HAL
open science

Rejuvenation and the Spread of Epidemics in General Topologies

Eitan Altman, Alberto Avritzer, Rachid El-Azouzi, Leandro Pfleger de Aguiar, Daniel Sadoc Menasche

To cite this version:

Eitan Altman, Alberto Avritzer, Rachid El-Azouzi, Leandro Pfleger de Aguiar, Daniel Sadoc Menasche. Rejuvenation and the Spread of Epidemics in General Topologies. 6th International Workshop on Software Aging and Rejuvenation (WoSAR 2014), Nov 2014, Naples, Italy. hal-01087263

HAL Id: hal-01087263

https://inria.hal.science/hal-01087263v1

Submitted on 25 Nov 2014

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

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

Rejuvenation and the Spread of Epidemics in General Topologies

Eitan Altman*, Alberto Avritzer¶, Rachid El-Azouzi§, Daniel S. Menasche‖ and Leandro Pfleger de Aguiar¶
*INRIA, Sophia-Antipolis, France
¶Siemens Corporation, Corporate Technology, Princeton, USA
‖Federal University of Rio de Janeiro, Rio de Janeiro, Brazil
§University of Avignon, Avignon, France

***Abstract*—Epidemic models have received significant attention in the past few decades to study the propagation of viruses, worms and ideas in computer and social networks. In the case of viruses, the goal is to understand how the topology of the network and the properties of the nodes that comprise the network, together, impact the spread of the epidemics. In this paper, we propose rejuvenation as a way to cope with epidemics. Then, we present a model to study the effect of rejuvenation and of the topology on the steady-state number of infected and failed nodes. We distinguish between a state in which the virus is incubating and in which symptoms might not be visible and yet they may be contagious and infecting other nodes, and a state of failure where symptoms are clear. Sampling costs might be incurred to examine nodes in search for viruses at an early stage. Using the proposed model, we show that the sampling rate admits at most one local minimum greater than zero. Then, we numerically illustrate the impact of different system parameters on the optimal sampling rate, indicating when rejuvenation is more beneficial.**

***Keywords*-Performance analysis, Reliability, Security**

I. Introduction

Epidemic models have received significant attention in the past few decades to study the propagation of viruses, worms and ideas in computer and social networks [1], [2]. In the case of viruses, the goal is to understand how the topology of the network and the properties of the nodes that comprise the network, together, impact the spread of the epidemics [1], [3]. If different treatments are available, at different costs, it is important to understand their impact on each individual node and on the network as a whole.

In this paper, we propose rejuvenation as a way to cope with epidemics. Rejuvenation consists of periodically sampling the nodes for early detection of viruses and preemptively running applications to prevent failures in the future. Full repair, in contrast, occurs after failure and requires significantly more effort.

We distinguish between a state in which the virus is incubating and in which symptoms might not be visible and yet they may be contagious and infecting other nodes, and a state of failure where symptoms are clear. One of the visible symptoms might be the decrease of capacity of nodes, also referred to as node degradation or failure. Aging of infected nodes produces degradation and allows the identification of failed nodes. Alternatively, nodes that carry the virus might be identified through sampling.

Sampling a node one can detect viruses and intrusions through their signatures [4]. In particular, viruses and security intrusions may be detected through identification of subtle changes in performance metrics (CPU, Memory, IO), also referred to as *performance signatures*, and the performance impact might depend on the sophistication of the attack. In some cases, it might be possible to detect a virus by analyzing performance signatures, even though the service capacity of the node as a whole does not show any level of degradation. In any case, sampling costs must be incurred if one wants to examine nodes to search for viruses and intrusions at an early stage. Using the proposed model, we numerically show the impact of sampling and rejuvenation on the steady-state population dynamics.

The two key questions posed in this paper are the following,

1) what is the impact of rejuvenation on the spread of epidemics and how does the effect of rejuvenation compare to other strategies such as relying only on full restoration after failures?
2) what is the effect of incubation time and how to quantify the effort for detecting viruses at an early stage?

Note that there is a vast literature on these topics [2], but not on general topologies (graphs). Our objective is to understand the impact of topology on the spread of epidemics.

To answer these questions, we propose an analytical model whose main components are a graph, wherein nodes are connected to their neighbors through edges, and each node can be in one of five states: a) vaccinated, b) healthy, c) prone to failure (infected), d) under rejuvenation and e) failed. Infected nodes have the virus and can infect others, without showing immediate symptoms. Infected nodes that are not sampled remain infecting other nodes until one discovers that they are infected and a treatment strategy is applied. The discovery that a node is infected might occur either because the node is sampled or because it fails, in which case its performance degradation signalizes that the node requires treatment.

