Killing Nodes as a Countermeasure to Virus Expansion

François Bonnet[1], Quentin Bramas[2], Xavier Défago[3], and Thanh Dang Nguyen[4]

[1] Graduate School of Engineering, Osaka University, Japan
francois@cy2sec.comm.eng.osaka-u.ac.jp
[2] Sorbonne Universités, UPMC Univ Paris 06, CNRS, LIP6 UMR 7606, 4 place Jussieu 75005 Paris.
quentin.bramas@lip6.fr
[3] School of Computing, Tokyo Institute of Technology, Japan
defago@c.titech.ac.jp
[4] University of Chicago, USA
thanhnd@uchicago.edu

Abstract. The *spread of a virus* and the *containment of such spread* have been widely studied in the literature. These two problems can be abstracted as a two-players stochastic game in which one side tries to spread the infection to the entire system, while the other side aims to contain the infection to a finite area. Three parameters play a particularly important role: (1) the probability p of successful infection, (2) the topology of the network, and (3) the probability α that a strategy message has priority over the infection.

This paper studies the effect of *killing strategies*, where a node sacrifices itself and possibly some of its neighbors, to contain the spread of a virus in an infinite grid. Our contribution is threefold: (1) We prove that the simplest killing strategy is equivalent to the problem of site percolation; (2) when killing messages have priority, we prove that there always exists a killing strategy that contains a virus, for any probability $0 \leq p < 1$; in contrast, (3) when killing message do not have priority, there is not always a successful killing strategy, and we study the virus propagation for various $0 \leq \alpha < 1$.

Category: Regular paper.

Contact author: Quentin Bramas, quentin.bramas@lip6.fr

1 Introduction

Consider the propagation of a virus in a large distributed system, such as the Internet, a sensor network, or a social network. An epidemic starts with an arbitrary node being initially infected, and then continues with each newly infected node attempting to infect its neighbors. If nothing is done, all connected nodes are eventually infected, and the time it takes only depends on the infection rate (*i.e.*, the probability of success for each infection attempt), the topology of the network, and the location of the initial node.

Consider now that, upon an unsuccessful infection attempt, the targeted node has a chance to detect the attempt and react to it. One simple strategy consists in ordering that node to inform its neighbors of the presence of the virus and then to kill itself. The informed node can have the same behavior so that the infection can no longer spread through them. The hope is that, given enough nodes are killed, they could actually isolate the infected nodes from the sane ones, thus containing the spread of the virus. The chance of this happening depend mainly on the following factors: (1) the infection rate p, (2) the detection probability, (3) the probability α that killing messages have priority over the infection, and (4) the topology of the network. In this paper we suppose that a node always detect the virus when the infection fails (the detection probability equals $1-p$).

Such systems are notoriously known to exhibit a *critical threshold*, that is, the spread is almost surely contained if the infection rate is below the threshold, but is very unlikely otherwise. This is reminiscent of another problem, known as percolation [2], where the same *phase transition* occurs.

In this paper, we look at the following question: is there always a strategy to contains the spread of a virus? We answer this question when the topology of the network is a grid and for $0 \leq \alpha \leq 1$.

In our recent work [19], we have studied this question and the impact of several containment strategies in various topologies, by relying on extensive simulations. We found that containment strategies do have an impact on the critical threshold of the system, and so does the topology. In this paper we formalize and prove some of those observations.

Our Contributions The contribution of this paper is threefold. First, we show an equivalence between infection rate threshold with the simplest killing strategy and percolation threshold for the same topology. Second, we prove that, in the infinite grid and when killing message have priority, there is always a killing strategy such that the spread is contained with probability one, for any infection rate $0 \leq p < 1$. Third, we prove that, in the infinite grid and when killing message do not have priority, we give a lower bound on the threshold of the infection probability, above which the virus spreads infinitely with a positive probability

The rest of the paper is structured as follows. Section 2 gives a brief overview of the state-of-the-art on epidemic research. Section 3 defines the model, the topology, and the killing strategies considered in the paper. Section 4 proves the equivalence between the simplest killing strategy and percolation. Finally, Section 5 proves that containment is always possible in an infinite grid, and Section 6 shows a lower bound on the threshold of the infection probability.

2 Related Work

On the spread of a virus Starting from the epidemiology in human community, much research has been conducted on propagation. The first mathematical models appeared in the 18th century, but modern models were essentially developed in the middle of the 20th century (e.g., [1, 12, 17]). While original models did not consider geographic distributions, more recent epidemic models consider geographic topologies, such as an infinite grid [6]).

Kephart and White [10] propose a birth-death model to study the spread of computer viruses in homogeneous sparse graphs and conclude that a pandemic occurs only when the infection rate exceeds a finite threshold that depends on the connectivity of the network (phase transition). They also extend their model to allow doing a virus scan [11].

Later, many works improve the results on the birth-death model and compute new epidemic thresholds. Pastor-Satorras and Vespignani [20, 21] look at the dynamics of epidemics in power-law scale free networks for which they find the critical threshold. Chakrabarti et al. [4] study an epidemic model with recovery and find that the propagation threshold is related to the eigenvalues of the adjacency matrix of the network. Lately, Van Mieghem et al. [24, 25] use mean field approximation to transform from individual random infection rates into an average infection rate. Their model is called N-intertwined Markov chain.

In another direction, after the propagation of Code Red in 2001, many researches look for the most accurate model to reflect the spread of different kinds of viruses in the Internet. They propose different models from the *scanning worms* [23,27,30] to the *event-based* worms [28,31], where the question is to predict, as accurately as possible, the evolution of the expected number of infected entities in the network after the virus starts propagating.

