MeDrone: On the use of a medical drone to heal a sensor network infected by a malicious epidemic
Nicola Roberto Zema, Enrico Natalizio, Michael Poss, Giuseppe Ruggeri, Antonella Molinaro

To cite this version:

HAL Id: hal-01396858
https://hal.archives-ouvertes.fr/hal-01396858
Submitted on 17 Nov 2016

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L’archive ouverte pluridisciplinaire HAL, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d’enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.
MeDrone: On the Use of a Medical Drone to Heal a Sensor Network Infected by a Malicious Epidemic

Nicola Roberto Zema*, Enrico Natalizio*, Michael Poss†, Giuseppe Ruggeri†, Antonella Molinaro†

*Lab. Heudiasyc, UMR CNRS 7253, France.
{nica.zema, enrico.natalizio}@hds.utc.fr
†ARTS Laboratory, University “Mediterranea” of Reggio Calabria - DIIES Department, Italy.
{antonella.molinaro, giuseppe.ruggeri}@unirc.it
‡LIRMM, UMR CNRS 5506, France.
michael.poss@lirmm.fr

Abstract—The complexity increase in the software and hardware necessary to support more and more advanced applications for Wireless Sensor Networks conspicuously contribute to render them susceptible to security attacks. The nodes of most complex WSN applications sport desktop-level operating systems and this reliance on software make them ideal prey for traditional threats like viruses and general malware. To address these problems, in this paper we devise a system for a mobile node to locate, track, access and cure the infected nodes of a WSN threatened by a proximity malware infection. For this purpose we first devise a curing operation scheme for a dedicated mobile node and then we provide an implementation for it in form of a multiprocess mathematical formulation for the aforementioned operation scheme in order to provide an optimized solution. We perform extended simulations putting our proposal against different movement schemes in different network scenarios and we use the results of the mathematical formulation as a benchmark. Furthermore, we introduce a variation of our proposal capable to support the concurrent operation of multiple mobile actors and implement cooperation.

I. INTRODUCTION

Proximity malware consists in the injection of malicious code in a network and has become recently a serious issue in Wireless Sensor Networks [1]–[3].

The nodes of a WSN are devised traditionally as resource-constrained devices with hard-wired algorithms driving their behavior. Normally, this assumption has made them resilient to the modifications related to the diffusion of software malwares. However, some recent advanced sensing applications [4] use complex sensor nodes that require capabilities comparable to smartphones. Furthermore, as industrial applications expand, devices are developed from commercial, off-the-shelf equipment, which increases the risk of security vulnerabilities [5]. These capabilities usually translate into complex operating systems and communication protocols for the devices and networks operations. For instance, wireless systems for video surveillance are subject to the usual constraints of WSNs, such as the deployment in a difficult to access environment or energy limitations, but some additional complexity is needed to handle a streamed video and transmit it to a remote sink. Furthermore, the deployment of networks tailored for security that employ devices like arduino and android apps, further increases the chances of security threats. The feasibility of full-blown attacks on WSNs has been studied and, given the relative node isolation of these networks, the majority of threats come from proximity malwares [6]. This kind of malware spreads from one device to another using short-range communication interfaces like Bluetooth and WiFi. Examples of its virulence and dangerousness are available in literature, like the CABIR worm for Bluetooth [7] and the iKee for WiFi suites [8]. Even if several defense mechanisms have been proposed [9], [10], so far they rely on external network infrastructure or direct human intervention. They can, for instance, apply patches, isolate malicious nodes, run antiviruses or securing remote accesses. The aforementioned possibilities require direct node access but it is almost completely negated in a remote WSN deployment [11]–[13] and so different approaches and methods are needed to manage security risks. In this paper we aim at fulfilling this lack of solutions for WSNs and in the following list we explain our contributions:

1) We devise a counter-epidemic protocol that heals the network driving a mobile flying node (referred to as searcher in the rest of this paper) through the WSN deployment area after a in-network analysis of the malware spreading. The searcher is capable of autonomous operations and movement.

2) We propose both a heuristic and an optimization model for the Targeted Curing Problem (TCP). The solutions aim at respectively determining the sub-optimal and the best trajectory the searcher has to travel to both contain the malware diffusion and cure the nodes, minimizing the time they stay infected. The metrics of diffusion containment are the lowest delay to restore all nodes and the minimization of the number of infected nodes.

In order to show the effectiveness of our findings we use the model output as a benchmark and then we validate our heuristic by comparing it with other movement schemes suited for area coverage.

A preliminary contribution on this topic has been provided by authors in [14]. This paper represents a step forward with respect to [14]: (i) by offering a formal proof that the
analytical solution of the TCP in NP hard, (ii) by extending the performance evaluation to include a wider set of scenarios and competitors, and (iii) by providing an outlook on the simultaneous deployment of more than one searcher.

This paper is organized as follows: Section II introduces the background research on the various subjects this paper refers to, followed by the set of assumptions we used. The main contribution is described in Sections IV and V. In the former we describe the scenario, the main proposal and the mathematical model. In the latter we describe our heuristic-based approach to the same problem and the framework needed to support it. In Section VI we show the simulation and numerical comparative performance as well as we present some results for the coordination problem with multiple searchers. Section VII ends the paper.