Using the proposed model, we numerically show the impact of sampling, rejuvenation, vaccination and full repair on

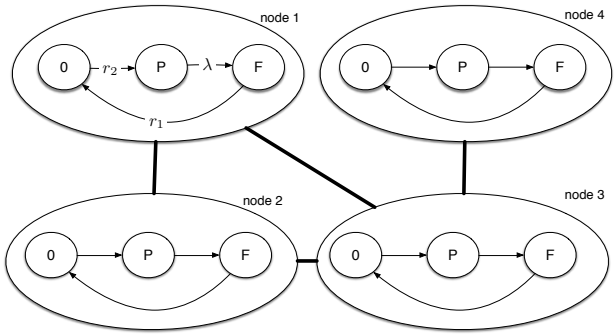

Figure 1. Simple model with 4 nodes without rejuvenation.

the steady-state population dynamics. Our key contributions are the following:

Comparison of rejuvenation and full repair: we show how rejuvenation and full repair can be used to cope with the spread of epidemics, indicating their advantages and disadvantages. We account for *sampling costs and benefits*, in order to identify whether nodes are infected at an early stage. Although sampling incurs costs, it might prevent even larger healing costs, as well as avoid the spread of the virus while infected nodes incubate the virus without symptoms.

Analytical model: we present an analytical model which can be used to study the impact of rejuvenation on the spread of epidemics. Our model is tractable and allows to compute the transient and steady-state number of nodes in each of the possible states as a function of different system parameters. Using the proposed model, we establish conditions for the uniqueness of the optimal sampling rate, and we numerically indicate how different system parameters affect the optimal sampling rate.

The rest of this paper is organized as follows. In Section II we present a simple instance of our model, which already gives insights about the methodology used in the general model introduced in Section III. In Section IV we present numerical results, obtained with the proposed model, which relate the optimal sampling rate to different system parameters. Section V contains related work and Section VI concludes.

II. Aging and Infection

To appreciate the key features of the proposed model, in this section we illustrate the proposed model in a simple setup accounting for aging, infection and full repair. Sampling costs and rejuvenation are not taken into account and the only treatment available is to fully repair failed nodes. In the next section, the simplifying assumptions are removed. The model presented in this section extends the model introduced in [2, Figure 1] to a network setting and general topologies.

Figure 1 illustrates the key components of our model. In state P the node is prone to failure but its performance remains the same as that of a healthy node, hence an external observer cannot distinguish between a healthy node and an infected node (except after careful examination also referred to as sampling). In state F, in contrast, the node has failed.

Each infected or failed node infects each of its neighbors at rate β. Let A_{ij} be equal to 1 if node j is a neighbor of i and 0 otherwise. A healthy node i becomes infected by its neighbors at rate $r_{i,2}(t)$, and by exogenous threats at rate κ_i. An infected node further develops the infection and fails at rate λ. A failed node submitted to treatment heals at rate r_1 (Table I contains the notation). We assume that all times between events are exponentially distributed.

Let $X_{i,0}(t)$, $X_{i,P}(t)$ and $X_{i,F}(t)$ be three indicator random variables equal to 1 if node i is healthy, prone to failure and failed, respectively, and 0 otherwise. Accordingly, let $\pi_{i,0}(t)$, $\pi_{i,P}(t)$ and $\pi_{i,F}(t)$ be the probabilities that node i is healthy, prone to failure (infected) and failed at time t, respectively. Note that $E[X_{i,0}(t)] = \pi_{i,0}(t)$, $E[X_{i,P}(t)] = \pi_{i,P}(t)$ and $E[X_{i,F}(t)] = \pi_{i,F}(t)$.

The time change of $E[X_{i,P}(t)]$ is given by

$$\frac{dE[X_{i,P}(t)]}{dt} = E[-\lambda X_{i,P}(t)] + \\ + \quad E\left[X_{i,0}(t)\left(\sum_{j=1}^{N} A_{ji}\beta(1-X_{j,0}(t)) + \kappa_i\right)\right] \quad (1)$$

The direct and exact solution of (1) requires the characterization of the expected value of $X_{i,0}$ conditioned on all possible combinations of states of neighbors of node i, for $i = 1, 2, \ldots, N$. The number of variables to be determined grows exponentially with respect to N. For this reason, in the remainder of this paper we rely on a mean-field (MF) approximation referred to as N-interwinded MF approximation (NIMFA) [5] to compute the fraction of infected nodes. The approximation consists of replacing the expectation of the product of random variables by the product of their expectations.

