient network security mechanisms [12], [13], the compromise itself and in particular the propagation of node compromise (possible epidemics) have attracted little attention.

While node compromise, thanks to physical capture and succeeding analysis, is naturally constrained by the adversary's capability, software originated compromises can be much more damaging. Specifically, the recently surfaced virus $Cabir^1$ that can spread over the air interface has unveiled a disastrous threat for wireless sensor networks. Inescapably, viruses targeting wireless sensor networks will emerge. Consequently, node compromise by way of virus spreading (over the air interface) can effortlessly devastate the entire network in a short period of time. With recent advancements on sensor design empowering nodes such as MICA2 motes with over-the-air programmability, the network becomes vulnerable to the above described attack. Even worse, the inherent dense, large scale nature of sensor networks undoubtedly further facilitates the virus propagation.

While virus spreading over the internet has been widely studied, and notably by means of epidemic theory [2], [3], the distance and pairwise key restricted communication pattern in wireless sensor networks uniquely distinguish the phenomena from those on the Internet.

B. Our Contribution

In this paper, we investigate the spreading process of node compromise in large scale wireless sensor networks.

¹http://www.f-secure.com/v-descs/cabir.shtml

Starting from a single point of failure, we assume that the adversary can effectively compromise neighboring nodes through wireless communication and thus can threat the whole network without engaging in full scale physical attacks. In particular, due to security schemes employed by the sensor networks, we assume that communication can only be performed when neighboring nodes can establish mutual trust by authenticating a common key. Therefore, node compromise is not only determined by the deployment of sensor nodes which in turn affects node density, but also determined by the pairwise key scheme employed therein. By incorporating these factors of the networks, we propose an epidemiological model to investigate the probability of a breakout (compromise of the whole network) and if not, the sizes of the affected components (compromised clusters of nodes). Furthermore, we analyze the effect of node recovery in an active infection scenario and obtain critical values for these parameters that result in an outbreak. Through extensive simulations, we show that our analytical results can closely capture the effects in a wide range of network setups.

The remainder of the paper is organized as follows. In Section II we present the preliminaries, including the threat model, random key pre-distribution, and epidemic theory. In Section III, we study the compromise propagation without node recovery and with node recovery, and detail our analytical results. We perform experimental study in Section IV. Related work is presented in Section V and we conclude in Section VI.

II. Preliminaries

In this section, we present our threat model and briefly overview pairwise key distribution in wireless sensor networks and epidemic theory.

A. Threat model

We assume that a compromised node, by directly communicating with a susceptible node, can spread the infection and conduce to the compromise of the susceptible node. Communication among sensor nodes is not only constrained by their distances, but also shall be secured and thus determined by the probability of pairwise key sharing. Therefore, the spreading of node compromise is dependent on the network deployment strategy and the pairwise key scheme employed therein. We assume that the "seed" compromise node could be originated by an adversary through physical capture and analysis of that node or by other similar means.

The spread of node compromise in a wireless sensor network, particularly thanks to its dense nature, can lead to an epidemic effect where the whole network will get infected. We consider this epidemic effect as the key threat to the network and hence the investigation target of this paper.

B. Pairwise Key Pre-distribution

As the pairwise key scheme affects the communication and hence the propagation of the node compromise, we provide below, a brief overview of the key distribution schemes in wireless sensor networks.

Due to the severe resource constraint of wireless sensor networks and limited networking bandwidth, proposed pairwise key schemes have commonly adopted the predistribution approach instead of online key management schemes with prohibitive resource consumption. The concept of pre-distribution was originated from [11], where the authors propose to assign a number of keys, termed *key ring* randomly drawn from a key pool. If two neighboring nodes share a common key on their key rings, a shared pairwise key exists and a secure communication can be established. Pre-distribution schemes that rely on bivariate polynomials is discussed in [13]. In this scheme, each sensor node is pre-distributed a set of polynomials. Two sensor nodes with the same polynomials can respectively derive the same key.