II. RELATED WORK

The definition of strategies maximizing the movement efficiency of a tracking entity against an adversary has traditionally been addressed under the area of the search-and-pursuit research [15]. According to this topic, a combination of tracker entities try to define the best trajectories in respect to an evader entity in order to, at the same time, occupy the same space and capture it. Considering a malware that spreads among nodes as the aforementioned evader, and the searcher node as the tracker different methodologies can be applied in order to extract the set of best trajectories to reach not only the actual infected nodes but also anticipate the spreading and block the epidemic. However, the solutions already available focus on the tracker motion plan analysis to bound its movement performances [16], [17]. In this paper we define an actual movement optimization strategy, coupled with a heuristic. Anyway for the peculiar environment our system is put into, its class of dynamics has not been yet treated.

The research on immunization and networks epidemics, which can also be seen as a site percolation problem [18], has been in general [19]–[21] a popular topic. With reference to WSNs [9], [22] it has recently caught the interest from the networking research community. According to the research, a useful strategy limiting the effects of an epidemic is to immunize a sufficiently large subset of the original population [23]. However, according to [24], a computer epidemic is more viral than real-world counterparts as it is engineered to spread fast and it needs almost the 100% of the node population to be immunized to limit the diffusion. The intolerable delays necessary in order to immunize all the nodes in a WSN make this solution impractical [24]. According to [23], in WSNs, it has found that the most effective way to restore a node from any condition is to wipe and reload the non-protected (userland, data) part of its operating system and this solution is the one we devised to be implemented by the searcher. More than often, in WSNs and ad-hoc networks, there is a strict bond between the detection of node failures and the detection of outbreaks [1], [26] but, as described in [27], [28] a malware could be also programmed to not disable the nodes during an incubation phase in order to maximize the damage once all the network is infected. As a final note, even if the research on models of malware diffusion on social and proximity network is widespread [29], [30], in this work we focus on a curing trajectory setting upon an isolated and difficult to access wireless sensor network rather to model the actual diffusion of a multi-environmental malware.

III. MAIN ASSUMPTIONS AND DESCRIPTION

In this Section we describe the main assumptions on which our work is based, and we give some details regarding the epidemic dynamics and the network model. We suppose that sensors have an operating system, computational and communication resources that make them capable to perform advanced applications [4]. Considering the WSN deployment in harsh and remote WSNs scenarios, we have devised to use a flying robot to distribute the cure. Furthermore, we assume that:

1) in the remote WSN installation the only mean of information exchange is over short-range radio links and there is no infrastructured access to an external network;
2) after a variable time a node has been in contact with an infected neighbor, it will become infected itself with the same malware;
3) The identification of malware outbreaks in different kind of networks has been widely studied and given the various implementations of Intrusion Detection Systems (IDS) [26], [31]–[34] available, we assume that an IDS is active in the network.
4) It is supposed that the geographical localization of nodes is available at least in form of predefined coordinates entered by a human operator when a node is deployed;
5) a node can be cured by uploading a new image of the operating system;
6) all the messages among nodes, and between nodes and searchers are directly embedded in layer-2 frames, as the primitive used for their exchange are usually embedded in the physical network devices and thus more robust to attacks;
7) we suppose the presence of a low-level secured interface [35] to the nodes operating system as previously described and this feature concretizes in a secure computing interface that is easily to provide as stated in [26]. Various works [36], [37] present the possibility of a secure boot of a node after the retrieval of a new part of the operating system as specified by the guidelines of Trusted Computing [38]. In this way we can assume that nodes could replace parts, modules and possibly their whole operating system by retrieving the necessary software by the means of their wireless interface.

IV. MATHEMATICAL FORMULATION

In this paper, we introduce the Targeted Curing Problem (TCP) determination, as the determination of the positions that a searcher should travel to within an infected WSN, in order to diffuse the cure, to limit the malware spreading and to clear the infected nodes. In this Section, we address the formulation of the TPC problem that, for a given network, is capable of
Theorem 1: The TCP problem is \(NP\)-hard.

Proof: Given a graph
\[
G = (V, A)
\]
and an integer \(K\), the DOMINATING SET PROBLEM [39] tests whether it is possible to find a subset \(W\) of \(V\) that contains at most \(K\) nodes and such that every other node of \(V\) belongs to an edge adjacent to \(W\). Given an instance to the DOMINATING SET PROBLEM, it is defined a corresponding instance for the TCP as follows:

- \(N = V \cup \{i^*\}\) The set of nodes
- \(E = A\) The set of wireless links
- \(E_c \supseteq\) all pairs of nodes in \(N\) The set of physical paths
- \(N^* = V\) The set of initially infected nodes
- \(d_e = 1 \forall e \in E_c\) Edge lengths
- \(T_m = T_i = T_a = 0\) The time intervals extracted from the simulated results
- \(d_i = 0 \forall i \in N\) The cure travel times

Then, it can be seen immediately that the answer to the DOMINATING SET instance is YES if and only if the optimal solution cost to the TCP instance is less than \(K\).