Let the endogenous infection rate of node i by its neighbors, at time t, be

$$r_{i,2}(t) = \sum_{j=1}^{N} A_{ji}\beta(1-\pi_{j,0}(t)) \quad (2)$$

Applying the NIMFA approximation to (1) and making use of (2) we obtain

$$\frac{d\pi_{i,P}(t)}{dt} = -\lambda\pi_{i,P}(t) + \pi_{i,0}(t)\left(r_{i,2}(t) + \kappa_i\right) \quad (3)$$

$$\frac{d\pi_{i,F}(t)}{dt} = \lambda\pi_{i,P}(t) - r_1\pi_{i,F}(t) \quad (4)$$

$$\frac{d\pi_{i,0}(t)}{dt} = -\left(\frac{d\pi_{i,P}(t)}{dt} + \frac{d\pi_{i,F}(t)}{dt}\right) \quad (5)$$

$$\pi_{i,0}(t) = 1 - \pi_{i,P}(t) - \pi_{i,F}(t) \quad (6)$$

Equation (3) characterizes the probability that node i is infected at time t, which decreases at failure rate λ in case

node i is infected and increases at infection rate $r_{i,2}(t)+\kappa_i$ in case node i is healthy. Equation (4) characterizes the probability that node i is failed, which increases at rate λ in case node i is infected and decreases at repair rate r_1 in case node i is failed. Equations (5) and (6) capture the fact that each node must be either healthy, infected or failed.

Let $\pi_{i,0}$, $\pi_{i,P}$ and $\pi_{i,F}$ be the steady state probabilities that node i is susceptible, infected and failed, respectively. Next, consider the special case of a connected graph where all nodes have the same degree k. In addition, assume $\kappa_i = \kappa$ for $i = 1, \ldots, N$. Then, by symmetry we have $\pi_{i,F} = \pi_F$, $\pi_{i,0} = \pi_0$ and $\pi_{i,P} = \pi_P$ for $i = 1, \ldots, N$, where π_0, π_P and π_F are the steady state probabilities that a node is susceptible, infected and failed, respectively, in the symmetric topology. It follows from (3)-(6) that the non-trivial equilibrium for the symmetric topology is given by the value of π_F which solves the following equation,

$$\left(1+\frac{r_1}{\lambda}\right)^2 \beta k \pi_F^2 + \left(r_1 + (\kappa - \beta k)\left(\frac{r_1}{\lambda}+1\right)\right)\pi_F = \kappa \quad (7)$$

In the absence of exogenous infection rates ($\kappa = 0$) and in case infected nodes immediately degrade ($\lambda = \infty$), the solution of (7) is given by [6, equation (2)].

III. Sampling and Rejuvenation

In this section we generalize the model presented in the previous section to account for sampling costs as well as rejuvenation. Our goal is to show that a simple adaptation of the previously presented model allows us to account for these additional features. To this aim, we extend [2, Figure 2] to general topologies.

We start presenting the basic balance equations which capture the dynamics presented in Figure 2. According to the NIMFA approximation [5] we have

$$\begin{aligned}
\frac{d\pi_{i,P}(t)}{dt} &= -(\lambda + r_4)\pi_{i,P}(t) + && (8)\\
&+ \pi_{i,0}(t)\left(\sum_{j=1}^{N} A_{ji}\beta(1-\pi_{j,0}(t)) + \kappa_i\right) \\
\frac{d\pi_{i,R}(t)}{dt} &= r_4\pi_{i,P}(t) - \pi_{i,R}(t)(r_3+\gamma) && (9)\\
\frac{d\pi_{i,F}(t)}{dt} &= \lambda\pi_{i,P}(t) + \gamma\pi_{i,R}(t) - r_1\pi_{i,F}(t) && (10)\\
\frac{d\pi_{i,0}(t)}{dt} &= -\left(\frac{d\pi_{i,P}(t)}{dt} + \frac{d\pi_{i,R}(t)}{dt} + \frac{d\pi_{i,F}(t)}{dt}\right) && (11)\\
\pi_{i,0}(t) &= 1 - \pi_{i,P}(t) - \pi_{i,R}(t) - \pi_{i,F}(t) - \pi_{i,V}(t) && (12)
\end{aligned}$$

Comparing Figures 1 and 2, the two new states represented in Figure 2 are R and V which correspond to nodes under rejuvenation that were sampled, wherein infection was detected at an early stage, and vaccinated nodes, respectively. We assume that sampling for early infections occurs according to a Poisson process with rate r_4, and failures after

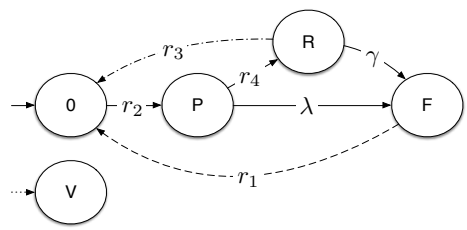

Figure 2. A node of the complete model with rejuvenation. Edges representing vaccination, rejuvenation, full repair and aging are represented by dotted, broken-dotted, broken and solid lines, respectively

infection occur with rate λ. Therefore, we have transitions from P to R and F with rates r_4 and λ, respectively. We have a transition from R to F with rate γ, characterizing failures that might occur during rejuvenation. The effect of γ is subject for future work, and in this paper we assume $\gamma = 0$ as in [2]. Under normal conditions, a node that is prone to failure can be rejuvenated. Once started, the rejuvenation takes mean time $1/r_3$. Finally, full repairs occur from state F, and yield a transition to state 0.