On defense against a virus propagation In virus defense area, there are many works studying how to contain or quarantine the virus or worms in different network environments [18, 29, 31]. The containment strategies can be classified into two main classes; *proactive* or *reactive*. In the first one, some nodes are initially immune to the virus and only other nodes can be infected. In the latter, all nodes are initially susceptible, but eventually any node may become immune if it detects the virus (or receive some informations from other nodes).

Several work [29, 31] consider the spread viruses against proactive containment strategy, where the goal is to study the impact of the choice of the set of immune nodes on the spread of the virus.

Moore et al. [18] proposed a model for scanning worms in complete graph topology and give a comparison between two reactive strategies; (1) *blacklisting*, upon detection of an infection, a node adds the attacker into a blacklist; and (2) *filtering-content*, upon detection of a virus, a node transmits its signature to all other nodes. They assume that when a node detects an infection, the information (blacklisted IP address, or virus' signature) will be available to all other nodes after some time. They study the efficiency of both strategies when this delay varies.

In all these work, immune nodes can always detect virus attack successfully. However, with a polymorphic (such as Sality [5]) or metamorphic virus, it may not be always possible to detect the virus correctly. In this context, several detectors are introduced in [3, 13, 15]. We call *imperfect detection* the ability to detect a virus but not always successfully. The problem of the containment of a virus, when nodes are provided with an imperfect detector, remains open.

Flooding and percolation This containment problem is related to probabilistic broadcast (or information flooding) albeit with an opposite objective. Sasson et al. [22] and later Hu et al. [7] both study the question and show the relationship with the theory of percolation [2].

3 Model and Definitions

Let $G = (V, E)$ be a connected undirected infinite graph, where V is the set of vertices, and E the set of edges. Vertices model the nodes in the system, and edges the bidirectional communication links between a pair of nodes. Thus, edge $e_{ij} \in E$ represents a communication link between the two nodes $i, j \in V$.

The system evolves in numbered synchronous rounds, also called *timeslots*. During a round, every node can exchange messages with its direct neighbors. Time is discrete and measured according to the round number.

A node can be in one of four states:

- *infected*: the node is compromised by the virus and acts as an infectious agent.
- *killed*: the node no longer sends or receives any message.
- *susceptible*: the node is neither infected nor killed, but can still be affected in the future.
- *sane*: after the spread of the virus if over and no more nodes can be infected, the remaining susceptible nodes are said to be sane.

Initially ($t = 1$), all nodes are susceptible, except for a single node which is initially infected. Then, at each timeslot, the virus can propagate via communication links, from every infected nodes to its susceptible neighbors.

Let p be a parameter of the system denoting the *infection probability*. At each round, an infected node attacks and attempts to infect each of its susceptible neighbors. For each attack, the susceptible node targeted is infected with probability p. Conversely, with probability $1 - p$, the node detects the attack, in which case it starts a containment strategy.

Killing strategies When a susceptible node is victim of a failed attack, it detects the attempt and can initiate a containment strategy. This paper considers a family of killing strategies, in which a certain number of nodes are killed to act as an obstacle to the spread.

We consider killing strategies that differ in the extent to which neighboring nodes are deactivated/killed. Strategy Kh-Hop is defined for any non-negative value of h, such that the detector (i.e., the node detecting the infection attempt) send messages to kill all nodes in its h-hops neighborhood, and then kills itself.

The behavior of a node receiving such a message depends on its current state:

- A susceptible node sacrifices itself as requested by the message, after having potentially relayed the message as requested by the strategy.
- An infected nodes ignores the message; it neither forward the message nor sacrifice itself.
- A killed node neither receives nor forwards the message, and cannot be infected.

When a node receives both a killing message and an infection message, there is a probability α that the strategy messages take precedence over infection messages. When $\alpha = 1$ we say that killing messages have priority.

An example of virus propagation and its containment by K1-Hop strategy is given in Appendix A.

Infinite grid The infinite grid corresponds to the graph $G = (V, E)$ where nodes are located on the square lattice. Each node is connected to the nodes at distance one in each of the four cardinal directions. $V = \mathbb{Z} \times \mathbb{Z}$ and $E = \{e_{i,j} \mid i \in V \wedge j \in V \wedge \mathit{dist}(i,j) = 1\}$

4 Equivalence of Site Percolation and K0-Hop Strategy

In this section we show that site percolation and strategy K0-Hop have the same threshold, below which propagation is contained with probability one and above which it is not contained with non zero probability.

Site percolation Given an infinite graph and a probability p such that any site (*i.e.*, node) is *occupied* with probability p (resp., *empty* with probability $1-p$), let us consider the random variable X_p that equals 1 if there is an infinite connected component of occupied nodes and 0 otherwise.

There exists a threshold τ (called the percolation threshold) such that [2,14]:

$$\begin{cases} \mathbb{P}[X_p = 1] = 0 & \text{if } 0 \leq p < \tau \\ \mathbb{P}[X_p = 1] = 1 & \text{if } \tau < p \leq 1 \end{cases}$$

In other words, with probability 1, there is an infinite connected component of occupied nodes when the probability p is above τ. Respectively, there is no such component when p is below the threshold. When p exactly equals the threshold, the situation is unclear.

Strategy K0-Hop Given a probability p of infection and an infinite graph, consider a spread against the K0-Hop containment strategy. Let Y_p be a random variable that equals 1 if the spread never stops and 0 otherwise.