Hence, no efficient combinatorial algorithm exists for the problem and it is reasonable to address it by using binary integer linear programming. The detailed formulation of the problem in terms of binary variables and linear constraints is given in appendix A. Unfortunately, the numerical solution determined by using binary integer linear programming is not directly applicable in real cases, because it requires the complete \(a\)-\(priori\) knowledge of the diffusion dynamics of nodes and epidemics. For this reason in Section V we developed a heuristic and we retained the integer linear programming results as an upper bound on the quality of the solution compare the results obtained with other movement schemes. The results of the comparison are showed in Section VI.

V. DISTRIBUTED CURING ALGORITHM

The mathematical approach previously described produces a set of waypoints for the searcher that, if followed, allows to contain the infection and to minimize the malware spreading through the network. Unfortunately, it is required from the model to know precisely the whole infection dynamic of the malware in terms of which node is going to be infected at what time. In a real WSN deployment this last assumption is not feasible and hence we devise a heuristic powerful enough to provide a sub-optimal movement scheme using only the information available locally at each node in the network. The proposed idea is to move the searcher continuously towards the regions with the highest density of infected nodes. In this paper, the density of infected nodes perceived by one of them is intended as the number of its infected neighbors (separated from the uninfected ones thanks to the IDS [26]) versus its communication range. Moreover, if all nodes are of the same type in terms of transmission capabilities, the total number of infected neighbors is enough to estimate the aforementioned density in a circular area of which the node is the center.

<table>
<thead>
<tr>
<th>Transition</th>
<th>States</th>
<th>Happens when</th>
</tr>
</thead>
<tbody>
<tr>
<td>A</td>
<td>1(\Rightarrow)2</td>
<td>a node in state 1 will switch to state 2 if it has been connected for (T_a) time units to any number of nodes in state 3</td>
</tr>
<tr>
<td>B</td>
<td>2(\Rightarrow)3</td>
<td>a node in state 2 will switch to state 3 if any number of nodes in state 3 have been connected to it for, at least, (T_2) time units</td>
</tr>
<tr>
<td>C</td>
<td>3(\Rightarrow)4</td>
<td>a node in state 3 will switch to state 4 if it has been connected for (T_a) time units to the searcher</td>
</tr>
<tr>
<td>D</td>
<td>4(\Rightarrow)5</td>
<td>a node in state 4 will switch to state 5 if any number of nodes in state 1 or 2 or the searcher are connected to it for, at least, (T_1) time units</td>
</tr>
<tr>
<td>E</td>
<td>2(\Rightarrow)5</td>
<td>a node in state 2 will switch to state 5 if it has been connected for (T_a) time units to the searcher</td>
</tr>
<tr>
<td>F</td>
<td>1(\Rightarrow)5</td>
<td>a node in state 1 will switch to state 5 if it has been connected for (T_a) time units to the searcher</td>
</tr>
</tbody>
</table>

TABLE I: State Transitions

determining the optimal set of positions in time and space for the searcher to minimize the healing time.

In our model, each node can be in only one of five different states at a time. They are illustrated in the diagram of Figure 1 and are:

1) Uninfected
2) In Transition to Infected
3) Malicious (and Infected)
4) In Transition to Cured
5) Cured

The transitions between each couple of states are described in Table 1. Unfortunately, the TCP problem is \(NP\)-hard, because it generalizes the problem of looking for a minimum dominating set.
Two concurrent processes compose the Distributed Curing Algorithm (DCA) proposed in this paper: (i) an Alerting Process, in which there is an exchange of the malware diffusion information between the nodes and a (ii) Healing Process. The last process is carried out by the searcher and is composed of two continuously repeated phases. The first one is the Positioning phase, in which the searcher, aware of the epidemic from the Alerting Process plans its movements. The second one is the Curing phase, where the searcher accesses the infected nodes and restores them.

The processes can be handled by internal timers present into the nodes without needing any synchronization.

A. The Alerting Process

As mentioned in Section III all messages for the cure are carried in Layer-2 frames and broadcasted by nodes through their short-range radio interface.

Whenever the IDS discovers the presence of a compromised component of its node, it begins the Alerting Process and sends a SOS message every $\Delta t_{SOS}$ seconds on the node’s short range interface.

Any SOS message contains:

1) Sender hash ($Sid$): A variable-length label which identifies both the node sending the message locally and the original issuer.

2) Sender position (SP): 64 bit long field containing the absolute position (GPS) of the node generating the SOS. Since in our scenario the nodes are stationary this information can be stored in the secure memory at time of node installation. According to [40], the absolute position can be substituted by a relative one using the $nKIN$ field.

3) The number of known infected nodes ($nKIN$): representing the perceived density of infected neighbors. The number of SOS messages heard in a reference time interval $\Delta t_{observ} = 2\Delta t_{SOS}$ can be an indication of this density. If no other SOS message has been received, the field is set to one.

