Computer Ecology: Responding to Mobile Worms with Location-Based Quarantine Boundaries

Baik Hoh
WINLAB, ECE Department
Rutgers, The State University of New Jersey
baikhoh@winlab.rutgers.edu

Marco Gruteser
WINLAB, ECE Department
Rutgers, The State University of New Jersey
gruteser@winlab.rutgers.edu

Abstract

The emerging malware that can spread through local wireless networks among mobile devices has so far received less attention than computer worms in the Internet. The local wireless links provide an alternative propagation path that circumvents intrusion detection at the service provider gateways. On the mobile nodes, conventional intrusion detection and intrusion response techniques such as address blacklisting and content filtering are more difficult to deploy due to the lack of central entities and the resource constraints of mobile nodes.

We propose a new architecture for an intrusion response system that takes advantage of an infrastructure network (e.g., cell phone network) to manage security of the mobile nodes. Infection patterns in ad hoc networks are highly correlated with geographic proximity. Thus an ecologically inspired diffusion-reaction and advection models can provide estimates for the current spread of the worm. These estimates allow the service provider to precisely target a containment response.

1 Introduction

A current trend in pervasive devices is towards multi-radio support, allowing direct local interaction between devices in addition to maintaining long-haul links to infrastructure networks. Many current cell phones already contain Bluetooth radios that enable peer-to-peer exchange of files and usage of services from nearby devices. Bluetooth is also available in some automobiles and the US Federal Communications Commission has reserved spectrum for Dedicated Short Range Communications (DSRC), a wireless communications standard for inter-vehicle networks based on the IEEE 802.11 medium access protocol [10]. Example applications are collaborative crash warning and avoidance, dynamic traffic light control, or ad hoc forwarding of traffic probe information [26, 27].

Unfortunately, peer-to-peer interaction between devices provides an alternative propagation path for worms and virus. The Internet experience illustrates that worm attacks are a significant concern and a proof-of-concept Bluetooth worm, Cabir, has already been implemented.[1] More aggressive worms that exploit bugs and make unwanted phone calls are not hard to imagine [5, 24], and likely as financial incentives increase.

Regardless of the sophistication of the prevention strategies, in an environment with high reliability requirements it is only prudent to also plan for outbreaks with appropriate containment strategies. Peer-to-peer replication over short-range wireless networks creates a challenge for intrusion detection and response, because the worm cannot be observed and blocked by intrusion detection and response systems (IDSs) in the cellular service provider's core network. Instead intrusion detection must be deployed on resource-constrained mobile devices or on specialized honeypot devices distributed in high-traffic zones [28, 2]. Re-

[1] In fact, a Cabir outbreak was recently reported during a sporting event at the Helsinki Olympic Stadium [21] and rumors are abound that it could spread to in-car computers of a luxury sport utility vehicles.

gardless of the employed intrusion detection method, these constraints will lead to a delay between the time of outbreak and alarm because of distributed processing delays and human analysis. Thus, the intrusion response system only has at best an outdated few of the current worm propagation.

In this work, we consider an intrusion response architecture where a service provider remotely administers mobile nodes over the wide-area infrastructure wireless network. Using ecologically inspired [18] location-based quarantine boundary estimation techniques, the service provider can estimate a set of likely infected nodes. This allows the service provider to concentrate efforts on infected nodes and minimize inconvenience and danger to non-affected parties.

The remainder of this paper is structured as follows. Section 2 clarifies threat model and system assumptions. It also defines the estimation problem that this paper addresses. Section 3 develops a quarantine boundary estimation algorithm from ecological diffusion-reaction and advection models. We evaluate our proposed algorithm by applying it to two ad hoc network scenarios: a pedestrian random-walk and an a vehicular network on a highway. These results are reported in section 4. In section 5, we analyze the simulation results and discuss the effectiveness of the approach. Section 6 compares our work with directly related prior works before we conclude.

2 Threat assessment

We consider a network system that comprises mobile radio nodes with ad hoc networking capabilities and a wide-area wireless infrastructure network with central network management by a service provider. Each mobile node is connected to the infrastructure network, provided that radio coverage is available, and can directly communicate with other mobile nodes over a short-range radio interface. Examples of such a system are a CDMA/GSM cell-phone network with Bluetooth handsets or an automotive telematics system supporting CDMA and DSRC . We also assume that the service provider can locate each mobile node. This could be implemented through Assisted GPS on the nodes or triangulation technology in the infrastructure. Hybrid approaches are also possible.

In this network system, worms and viruses may spread through ad hoc connections over the short-range interface, rather than the infrastructure network. Mobile nodes can be infected if they are susceptible and a neighbor, meaning in the communication range C_r, of an already infected node. Typically, an infected node is able to identify all its neighbors through a network discovery mechanisms (e.g., IEEE 802.11 probe request, probe response protocol). In a sparse, mostly disconnected network, nodes will be infected when they first enter the communication range of an infected node. While most current Bluetooth worms need to be manually accepted by the receiving devices owner, the increasing complexity of the software stack software vulnerabilities more likely. We assume that future worms will exploit software vulnerabilities and do not require user interaction to spread.[2]

Malware spreading over the ad hoc network is more difficult to detect and contain than malware spreading over an infrastructure network, because the network does not contain concentration points (chokepoints) where centralized network-based intrusion detection and traffic filtering techniques can be applied. Instead detection and response techniques must be implemented in a highly distributed architecture on the mobile nodes themselves. While it is plausible that malware propagates over both the short-range *and* the infrastructure network, we ignore this case here because the infrastructure connections can be prevented with traditional defenses.