For the two extreme values of p, the distribution of Y_p is trivial: when $p = 0$, the spread is immediately contained and $\mathbb{P}[Y_0 = 1] = 0$; conversely, when $p = 1$, all nodes are infected and $\mathbb{P}[Y_1 = 1] = 1$. Moreover, by a simple coupling argument, it is also clear that the function $p \mapsto P[Y_p = 1]$ is non-decreasing. Therefore, there exists a unique threshold τ_0 such that:

$$\begin{cases} \mathbb{P}[Y_p = 1] = 0 \text{ if } 0 \leq p < \tau_0 \\ \mathbb{P}[Y_p = 1] > 0 \text{ if } \tau_0 < p \leq 1 \end{cases}$$

Of course, it is also possible that either $\mathbb{P}[Y_p = 1] > 0$ for all $p > 0$ ($\tau_0 = 0$) or $\mathbb{P}[Y_p = 1] = 0$ for all $p < 1$ ($\tau_0 = 1$). It is important to note the inequality. Contrary to the random variable X_p associated to site percolation, Y_p does not "jump directly from 0 to 1". In fact, Y_p equals 1 if and only if $p = 1$. For any probability of infection p strictly lower than one, there is a non-zero probability that strategy $K0$-Hop contains the spread. Indeed, if all neighbors of the initially infected node detect the infection, the propagation stops immediately. This case happens with probability $(1-p)^\delta$, where $\delta > 0$ is the degree of the initial node, and thus $\mathbb{P}[Y_p = 1] \leq 1 - (1-p)^\delta$.

Theorem 1. *Site percolation and strategy $K0$ have the same thresholds: $\tau = \tau_0$.*

We can observe that, if the probability of infection of a node depends on the state of its neighbors (if the events are not independent), then the result may be different, but, with some bound on the probability of infection, we still obtain the following corollary.

Corollary 1. *Given a graph $G = (V, E)$, and let τ be the site percolation threshold on G. If, for all $u \in V$ and $t \in \mathbb{N}$, the probability $p_{u,t}$ that a susceptible node u is infected at time t is bounded by τ, then the $K0$-Hop strategy contains the propagation of the virus with probability one.*

5 Virus Containment with Priority Killing Messages

In this section, we prove that it is always possible to contain the propagation in an infinite grid, provided that we use a killing strategy that sacrifices enough nodes (*i.e.*, a strategy Kh-Hop with h large enough). The proof is done by reduction to strategy $K0$-Hop so that we can apply Theorem 1 and use the existence of a percolation threshold for a specific topology.

Theorem 2. *Given a virus spread with infection probability p, where $0 \leq p < 1$, there exists a killing strategy that contains the virus spread.*

5.1 Definitions and explanations

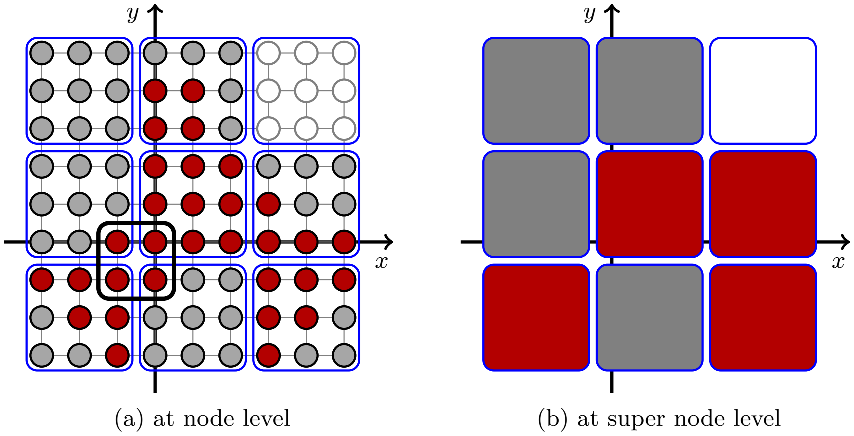
General idea of the proof Based on the infinite grid, we define the notion of *super nodes* to encompass a squared subset of nodes. Super nodes are considered as infected, killed, or sane depending on the states of their internal nodes. By choosing an appropriate size for the super nodes and an adequate containment strategy, we can show that, at the super node level, the propagation of the virus behaves similarly as when strategy $K0$-Hop is executed. From there, we can deduce that the infection is contained at the super node level, which then implies containment at the lower level.

Super node Given a strictly positive integer h and the infinite grid $G(V, E)$ with $V = \mathbb{Z} \times \mathbb{Z}$, we partition the nodes into *super nodes of size* h. Each super node is a subset of h^2 nodes organized in a square of side h. The super node at column x and row y, denoted $g(x, y)$, corresponds to the following subset of V:

$$g(x, y) = \{(hx + i, hy + j) \in V \mid 0 \leq i, j < h\}$$

Similarly to nodes, super nodes may be infected, killed, or susceptible. We say a super node is infected if all nodes in at least one of its sides are infected. If a super node is not infected, we say it is killed if it contains at least a killed node, otherwise, we say it is susceptible (or sane if the spread is over). For simplicity, we assume that the super node containing the initial infected node (at $(0, 0)$ wihtout loss of generality) is also infected.

Containment strategy Given a super nodes of size h, we study the strategy $K2h$-Hop. For the analysis we consider a slightly modified version of the $K2h$-Hop strategy: nodes do not send killing messages to neighbors belonging to different super nodes. Thus, the killing strategy affects only nodes contained in the same super node as the detector. This modification favors the propagation; it weakens the containment strategy and helps the virus propagate and therefore has no impact on our result. Indeed, if the spread is contained with the modified version, it is also contained with the standard $K2h$-Hop strategy.



(a) at node level

(b) at super node level

Fig. 1. Example of infection

Figure 1 provides an example of infection at the nodes level and at the super nodes level. At the super node level, it is worth noting that, unlike what happens at node level, an infection may be transmitted in "diagonal": in the figure, the central super node has infected the super node located at the lower left corner.