Regardless of the specific key distribution scheme, a common parameter capturing the performance is the probability that two neighbors can directly establish a secure communication. We denote this probability by q. As it shall be revealed later, q plays an important role in the spreading of node compromise, because direct communication, as explained in the threat model, can result in propagation of malicious code.

C. Node Recovery

In the event that a node is compromised, its secrets will be revealed to the attacker. The network may attempt to *recover* the particular node. Recovery might be realized in several possible ways. For example, the keys of the nodes might be revoked and the node may be given a fresh set of secret keys. In this context, key revocation, which refers to the task of securely removing keys that are known to be compromised, has been investigated as part of the key management schemes, for example in [5]. Moreover, recovery can also be achieved by simply removing the compromised node from the network, for example by announcing a blacklist, or simply reload the node's programs. More sophisticated methods may include immunizing a node with an appropriate antivirus patch that might render the node immune from the same virus attack.

Regardless, in our analysis, we will study virus spreading under the two cases respectively depending on whether a node can be recovered or not.

D. Epidemic Theory

Originally, epidemic theory concerns about contagious diseases spreading in the human society. The key feature of epidemiology [2], [7] is the measurement of infection outcomes in relation to a *population at risk*. The population at risk basically comprises of the set of people who possess a susceptibility factor with respect to the infection. This factor is dependent on several parameters including exposure, spreading rate, previous frequency of occurrence etc., which define the potential of the disease causing the infection. Example models characterizing the infection spreading process include the Susceptible Infected Susceptible (SIS) Model, Susceptible Infected Recovered (SIR) Model etc. In the former, a susceptible individual acquires infection and then after an infectious period, (i.e., the time

the infection persists), the individual becomes susceptible again. On the other hand, in the latter, the individual recovers and becomes immune to further infections.

Of particular interest is the phase transition of the spreading process that is dependent on an epidemic threshold: if the epidemic parameter is above the threshold, the infection will spread out and become persistent; on the contrary, if the parameter is below the threshold, the virus will die out.

Epidemic theory indeed has been borrowed to the networking field to investigate virus spreading. In this paper, we will mainly rely on a random graph model to characterize the unique connectivity of the sensor network and perform the epidemic study [8], [10].

III. Modelling and Analysis of Compromise Propagation

In this section, we analyze the propagation of node compromise originating from a single node that has been affected. Our focus is to study the outbreak point of the epidemic effect where the whole network will fall victim to the compromise procedure.

Our key method is to characterize the sensor network, including its key distribution, by mathematically formulating it as a random graph whose key parameters are precisely determined by those of the sensor network. Therefore, the investigation of epidemic phenomena can be performed on the random graph instead. Following this approach, we observe the epidemic process under two scenarios: without node recovery and with node recovery, depending on whether infected nodes will be recovered by external measures like key revocation, immunization, etc.

A. Network Model as Random Graph

Assume that sensor nodes are uniformly deployed in a disc area with radius R. Let $\rho = \frac{N}{R^2}$ denote the node density of the network where N is the total number of the nodes. For a sensor node with communication range r, the probability that l nodes are within its communication range is given by

$$p(l) = \binom{n}{l} p^l (1-p)^{n-l} \tag{1}$$

where p is defined by

$$p = \frac{r^2}{R^2} = \frac{r^2 \rho}{N}.$$
 (2)

Thus p is the probability of a link existing at the physical level, i.e., whether the two nodes fall within their respective communication ranges.