Vaccinated nodes remain in state V, which is an absorbing state, as we assume that the vaccine completely imunizes the nodes and they are removed from the graph. In the remainder of this paper, except otherwise noted we focus on nodes that are not vaccinated.

Note that if we do not allow for sampling and rejuvenation ($r_4 = r_3 = 0$) the model introduced in this section degenerates to the one presented in Section II. The proposed model is general enough to allow for different recovery strategies at different nodes (i.e., some nodes might be amenable to rejuvenation and early sampling whereas others might count only with rejuvenation). The methodology to solve the model above is the same as the one used in the previous section.

A. Regular Topology

Next, we let $\kappa_i = 0$ and consider the special case of a connected graph where all nodes have the same degree k. In addition, we restrict ourselves to nodes that are not vaccinated (recall that vaccinated nodes are removed from the infection network). Due to symmetry, we drop the subscript i from the state variables to obtain,

$$\begin{aligned}
\pi_0 &= \frac{\lambda + r_4}{\beta k(\frac{\lambda}{r_1} + 1 + \frac{r_4}{r_3})} && (13)\\
\pi_P &= \frac{1-\pi_0}{\frac{\lambda}{r_1} + 1 + \frac{r_4}{r_3}} && (14)\\
\pi_R &= \frac{r_4}{r_3}\pi_P, \qquad \pi_F = \frac{\lambda}{r_1}\pi_P && (15)
\end{aligned}$$

where we assume that $\beta k > (\lambda + r_4)/((\lambda/r_1) + 1 + (r_4/r_3))$. Otherwise, only the trivial solution $(\pi_0, \pi_P, \pi_R, \pi_F) = (1, 0, 0, 0)$ exists.

B. Costs: Infection Probabilities and Sampling Rates

Next, we consider the infection and sampling costs incurred by the nodes. The setup is the one considered in

variable	description
$\pi_{i,0}(t)$	probability that node i is healthy at time t
$\pi_{i,P}(t)$	probability that node i is prone to failure (infected) at time t
$\pi_{i,R}(t)$	probability that node i is under rejuvenation at time t
$\pi_{i,F}(t)$	probability that node i is failed at time t
κ_i	exogenous infection rate at node i
r_1	full repair rate
$r_{i,2}$	endogenous infection rate at node i
r_3	repair rate (rate of recovery once rejuvenation started)
r_4	sampling rate
λ	failure rate
β	rate of infection of infected node towards each neighbor
A_{ij}	equals 1 if node j is a neighbor of i and 0 otherwise
k	number of neighbors of each node in a regular graph
N	number of nodes in the network
c_v	cost of infection per time unit
c_s	cost of sampling per sample
$C(r_4)$	mean total cost

Table I
TABLE OF NOTATION. NOTE THAT WE DROP THE DEPENDENCY ON t WHEN CONSIDERING STEADY-STATE, AND THE DEPENDENCY ON i WHEN CONSIDERING SYMMETRIC TOPOLOGIES.

Section III-A. The infection cost is the cost due to a machine having a virus. Let c_v be the average cost per time unit incurred when a machine is infected. The cost c_v is proportional to the severity of the infection. The average infection cost per time unit incurred by a node is proportional to c_v and to the probability that a node is infected, $1-\pi_0$, and is given by $C_v(r_4) = c_v(1-\pi_0)$.

The sampling cost is the cost to examine the machines (e.g., to run a comprehensive checkup). Let c_s be the sampling cost (per sample) to identify an infection at an early stage. Samples are taken at states 0 and P. The fraction of time spent at these states is $(\pi_0+\pi_P)$. Therefore, the expected number of samples taken per time unit is $(\pi_0+\pi_P)r_4$ and the expected sampling cost per time unit is $C_s(r_4) = (\pi_0+\pi_P)r_4c_s$.