Every node that receives the SOS, rebroadcasts it once at each reception from the original sender, using the hashes to discern duplicates within a listening $\Delta t_{observ}$ time frame. If during this interval more than one message is received, only the one with the highest $nKIN$ is stored for forwarding. The local $nKIN$ is considered also in the ranking.

B. The Healing Process

We suppose that the searcher roams the area where the nodes are deployed prior to the beginning of any signaling, constantly monitoring its network interfaces, listening to any incoming SOS packet. Whenever one of these messages is received, it extracts and ranks the SP location and starts its own Positioning phase (phase i), moving towards it. As at any moment the position of the searcher is unknown to the network, in order to allow the information about the most dense area of infected nodes to diffuse properly, before each Positioning phase the searcher waits for a movement interval. This interval is the same mentioned in the aforementioned model of Section IV set in order to allow distant SOS messages carrying the highest values to reach every other node.

Previous work results [41] shows that for the information about the most dense point, originating from a node $i$, to arrive to another node $j$, the information has to travel twice the number of hops between $i$ and $j$. According to this result, we have set the movement interval equal to the network diameter multiplied by $\Delta t_{SOS}$. After the movement interval expires, the searcher selects the SP with the highest $nKIN$ and starts the movement. In the following we name this movement scheme as MeDrone (which stands for Medical Drone). The infected nodes are restored and the cure delivered during the Curing phase, phase (ii). The cure diffusion scheme can be seen as an exchange of frames between the protected section of the infected node operating the IDS systems with the searcher. The last one delivers a clean image of the operating system of the nodes to them and their IDS. This copy is used to overwrite the compromised parts of the operating system. We have chosen to fragment the image of the operating system diffused by the searcher in variable-length chunks at the Application Layer. This choice permits to change the chunk dimensions on-the-fly according to the L2 technology used. For instance when IEEE 802.11 is used, a proper chunk size could vary from 1024 to 1500 bytes according to the various versions of the standard. At the end of the secured transfer the infected node is rebooted as healed. The transfer is called secure because each message is digitally signed in order to be recognized as valid by the IDS on the infected nodes [35]. For preventing the restored elements from being infected again, the searcher instructs all the nodes it contacts to drop any incoming unsecured communication until further notice. However, this last specific event exceeds the time frames considered in this paper.

To start the Curing phase, the searcher, which is in the range of one or more infected nodes, advertises its presence through Hello packets every $\Delta t_{Hello}$ seconds.

The Hello message consists of a set of identifiers for the IDS of the nodes to recognize the searcher and, upon the reception these information are stored by the node.

The healed OS’s image is transferred in signed chunks. The process is driven by the exchange of Interest and Data packets between the infected node and the searcher, respectively. The Interest is broadcasted by the infected node upon receiving a Hello message to request a specific set of chunks.

Nominally the searcher waits for a variable interval (the Interest Interval, set according to the number of sensed neighbors) to collect all the interests, sort the requests and then send only one copy of the information requested in a burst of chunks during the variable Chunk Interval.

The same scheme can be applied to any node that has already received the new module. Upon the reception of the set of Interest any node that received the information, broadcasts the corresponding set of chunks within a Data packet.

In case of a packet loss, the Interest is retransmitted with its values properly adjusted to request the missing pieces, until all pieces are received. To reduce the collision probability from multiple simultaneous transmissions (that would happen, for instance, when multiple nodes respond to the same Interest
request) each node $i$ waits a random defer time $T_D$ before transmitting each burst of chunks through Data Packets. During this interval, the node $i$ checks the channel to detect other nodes’ transmission that are responding to any other requests. If no other node is transmitting, then $i$ sends the burst.

When the download is complete the searcher returns in phase (i) and starts again to wait for SOS messages.

VI. PERFORMANCE EVALUATION

In this Section, we compare the movement scheme computed by our algorithm, MeDrone, to four other movement schemes. In all five schemes the curing phase, described in Section IV is the same, whereas they differ in the path chosen for the searcher to move through the WSN.

The first scheme of movement is to let the searcher move according to the Random Way Point model (RWP) \cite{142}. This approach represents a simple yet effective scheme, where no effort is made to optimize the movements. The second and third alternatives consist in moving the searcher according to two predefined patterns, often used in the trajectory setting of planes and helicopters in order to cover the searching area during rescue operations: the Spiral and Billiard schemes \cite{143}. In the Spiral scheme, the searcher, starting from any position in the network lattice, will move to the lattice’s center and then start a rectangular spiral-like counterclockwise movement pattern, where the separation distance between the spiral arms is constant and equal to twice the estimated node transmission range. In the Billiard scheme, the searcher node will initially position itself in a corner of the lattice and then move and rebound against the lattice’s external boundaries. The angles of incidence with the boundaries are set in order to cover the whole area. Using this scheme, the searcher occupies the network lattice’s boundaries longer than its center. The fourth alternative is to use the set of waypoints that come from the model of Section IV and its implementation in Appendix A which is used as an upper bound for the performance evaluation. In the following we will refer to it as the MM (Mathematical Model) scheme. The Billiard and Spiral proposals just require the searcher to know a rough estimate of the network lattice extension, whereas the MM requires the complete a-priori knowledge of the network and nodes’ characteristics, as well as the complete description of the diffusion dynamics.