Since an infection can be propagated via the diagonals, the infection in super nodes can be seen as an infection in the lattice where every node has 8 neighbors. This lattice is called Moore's neighborhood [16]. In this section, the neighbors of a super node are the neighbors according to Moore's neighborhood and the neighbors of normal node are the four neighbors in the default model (Von Neumann's neighborhood).

Also, one can observe that when a node in a super node U detects an attack, then it broadcast killing messages in the whole super node (because all the nodes in the super node are at distance at most $2h$ from one another). So that if one node is attacked, then all the other nodes in its super node are either killed or infected. Therefore, a super node is either sane, or contains no sane node.

Small-Four Based on the previous observation, we introduce the notion of "Small Four". For any group of four super nodes arranged in a 2×2 grid, let us call "Small-Four" the group of four normal nodes that connect these four super nodes by their corners. In Figure 1a, the group of four nodes $(0,0),(0,-1),(-1,0),(-1,-1)$ located in the small black square is an example of the "Small-Four" that connects the four super nodes $g(0,0),g(0,-1),g(-1,0),g(-1,-1)$.

Rectangle In the remaining of the section, the distance $d(u,v)$ between two nodes u and v, denotes the length of a shortest path between u and v. For two nodes u and v, $Rect(u,v)$ denotes the set of nodes in the rectangle delimited by u and v: $Rect(u,v) = \{w \in V \mid d(u,w)+d(w,v)=d(u,v)\}$.

5.2 Proof

Now, we prove several Lemmas to analyze the propagation of the infection within a super node (and its origin), then we prove that the propagation of the virus at the super node level is a simple virus propagation in the square lattice with Moore's neighborhood against the $K0$-Hop strategy. Finally, we conclude by a proof of the theorem using the equivalence proved in the previous Section.

Lemma 1. *Let u be a node, in a super node U, infected for the first time at time t. Let v be a node in U, at distance $d_v \geq 0$ from u. Then, v is susceptible at time $t' < t - d_v$, and cannot be killed at time $t - d_v$.*

Proof. First, suppose by contradiction that there exists a node v in U and at distance d_v from u that detects a failed attack at time $t-d_v$. Then, the killing messages sent by v reach u at time t, killing node u before it gets infected. So, each node in U at distance d from u cannot detect an attack at time $t-d$ or before. Moreover, if a v is infected at time $t-d_v-1$, then the infection reaches u before time t, a contradiction, so v is susceptible at time $t' < t-d_v$.

Lemma 2. *Let u be a node, in a super node U, infected for the first time at time t. Let v be a node in U, at distance $d_v \geq 0$ from u, that is infected at time $t-d_v$. Then, each node in $Rect(u,v)$ and at distance d from u is infected at time $t-d$.*

Proof. Let v be a node in U at distance d_v from u that is at infected at time $t - d_v$. Let w_1 be a node in the rectangle $Rect(u, v)$ that is at distance 1 from v (thus at distance $d_v - 1$ from u). By Lemma 1, we know that w does not detect an infection at time $t - (d_v - 1)$, then the infection from v is successful and w_1 is infected at time $t - (d_v - 1)$. Again, if w_2 is a node in $Rect(u, v)$ at distance 2 from v then w_2 has a neighbor that is infected at time $t - (d_v - 1)$ (a node that is in $Rect(u, v)$ and at distance 1 from v). Therefore, w_2 is infected at time $t - (d_v - 2)$. Recursively, each node in $Rect(u, v)$ at distance d from u (and at distance $d_v - d$ from v) is infected at time $t - d$.

According to the previous Lemma, we can define formally what we mean by the origin of the infection of a node: the infection of a node u *originates from a node* v (in the same super as u) if u is infected for the first time at time t and each node in $Rect(u, v)$ and at distance d from u is infected at time $t - d$.

One can observe that the origin of the infection may not be unique as we can say the infection of u may originates from any node in the rectangle $Rect(u, v)$. The interesting origin of the infection (in the super node) may be the first infected node with such property. Its clear that the first infected origin of the infection is on one side of the super node, and the next lemma proves that it is actually always a corner.

Lemma 3. *The infection of any node u, in a super node U, originates from a corner of U.*

Proof. The Lemma is proved by induction on the time t of the first infection of u. At time 0, the lemma is proved for node $(0, 0)$ as it is in the corner itself.

Now let u be a node infected for the first time at time t and suppose the Lemma proved for every node that is infected before time t. Let U be the super node of u. Let v be one of the first infected node on the side of U such that the infection of u originates from v. If v is first infected at time t_v, then it is a distance $t - t_v$ from u. Thus, all the neighbors of v in U are not infected at time $t_v - 1$ (otherwise, such neighbor would be an origin of the infection of u in U, which contradicts the fact that v is the first one). So that v has a neighbor outside U, call it w, that is infected at time $t_w = t_v - 1$. By hypothesis, the infection of w in its super node W originates from a corner of W. For all possible corner c of W, the rectangle $Rect(w, c)$ contains a corner c_0 of W that is also a neighbor of a corner b_0 of U (see Figure 2). Let d_b be the distance from u to b_0 and d_c be the distance from c_0 to w (which is also the distance from v to b_0). See Figure 2 for an illustration of the configuration.

By Lemma 2, c_0 is infected at time $t_w - d_c$. So that at time $t_w - d_c + 1 = t_v - d_c$, node b_0 is either killed (it has detected the attack from c_0) or infected.