Let $C(r_4)$ be the mean total cost per time unit. It is given by the sum of the mean sampling and infection costs,

$$C(r_4) = C_s(r_4) + C_v(r_4) = c_s r_4(\pi_0+\pi_P) + c_v(1-\pi_0) \tag{16}$$

Theorem 1: If sampling costs are negligible ($c_s = 0$), the optimal sampling strategy consists of sampling the system at the highest possible rate (i.e., setting $r_4 = \infty$) if

$$\frac{1}{\lambda} + \frac{1}{r_1} > \frac{1}{r_3} \tag{17}$$

and not sampling (i.e., setting $r_4 = 0$) otherwise.

Proof: Increasing the sampling rate is beneficial if the derivative of (13) with respect to r_4 is positive, i.e., $d\pi_0/dr_4 \geq 0$. It is easily verified that $d\pi_0/dr_4 \geq 0$ if (17) is satisfied. ■

Let $r_4^\star$ be the optimal sampling rate. In the presence of sampling costs ($c_s > 0$), the following theorem establishes that either $r_4^\star = 0$ or $r_4^\star$ is the unique local minimum of $C(r_4)$ in $(0,\infty)$.

Theorem 2: If $c_s > 0$, the mean total cost $C(r_4)$ admits at most one local minimum in $(0,\infty)$.

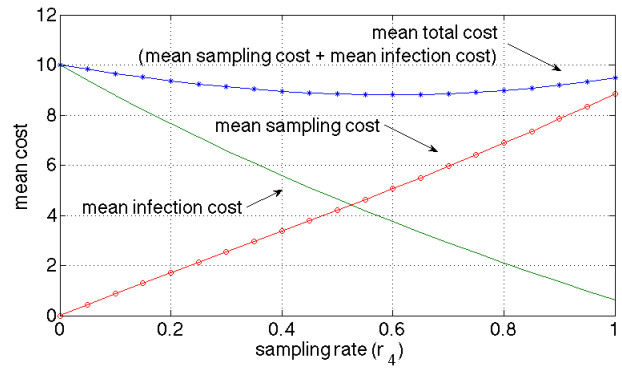

Figure 3. Finding the optimal sampling rate

Proof: Replacing (13) and (14) into (16) we obtain an expression for $C(r_4)$ that depends only on system parameters. Taking the second derivative of the resulting expression with respect to r_4, we note that $\frac{d^2C(r_4)}{dr_4^2} = 0$ admits at most one real root. It follows that $C(r_4)$ is either $i)$ strictly concave, $ii)$ strictly convex or $iii)$ has one single inflection point. But $\lim_{r_4\to\infty}\frac{dC(r_4)}{dr_4} = c_s r_3/(k\beta)$ is strictly positive, meaning that increasing r_4 beyond a threshold increases the mean total cost unboundedly. Therefore, $i)$ $C(r_4)$ cannot be strictly concave, $ii)$ if $C(r_4)$ is strictly convex it admits only one local minimum in $[0,\infty)$, which is also the global minimum, and $iii)$ if $C(r_4)$ has a single inflection point it has at most two local minima, one at $r_4 = 0$ and another in $(0,\infty)$. The result follows from $i), ii)$ and $iii)$. ■

IV. NUMERICAL RESULTS

In this section we numerically illustrate some of the conclusions derived in the previous sections. Our goals are a) to quantify the potential gains of rejuvenation, b) to show how different system parameters impact the mean total cost and the optimal sampling rate. Our base setup is the one considered in [2, Figure 3]: $\lambda = 1/(12\times 30\times 24)$, $r_1 = 2$ and $r_3 = 3$. We consider a full mesh topology (complete graph) and let $k = 10$, $\beta = 0.08$, $c_s = 9$ and $c_v = 10$. Given the initial setup, we vary the system parameters according to our experimental goals.

Figure 3 shows the tradeoff involved in the choice of the sampling rate. Initially, when $r_4 = 0$ (no rejuvenation), the mean total cost equals infection costs ($C(0) \approx 10.00$) and it is beneficial to increase the sampling rate. When $r_4 = 0.60$, the mean total cost reaches its minimum of $C(0.60) = 8.8$. Further increasing r_4 increases the mean total cost. When $r_4 > 0.6$ sampling costs dominate the total cost.

Figure 4 shows how the sampling cost per sample, c_s, affects the optimal sampling rate. As c_s increases, the optimal sampling rate decreases. The optimal sampling rate r_4 is roughly $0.85, 0.60, 0.35$ for $c_s = 7, 9, 11$ respectively.