We further assume that the probability that two neighboring nodes sharing at least one key in the random predistribution pairwise key is q. Notice that q is determined by the specific pairwise key scheme employed. For a particular node having l neighboring nodes, the probability that there are k nodes, $k \leq l$, sharing at least one key with it is given by

$$p(k|l) = \binom{l}{k} q^k (1-q)^{l-k}$$
(3)

Therefore, the probability of having k neighboring nodes sharing at least one key is

$$p(k) = \sum_{\substack{l=k\\\infty}}^{\infty} p(l)p(k|l) \tag{4}$$

$$= \sum_{l=k}^{\infty} \binom{n}{l} p^{l} (1-p)^{n-l} \binom{l}{k} q^{k} (1-q)^{l-k}$$
(5)

Thus, based on both physical proximity and the probability of key sharing between neighbors, we get a degree distribution p(k). Notice that this degree distribution can be employed to generate a random graph G. Since G possesses the same property in terms of secure communication pattern as the sensor network of concern, we will next perform the analysis on G instead.

B. Compromise Spread Without Node Recovery

Given the random graph construction, we now analyze the case of compromise spread when no node recovery is performed. In other words, a compromised sensor node will remain infectious indefinitely.

Let $G_0(x)$ be the generating function of the degree distribution of a randomly chosen vertex in G and is defined by

$$G_0(x) = \sum_{k=0}^{\infty} p(k) x^k \tag{6}$$

Moreover, with $G_1(x)$ given by

$$G_1(x) = \frac{1}{G_0'(1)}G_0'(x) \tag{7}$$

and with λ denoting the infection probability of a node being infected by communicating with a compromised node, then following the analysis presented in [8], the average size of the outbreak is derived as

$$s = 1 + \frac{\lambda G_0'(1)}{1 - \lambda G_1'(1)}.$$
(8)

Infection probability λ essentially captures the spreading capability of the virus that could compromise the network: the larger it is, the stronger the virus is. We assume that its value can be obtained by means of measurement or analysis.

Given the above result, we can see that the outbreak point for the network is $\lambda = 1/G'_1(1)$ which marks the onset of an epidemic. For $\lambda > 1/G'_1(1)$ we have an epidemic in the form of a giant component in the random network and the size S of the epidemic, where S denotes the expected fraction of the network that will be compromised if an outbreak happens, is given by

$$S = 1 - G_0(u)$$

Here u is the root of the self-consistency relation

$$u = G_1(u).$$

Intuitively, the above conclusion reveals that if $\lambda \leq 1/G'_1(1)$, the component of compromised nodes is finite



(a) Non-epidemic cluster size vs. infection probability (λ)



(b) Epidemic size vs. infection probability (λ)

Fig. 1. Size of compromised node clusters: (a) depicts the average size of infected clusters when there is no epidemic and (b) shows the epidemic size as the fraction of the entire network. The point where non-zero value appears indicates the transition from non-epidemic to epidemic

in size regardless of the size of the network and each node's probability of being compromised is zero for large networks. On the contrary, if $\lambda > 1/G'_1(1)$, there always exists a finite probability for a node to be compromised.

Fig. 1 depicts this effect for a network with N = 1000nodes with different key sharing probabilities q. The underlying physical topology, determined by the communication range and node density, has an average edge probability of p = 0.25. Given the physical deployment, we vary the probability of direct pairwise key sharing (q) and study the point of outbreak. As we can see in Fig. 1, while undoubtedly increasing q can facilitate communication in the network, the network also becomes more vulnerable to virus spreading. Specifically, when q = 0.01, network wide breakout is only possible when a compromised node has an infection probability (λ) larger than 0.4 to infect a neighbor. We note that in this case, we have an average node degree of 2.5. On the contrary, this probability only needs to be around 0.05 when q = 0.1 which subsequently makes the node degree 25. Fig. 1(b) illustrates the fraction of the network that is ultimately infected as the infection probability is increased beyond the critical point of the onset of outbreak. For instance, we observe that when q = 0.1, the whole network is compromised with a λ value of less than 0.2. On the contrary, with q = 0.01, 80% of the network could be compromised with only a high value of $\lambda = 0.8$.

In summary, Fig. 1 clearly indicates the tradeoff between key sharing probability among sensor nodes and the vulnerability of the network to compromise.