By the triangular inequality, we have $d_b \leq d_v + d_c$ so that $t_v - d_c = t - d_v - d_c \leq t - d_b$ and by Lemma 1, b_0 cannot be killed at time $t_v - d_c$. Therefore, the corner b_0 of U is infected at time $t_v - d_c$. Since b_0 is at distance at most $d_v + d_c$ from u then, by Lemma 2, the infection of u originates from b_0. By hypothesis, b_0 cannot be infected before v so $t_v - d_c = t_v$ i.e., $d_c = 0$, which means $v = b_0$ and v is in a corner of U.

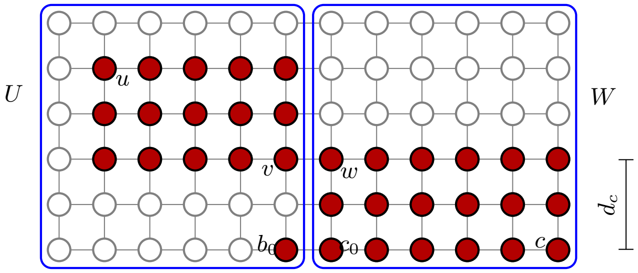


Fig. 2. The infection of a node u originates from a corner of its super node.

Lemma 4. *The first infected node in a "Small-Four" belongs to an infected super node, or is the node* $(0,0)$ *and the source of the infection.*

Proof. Consider a "Small-Four" composed of the following set of nodes $S = \{(hx, hy), (hx-1, hy), (hx-1, hy-1), (hx, hy-1)\}$. Assume that $u \in S$ is the first infected node among the four nodes and belongs to the super node U. Because the three other nodes of S are infected after node u, either $u = (0,0)$ is the source of the infection or u is infected by another node inside U. In the latter case, according to Lemma 3, the infection of node u originates from a corner of U distinct from u (because u is infected by a node inside U). One can observe that the infection originates from a corner that is adjacent (i.e., on the same side of the super node) to the corner u. Indeed, if it originates from the opposite corner, it also originates by definition from any node in the super node, including the adjacent corners. This means that, at least one side of super node U is entirely infected i.e., U is infected.

Lemma 5. *A super node (distinct from* $g(0,0)$*) may be infected only if it is a direct or diagonal neighbor of a previously infected super node. Moreover, in this situation, the probability of being infected is bounded by* $4p^h$*.*

Proof. According to Lemma 3, if a super node U is infected, the attack originates from a corner. The node u in this corner was infected by a neighbor outside U, thus by a node that belongs to the same "Small-Four". By Lemma 4, the first infected node of this "Small-Four" belongs to an infected super node, or to $g(0,0)$ (which is considered infected). Hence, one neighbor (direct or diagonal) of super node U is infected.

Let us recall that p is the probability that a normal node in the grid is infected. To become infected, a super node must have one of its sides entirely infected. For a given side, the probability that it becomes infected is exactly by p^h since any of the h nodes has a probability p of being infected. The events corresponding to the entire infection of each side are not independent (two adjacent sides have a node in common), but the probability that a super node has at least one side entirely infected is bounded by $4p^h$.

In more details the probability that a super node is infected depends on the way its neighbors are infected. Indeed, when considerging a super node U, if only one side of a neighbor is infected, and if this side is not connected to U then U is not attacked by this neighbor and the probability that U is infected is 0. But, if a unique node effectively attacks U, then the probability that U is infected is exactly $2p^h - p^{2h-1}$, which is the probability that at least one of the two sides of U adjacent to the attacked corner gets entirely infected. Since U can be attacked from different corner at different time, we simply bound the probability that U is infected by $4p^h$.

Proof (Proof of Theorem 2). According to Lemma 5, the set of infected super nodes forms a single connected component, with the Moore's neighborhood (two diagonal super nodes are considered connected).

Since every neighbor of an infected super node has a bounded probability to be infected, the infection spreading through super nodes can be seen as strategy $K0$ on an infinite lattice with Moore's neighborhood.

According to Corollary 1, if we choose p such that the probability that a node gets infected is smaller than the percolation threshold, then the propagation of the virus is contained by the $K0$-Hop with probability one. According to Malaz and Galam [16], there exists a percolation threshold in an infinite lattice with Moore's neighborhood. Let ρ_c^{Moore} be the percolation threshold in an infinite lattice with Moore's neighborhood.

Let $h \in \mathbb{N}$ be such that $h > \frac{\log(\rho_c^{\text{Moore}}/4)}{\log p}$. The virus spreading with infection probability p against the $K2h$-Hop strategy can be seen as the propagation of virus via super nodes of size h with probability bounded by $4p^h < \rho_c^{\text{Moore}}$ against strategy $K0$-Hop. By Corollary 1 such propagation is contained, so that the virus propagation with infection probability p is contained by strategy $K2h$-Hop in an infinite grid.

6 The Case of Non-Priority Killing Message

When the killing messages do not have priority ($\alpha < 1$) then the analysis of the previous Section does not hold. In this Section we study the behavior of the best possible strategy, which is $K\infty$-Hop. Even though this strategy may not be applied in practice, as all nodes in the network are killed if a virus is detected, it has theoretical interest. Indeed, if the virus is not contained when using the $K\infty$-Hop strategy, then, no strategy contains the virus.

First, one can observe that, since we consider the $K\infty$-Hop strategy, at timeslot $t \in \mathbb{N}$ all the nodes at distance t from the origin are either infected or killed. Those nodes form a square whose vertices are nodes $(t, 0)$, $(0, t)$, $(-t, 0)$ and $(0, -t)$. Therefore, we can restrict our analysis on a quarter of the grid: if the virus is contained with probability one (resp. smaller than one) on a quarter of the grid, it is contained with probability one (resp. smaller than one) on the whole grid. Also, we can order the nodes and say that the two predecessors of a node are its neighbors that are closer to the origin. Figure 3 represents an execution of $K\infty$-Hop starting from an initial infected node at $(0, 0)$.