We are especially concerned with unknown malware, which signature-based intrusion detection systems cannot yet detect. The service provider may learn of a new epidemic through different mechanisms ranging from mundane user calls to its service hotline to a sophisticated anomaly detection system. We observe that any of these mechanisms suffer from a high false-alarm probability and thus require the intervention of human analyst to verify that an actual outbreak exists. This leads to a detection delay of minutes in the best case. Even in a fully automated system, a distributed intrusion detection system would add delay due to the distributed detection processing and the latency overhead of delay-tolerant communication. During this

[2]The Bluesmacking attack on Bluetooth devices already exploits a buffer overflow vulnerability present on some systems. In addition, a vulnerability that allows arbitrary code execution has been discovered in the WIDCOMM software stack in 2004.

time the malware can spread further (and anomaly reports from new nodes may again require verification) leaving the analyst with an incorrect, delayed view of the epidemic.

This work assumes, however, that the analyst can accurately locate *patient 0*, the initially infected node. If every node runs an intrusion detection system with sufficient memory for logging events, the infection can generally be traced to its origin. An inaccurate estimate of patient 0's position will lead to degraded system performance. We leave making the system more robust to the patient 0 estimate for future work.

In summary, the service provider will determine from a range of clues whether an intrusion took place. The service provider characterizes an intrusion by a tuple $(pos_x, pos_y, time)$ that describes the time and position of patient 0 at the start of the outbreak.

2.1 Intrusion Response

Given that an intrusion event occurred, a service provider's main interest lies in minimizing inconvenience and potential danger (e.g., users may depend on cell phones for 911/112 emergency calls or distractions from an infected in-vehicle system may cause car accidents) to customers.

Responding effectively requires a secure management interface to the mobile nodes that allows service providers to remotely regain control of a compromised mobile node. Remote management interfaces are common practice for managing servers in larger data centers and have become increasingly prevalent in the cell phone world. For example, the Open Mobile Alliance Client Provisioning Architecture [1] allows over-the-air configuration of mobile nodes. It also specifies a privileged configuration context, whose settings cannot be modified by users or applications. Such interfaces could be further hardened to ensure availability when malicious code controls the phone.

Given an over-the-air provisioning architecture, possible responses to an intrusion event include:

1. Sending a warning to users of the mobile nodes
2. Deactivating mobile nodes
3. Disable the short-range network interface on mobile nodes
4. Installing port or content-based filters
5. Installing patches to remove exploits
6. Provisioning patches to remove the worm

All of these responses can slow or stop the spread of the virus, however, they also incur user inconveniences of its own. For example, frequent use of response 1 may reduce its effectiveness, response 2 may prevent emergency calls, and response 3 may prevent the use of handsfree operation by drivers. Responses 3-6 require a more detailed understanding of the worm implementation and so may allow the worm to spread unrestricted for a period of hours or days. Even then, installing hastily developed patches often leads to failures on a subset of phones.

We define the *intrusion response planning problem* as identifying an optimal set of nodes to minimize the impact of the worm *and* the inconvenience and dangers cause by (partial) service outages due to the response. An optimal response plan only targets nodes that have already been infected or will be infected until the provisioning process is completed.

3 Quarantine Boundary Estimation

The optimal response set can be best found through an estimation technique because the service providers knowledge about the spread of the mobile worm is incomplete. Anomaly reports usually trickle in only after nodes are infected and may be severely delayed in areas of sparse coverage from the infrastructure wireless network.

3.1 A Macroscopic Model of Worm Propagation

Diffusion-reaction and advection models [18] have been successfully applied to describe the spatial and temporal distributions of diverse phenomena ranging from animal dispersion [3] to groundwater contamination.

The diffusion-reaction model comprises a diffusion process and a reproduction process. The diffusion process describes random movements and is characterized by the diffusion coefficient D. The reproduction process describes the exponential population growth and is specified by parameter α. Equation 1 specifies

[3] An early notable application of diffusion-reaction model was designing a hostile barrier for stopping the dispersal of Muskrats. In 1905, Muskrat was imported to Europe but some of them escaped and started to reproduce in the wild [6]. Skellam [23] later modeled the dispersal of Muskrats though a diffusion-reaction equation.

Model Parameter	Correspondence in automotive scenario
Diffusivity	Models minor roads and collector streets or pedestrian movements
Growth rate	Rate of new infections depends on density and distribution of susceptible nodes, communication range, and node velocity
Origin	Positions of initially infected nodes

Table 1. Mapping of model parameters to automotive networking scenario.

the diffusion-reaction model. It assumes polar coordinates centered at the position of an initially infected node, isotropic dispersal with constant diffusivity D, and growth proportional to the population density S.

$$\frac{\partial S}{\partial t} = \frac{D}{r}\frac{\partial}{\partial r}\left(r\frac{\partial S}{\partial r}\right) + \alpha S \tag{1}$$

This model has a closed form solution by solving under the initial condition that at time $t = 0$, m infected nodes are concentrated at location of patient 0 ($r = 0$). From this solution shown in equation 2, the radius R of the frontal wave can be calculated from the propagation speed which depends on α and D as described in equation 3.

$$S = (m/4\pi Dt)\exp(\alpha t - r^2/4Dt) \tag{2}$$

$$R = 2\sqrt{\pi\alpha D}t \tag{3}$$

Thus the propagation boundary is proportional to the time since the outbreak t and the boundary moves with velocity $v = 2\sqrt{\pi\alpha D}$. The parameter α and D are depended on the exact scenario. Table 1) identifies the parameter dependencies in an automotive scenario.