C. Compromise Spread With Node Recovery

In this case, we assume that the network has the capability to recover some of the compromised nodes by either immunization or removal from the network. To capture this recovery effect, we assume that an infected node recovers or is removed from the network after an average duration of infectivity τ . In other words, a node in the sensor network remains infective for an average period τ after which it is immunized. During this infective period, the node transmits the epidemic to its neighbors with the infection rate β , denoting the probability of infection per unit time. Evidently, the parameter τ is critical to the analysis as it measures how soon a compromised node recovers. Naturally, we will perform our analysis following the SIR model in epidemic theory [10], [8].

First, consider a pair of adjacent nodes where one is infected and the other is susceptible. If T denotes the compromise transmission probability, given the above definitions for β and τ , we can say that the probability that the disease will not be transmitted from the infected to the susceptible is given by

$$1 - T = \lim_{\delta t \to 0} (1 - \beta \delta t)^{\tau/\delta t} = e^{-\beta \tau}.$$
 (9)

Subsequently, we have the transmission probability

$$T = 1 - e^{-\beta\tau}.$$

In other words, the compromise propagation can be considered as a Poisson process, with average $\beta\tau$. The outcome of this process is the same as bond percolation and T is basically analogous to the bond occupation probability on the graph representing the key sharing network. Thus, the outbreak size would be precisely the size of the cluster of vertices that can be reached from the initial vertex (infected node) by traversing only occupied edges which are occupied with probability T. Notice that T explicitly captures node recovery in terms of the parameter τ .

Replacing λ with T in Equation 8, and following similar steps, we get the size of the average cluster as

$$s = 1 + \frac{TG'_0(1)}{1 - TG'_1(1)}.$$
(10)

and the epidemic size is obtained by

$$S = 1 - G_0(u;T).$$
(11)

where u is obtained by

$$u = 1 - G_1(u;T), (12)$$



(a) Non-epidemic cluster size vs. infectivity duration τ



(b) Epidemic size vs. infectivity duration τ

Fig. 2. Size of compromised node clusters: (a) depicts the average size of infected clusters when there is no epidemic and (b) shows the epidemic size as the fraction of the entire network. The point where non-zero value appears indicates the transition from non-epidemic to epidemic

and $G_0(u;T)$ and $G_1(u;T)$ are given respectively by

$$G_0(u;T) = G_0(1 + (u-1)T), \tag{13}$$

and

$$G_1(u;T) = G_1(1 + (u-1)T).$$
(14)

Fig. 2 summarizes this effect, depicting the epidemic outbreak against the average recovery time τ for the respective infection rates β . The plots are for a sensor network with typical average degree of 10. In Fig. 2(a), we can identify the average duration that an infected node is allowed to remain infective before an epidemic outbreak occurs. We notice that, when the infection rate is 0.01, infected nodes have to be recovered/removed on the average in less than 100 time units in order to prevent an epidemic. As expected, this time is much lower when the infection rate is 0.2. Fig. 2(b) depicts the epidemic outbreak point for different infection rates β in terms of the average duration of infectivity of a node.

We remark that both the analytical and experimental results have significant implication for security scheme design in terms of revoking/immunizing compromised nodes in wireless sensor networks: it dictates the speed at which the network must react in order to contain/prevent the effect of network wide epidemic.

IV. Simulation

We employ a discrete event-driven simulation to accurately simulate the propagation of the infection spreading process. In this section, we first outline our discrete-event driven simulation model for the gradual progress of the spread of node compromise. Then we use this model to capture the time dynamics of the spread of the compromise in the whole population.

A. Simulation Setup

In our simulation, we assume the number of sensor nodes in the network to be 1000. The sensor network is produced by uniformly distributing the sensors in a $1200 \times 1200 \text{ unit}^2$ area. The communication range of each node is assumed to be 100 units. Our goal is to make the physical network fairly connected with an average node degree of around 20 to 25. We use the key sharing probability on top of this network to further reduce the average node degree of the final key sharing network to typical values of 3 and 10.

