

Leapfrogs and Honey pots: How dynamic distributed immune systems makes the difference

Jacob Goldenberg¹, Yuval Shavitt², Eran Shir², Sorin Solomon¹

¹*Hebrew University,*

²*Tel-Aviv University*

Although computer viruses are responsible for tremendous economic loss¹⁻³, defense mechanisms fail to adapt commensurately to their rapid evolution. Previous immunization strategies were characterized by being static and centralized⁴⁻⁷. Such strategies do not take into account the possibility of transmitting the cure by epidemic mechanisms similar to the disease itself. This makes virus containment difficult and - in certain network topologies - impossible⁸.

We analyze the dynamics of modern viruses and offer a new combative (Dynamic Distributed Immunization) strategy. We introduce the concept of "partially correlated networks", namely a manipulation of the effective underlying topology which allows the immunization agents to spread and contain the epidemics efficiently. We illustrate the incorporation of these concepts into a "honey pot" based architecture which is highly effective in controlling and containing viral epidemics. We present analytic, as well as simulation results for a set of realistic topologies⁹⁻¹¹. The presented analysis can be extended for any epidemic process where the immunization agent can also be spread in an epidemic manner. Employing this strategy, networks immunization through a dynamic, infectious process is finally feasible.

The realization that network models possess non-trivial geometrical properties such as a logarithmic diameter¹²(i.e. a diameter which grows logarithmically with the size of the network) and a non-existent percolation threshold¹³¹ impinged upon the dynamic properties of epidemic virus propagation models. The lack of a percolation threshold, means ⁷ that for predominant epidemic models, there is no critical infection rate.

Current immunization strategies ^{5,14,15} focus on a priori removing nodes from the network by immunizing them before the epidemic outburst. However, in the absence of complete knowledge of the network topology, these strategies are confined to a random character. These strategies require in most cases the removal of almost all of the nodes, and in all cases⁵ the removal of at least a quarter of the nodes.

In contrast, we introduce a dynamic, distributed immunization strategy, where the vaccine development and immunization processes depend on and interact with the virus dissemination process itself, creating co-dependency between virus dissemination and immunization.

In the context of traditional biological epidemiology there was little sense in considering dynamic, distributed immunization strategies. This is mainly due to the fact that the time scale gap between epidemic outburst and vaccine creation is very large, and that there is no 'infectious' delivery mechanism available for biological vaccines.

The world of computer viruses has diametrically different characteristics. First, new viruses emerge at an increasing pace. Secondly, computer viruses are much less complex than their biological counterparts, and are much easier to analyze and characterize ^{16,17}. Thus, vac-

¹By non-existent percolation threshold we mean that one can take out any portion of the network's nodes and the network will not disintegrate.

cine development can be achieved in a timescale comparable to that of the infection process. On the negative side, however, the viral process possesses an inherent lead time advantage : It appears before the vaccine, since a new vaccine can be created only after the new virus has started percolating the network. This fact, in itself, imposes strong constraints⁸ on the usability of dynamic approaches. However, as shown below, one can devise design principles which compensate the virus lead time advantage and support the deployment of efficient dynamic immunization systems.

We discuss the concrete example of the e-mail network. In this network, an e-mail account constitutes a node while the directed edges of the network are the entries in the accounts address book. Studies show that the outdegree of nodes in this network have a broad distribution^{4,18}. When one network node is infected by a virus, the virus spreads through the account address book. A sufficiently infectious virus will spread exponentially. The goal of our proposed immunization strategy is to react and adapt in real time to the emergence and propagation pattern of a new virus, in order to minimize the size of the ultimately infected cluster. The size of the virus cluster is the portion of infected nodes after a time period which we take to infinity. The size of the immunized clusters is defined consistently as the aggregate number of immunized nodes. The underlying assumptions of our model are:

1. A node can be in one of three modes: susceptible, infected, or immunized (removed). A node cannot change its mode once it is either infected or immunized. As such, we conform to the SIR epidemiologic model^{19,20}. The described model is in close agreement with the behavior observed on the Internet today, as an increasing number of viruses shut down security related software upon infecting a new node.
2. Some nodes, in accordance with some probability function, may recognize their own infection by a new virus, identify its characteristics, and create an immunization agent, as the

infection process progresses^{17,21}. In addition we define:

- Average infection delay v = the average time required by a virus in a given node to infect a neighboring node
- Average immunization delay m = the average time required by an immunization agent (propagating anti-virus) in a given node to immunize a neighboring node.
- Average development delay c = the average time required by an infected node to develop an immunization agent (propagating anti-virus).