We have implemented RWP, Spiral, Billiard and MeDrone in the ns-2 \cite{144} simulator, while the MM has been coded in the FICO Xpress Suite and Mosel language \cite{145}.

The reference simulated topologies are multiple: (i) a $300 \times 300m$ area where the nodes are placed according to a bidimensional Gaussian distribution originating from a corner (identified as Gaussian), with a standard deviation of $150m$; (ii) a $160 \times 160m$ square area where a variable number of nodes are placed according to a regular grid pattern (identified as Grid) and (iii) a variable dimension area where the nodes are placed according to the realistic topology (identified as Realistic) generated using a specific framework called NPART \cite{146}. NPART is capable of designing network topologies whose statistical characteristics are similar to the ones measured in real networks. The number of nodes in all the topologies varies as follows \cite{147, 148, 149, 150}. Hence, the distances between each couple of nodes in the Grid vary accordingly \cite{142, 148, 149, 150} m. The Gaussian topology is characterized by the presence of an agglomeration point for the nodes, i.e. the center of the distribution, which has a higher relative density than the rest of the area. In the Grid topology, instead, the nodes are evenly distributed. As specified in the literature \cite{146}, the Realistic topology is composed of a set of connected clusters of nodes around the center of the network area, with multiple areas of node concentration. These features influence how effectively the proposed algorithm can find the most dense areas of infected nodes. For instance, the Grid topology does not have areas with an increased relative density of nodes, being these infected or not, and the Gaussian topology has a single one.

All the nodes are equipped with a single IEEE 802.11 interface and a Ricean fading model takes into account the multi-path effects due to various obstacles. As shown in Table II Physical and MAC parameters are based on IEEE 802.11g. Transmission power and receiver sensitivity figures are taken from data-sheets of devices available on the market \cite{147}.

The searcher and the first set of infected nodes are initially placed at random positions within the network lattice using a uniform distribution. Once the simulation starts, the searcher moves according to the mobility pattern computed by the RWP, MeDrone, Spiral, Billiard or MM schemes. The searcher speed is set according to the literature in this field \cite{148}. The results of simulation campaigns are presented in the following subsection.

Our main objective is to evaluate the evolution and the steady state number of infected nodes to understand which scheme gives the best results, in a given topology, for the following two metrics: (i) the capability of a scheme to limit the overall infection, measured as the peak number of infected nodes (Peak Value) and (ii) the time needed to cure the network, i.e., the time needed to reduce the number of infected nodes to the 10% of the total network nodes after the infection peak has been reached (Time To Settle, TTS). To give a clearer insight on how these values vary with topologies and movement schemes, they are reported separately in the last subsection, aggregating all the previous results in a wider view.

A. Outcomes of the simulation campaigns

Several simulation campaigns, varying the number of nodes in the network, have been performed. The first set of results considers a network of only 25 nodes. This first simulation campaign is meant to provide an insight on how each movement pattern works, when scarcely connected networks are
considered. Successive simulation campaigns, which consider an increasing number of nodes in the network, have been conducted to investigate the scalability of the proposal as well as the impact of nodes density on the performance of the movement schemes. Results for higher number of nodes are summarized in figures 3 and 4, which show the average of time to settle (TTS) and of the Peak Value respectively.

As a general trend we may observe that, when the number of nodes in the network increases, the epidemic spreads quicker and reaches a greater Peak Value. This is due to the greater number of connection between adjacent nodes that the epidemic can use to propagate. Also, the TTS increases with the number of nodes. This is essentially due to: i) the increased number of times the searcher must iterate the healing process (see sec. V-B) because of the increased number of infected nodes, ii) the fact that the epidemic continues to spread while the searcher is applying the cure, and iii) the fact that the increasing number of requests issued by nodes which struggle to be healed (see sec. V-A) tend to congestionate the communication channel thus slowing the healing process.

Following our evaluation metrics, the presence of an area with a large concentration of nodes in the Gaussian topology (Fig. 2a) gives the MM and MeDrone schemes a conspicuous advantage over the others in terms of TTS, as they are capable of exploiting the density increase in a limited area. Driving the searcher across empty areas at the lattice boundaries the Billiard gives instead poor results. In this topology the Spiral scheme performances suffers highly in the 100-nodes scenario while for lowest density it is capable to outperform the Billiard (Fig. 3a).