When a toxic pollutant diffuses going along the groundwater paths, advection term describing a mean flow is added to the diffusion-reaction model [22]. Advection term is governed by the velocity u in x-axis and v in y-axis.

If we take an advection effect and ignore a diffusion process, equation 1 is changed into an advection equation model described by equation 4.

$$\frac{\partial S}{\partial t} = -\frac{\partial}{\partial x}(uS) - \frac{\partial}{\partial y}(uS) + \alpha S \tag{4}$$

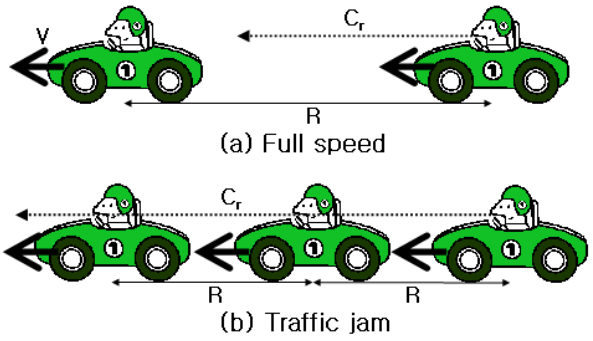

Figure 1. Different proportions of inter-vehicle distance to communication range lead to different worm propagation velocities.

This model can be used in modeling the behavior of mobile worms in highway networks (e.g., Southern New Jersey Highway Networks).

3.2 Algorithms

Given an initial position of each infected node i, (x_i, y_i) for all i at time T_o, the algorithms should estimate the frontal wave of propagation at

$$T_c = T_o + T_\Delta$$

, where T_o is the time of outbreak and T_Δ means time delay. We can divide the problem into estimating the worm propagation velocity and estimating the spatial distribution.

In an ad hoc network where mobile nodes move randomly in x-y coordinates, the propagation speed is governed by equation 3. Constant diffusivity D and reproductivity α guarantees constant propagation speed. As long as the same node density and velocity is maintained, the propagation velocity remains constant.

However, in the vehicular scenario, every road segment may have a different propagation velocity because vehicle speeds and inter-vehicle distances differ. Figure 1 illustrates how the relationship between communication range and inter-vehicle distance affects propagation velocity. In the case (a) the inter-vehicle distance R is greater than the communication range C_r, so that an infected car cannot communicate with neighboring cars. Thus, the propagation velocity V' is solely determined by the vehicle speed V. In case

(b) however, the communication range is greater than the inter-vehicle distance. Thus the worm can travel over the wireless medium to the foremost car in communication range in addition to the vehicle speed. If a worm manages n such hops per second, this leads to the following equation.

$$V' = \begin{cases} V + nR\left\lfloor \frac{C_r}{R} \right\rfloor & \text{if } R \leq C_r \\ V & \text{else} \end{cases}$$

Because a one hop communication can never go farther than C_r, an upper bound for V' can be obtained by substituting C_r for $R(C_r/R)$, yielding

$$V' = V + nC_r \tag{5}$$

The inter-vehicle distance R and mean vehicle speed V on each highway segment can be obtained from Department of Transportation inductive loop sensors on an hourly basis, for example. They could also be inferred from tracking the position of probe vehicles on the highway network.

Given this propagation velocity, a straightforward isotropic estimate for worm distribution can be obtained with the diffusion-reaction equations. For each independent outbreak this approach yields a circular boundary estimate centered at the location of patient 0 (at the time of the outbreak). The radius of the circle increases linearly with the time duration T_Δ since the outbreak.

This approach is suitable when nodes movements do not exhibit any directional trends, such as in a random walk. Estimation can be improved, however, when mobile nodes move on an underlying network of roads or walkways. We frame our discussion of this algorithm in the context of an automobile vehicular ad hoc network, but the concepts are generally applicable to nodes that follow a network of paths.

This algorithm assumes the availability of cartographic material so that the position of patient 0 at the initial outbreak can be mapped onto a road segement. The maps must contains road classifications and the geographical positions of roads and their intersections. For example, this data is available from the US Geological Survey, which published detailed transportation network information in the spatial data transfer standard. These maps also classify roads into expressways, arterial, and collector roads, according to their size and traffic volume. The algorithm also requires a mapping of the position of patient 0 at the time of outbreak onto a road segment. This mapping can be achieved by finding the road segment with the minimum euclidian distance to the patient 0 position.

The key idea of this algorithm is to build an advection model using the transportation network information. The underlying heuristic is that the maximum propagation speed will be observed along the road network—propagation across parallel road segments in communication range and along smaller roads is ignored by this heuristic. The algorithm 1 follows all possible propagation paths using a traversal of the road network graph and a propagation speed estimate for each road segment. It outputs a polygon that includes all (partial) road segments that a worm could have reached in the time since the outbreak.

For example, consider the section of the southern New Jersey highway network in figure 2. Assume that patient 0 lies on the link L_n between junction 3 (J3) and junction 4 (J4). If we know the propagation speed V_n on that link, we can calculate after how much time a mobile worm arrives at either junction. Let us denote T_3 and T_4 for the arrival time at J3 and J4. If the time since outbreak

$$T_\Delta = T_c - T_o$$

is greater than T_3, the mobile worm has already passed this junction and has most likely propagated along both the link J1-J3 and the link J2-J3. This process is repeated for each link until a junction with arrival time greater than T_Δ is found. This segment is then only partially infected and the infection boundary is known based on the estimated link propagation speed. The same process is also repeated in the opposite direction from patient 0, towards J4. The algorithm then encloses each fully infected link in a rectangle with length and width set to the road length and road width, respectively. Partially infected links are only enclosed up to the infection boundary. All rectangles are then merged into a polygon.[4] Once we get a polygon, we group nodes within a polygon into the optimal response set by using 'Point-In-Polygon Algorithm [7]'.