We know from the previous Section that, when $\alpha = 1$, $K\infty$-Hop strategy successfully contains the virus. However, when $\alpha = 0$, it depends on the probability

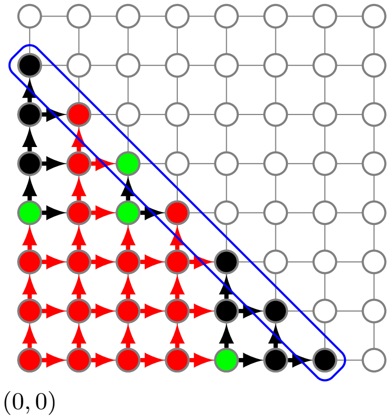


Fig. 3. An execution of $K\infty$-Hop on the upper right quarter of an infinite grid. Red, green, and black nodes are resp. infected, detector, and killed.

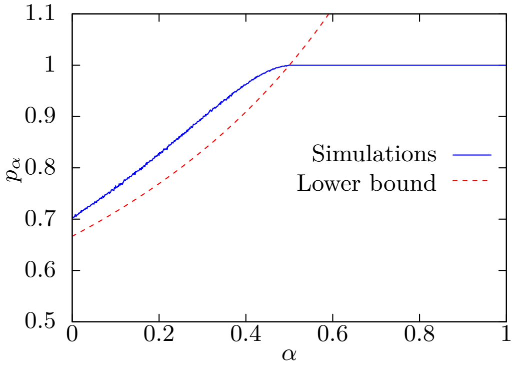


Fig. 4. Threshold of the probability p_α to contain the virus, depending on the value of α.

p. Indeed, when $\alpha = 0$, a node becomes infected if at least one of its two predecessors is infected. This model is exactly a directed site percolation on the square lattice. The percolation threshold for this model is around 0.7054852 [8, 9, 26] (computed by simulations). Thus $K\infty$-Hop strategy contains the virus if p is above this threshold. Now, the main result of this Section, is to give a lower bound on the the value of the percolation threshold p_α, depending on the value of α. Figure 4 depicts the lower bound and an approximation of the threshold that we computed by simulations.

Theorem 3. *If* $p < \min\left(1, \frac{2}{3-2\alpha}\right)$*, the virus is contained. In particular, the virus is always contained if* $\alpha \geq 0.5$ *and* $p \neq 1$.

Proof. On Figure 3, the blue box encompasses the nodes that were infected/killed at round 6. We are interested in the evolution of the nodes in the diagonal over the time, see Figure 5. One can observe that if there are k killed nodes at one extremity of the diagonal, then there will be also at least k killed nodes on the same extremity of the next diagonal.

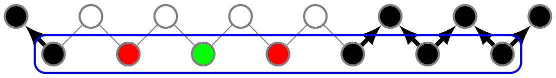


Fig. 5. The spread of the virus on the diagonal ($k_1 = 1$ and $k_2 = 3$)

Consider a diagonal of size d, which contains k_1 killed nodes at one extremity and k_2 at the other. Let us define the central part of the diagonal as the $n =$

$d - k_1 - k_2$ nodes between the two extremal blocks of killed nodes. At the next round of the execution, the new diagonal contains $d+1$ nodes and its central part contains $n-1$, n, or $n+1$ nodes.

The proof of the theorem is the result of the two following Claims, that are true for an arbitrary constant $X \in \mathbb{N}$:

– Claim 1: if the central part has at most $2X$ nodes, the probability the virus is contained in the next round is greater than a fixed constant $C_X > 0$.
– Claim 2: if the central part has more than $2X$ nodes and if $p < \frac{2}{3-2\alpha+f(X)}$ (with $\lim_{X\to\infty} f(X) = 0$), then eventually the central part has size at most $2X$.

Indeed, if $p < \frac{2}{3-2\alpha}$ we can choose X large enough so that $p < \frac{2}{3-2\alpha+f(X)}$. Then, by the second Claim, there cannot be an infinite execution without the condition of the first Claim being verified, so that we always have a strictly positive probability that the virus is contained i.e., the virus is always contained.

The proof of the first Claim is straightforward as there is a probability at least $C_X = (1-p)^{2X+1} > 0$ that all nodes on the next diagonal detect the virus.

The proof of the second Claim is a by studying the expectation E of the reduction of the size of the central part after one round, on one side of the diagonal (since the other side has the same behavior).

The central part decreases by k nodes when the node in the extremity is killed (not infected or receives the killing message from one of its neighbors), $k-1$ other nodes detect the attack and the next node is infected. This event occurs with probability $(1-(1-\alpha)p)(1-p)^{k-1}p$. We consider only the case $k \leq X$ because we are on one side of a diagonal of length at least $2X$. Then

$$E = (1-(1-\alpha)p)\ p\ \sum_{k=0}^{X} k(1-p)^{k-1}.$$

For $p < 1$, $\sum_{k=0}^{\infty} k(1-p)^{k-1} = \frac{1}{p^2}$, therefore $E = (1-(1-\alpha)p)\ p\left(\frac{1}{p^2} - g(X)\right)$

with $\lim_{X\to\infty} g(X) = 0$. The size d of the diagonal increases by 1 every round. Since we considered only one side of the diagonal, we need to solve $E > 1/2$ to determine when containment is possible, which gives the following bound:

$$p < \frac{2}{3-2\alpha+f(X)} \qquad \text{with} \lim_{X\to\infty} f(X) = 0$$

7 Conclusion

We considered the virus propagation problem where nodes that detect a failed infection attempt can broadcast a deactivation message to their neighbors to contain the propagation. We first show that the strategy consisting in deactivating itself when an attack is detected (without alerting its neighbors) has the same behavior as the site percolation, in the sense that both problems have the same critical probability threshold. Then, we prove that in an infinite grid graph the propagation of a virus can always be contained when the killing message have priority, for all value of $0 \leq p < 1$, by choosing an adequate strategy. Finally, when killing messages do not have priority, we showed a lower bound on the threshold of the infection probability, above which the virus spreads infinitely with a positive probability.