The immunization process begins when one potential immunization generator node in the network identifies its own infection and develops an immunization agent which bears the signature of the new virus. The node then relays the agent to a set of neighboring contacts. Each susceptible node receiving the agent becomes immunized, and consequently, relays the agent to its own net of neighboring contacts. In essence, what we describe is a competition between two types of branching processes on a network²², where the first type creates a connected virus cluster. The second type creates a collection of immunized clusters, each rooted in one of the immunization developers. We consider the deterministic case, where the various delays associated with infection, agent creation and immunization are all constant, and all neighboring nodes become infected / immunized simultaneously. The resulting dynamics exhibits a sharp transition at the point where $v = c + m$. Whenever $v > c + m$, no node becomes immunized, because the virus infects all neighbors before the immunizing agent is deployed and effectively traps the immunization agent. When $v < c + m$, the dynamics is governed by agent development dynamics and the structure of the network. Thus, in this parameter range, it is possible for some nodes to become immunized. When immunization deployment time scale considerably exceeds

the infection time scale, the virus captures a major portion of the network. The transition is razor sharp in the deterministic case and is very dramatic even in the stochastic case, where all neighbors neither are simultaneously infected nor have the same probability of infection.

To unleash the potential of the immunization system, we offer a slight modification of the problem by introducing a relatively small number of links to the network topology. These immunization links, which are used exclusively by the immunization agents, have a dramatic effect on the ability of a dynamic immunization scheme to contain the virus spread by offering access to a parallel network which has identical nodes and almost identical links set as the original network. In our example, the parallel network is the phone book network, which has a strong correlation with the e-mail network. These links connect the node that produced the immunization agent to nodes that are beyond its immediate neighborhood as defined by the the initial network. Thus, once the immunization agent is produced, it can be deployed "behind enemy lines", unconstrained by the boundaries of the surrounding virus cluster. In figure 1 we illustrate the difference between a network with no extra immunization edges and a network which posses some edges of this type.

The effect of introducing additional immunization links amounts to the generation, together with the original network, of a pair of *Partially Correlated Networks*^{23,24}, which is defined as follows:

Two given networks $G_1 = (V, E_1), G_2 = (V, E_2)$ are partially correlated with overlap p if

$$p = \frac{|E_1 \cap E_2|}{\max(|E_1|, |E_2|)}.$$

Starting with our initial network G_1 , we created a new network G_2 for the immunizing

agent by adding to G_1 a set of edges e_1 which do not belong to G_1 . Using the relative edges addition, $q = |e_1|/|E_1|$, the overlap will be:

$$p = \frac{|E_1|}{|E_2|} = \frac{|E_1|}{|E_1 \cup e_1|} = \frac{1}{1 + q}. \quad (1)$$

Then, we alter the Distributed Immunization Dynamics in the following way. The virus spreads on the original network G_1 , while the immunizing agent is deployed over the partially correlated network G_2 , obtained by adding $q|E_1|$ edges randomly to G_1 . By doing so, we achieve our aim to allow the immunizing agent to break through the virus cluster and immunize the network.

In the methods section we show analytically that for the discrete time, deterministic model, the relative size of the infected cluster (i.e. the ratio of infected to immunized clusters) as a function of the relative edge addition q has a power law upper bound with power exponent equal to -1 .

Additionally, we have studied the problem through network simulations. In Figure 2, we present simulation results which show that the ratio dependence indeed exhibits a power law dependence with an exponent close to $-4/3$. The virus cluster is reduced by 44% by a relative edge addition of six thousandth.