In the Grid topology the MM scheme retains again the lowest Peak Value and the lowest TTS, while MeDrone still obtain the second lowest TTS apart from the one obtained by MM, thanks to the capability to track down infections (Fig. 2b). Nonetheless, this is the scenario where the proposed scheme shows less improvements compared to its competitors. This is due to the fact that MeDrone cures the infection by starting from the most dense region in the network while a Grid topology does not present such a region. Hence, the performance of the other movements schemes, particularly those of the Spiral and RWP schemes, are close to that of the proposed scheme and sometimes present a lower Peak Value. It could be further noticed, that the Billiard scheme behaves particularly bad in this scenario. Specifically, as the number of nodes increases the performance of the Billiard scheme quickly worsen at such an extent that, for the highest number of nodes, it fails to completely eradicate the infection. This behavior is due to the intrinsic movement features of the scheme. In fact, in this case the searcher, travelling back and forth between the network area boundaries, occupies the borders of the lattice for longer than the lattice’s center. In this way it has less time to contact the nodes in the middle. Thus, in that area the infection is less constrained. In this topology, (Fig. 3b) the Spiral output is almost independent from the density whereas in the Realistic one (Fig. 5c) the TTS decreases with the number of nodes to increase again for the 100 nodes case. The RWP scheme instead shows an improvement of the TTS as density increases and a degradation in the Gaussian one as the same value decreases. In the Realistic topology (Figure 2c, in respect to Spiral, Billiard and RWP schemes, the MeDrone performs the best in terms both of Peak Value and of TTS. The searcher can move among the set of connected clusters of this topology, and briefly distribute the cure.

Figure 4 summarizes the trends illustrated so far. Analyzing the Peak Value, we show that, at least for the Gaussian and Realistic topologies, the MeDrone scheme is the best option as it is capable of reaching the various areas dense of infected nodes and quickly cure them. The only exception to this behavior, also shown in the previous set of figures, is for high total number of nodes, where the tracking algorithm is slow in identifying the areas with the highest number of infected nodes. In the Grid topology (Fig. 4b), the epidemic diffuses among evenly-placed nodes, thus the MeDrone cannot easily detect the target areas for low densities and also, when density is higher, the sheer number of nodes makes the very general epidemic diffusion difficult. In the same topology the Billiard and Spiral schemes fare reasonably the same. Results are different for the Gaussian topology (Fig. 4a), where the presence of a unique large concentration point close to the center supports better the MeDrone and Spiral schemes. For the Realistic topology (Fig. 4c) the best choice, apart from the MM, is again the MeDrone. The peculiar features of the node placement scheme, and the presence of various clusters, favors the RWP scheme instead of the Billiard one. In this last topology, the greatest concentration of nodes is near the center of the lattice, therefore the Spiral scheme outperforms the Billiard and the RWP schemes that waste time travelling through empty areas. The Billiard is again the worst performer. As a result, none of the Billiard, RWP and Spiral movement schemes can be considered as a valid alternative to MeDrone because they are all unable to guarantee a constant and good performance over a wide range of topologies and nodes densities.

VII. AN INTRODUCTION ON MULTIPLE SEARCHER COORDINATION

Up to this point, in this paper, we have considered just one searcher node. Increasing the number of mobile elements and introducing coordination is an important step in order to create a more efficient and complete framework. Unfortunately, the model developed so far for the TCP only includes the possibility of finding the optimal solution for a single searcher and an actual coordination environment between entities devoted to cure a network is beyond the scope of this paper. However, it is possible to initially evaluate how the network and the epidemic spread will react in presence of a very simple coordination scheme. For achieving this feature, we modified our previous algorithm for the identification of the most dense areas of infected nodes in order to keep track of a set of points instead of a single one. Specifically, the framework can keep track of the two most dense geographical areas and diffuse the information about them. We further considered two searchers. At each time the first searcher is committed to cure the most dense area while the second searcher will take care of the
second most dense area (see fig. 5). If a single aggregation of infected node is detected both the searchers converge on that region. Following this approach we got a first promising outlook about the feasibility and the convenience to extend MeDrone to a scenario where a platoon of drones is deployed to cure an infected network. To evaluate the results we used the same simulation environment of the Section [VI] and we focused on the scenario with 100 nodes, in order to have a dispersed network where a single searcher has demonstrated some difficulties in following the epidemic. Figures from [5] to [8] show the evolution of the epidemic comparing the single and the multi-node solutions for respectively the Gaussian, the Grid and the Realistic topologies. In detail, for each figure it is displayed the mutation dynamics of the MeDrone, together with each one of the other schemes for both one and two searchers. For the 2-searcher version the other schemes are modified as follows: for the RWP, a second set of random waypoints are followed by the second searcher; for the Billiard, the starting point for the second searcher is set on the other side of the network area and the rebound points are mirrored in respect to the original trajectory; finally, for the Spiral, the second searcher travels clockwise starting from the same point of the first one. For all the topologies, the MeDrone scheme shows an improvement in both in the time necessary to eradicate the infection (TTS) and the Peak Value. Important results can be obtained analyzing the behaviors of different schemes in the different topologies. In Figure [6] the results for the Gaussian topology are shown. The presence of multiple searchers driven
by the MeDrone scheme decreases the spread of the malware better than the other schemes. This happens because the two searchers are both driven to the center of the Gaussian where the infected node density is the highest.

This maximizes the initially covered area and quickly clean the infection. The improvements introduced by the 2-searcher version of the other schemes in the same topology are not as significant. In Figure 6a and 6c it is shown that, for the Billiard and Spiral schemes, when the two searchers move at the same time in the very same area, the values of TTS and Peak Value do actually increase.

Instead if the movement of the two searchers is loosely decoupled as in the RWP (Figure 6b), there is an improvement.