[4]This can be implemented using well-known algorithms such as provided by the *polybool* function [12] in MATLAB

Algorithm 1 $QuarantineBoundaryEstimation$ generates a polygon which estimates the frontal wave of mobile worms at T_r given $Patient0$ at T_0.

{Inputs: $Patient0$, the position of initially infected node; T_0, the time of outbreak; T_c, the time of intrusion response; v_n, the average car speed on nth road segment; R_n, the average distance beween adjacent cars on nth road segment;
Parameters: J_n, nth junction's x and y coordinates and every junction should have information on its neighbor junctions; C_r, Communication range
Outputs: Quarantine polygons}
(A) Estimate the worm propagation speed, V_n for all n with v_n and R_n
if $R \geq C_r$ **then**
  $V_n = v_n$
else
  $V_n = v_n + \alpha * C_r$
end if
(B) Estimate the spaitial distribution
Calculate $T_\Delta[0][0] = T_c - T_0$.
Locate the link (L_n) which $Patient0$ lies on.
Set $Patient0$ as the starting points of traversal and push it into queue, $Q[0]$
Keep pushing all junctions in two ways to be visited next in Q until the last level
$i = 0$;
while Any $T_\Delta[i][] \geq 0$ **do**
  $i++$
  K = the number of elements in $Q[i][]$
  for $j = 1$ to K **do**
    Save the parent junction of $Q[i][j]$ into $Prev$
    $T_j = \frac{D(Prev,Q[i][j])}{V_n}$ where n is the link index between Prev and $Q[i][j]$
    $T_\Delta[i][j] = T_\Delta[i-1][parent]$
    if $T_\Delta[i][j] \geq T_j$ **then**
      Generate a rectangular boundary from $Prev$ to $Q[i][j]$
    else
      Generate a rectangular boundary from $Prev$ to $T_\Delta[i][j] * V_n$
    end if
    $T_\Delta[i][j] = T_\Delta[i][j] - T_j$
  end for
end while
Merge all rectangular boundaries into polygon.

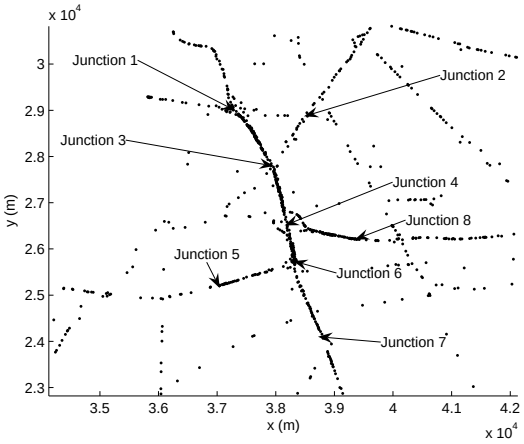

Figure 2. In our target map, there are 8 junctions and 7 links between them. This region is the part of Southern New Jersey Highway Networks. Every black dot depicts the position of individual car at specific time.

4 Evaluation

This evaluation studies the performance of the quarantine boundary estimation algorithms in a random walk and a vehicular ad hoc network scenario. We compare the accuracy of the macroscopic quarantine boundaries against infection patterns generated by a microscopic simulation model.

4.1 Metrics and Measures

Informally, the algorithm should maximize the number of infected nodes within the boundary and minimize the number of clean (uninfected) nodes within it. We measure the accuracy of the quarantine boundary estimation through detection, false-alarm probability, and Jaccard similarity.

The *detection probability* is defined as the ratio of infected nodes within the boundary to all infected nodes. More formally, $P_d = \frac{i}{I}$, where P_d is the detection probability, i is the number of infected nodes within the boundary and I is the total number of infected nodes. We define the *false-alarm probability* as the ratio of clean nodes within the boundary to all clean nodes. Accordingly, $P_f = \frac{c}{i+c}$, where P_f is the false alarm probability, c is the number of clean nodes

within the boundary and C is the total number of clean nodes. Notice that $c + i$ is the number of nodes within the quarantine boundary and $C + I$ is the total number of nodes in the scenario. A perfect quarantine boundary has a detection probability of 1 and a false-alarm probability of 0.

The *Jaccard similarity* J provides a convenient way to combine both probabilities. It is defined as shown in equation (6), where X is the optimum quarantine boundary in x-y coordinates and Y indicates an estimated quarantine boundary.

$$J = \frac{2\left(|X \bigcap Y|\right)}{|X| + |Y|} \tag{6}$$

It can be computed from detection and false alarm probabilities by substituting $X = I$ and $Y = i + c$, yielding equation (7).

$$J = \frac{2P_d(1 - P_f)}{1 + P_d - P_f} \tag{7}$$

The Jaccard similarity lies in the interval $[0, 1]$ with 1 indicating a perfect estimate, corresponding to detection probability 1 and false-alarm probability 0.

4.2 Simulation Model

We use the SIR model [3] for implementing the dynamics among susceptible nodes, infected nodes and recovered nodes. This model is characterized by the fraction of nodes that are susceptible to infection, the infection probability when a susceptible node is in contact with an infected node, and a recovery probability. In our model a susceptible node is in contact with an infected node, if they are in communication range C_r of each other.

Generally, we chose aggressive parameters for our simulations to evaluate a near worst-case worm. We set the infection probability to 1, which assumes the absence of any communication errors. In other words if a susceptible node is within the communication range of an infected node it becomes infected. We assume that infected nodes can only be recovered by the service provider, that is they must be within the quarantine boundary. Worm propagation then depends on the communication range and the exact mobility model.