Figure 5 shows how the infection rate per contact, β, affects the optimal sampling rate. As β increases, the optimal sampling rate decreases. The optimal sampling rate r_4 is

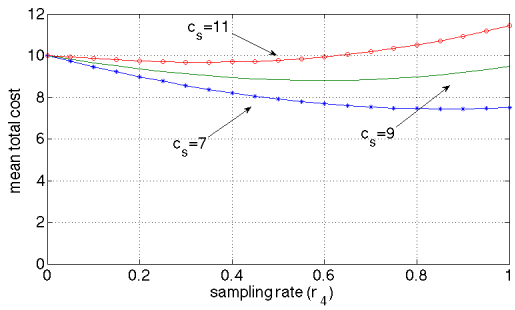

Figure 4. Varying the sampling cost per sample, c_s.

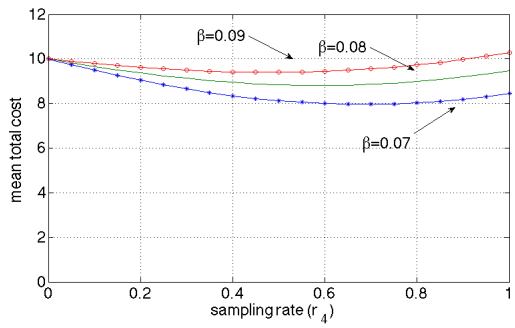

Figure 5. Varying the infection rate per contact, β.

roughly $0.7, 0.6, 0.5$ for $\beta = 0.07, 0.08, 0.09$ respectively. This is because as nodes are more prone to infection, sampling must occur at a higher rate to cope with greater infection costs.

Figure 6 shows how the mean time to complete rejuvenation, $1/r_3$, affects the optimal sampling rate. As $1/r_3$ decreases, the mean total cost decreases as the system repair time is reduced. In addition, the optimal sampling rate increases as r_3 increases. As the time to execute rejuvenation decreases, it becomes more advantageous to detect infections at very early stages and immediately treat the nodes as opposed to waiting for them to fail and then execute a full repair. The optimal sampling rate r_4 is roughly $0.35, 0.5, 0.6$ for $r_3 = 1, 2, 3$ respectively.

Infection costs might differ between nodes, as more

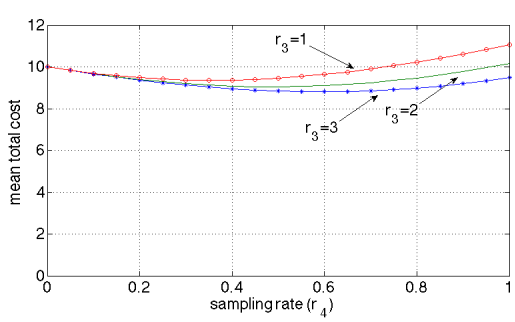

Figure 6. Varying the mean time to complete rejuvenation, $1/r_3$.

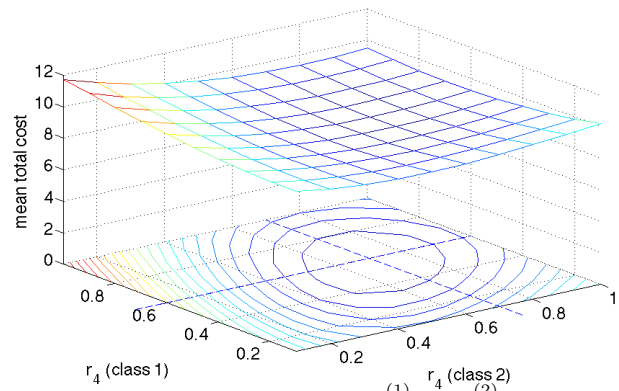

Figure 7. Asymmetric setup with $c_v^{(1)} = 7, c_v^{(2)} = 14$.

critical nodes might incur higher costs than others. Consider an asymmetric setup where nodes are divided in two classes. The population is equally divided between the two classes. Nodes of classes 1 and 2 incur costs $c_v^{(1)} = 7$ and $c_v^{(2)} = 14$, respectively, due to infection, when infected by viruses. The model presented in the previous section naturally extends to multiple classes. We use superscripts (1) and (2) to denote variables associated to classes 1 and 2. Figure 7 shows the mean total cost $C(r_4^{(1)}, r_4^{(2)})$ as a function of the sampling rates of nodes in the two classes. The minimum mean total cost is 8.9 and is attained when $(r_4^{(1)}, r_4^{(2)}) = (0.6, 0.7)$. Nodes of class 2 need to be sampled at higher rates than those of class 1 because $c_v^{(2)} > c_v^{(1)}$. In contrast, if $c_v^{(1)} = c_v^{(2)} = 10$ (not shown in the figure) the minimum cost is 8.8 and is attained when $(r_4^{(1)}, r_4^{(2)}) = (0.6, 0.6)$ (in accordance to Figure 4, for $c_s = 9$).