As in the case of the single searcher, the Grid topology is the most challenging one for the MeDrone scheme (see fig. 7). Even if it is possible to highlight an improvement when two searchers are deployed, the actual gains are limited.

This is mainly due to the dispersion of the nodes and the time spent by the searchers waiting for an area to be cured before moving to the next one. As shown in the previous sections, the Billiard scheme is extremely dispersive for the Grid topology (Figure 7a). In this case the RWP and Spiral schemes show similar results as the MeDrone with 1 and 2 searchers (Figure 7b).

The results relevant to the the Realistic are shown in Figure 8. The presence of multiple aggregation points with different densities (i.e. different clusters of nodes) allows the 2-node MeDrone scheme to achieve the best improvements among all the topologies. Selecting different areas and serving them in parallel shows that the Peak Value and TTS are extremely reduced (roughly one third of nodes becomes infected and half the time is necessary to cure 90% of nodes) using the simple scheme just described. The MeDrone scheme outclasses the Billiard with both 1 and 2 searchers (Fig. 8a), the RWP (Fig. 7b) with two searchers and it is slightly better than the Spiral in the same situation (Fig. 7c). We underline again that the Spiral, Billiard and RWP schemes need the information about the network area extension in order to calculate their starting points while, in the same situation, the MeDrone just needs to be in the same connected cluster. These results open the way to research on a practical implementation of a coordination system between nodes that can bring significant benefits to the management of proximity infections in WSNs.

VIII. CONCLUSIONS AND FUTURE WORKS

In this paper we addressed the issue of facing a malware spreading in a WSN. Using the concepts of controlled mobility and information diffusion we defined an algorithm (i) to notify the spreading of the malware, (ii) to lead an autonomous flying robot, which can cure infected nodes, along a path through the WSN and (iii) to heal the infected nodes. To benchmark the proposed algorithm, we developed a mathematical model that defines the best path to be followed by the flying robot to cure the infected nodes as quick as possible. Using the results obtained from the mathematical model as the upper bound and the ones from the application of random way-point movement as the lower bound, we have assessed the satisfactory performance of the proposed algorithm. Finally, we have given an outlook on the directions to extend the proposed algorithm by using multiple searchers. As a future work we intend to further address the use of multiple searchers and to develop a larger and more comprehensive mathematical model, in order to provide more generalized solutions, which can be applied in real scenarios. We expect also to address the creation of a coordination framework based on solutions that do not rely on absolute coordinate systems.

APPENDIX A

MATHEMATICAL FORMULATION

In the following subsections we describe a formulation for the mathematical programming of the problem that only involves binary variables and linear constraints. In a first place the various states a node can be into and the transitions among them are characterized. The parameters used in the model are then introduced and defined as long with the optimization variables and their roles. In the end the model constraints and the objective functions are detailed.

Parameters

All the parameters of the model are described in the following list.

Regarding the last parameter, it is to note that in the model, possessing the cure is a property of a node and means that, at certain time, the searcher is in the same place as the node possessing the property.

The content of the set of wireless links and other parameters come from a preliminary simulation campaign in which were analyzed the behaviors of the nodes. The aforementioned set was filled from the adjacency matrix of the various topologies considered in Section VI while the times necessary for the searcher to physically travel are computed according to the average speeds of flying robots.

As the modification of the infection dynamics occurs as the searcher moves, the parameter $T_m$ (movement interval) is the smallest time unit considered for the model and we define the others using the former as a measurement unit. In this way...
the parameter $T_1$, comes from an analysis of the time needed to complete the transfer in a preliminary set of simulations, quantized in $T_m$ units. The parameter $T_m$, that describes the time needed for the searcher to access the secured part of an infected node S.O, is set again accordingly but, as in the simulations it was almost negligible, we set it as the same as $T_m$.

**Variables**

We classify the optimization variables in two groups. The first group (tab. III) contains the variables of the problem, describing, for each time unit, the state of each node and the searcher movements. The second group (tab. IV) describes the variables imposing conditions in the model.

**Objective function and constraints**

The purpose of the objective function is to minimize the incurrence of the time unit at which there are no more malicious or infected nodes left.

$$\min \sum_{t \in T} x^t.$$  

Two groups can be used to classify the actual formulation constraints:

- Constraints of Equation 1 ensure that the problems variables take the admissible values.
- Constraints of Equation 2 are a mathematical translation of the transitions of Fig. 1.
Data

$T_m$ movement interval, the running time needed to choose a new direction for the searcher

$T_i$ running time for a node to download a block of software representing either the cure or the malware

$T_a$ running time for the searcher to access the secured part of a node operating system

$N$ set of all nodes

$S$ set of possible states for each node

$E$ set of wireless links, bidirectional

$E_c$ set of physical paths, bidirectional

$d_e$ time needed to travel along the edge $e$, in time units

$d_i$ total time needed for a node $i$ in time units to complete the transfer of a module

$N^*$ subset of nodes initially infected

$t^*$ time unit at which the searcher enters the network

$i^*$ node at which the searcher is inserted

State and decision variables

$x_{si}^t$ equal to 1 if node $i$ is in state $s \in \{1, 3, 5\}$ at the beginning of period $t$