Thus, we conclude that dynamic immunization employed over partially correlated networks can considerably reduce the size of a virus cluster at a negligible price.

More systematic architectures for the immune system may be envisioned. We present a *Honey Pot Architecture*, with a fundamental aim to introduce a virtual super-hub, which trans-

forms the shortcomings of a scale free network (which is considerably impaired when its largest hubs are removed)¹¹ into an advantage.

The Honey Pot Architecture is constructed in the following manner: We implant the exclusive ability to develop an immunizing agent to a set of randomly distributed nodes in our network (*the Honey Pots*). *The Honey Pots* are embedded randomly in our network, such that any virus spreading through the network will reach them very soon with a high probability. Finally, all Honey Pots are connected in a complete graph topology using special edges which permit passage of the immunizing agents only.

The dynamics of this architecture is as follows. Initially, the virus spreads unhindered, until it infects the first *Honey Pot* and triggers an immunization agent development process. By this time, the size of the virus cluster on average equals the network size divided by the number of *Honey Pots* it contains. As the virus continues to spread, all the *Honey Pots* are notified of the new virus, and each *Honey Pot* then functions as the root of a separate infectious immunization process. The *Honey Pots* have the effect of a super-hub, with a degree which is the sum of the degrees of the separate *Honey Pots*, and links spanning all areas in the graph.

In the methods section we calculate an upper bound for the relative virus cluster under the honey pot architecture. It is shown that if the amount of *Honey Pots* as a function of the network size N , $f(N)$, grows faster than $N^{\frac{1}{2}}$ the size of the virus cluster will approach zero portion of the network, as the network size approaches infinity. In the case where $f(N) = \beta N$, we get

$$\frac{V_n}{A_n} = \frac{1}{\beta^2 N} \cdot (\alpha - 1) \quad (2)$$

which means we have a power law relation between this ratio and the relative amount of honey pot nodes, β , with exponent equals to -2 , as expected, since $f(N) = N^{\frac{1}{2}}$ is the function for

which the relative special edge addition due to the *Honey Pots* is kept constant in the infinite size limit. This analytic estimation is validated in simulations, and presented in figures 3 and 4.

In the face of the systematic defeats in the war against computer viruses, a paradigm shift may be required. We propose such a shift from the current, static, centralized immunization strategies to a dynamic distributed immune system approach. The effectiveness of such architecture in protecting large networks, both when built randomly or when designed artificially is demonstrated. Naturally, the proposed architectures are only first examples of a potentially new type of network-wide immunization systems which focus on network survivability, rather than the survivability of the individual network nodes.

Methods

Analysis of the random edges effect Given a network with degree distribution $P(k)$, let us calculate the rate of growth of the virus cluster: let us look on the portion of the $n + 1$ time layer with degree k .

$$l_{n+1}(k) = P(k) \cdot \sum l_n(k') \cdot C \cdot (k' - 1) \quad (3)$$

where C holds the topological clustering properties of the network (which reduces the number of effective neighbors). Since the sum does not depend on k , we can calculate the sum independently and call it a_{n+1} . Then, $l_{n+1}(k) = a_{n+1}P(k)$. Substituting in (1), gives us

$$l_{n+1}(k) = l_n(k) \cdot \sum P(k') \cdot C \cdot (k' - 1) \quad (4)$$

We call the outcome of the new sum α . Since it does not depend on k , we get that $l_{n+1} = l_n \cdot \alpha$. If α is larger than 1 we get an exponential growth.

Let us turn to the immunized cluster(s). Given a relative edge addition q , and an average degree m , we get that the probability that a node will have an immune specific edge is qm . each infected node (in the deterministic case) has a probability qm to initiate an immunized cluster. Once started, these clusters grow with ratio α also. The immunized set of clusters grow like:

$$A_n = qm \left[(n-1) \cdot \alpha^{n-2} + (n-2) \cdot \alpha^{n-3} + \dots + 1 \right]. \quad (5)$$

which can be compacted:

$$A_n = qm \left[\frac{n\alpha^{n-1}}{\alpha-1} - \frac{\alpha^n - 1}{(\alpha-1)^2} \right] \quad (6)$$

The ratio between the size of the virus cluster and immunized clusters is:

$$\frac{V_n}{A_n} = \frac{1}{qm} \left[\frac{(\alpha^{n+1} - 1)(\alpha - 1)}{(n-1)\alpha^n - n\alpha^{n-1} + 1} \right] \quad (7)$$

From which we get an upper bound (since all our assumptions were in favor of the virus cluster) power law dependence with exponent -1 on q for the discrete time, deterministic model.