V. Related Work

There is a vast literature on epidemic models to analyze the spread of diseases [1], viruses in networks [7] and ideas in social networks [8]. Rejuvenation as a possible strategy to mitigate the effects of worm epidemics was introduced in [3]. In this paper, we propose a novel model to account for the effects of rejuvenation in general topologies.

From an industrial viewpoint, the use of rejuvenation in telecommunication systems was discussed in [9], [10]. The typical approach consists of monitoring OS-level indicators to detect performance degradation and anomalies [11], [12]. Avritzer and Weyuker [13] propose heuristics to decide if a node requires treatment as a function of its performance degradation. In this paper, in contrast, we assume that the performance degradation of infected but non-failed nodes is negligible, and the performance degradation of failed nodes is immediately observable. Future work consists of considering multiple performance degradation levels and the effort to detect them.

We show that approximations to compute steady-state metrics on networks, such as the NIMFA model [5], can be extended to account for rejuvenation and sampling. Our

model extends the standard susceptible-infected (SI) model in a number of ways. In the standard SI model, nodes are either susceptible or infected, and transition among these two states until they all reach the susceptible state, which is an absorbing state. Previous works considered the time till absorption [14] or the quasi-stationary regime [15], [7], [5], [6], [16]. As we show in Section II, some of the results presented in [5], [6] are special cases of ours, obtained by setting the exogenous rate to zero.

Newman [1, Section 18.3] studies dynamics on graphs, with more than one state variable per vertex. Our work extends [1] by making use of the N-Interwined Mean-Field Approximation (NIMFA) [5]. In addition, we also discuss sampling and rejuvenation strategies which are out of the scope of [1].

Asavathiratham proposes the influence model [17] as a tractable way to handle pairwise interactions of nodes, each node having its associated Markov chain to represent its internal state. Garetto et al. [18] study the spread of viruses in a network under this model. According to the influence model, a susceptible node that has one single neighbor, which turns out to be infected, will become infected at the same rate as a node with multiple neighbors all of which are infected. This is because interactions occur in a pairwise manner. In the model proposed in this paper, in contrast, the larger the number of infected neighbors incident to a susceptible node, the larger the infection rate of the susceptible node [14], [5].

VI. Conclusion

In light of security threats posed by viruses and worms to cyber infrastructures, the study of the spread of epidemics has gained significant attention in the past few years. In this paper, we discussed the role of sampling and rejuvenation as solutions to mitigate epidemics. We showed when and how rejuvenation is beneficial, and we established conditions under which the sampling rate admits one global minimum. Then, we numerically illustrated how different system parameters impact the optimal sampling rate. We believe that the models proposed in this paper are the first step towards a better understanding of how to sample and rejuvenate nodes accounting for security threats. This work opens up several avenues for future exploration, including control and game theoretical aspects involved in the choice of when and how to sample, vaccinate, rejuvenate and heal nodes.

Acknowledgment: Research conducted within the context of the THANES Associate Team, jointly supported by Inria (France) and FAPERJ (Brazil). This work has been partially supported by the European Commission within the framework of the CONGAS project FP7-ICT-2011-8-317672 and by Siemens Corporation, Corporate Technology.

References

[1] M. Newman, *Networks: an introduction*. Oxford University Press, 2010.

[2] Y. Huang, C. Kintala, N. Kolettis, and N. D. Fulton, "Software rejuvenation: Analysis, module and applications," in *FTCS-25*. IEEE, 1995, pp. 381–390.

[3] A. Avritzer, R. G. Cole, and E. J. Weyuker, "Using performance signatures and software rejuvenation for worm mitigation in tactical manets," in *WOSP*, 2007, pp. 172–180.

[4] A. Avritzer, R. Tanikella, K. James, R. G. Cole, and E. Weyuker, "Monitoring for security intrusion using performance signatures," in *WOSP/SIPEW*. ACM, 2010, pp. 93–104.

[5] P. Van Mieghem, J. Omic, and R. Kooij, "Virus spread in networks," *ToN*, vol. 17, no. 1, pp. 1–14, 2009.

[6] Y. Hayel, S. Trajanovski, E. Altman, H. Wang, and P. Van Mieghem, "Complete game-theoretic characterization of sis epidemics protection strategies," in *CDC*, 2014.

[7] Y. Wang, D. Chakrabarti, C. Wang, and C. Faloutsos, "Epidemic spreading in real networks: An eigenvalue viewpoint," in *Reliable Distributed Systems*. IEEE, 2003, pp. 25–34.

[8] P. Netrapalli and S. Sanghavi, "Learning the graph of epidemic cascades," in *ACM SIGMETRICS Performance Evaluation Review*, vol. 40, no. 1. ACM, 2012, pp. 211–222.