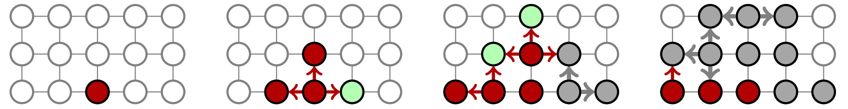
References

1. Anderson, R.M., May, R.M.: Population biology of infectious diseases: Part i. Nature 280(5721), 361–367 (Aug 1979)
2. Bollobás, B., Riordan, O.: Percolation. Cambridge University Press (2006)
3. Brumley, D., Newsome, J., Song, D., Wang, H., Jha, S.: Towards automatic generation of vulnerability-based signatures. In: Proceedings of the 27th IEEE Symposium on Security and Privacy (S&P'06). pp. 2–16 (May 2006)
4. Chakrabarti, D., Wang, Y., Wang, C., Leskovec, J., Faloutsos, C.: Epidemic thresholds in real networks. ACM Trans. on Inf. and System Security 10(4) (2008)
5. Falliere, N.: Sality: Story of a peer-to-peer viral network. http://www.symantec.com/connect/sites/default/files/sality_peer_to_peer_viral_network.pdf (Jul 2011)
6. Grassberger, P.: On the critical behavior of the general epidemic process and dynamical percolation. Mathematical Biosciences 63(2), 157–172 (Apr 1983)
7. Hu, R., J., S., Arantes, L., Sens, P., Demeure, I.: Fair comparison of gossip algorithms over large-scale random topologies. In: Proceedings of the 31st IEEE Intl. Symp. on Reliable Distributed Systems (SRDS'12). pp. 331–340 (Oct 2012)
8. Jensen, I.: Low-density series expansions for directed percolation: Ii. the square lattice with a wall. Journal of Physics A: Mathematical and General 32(33), 6055 (1999), http://stacks.iop.org/0305-4470/32/i=33/a=304
9. Jensen, I.: Low-density series expansions for directed percolation: Iii. some two-dimensional lattices. Journal of Physics A: Mathematical and General 37(27), 6899 (2004), http://stacks.iop.org/0305-4470/37/i=27/a=003
10. Kephart, J.O., White, S.R.: Directed-graph epidemiological models of computer viruses. In: Proceedings of the 12th IEEE Symposium on Security and Privacy (S&P'91). pp. 343–361 (1991)
11. Kephart, J.O., White, S.R.: Measuring and modeling computer virus prevalence. In: Proceedings of the 14th IEEE Symposium on Security and Privacy (S&P'93). pp. 2–15 (1993)
12. Kermack, W.O., McKendrick, A.G.: A Contribution to the Mathematical Theory of Epidemics. Proceedings of the Royal Society of London. Series A 115(772), 700–721 (Aug 1927)
13. Kruegel, C., Kirda, E., Mutz, D., Robertson, W., Vigna, G.: Polymorphic worm detection using structural information of executables. In: Proceedings of the 8th International Conference on Recent Advances in Intrusion Detection (RAID'05). pp. 207–226. Springer-Verlag, Berlin, Heidelberg (Sep 2006)
14. Lee, M.J.: Pseudo-random-number generators and the square site percolation threshold. Physical Review E 78(3), 031131–1–11 (2008)