Analysis of the Honey Pot architecture effect We would like to calculate an upper bound for the ratio $\frac{V_N}{A_N}$ where N approaches infinity.

Let us assume that there are $f(N)$ honey pots distributed randomly in the network, all connected in a complete graph using immunization edges. Then clearly the average virus cluster size when a honey pot is infected with a virus for the first time is: $\frac{N}{f(N)}$ where N is the system size. At the next time step, there will be on the boundary of the virus cluster $\frac{N}{f(N)} \cdot (\alpha - 1)$ nodes. At the same time, there will be $f(N)$ nodes ‘infected’ with the immunization agent. From this point forward, we assume that (in the deterministic case) the virus cluster and immunized clusters grow

as a geometric series uninterrupted. Then their ratio approaches:

$$\frac{V_n}{A_n} = \frac{\frac{N}{f(N)} \cdot (\alpha^n - 1)}{f(N) \cdot \frac{\alpha^n - 1}{(\alpha - 1)}} = \frac{N}{f(N)^2} \cdot (\alpha - 1) \quad (8)$$

From this equation we can see that whenever $f(N)$ grows faster than $N^{\frac{1}{2}}$ the size of the virus cluster will approach zero portion of the network, as the network size approaches infinity. In the case where $f(N) = \beta N$, we get

$$\frac{V_n}{A_n} = \frac{1}{\beta^2 N} \cdot (\alpha - 1) \quad (9)$$

which means we have a power law relation between this ratio and the relative amount of honey pot nodes, β , with exponent equals to -2 . This result is not surprising, since $f(N) = N^{\frac{1}{2}}$ is the function for which the relative special edge addition due to the *Honey Pots* is kept constant in the infinite size limit.

1. Beard, A. & Potter, C. information security breaches survey 2004. Survey, Department of Trade and Industry, UK (2004).
2. Friedrichs, O. Symantec internet security threat report volume v. techreport, Symantec Inc. (2004). URL <http://www.symantec.com/threatreport>.
3. Leyden, J. Sobig-f is dead. news article, The Register (2003).
4. Newman, M., Forrest, S. & Balthorp, J. Email networks and the spread of computer viruses. *Physical Review E* **66**, 035101 (2002).
5. Havlin, S., Cohen, R. & Ben-Avraham, D. Efficient immunization strategies for computer networks and populations. *Phys. Rev. Lett.* **91**, 247901 (2003).
6. 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* (2002).

7. Pastor-Satorras, R. & Vespignani, A. Epidemic spreading in scale-free networks. *Phys. Rev. Lett.* **86**, 3200 (2001).
8. Moore, D., Shannon, C., Voelker, G. & Savage, S. Internet quarantine: Requirements for containing self-propagating code. In *Proceedings of the 2003 IEEE Infocom Conference* (2003).
9. Watts, D. J. & Strogatz, S. H. Collective dynamics of 'small-world' networks. *Nature* **393**, 440–442 (1998).
10. Albert, R., Jeong, H. & Barabasi, A.-L. Emergence of scaling in random networks. *Science* **286**, 509 (1999).
11. Albert, R., Jeong, H. & Barabasi, A.-L. Error and attack tolerance of complex networks. *Nature* **406**, 378 (2000).
12. Chung, F. & Lu, L. The average distances in random graphs with given expected degrees. *PNAS* **99**, 15879 (2002).
13. Cohen, R., Erez, K., ben Avraham, D. & Havlin, S. Resilience of the internet to random breakdowns. *Phys. Rev. Lett.* **85**, 4626 (2000).
14. Pastor-Satorras, R. & Vespignani, A. Immunization of complex networks. *Physical Review E* **65**, 036104 (2002).
15. Dezsó, Z. & Barabasi, A.-L. Halting viruses in scale-free networks. *Physical Review E* **65**, 055103 (2002).
16. Kephart, J., Sorkin, G., Swimmer, M. & White, S. Blueprint for a computer immune system. In *Proceedings of the 1997 Virus Bulletin International Conference* (1997).