We employ the random key pre-distribution scheme described in [11] to establish the pairwise key among sensor nodes. By tuning the parameters of the scheme, we can achieve any specific values for the probability of any two neighbors to share at least one key.

Our simulation works in two phases. In the first phase, we form the network where each node identifies its set of neighbors and entries are made into a neighbor table. The average degree of the key sharing network is controlled by changing the value of the key sharing probability between neighbors. The entry for each node in the neighborhood table can indicate whether a node is susceptible, infected or recovered. We use typical values obtained for the average node degree of the network, namely, 3 and 10.

In the second phase, we simulate actual virus propagation. Initially, at t = 0, the number of infected nodes, denoted by I(0) is set to be 1. At any time point t, the population is divided into the group of susceptible nodes, S(t), and the group of infected nodes, I(t). In the situation where we have nodes that are immunized and thus recovered, we denote that this set of recovered nodes by R(t). The sub-population dynamics is obtained by observing the population counts after fixed simulation intervals of 1 time unit. We assume that the time it takes for an infected node to infect its susceptible neighbor is negative exponentially distributed with a mean of 1 unit time.

There are two simulation scenarios corresponding to our analysis.

B. Simulation Results and Discussion

1) Simulation Results for No Recovery Case: The simulation results for the case without recovery are shown



Fig. 3. System dynamics without recovery

in Fig. 3. We vary the value of the infection probability β under different network connectivities and study the time dynamics of the infected population. We notice, as expected, that an increase in the average node degree from 5 to 10 has an impact on the rate of compromise of the network. For instance, the curve with the lowest β value(0.05) has compromised the entire network by simulation time 700 when the average node degree is 10. However, with the node degree at 5, a β value of 0.05 could compromise upto 70% of the network by that same simulation time. Thus, we find that in the no-recovery case the two key parameters affecting the network compromise rate are infection probability β and the average node degree.

2) Simulation Results for Recovery Case: Fig. 4 and 5 show the simulation results for the three sub-populations (infected, immunized, and susceptible) in the situation where nodes do recover.

In Fig. 4 we see the effects of the infectivity duration τ and infection rate β on the dynamics of the epidemic. In Fig. 4(c), the highest point is reached very fast because of the high value of β . Thereafter, its recovery also takes less time. However, in Fig. 4(a), β is smaller but τ_0 is higher

(i.e., 30), the infection rises slowly and also falls slowly because of the high recovery time.

In comparison, Fig. 5 has better connectivity of average node degree of 5 which in turn increases the rate of infection significantly. Comparing Fig. 4(c) and Fig. 5(c), we observe that infection penetration is higher in the latter even in presence of a smaller value of β . In Fig. 5(c), it shows that even with a low value of β , the infection still rises to above 60%.

Therefore, we observe that network connectivity has a high impact on the infection propagation and on the speed of reaching the maximal point of outbreak. However, thereafter during the recover phase, τ_0 affects aggressively the time it takes to recover the whole network.

V. Related Work

The mathematical modeling of epidemics is well documented [2], [7]. In fact, visualizing the population as a complex network of interacting individuals has resulted in the analysis of epidemics from a network or graph theoretic point of view [8], [9], [10].

Node compromise in sensor networks and the need for their security has also received immense attention [4]. A large portion of current research on security in sensor networks has been focused on protocols and schemes for securing the communication between nodes [12], [13]. Revocation of keys of compromised nodes has been studied in [14]. In [4], the authors demonstrate the ease with which a sensor node can be compromised and all its information extracted. Unfortunately, little work has been done on the defense strategies when the compromise of a single node could be used to compromise other nodes over the air. In this paper, we take the first step to model this potential disastrous propagation. In [6], the authors used an epidemic modeling technique for information dissemination in a MANET. However, they assumed homogeneous mixing which is not possible in a static sensor network as ours.