We choose the initially infected nodes randomly among all nodes in random walk scenario. However, in VANET scenario, we choose them only on the link between J3 and J4, which is the center of our concerned map in figure 2. The initial positions of initially infected nodes is independent from the performance of our quarantine boundary algorithm, but placing them on that link enables us to extend the simulation duration.

For a random walk scenario, I choose 5 seconds as T_Δ. After T_Δ elapsed in pedestrian scenario, the number of infected nodes amounts up to 40-50% of whole nodes and the propagation for each initially infected node covers up to the circle with about 13m radius. Because our network is 50m by 50m, this amount of T_Δ is appropriate to measure detection, false alarm probabilities. In VANET case, I vary a time delay, T_Δ from 25 seconds to 45 seconds. In the case of T_Δ =45 seconds, the propagation approaches almost 5 links out of all 7 links.

For the *random walk model*, we chose parameters to reflect dense pedestrian movements with short-range (e.g., Bluetooth) communications. Node density is varied from 100 to 300 in a 50m by 50m area with node velocity ranging between 1m/s to 3m/s. Communication range is set to 5m, 10m, and 20m, to represent different path loss and interference environments.[5]

For the *vehicular scenario*, we obtained location traces from a microscopic traffic model for the PARAMICS transportation system simulator [14]. The model is calibrated to real traffic observed in a section of the southern New Jersey highway network. [19] The full simulation model contains 2162 nodes, approximately 4000 links and 137 demand zones, from which serve as origins and destinations for vehicles. Out of all vehicles in the simulation model a fraction of susceptible vehicles are selected randomly during the simulation process as they leave their respective origin zones. This ensures that the overall traffic patterns remain realistic even though we assume that only a percentage of cars is equipped with susceptible communications equipment. At each time step of the simulation (0.5 seconds), the x and y coordinates of the susceptible vehicles are recorded until they reach their

[5]These parameters approximate a sport event environment such as the one in the Helsinki Olympic Stadium, where an outbreak of the Cabir virus was reported [21].

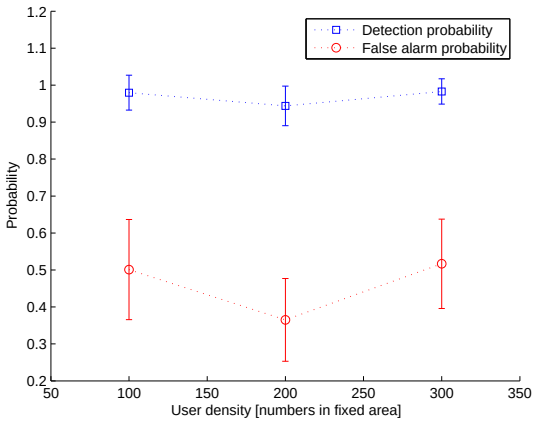

Figure 3. Estimation accuracy of diffusion-reaction model for random-walk scenario.

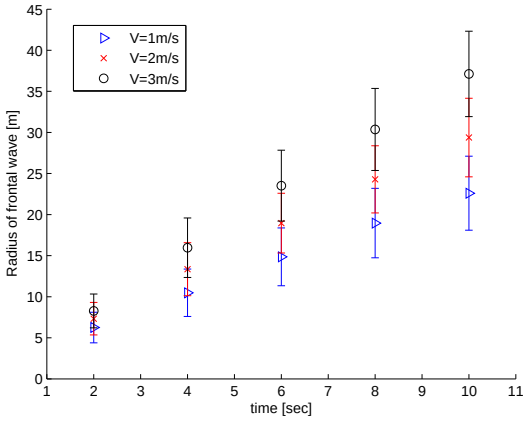

Figure 4. Distance of the farthest infected node from the outbreak position over time. Increasing node velocity has an additive effect on propagation speed. Propagation speed remains constant over time.

destination zones. For a low susceptibility scenario we selected 200 vehicles and for a moderate susceptibility scenario we chose about 1800 random cars. This represents about 5% of total traffic during the simulation which was restricted to 4min 10s, for computational tractability. The communication range is set to 50m, 100m and 200m in this scenario. 200m approximates free space propagation of a DSRC system [11, 20], while the shorter ranges model higher path loss environments, such as in congested traffic.

4.3 Pedestrian Scenario Results

To gain a better understanding of the effect of different model parameters we first discuss results from the less complex diffusion-reaction estimation model. The estimator's worm propagation speed is set to 2.56 m/s and the time delay T_Δ is set to 5 seconds for these experiments.

Figure 3 shows estimation accuracy of the diffusion-reaction estimator for different node densities. Mean and standard deviation for one hundred trials are shown. A mean detection probability between 95%-100% can be achieved with a false alarm rate of approximately 40%-50%. Our quarantine method behaves slightly more effective in the 200 node network because the worm propagation speed best matched this case. A change of +/-100 nodes increases the false alarm probability by about 10%.

The following results analyze the worm propagation speed in more detail. The speed is affected by node density, communication range, and node mobility. Figure 4 shows the distance of the farthest infected node from original position of patient 0 over differnt node velocities. Node density is set to 200 in the 50m by 50m region and communication range is 10m. Again, the graph show mean and standard deviation over one hundred trials. As expected, propagation speed increases with node velocity. An increase in node velocity has an additive effect on propagation speed. The graph also exposes that propagation speed remains constant over time, further supporting that a linear model fits well. A linear regression for v=2m/s yields intercept 2.1 and slope 2.8m/s.