15. Li, Z., Sanghi, M., Chen, Y., Kao, M.Y., Chavez, B.: Hamsa: fast signature generation for zero-day polymorphic worms with provable attack resilience. In: Proceedings of the 27th IEEE Symposium on Security and Privacy (S&P'06). pp. 32–47 (May 2006)
16. Malarz, K., Galam, S.: Square-lattice site percolation at increasing ranges of neighbor bonds. Phys. Rev. E 71, 016125 (Jan 2005), `http://link.aps.org/doi/10.1103/PhysRevE.71.016125`
17. May, R.M., Anderson, R.M.: Population biology of infectious diseases: Part ii. Nature 280(5722), 455–461 (Aug 1979)
18. Moore, D., Shannon, C., Voelker, G., Savage, S.: Internet quarantine: requirements for containing self-propagating code. In: Proceedings of the 22nd Annual Joint Conference of the IEEE Computer and Communications Societies (INFOCOM'03). vol. 3, pp. 1901–1910 (Mar 2003)
19. Nguyen, T.D., Bonnet, F., Défago, X.: Analyzing the impact of mitigation strategies on the spread of a virus. Research Report IS-RR-2014-002, Japan Advanced Institute of Science and Technology (JAIST) (May 2014)
20. Pastor-Satorras, R., Vespignani, A.: Epidemic dynamics and endemic states in complex networks. Physical Review E 63(6), 066117–1–8 (2001)
21. Pastor-Satorras, R., Vespignani, A.: Epidemic spreading in scale-free networks. Physical Review Letters 86, 3200–3203 (Apr 2001)
22. Sasson, Y., Cavin, D., Schiper, A.: Probabilistic broadcast for flooding in wireless mobile ad hoc networks. In: Proceedings of the IEEE Wireless Communications and Networking (WCNC'03). pp. 1124–1130 (2003)
23. Staniford, S., Paxson, V., Weaver, N.: How to own the internet in your spare time. In: Proceedings of the 11th USENIX Security Symposium (USENIX-Security'02). pp. 149–167. USENIX Association, Berkeley, CA, USA (Aug 2002)
24. Van Mieghem, P.: The N-intertwined SIS epidemic network model. Computing 93(2-4), 147–169 (Dec 2011)
25. Van Mieghem, P., Omic, J., Kooij, R.E.: Virus spread in networks. IEEE/ACM Transactions on Networking 17(1), 1–14 (2009)
26. Wang, J., Zhou, Z., Liu, Q., Garoni, T.M., Deng, Y.: High-precision monte carlo study of directed percolation in $(d+1)$ dimensions. Phys. Rev. E 88, 042102 (Oct 2013), `https://link.aps.org/doi/10.1103/PhysRevE.88.042102`
27. Xia, J., Vangala, S., Wu, J., Gao, L., Kwiat, K.: Effective worm detection for various scan techniques. Journal of Computer Security 14(4), 359–387 (Jul 2006)
28. Xu, W., Zhang, F., Zhu, S.: Toward worm detection in online social networks. In: Proceedings of the 26th Annual Computer Security Applications Conference (ACSAC'10). pp. 11–20. ACM, New York, NY, USA (Dec 2010)
29. Zhou, L., Zhang, L., McSherry, F., Immorlica, N., Costa, M., Chien, S.: A first look at peer-to-peer worms: Threats and defenses. In: Proceedings of the 4th International Conference on Peer-to-Peer Systems (IPTPS'05). pp. 24–35. Springer-Verlag, Berlin, Heidelberg (2005)
30. Zou, C.C., Gong, W., Towsley, D.: Code red worm propagation modeling and analysis. In: Proceedings of the 9th ACM Conference on Computer and Communications Security (CCS'02). pp. 138–147. ACM, New York, NY, USA (2002)
31. Zou, C.C., Towsley, D., Gong, W.: Modeling and simulation study of the propagation and defense of internet e-mail worms. IEEE Transactions on Dependable and Secure Computing 4(2), 105–118 (Apr 2007)

A *K*1-Hop Strategy Example

Figure 6 depicts a step-by-step execution of strategy $K1$-Hop in a 3×5 grid. An infected node attacks its neighbors at timeslot t; the right neighbor detects the attack, the left and the top ones become infected. Following strategy $K1$-Hop, the right neighbor (detector) sends kill messages to its own neighbors and kills itself; its two neighbors (top and right) react accordingly, even though the top one is attacked at the same time. Two other nodes detect the attack at time $t+1$. At time $t+2$, one of the kill message reaches an infected node and is ignored. The state of the nodes at time $t+2$ is final, and the network is partitioned i.e., the spread is contained.



(a) Timeslot $t-1$ (b) Timeslot t (c) Timeslot $t+1$ (d) Timeslot $t+2$

Fig. 6. Example of the spread of a virus against the strategy $K1$-Hop

B Omitted Proof

Theorem 1 (restated). *Site percolation and strategy $K0$ have the same thresholds: $\tau = \tau_0$.*

Proof. The proof consists in showing that Y_p and Z_p^x follow the same distribution for any p and any occupied node x. To prove the claim, consider the following (intuitive) link between both models in Table 7a.

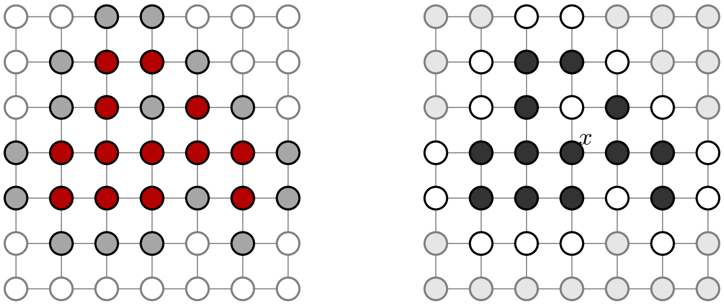
With site percolation, *all* nodes are randomly set to the occupied state (with probability p) or to the empty state (with probability $1-p$). With strategy $K0$-Hop, *some* nodes are randomly set to the infected state (with probability p) or to the killed state (with probability $1-p$) , but *some other* nodes are not set in one of these states. Indeed, any node that is neither an infected node, nor a neighbor of an infected node, remains in a sane state (see Section 3).

However, sane nodes do not have any effect on the random variable Y_p since they are disconnected from the initially infected node. The value of Y_p depends only on the infected (red nodes on Figure 7b) and killed nodes (gray nodes): all infected nodes are connected and therefore Y_p reflects the infiniteness of this infected component.

With site percolation, for a given occupied vertex x, the same observation applies to the random variable Z_p^x. It depends only on the occupied nodes (black nodes on Figure 7c) that are connected to x or the empty nodes (white nodes) that are around these occupied nodes. Other nodes (light gray nodes) do not have any effect on Z_p^x; they could be occupied or empty; it will not change the value of Z_p^x.

strategy $K0$	site-percolation
infected node	occupied node
killed node	empty node
sane node	— *(no equivalent)*
initial infected node	any occupied node x

(a) Intuitive link between $K0$ and site-percolation



(b) Strategy $K0$-Hop

(c) Site-percolation

Fig. 7. "Equivalence" between strategy $K0$-Hop and site-percolation model

Since (1) nodes are infected or occupied with the same probability p starting from an initial infected node or an arbitrary occupied node x, and (2) $Y_p = Z_p^x$, it follows that $\tau = \tau_0$.