17. Kreibich, C. & Crowcroft, J. Honeycomb-creating intrusion detection signatures using honeypots. *Computer Communication Review* **34(1)**, 51–56 (2004).
18. Ebel, H., Mielsch, L.-I. & Bornholdt, S. Scale-free topology of e-mail networks. *Physical Review E* **66**, 035103 (2002).
19. Newman, M. Spread of epidemic disease on networks. *Physical Review E* **66**, 016128 (2002).
20. May, R. M. & Lloyd, A. L. Infection dynamics on scale-free networks. *Physical Review E* **64**, 066112 (2001).
21. Kephart, J. & Arnold, W. C. Automatic extraction of computer virus signatures. In *Proceedings of the 4th Virus Bulletin International Conference 1994* (1994).
22. Huang, Z.-F. Self-organized model of information spread in financial markets. *Eur. Phys. J. B* **16**, 379 (2000).
23. K.Malarz. Social phase transition in solomon network. *Int. Journal of Mod. Phys. C* **14**, 561 (2003).
24. Chen, L.-C. & Carley, K. M. The impact of countermeasure propagation on the prevalence of computer viruses. *IEEE Transactions on Systems, Man and Cybernetics, Part B: Cybernetics* **34(2)**, 823 (2004).

Figure 1 Comparing infection process evolution with (bottom) and without (top) immunization edges. On the top row we see the network being infected fully by the virus. On the bottom row the virus cluster was reduced by half by introducing immunization edges. Blue, orange and green nodes represent susceptible, infected and immuned network nodes, respectively. The blue edges represent the original network edges, while the green edges represent the additional immunization links. On the top row we see three snapshots of a network in three different time coordinates, illustrating the virus dissemination over the entire network. On the bottom row we see the effect of introducing additional immunization edges, which reduces the size of the virus cluster considerably.

Figure 2 Relative virus cluster size as a function of immunization links density (log-log scale). The dependence of the relative infected cluster size on the relative edge addition q , as resulting from simulations over uncorrelated, scale free networks with power exponent -3 , mean degree equal to 4, and network size ranging from 100000 to 170000 nodes. Ratio dependence exhibits a power law form, with an exponent close to $-4/3$. The virus cluster is reduced by 44% by a relative edge addition of six thousandth.

Figure 3 Relative virus cluster size as a function of system size for different honey pots densities. The simulations were ran over uncorrelated, scale free networks with power exponent -3 , mean degree equal to 4, and network size ranging from 30000 to 150000 nodes. The curves show a power law dependence with exponent close to -1 . Illustrated is the inverse proportionality between the relative infected cluster size and the system size (N) for different values of density β . While, naturally, the

relative size differs for different densities, the dependence on system size have a similar power law nature, with exponents 0.98, 0.98 and 0.87 for the three presented density values: 0.002, 0.003, 0.004. In the simulation of a 100000-node scale-free network, designating less than one thousandth of the nodes as *Honey Pots* reduces the virus cluster portion of the network from 1 to less than 0.4.

Figure 4 Relative virus cluster size (multiplied by the system size - N) as a function of honey pots density for different system sizes. The simulations were ran over uncorrelated, scale free networks with power exponent -3 , mean degree equal to 4, and network size ranging from 30000 to 150000 nodes. Illustrated is the relation between the relative infected cluster size and the honey pot density for different system sizes, when the effect of the system size is being normalized out. A very similar dependence is observed, for largely different system sizes. The power exponents range from -1.7 to -1.4 , somewhat differing from the analytic -2 result, mainly due to finite size effects of the simulations.