In our work, we adopted some of the results presented in [8] where the author proposes a percolation theory based evaluation of the spread of an epidemic on graphs with given degree distributions. However, little has been shown there on the temporal dynamics of the epidemic spread and the authors only studied the final outcome of an infection spread.

VI. Conclusion

In this paper, we investigate the potential threat for compromise propagation in wireless sensor networks. Based on epidemic theory, we model the process of compromise spreading from a single node to the whole network. In particular, we focus on the key network parameters that determine a potential epidemic outbreak in the network. Due to the unique distance and key sharing constrained communication pattern, we resort to a random graph model which is precisely generated according to the parameters of the real sensor network and perform the study on the graph. Furthermore, we introduce the effect of node recovery after compromise and adapt our model to accommodate this effect. Our results reveal key network parameters in defending and containing potential epidemics. In particular,







Fig. 5. The dynamics of the population with recovery for average degree of 10

the result provides benchmark time period for the network to recover a node in order to defend against the epidemic spreading. Our extensive simulation results validate our analyses and moreover, provide insights of the dynamics of the system in terms of temporal evolution.

References

- I Akyildiz, W. Su, Y Sankarasubramaniam, and E. Cayirci, "A Survey on sensor networks," *IEEE Communications Magazine*, vol. 40, no. 8, 2002.
- [2] R. M. Anderson and R. M. May, "Infectious Diseases of Human: Dynamics and Control" (Oxford Univ. Press, Oxford, 1991).
- [3] S. Staniford, V. Paxson, and N. Weaver. "How to Own the Internet in Your Spare Time". In *11th Usenix Security Symposium*, San Francisco, August, 2002.
- [4] C. Hartung, J. Balasalle, and R. Han, "Node Compromise in Sensor Networks: The Need for Secure Systems", *Technical Report CU-CS-990-05* (2005).
- [5] H. Chan, V. D. Gligor, A. Perrig, G. Muralidharan, "On the Distribution and Revocation of Cryptographic Keys in Sensor Networks", *IEEE Transactions on Dependable and Secure Computing* 2005.
- [6] A. Khelil, C. Becker, J. Tian, K. Rothermel, "An Epidemic Model for Information Diffusion in MANETs", *MSWiM* 2002, pages 54-60.
- [7] N. T. J. Bailey, "The Mathematical Theory of Infectious Diseases and its Applications". *Hafner Press*, New York (1975)

- [8] M. E. J. Newman, "Spread of epidemic disease on networks", *Phys. Rev. E*, **66** (2002), art. no. 016128.
- [9] C. Moore and M. E. J. Newman, "Epidemics and percolation in small- world networks". *Phys. Rev. E* 61, 5678-5682 (2000)
- [10] P. Grassberger, "On the critical behavior of the general epidemic process and dynamic percolation", *Math. Biosc.* 63 (1983) 157.
- [11] L Eschenauer and V. D. Gligor. "A key-management scheme for distributed sensor networks", in Proc. of the 9th Computer Communication Security - CCS '02, pages 41–47, Washington D.C., USA, November 2002.
- [12] H. Chan, A. Perrig, and D. Song, "Random key predistribution schemes for sensor networks", *in Proc. of the IEEE Symposium on Research in Security and Privacy - SP '03*, pages 197–215, Washington D.C., USA, May 2003.
- [13] Donggang Liu and Peng Ning, "Establishing pairwise keys in distributed sensor networks", in Proc. of the 10th ACM Conference on Computer and Communications Security - CCS '03, pages 52– 61, Washington D.C., USA, October 2003.
- [14] H. Chan; V.D. Gligor, A. Perrig, G. Muralidharan, "On the Distribution and Revocation of Cryptographic Keys in Sensor Networks", *IEEE Transactions on Dependable and Secure Computing*, Volume 2, Issue 3, July-Sept. 2005
- [15] A. Chadha, Y. Liu. and S. Das, "Group key distribution via local collaboration in wireless sensor networks," in *Proceedings of the IEEE SECON 2005*, Santa Clara, CA, Sept. 2005.