$x_{pi}^t$ equal to 1 if node $i$ is in state $s \in \{2, 4\}$, since $p$ time units, at the beginning of period $t$

$y_i^t$ equal to 1 if node $i$ carries the searcher at the beginning of period $t$

$z_{ij}^t$ equal to 1 if the searcher is starting to travel from $i$ to $j$ at the beginning of period $t$

TABLE III: State and decision variables

Namely, the constraints translate the aforementioned logical conditions into linear constraints. Taking for instance conditional variable $Y_{i}^t$, the first two constraints imply that $Y_{i}^t$ takes value one if the searcher is located at node $i$ in period $t$ or connected to $i$ through an edge in $E$. Then, the third constraint implies that the variable takes value zero if the searcher is neither located at node $i$ nor connected to it through an edge.

The equations of $Y_{i}^t$ are non-linear. However, as the involved variables are all binary, we can use a well-known technique to linearize their products. Each product between any binary variable $v_1, v_2$ in the above equation can yield a new real variable $\text{res}$ that satisfies equation 3

$$\begin{align*}
\text{res} & \leq v_1 \\
\text{res} & \geq v_1 + v_2 - 1.
\end{align*}$$

For the sake of simplicity, it is avoided to rewrite all the linearized constraints.

The formalization of the searcher’s movement is formalized

Conditional variables

$Y_i^t$ equal to 1 if node $i$ is connected to the searcher

$M_i^t$ equal to 1 if node $i$ is connected to at least a node in state 3 and not the searcher

$C_i^t$ equal to 1 if node $i$ is connected to at least a node in state 1, 2 or to the searcher

$S_i^t$ equal to 1 if node $i$ is not connected to nodes in states 3 and not connected to the searcher

$X_i^t$ equal to 1 if at least one node is infecting or malicious at the beginning of period $t$

TABLE IV: Conditional variables

$$\begin{align*}
Y_i^t & \geq y_i^t & i \in N, t \in T^* \\
Y_i^t & \geq y_i^t & ij \in E, t \in T^* \\
Y_i^t & \leq y_i^t + \sum_{ij \in E} y_{ji} & i \in N, t \in T \\
M_i^t & \geq x_{3j}^t - y_i^t - \sum_{ij \in E} x_{ji}^t & ij \in E, t \in T \\
M_i^t & \leq \sum_{ij \in E} x_{ji}^t & t \in T \\
M_i^t & \leq 1 - y_i^t - \sum_{ij \in E} y_{ji}^t & t \in T \\
C_i^t & \geq x_{3j}^t & ij \in E, t \in T \\
C_i^t & \geq x_{2j}^t & ij \in E, t \in T, p = 0, \ldots, d_j - 1 \\
C_i^t & \geq x_{3j}^t & ij \in E, t \in T \\
C_i^t & \geq y_i^t & i \in N, t \in T \\
C_i^t & \geq y_i^t & ij \in E, t \in T \\
C_i^t & \leq y_i^t + \sum_{ij \in E} (y_{ji}^t + x_{1j}^t + x_{3j}^t + \sum_{p=0}^{d_j-1} x_{2j}^p) & t \in T \\
S_i^t & \leq 1 - y_i^t & i \in N, t \in T \\
S_i^t & \leq 1 - y_i^t & ij \in E, t \in T \\
S_i^t & \leq 1 - x_{3j}^t & ij \in E, t \in T \\
S_i^t & \geq 1 - \sum_{ij \in E} (y_{ji}^t + x_{3j}^t) - y_i^t & t \in T \\
X_i^t & \geq \sum_{i \in N} \frac{x_{2i}^t + x_{3i}^t}{2|N|} & t \in T
\end{align*}$$
in Equation 4

\[\begin{align*}
y_i^t &= 0 & & i \in N, t \leq t^* - 1 \\
y_i^{t^*} &= 1 \quad & & i \in N \\
y_i^t &= 0 & & i \in N \setminus N^* \\
y_j^t &= \sum_{ij \in \mathcal{E}_c} z_j^t - d_j + y_i^{t-1} - \sum_{ij \in \mathcal{E}_c} z_j^{t-1} & & i \in N \setminus N^* \quad (4) \\
\sum_{ij \in \mathcal{E}_c} z_j^t &\leq y_i^t & & i \in \mathcal{E}_c, t \in \mathbb{T} 
\end{align*}\]

Ending the discussion, the constraints of Equation 5 describe the system’s starting conditions, and enforce that all optimization variables be \( \{0, 1\} \)-valued.

\[\begin{align*}
x_{ij}^0 &= 1 & & i \in N \setminus N^* \\
x_{ij}^0 &= 1 & & i \in N^* \\
x_{ij}^0 + x_{ji}^0 + x_{ij}^0 + \sum_{p=0}^{d_i-1} \left( x_{i1}^{p0} + x_{ji}^{p0} \right) &= 1 & & i \in N \\
Y, M, C, S, X, x, y, z &\in \{0, 1\} 
\end{align*}\]

REFERENCES