The effect of changes in communication range C_r to worm propagation speed are shown in figure 5. Node velocity is set to 1m/s and other parameters remain the same as before. Propagation speed increases with higher node velocity. A larger communication range increases the likelihood that susceptible nodes are in rage, which hastens the spread of the worm. Propagation speed remains near-constant over time for each communication range.

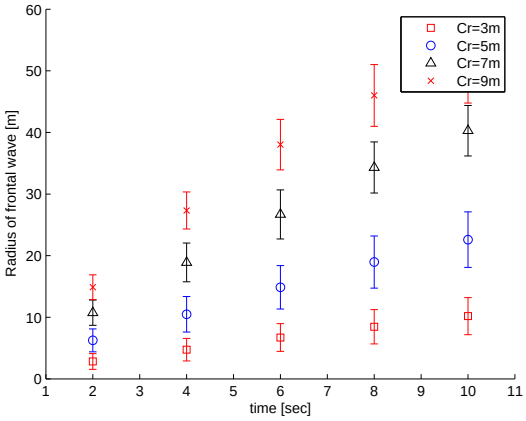

Figure 5. Dependency of propagation speed on communication range C_r. A larger communication range increases the likelihood that susceptible nodes are in rage, which hastens the spread of the worm.

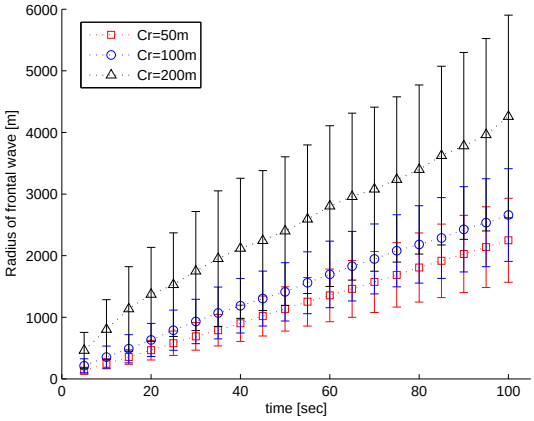

Figure 6. Worm propagation in highway model with 5% of vehicles susceptible

4.4 Vehicular Scenario Results

The first experiments measures the worm propagation velocity that can be expected in a highway outbreak. While prior works [25, 4, 9] have developed analytical equations for information propagation speed on road networks, these are not easily transferable to the worm scenario. The average radius of frontal wave is estimated by averaging 50 simulations and it is repeated for different communication ranges (50m, 100m and 200m). The estimated radius of frontal wave is shown in figure 6. The results show that for a communication range of 200m, the worm travels at a mean velocity of about 75m/s, significantly faster than typical highway traffic. Lower communication ranges result in reduced velocity.

The next experiment compares the estimation accuracy of the advection model over the diffusion-reaction model in the highway scenario. The communication range is set to 100m. Figure 7 and figure 8 show the detection and false alarm probability, respectively. The results from the advection algorithm described in section 3 are labeled "advection with analytical model". To allow a more detailed analysis, the graphs also contain two additional curves, which assume that a more precise estimate of worm propagation speed is available. In the "advection with same speed" approach, we use the average worm propagation speed (obtained from the previously described simulation) for all road segments. The "advection with different speed" approach, uses more detailed speed estimates, one per road segment, also derived from simulations.

The figures show that the advection models achieve superior detection probability over the diffusion-reaction model, while the false-alarm probability does not differ more than about 10% between advection and diffusion. The detailed knowledge about information propagation speed does not lead to a discernible improvement in detection probability. However, when worm propagation speed is known per road segment, the mean false alarm probability improves by up to 10%. This shows that at least slight improvements to the presented estimation techniques are possible.

5 Discussion

Estimation will necessarily lead to imperfect containment. Can this effectively slow worm propagation? We model the accuracy of quarantine boundary through an immunization probability P_{imm} between 0.8 and 1 and simulate worm propagation in the pedestrian random-walk scenario after such an imperfect containment. Figure 9 depicts the infection rates after one containment was performed at $T_c = 5seconds$. Detection probabilities greater than 0.95%, such as

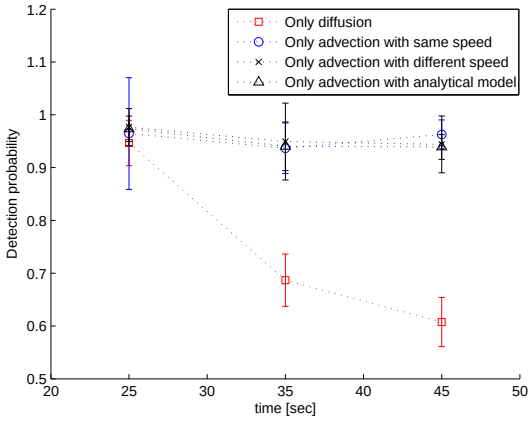

Figure 7. Detection probability on highway network. The advection models achieve superior accuracy over the diffusion-reaction model.

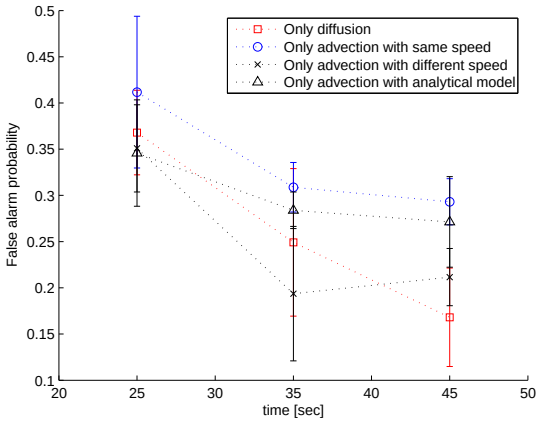

Figure 8. False-alarm probability on highway network. The advection model's better detection probability does not lead to a significant increase in false alarms.

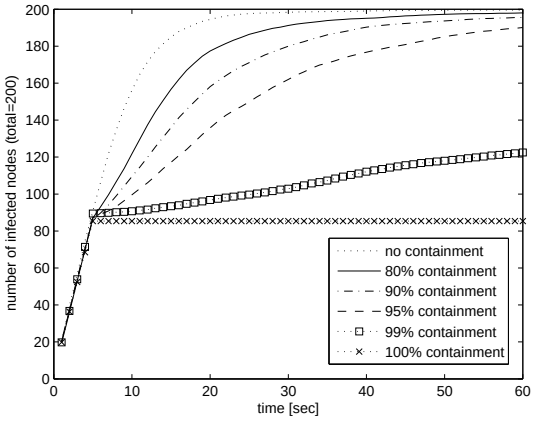

Figure 9. Effect of imperfect containment on worm propagation speed. Containment techniques with more than 95% detection probability can significantly slow worms.

achieved by the advection model, significantly slow the propagation of a worm, yielding additional analysis time for security engineers.

So far, we assumed that the intrusion response is only performed once. Repeated application, however, could further slow worm propagation. One approach would be to wait for any intrusion reports after the first response and then retry with an enlarged boundary. Another approach would treat every remaining infectious node as a new outbreak. However, this requires changes to the estimation model because the worm will continue to spread from multiple locations, rather than a single origin.

The current solution aims for a high detection probability, to effectively slow worms. In some scenarios a more balanced approach that also minimizes the false alarm probability may be desirable. Higher Jaccard similarity values, for example, can be obtained when small reductions in detection probability yield large reductions in false-alarm probability. To optimize Jaccard similarity we could choose a smaller radius $\hat{R} = \gamma R = \gamma 2\sqrt{\pi \alpha D t}$ for the random walk scenario. $\hat{R}$ denotes the effective radius which equals the square root of the propagation area enclosed by a real boundary (not a circle) against time. Our usage of R instead of $\hat{R}$ also explains the of our algorithm over different node densities.

The successful application of ecological models to estimating worm propagation raises the question about other potential synergies between the fields. Biologically inspired interdisciplinary work has long affected computer security. For example, computer immunology improves virus defenses [8]. Epidemiology enables us to investigate the spread of computer viruses on a hybrid networks that combine computer network and social networks, such as email [17]. In ecology the Allee effect (or reduced per capita reproduction when animals are scarce) may be useful for describe the dynamic change of the infection rate when we have disconnections in the ad hoc network. The effect of dispersal on competing populations (e.g., Predator-Prey model) also holds promise for modeling competition[6] or the cooperation of malicious codes [24].

6 Related Work

Moore and colleagues [16] investigated and compared the existing containment methods for Internet worms which can be implemented in gateway, firewall and router. The hierarchical structure of the Internet allows an administrator to partition and shut down a local subnetwork which is infected. In wireless networks, however, an infected node can move and communicate with a susceptible node via localized interaction such as Bluetooth. Our work instead focuses on estimating the geographic propagation pattern of short-range wireless worms. The notion of locality is less meaningful in wired networks where worms often use random probing.

Khayam and Radha [13] investigated the parameters governing the spread of active worms over VANET. They define the average degree of a VANET node and use a SIR model for the spread of worms. In our work, we provide a spatial and temporal distribution of the propagating worms rather than an infection rate over time.

Wu and Fujimoto [25] presented an analytical model for information propagation in Vehicle-to-Vehicle Networks. Worm propagation is very similar to information dissemination except that it has an malicious purpose and it lacks cooperation of neighboring nodes. Our work concentrated on practical estimation algorithms that are tractable for larger highway networks. We also presented simulation results from a calibrated highway simulation.

Several intrusion detection system for wireless ad hoc networks have been designed [28, 15]. Zhang and Lee present a collaborative intrusion detection system for ad hoc and assume that every node runs an IDS agent. Anjum and colleagues have investigated the optimal placement of intrusion detection nodes in an ad hoc network to reduce the need for one IDS agent per node [2]. This intrusion detection work concentrates mostly on external attacks such as distributing erroneous routing information. They do not address how to catch up with a propagating worm. Our work shows how to take advantage of a wireless infrastructure network and how to forecast the propagation of the worm.

7 Conclusions

Wireless ad hoc networks requires a new worm intrusion response architecture and mechanisms because it lacks central infrastructure choke-points such as routers, gateways and firewalls where network intrusion detection and address blacklisting or content filtering can take place. We have proposed an architecture in which a service provider manages the security of an hybrid (ad hoc with wide-area network) network over a low-bandwidth, wide-area infrastructure wireless network. This work concentrated on developing and analyzing location-based quarantine boundary estimation techniques. These techniques let service providers identify the current set of likely infected nodes when intrusion information is incomplete or delayed. Specifically, we found that

- a mobile worm could spread in a typical highway network with a mean velocity of about 75m/s even though only 5% of vehicles are susceptible to attack.

- advection-based estimation techniques can estimate the group of currently infected nodes with a detection probability greater than 95% and a false-alarm rate of less than about 35%. This provides a significant improvement over having to target a response at all nodes in a large geographic region.

[6] In 2001, the counterattacking CodeGreen appeared to disinfect CodeRed.

Future Work There are several directions for future work. First, designing algorithms that show robust accuracy if the geographic origin of the outbreak is not or only approximately known. Second, it appears valuable to develop techniques that effectively address partial outages of the wide-area wireless network. Finally, the system could take advantage of propagation speed information gained from the time difference in intrusion reports from different nodes.

Acknowledgment

The authors would like to thank Dr. Ozbay for providing a location trace file from Southern New Jersey Highway Network.

References

[1] O. M. Alliance. Provisioning architecture overview. `http://www.openmobilealliance.org/release_program/docs/ClientProv/V1_1-20050428-C/OMA-WAP-ProvArch-v1_1-20050428-C.pdf`, Apr 2005.

[2] F. Anjum, D. Subhadrabandhu, and S. Sarkar. Intrusion detection for wireless adhoc networks. In *Proceedings of Vehicular Technology Conference, Wireless Security Symposium*. IEEE, October 2003.

[3] N. T. Bailey. *The Mathematical Theory of Infectious Diseases and its Applications*. Hafner Press, New York, 1975.

[4] L. Briesemeister, L. Schafers, and G. Hommel. Disseminating messages among highly mobile hosts based on inter-vehicle communication. In *IEEE Intelligent Vehicles Symposium*, October 2000.

[5] D. Dagon, T. Martin, and T. Starner. Mobile phones as computing devices: The viruses are coming! *IEEE Pervasive Computing*, 3(4):11–15, 2004.

[6] C. S. Elton. *The Ecology of Invasions by Animals and Plants*. Methuen Co. Ltd., London, 1958.

[7] D. R. Finley. Point-in-polygon algorithm: Determining whether a point is inside a complex polygon. `http://www.alienryderflex.com/polygon/`, 1998.

[8] S. Forrest, S. Hofmeyr, and A. Somayaji. Computer immunology. *Communications of the ACM*, 40(10):88–96, 1997.

[9] S. Goel, T. Imielinski, and K. Ozbay. Ascertaining the viability of wifi based vehicle-to-vehicle network for traffic information dissemination. In *Proceedings of the 7th Annual IEEE Intelligent Transportation Systems Conference (ITSC)*, October 2004.

[10] T. D. G. S. Group. *Standard Specification for Telecommunications and Information Exchange Between Roadside and Vehicle Systems - 5GHz Band Dedicated Short Range Communications (DSRC) Medium Access Control (MAC) and Physical Layer (PHY) Specifications*. ASTM E2213-03, 2003.

[11] J. P. Hubaux, S. Capkun, and J. Luo. The security and privacy of smart vehicles. *IEEE Security and Privacy Magazine*, 2(3):49–55, June 2004.

[12] T. M. Inc. Overlaying polygons with set logic. `http://www.mathworks.de/access/helpdesk/help/toolbox/map/polybool.html`, 2005.

[13] S. A. Khayam and H. Radha. Analyzing the spread of active worms over vanet. In *Proceedings of the first ACM workshop on Vehicular ad hoc networks*, January 2004.

[14] Q. Limited. Paramics v4.0 - microscopic traffic simulation system. `www.paramics-online.com`.

[15] A. Mishra, K. Nadkarni, and A. Patcha. Intrusion detection in wireless ad hoc networks. *IEEE Wireless Communications*, 11:48–60, 2004.

[16] D. Moore, C. Shannon, G. M. Voelker, and S. Savage. Internet quarantine: Requirements for containing self-propagating code. In *INFOCOM*. ACM, 2003.

[17] M. E. J. Newman, S. Forrest, and J. Balthrop. Email networks and the spread of computer viruses. *Physical Review*, 66(035101), 2002.

[18] A. Okubo and S. A. Levin. *Diffusion and Ecological Problems: Modern Perspectives*. Springer, 2002.

[19] K. Ozbay and B. Bartin. South jersey real-time motorist information system. NJDOT Project Report, March 2003.

[20] M. Raya and J. P. Hubaux. The security of vehicular ad hoc networks. In *Proceedings of SASN'05*, November 2005.

[21] Reuters. Mobile phone virus infects helsinki championships: The cabir virus uses bluetooth to jump between cell phones. `http://www.computerworld.com/securitytopics/security/virus/story/0,10801,103835,00.html`, Aug 2005.

[22] M. Sadiq. *Toxic metal chemistry in marine environments*. New York : Marcel Dekker, New York, 1992.

[23] J. G. Skellam. Random dispersal in theoretical populations. *Biometrika*, 38(4):196–218, 1951.

[24] P. Szor. *The Art of Computer Virus Research and Defense*. Addison-Wesley Professional, symantec press, 2005.

[25] H. Wu, R. Fujimoto, and G. Riley. Analytical models for data dissemination in vehicle-to-vehicle networks. In *Proceedings of IEEE 2004-fall Vehicle Technology Conference (VTC)*, September 2004.

[26] Q. Xu, R. Sengupta, and D. Jiang. Design and analysis of highway safety communication protocol in 5.9 ghz dedicated short range communication spectrum. In *IEEE VTC Spring 2003*, April 2003.

[27] J. Yin, T. ElBatt, G. Yeung, B. Ryu, S. Habermas, H. Krishnan, and T. Talty. Performance evaluation of safety applications over dsrc vehicular ad hoc networks. In *VANET '04: Proceedings of the 1st ACM international workshop on Vehicular ad hoc networks*, pages 1–9, 2004.

[28] Y. Zhang and W. Lee. Intrusion detection in wireless ad-hoc networks. In *The Sixth International Conference on Mobile Computing and Networking (MobiCom)*. ACM, August 2000.