[9] J. Alonso, A. Bovenzi, J. Li, Y. Wang, S. Russo, and K. Trivedi, "Software rejuvenation: Do IT and telco industries use it?" in *ISSREW*. IEEE, 2012, pp. 299–304.

[10] A. Avritzer, R. G. Cole, and E. J. Weyuker, "Methods and opportunities for rejuvenation in aging distributed software systems," *Journal of Systems and Software*, vol. 83, no. 9, pp. 1568–1578, 2010.

[11] A. Bovenzi, F. Brancati, S. Russo, and A. Bondavalli, "An os-level framework for anomaly detection in complex software systems," *Trans. Dependable and Secure Computing*, 2014.

[12] S. Garg, A. Puliafito, M. Telek, and K. S. Trivedi, "Analysis of software rejuvenation using Markov regenerative stochastic Petri net," in *Software Reliability Engineering*. IEEE, 1995, pp. 180–187.

[13] A. Avritzer and E. J. Weyuker, "Monitoring smoothly degrading systems for increased dependability," *Empirical Software Engineering*, vol. 2, no. 1, pp. 59–77, 1997.

[14] A. Ganesh, L. Massoulié, and D. Towsley, "The effect of network topology on the spread of epidemics," in *INFOCOM*, vol. 2. IEEE, 2005, pp. 1455–1466.

[15] S. Méléard, D. Villemonais *et al.*, "Quasi-stationary distributions and population processes," *Probability Surveys*, vol. 9, pp. 340–410, 2012.

[16] V. Auletta, D. Ferraioli, F. Pasquale, and G. Persiano, "Metastability of logit dynamics for coordination games," in *SODA*. SIAM, 2012, pp. 1006–1024.

[17] C. Asavathiratham, "The influence model: A tractable representation for the dynamics of networked Markov chains," Ph.D. dissertation, MIT, 2000.

[18] M. Garetto, W. Gong, and D. Towsley, "Modeling malware spreading dynamics," in *INFOCOM*, 2003, pp. 1869–1879.

Appendix
Proof of Theorem 1

The probability that a node is healthy is π_0, given by (13),

$$\pi_0 = \frac{r_1 r_3(\lambda + r_4)}{\beta k(\lambda r_3 + r_1 r_3 + r_1 r_4)} \tag{18}$$

Therefore,

$$\frac{d\pi_0}{dr_4} = \frac{r_1 r_3(\lambda r_3 - \lambda r_1 + r_1 r_3)}{\beta k(\lambda r_3 + r_1 r_3 + r_1 r_4)^2} \tag{19}$$

As the denominator of (19) is always positive, $\frac{d\pi_0}{dr_4} > 0$ if $\lambda r_3 - \lambda r_1 + r_1 r_3 > 0$,

$$\lambda + r_1 > \frac{\lambda r_1}{r_3} \therefore \tag{20}$$

$$\frac{1}{r_1} + \frac{1}{\lambda} > \frac{1}{r_3} \tag{21}$$

and in this case it is worth increasing r_4 unboundedly ($r_4 = \infty$). Otherwise, set $r_4 = 0$.

Appendix
Proof of Theorem 2

Follows below the Matlab code used to symbolically obtain the results discussed in the proof of theorem 2.

```
syms sl sr4 sk sb spi0 sr1 sr3 sr2
syms cs cv sr1

% for theorem 1

pi0=(1/(1+sl/sr1+sr4/sr3+((sl+sr4)/sr2)))...
    *((sl+sr4)/sr2)
diff(pi0,sr4)
msol=simple(diff(pi0,sr4))
pretty(msol)

% for theorem 2

a=simple(spi0* ((sk*sb*(1-spi0)*...
    (1+sl/sr1+sr4/sr3)+sl+sr4)) - sl -sr4)
a=collect(a,spi0)
pretty(a)
healthyprob=(simple(solve(a,spi0)))
healthyprob=healthyprob(2)


sr2=sb*sk*(1-healthyprob)
pi0=simple(healthyprob)
piP=simple(healthyprob * sr2/(sl+sr4))
piR=simple(piP*sr4/sr3)
piF=simple((sl/sr1)*piP)

if (simple(pi0-(sl+sr4)/...
        (sb*sk*(sl/sr1+1+sr4/sr3))) ~= 0)
    error('error in definition of pi0');
end
if (simple(piP-(1-pi0)/...
        ((sl/sr1+1+sr4/sr3))) ~= 0)
    error('error in definition of piP');
end
% alternative and equivalent
% definition of piP
pprob=healthyprob*(sk*sb*(1-healthyprob))...
    / (sl+sr4)
if (simple(pprob-piP) ~= 0)
    error('error in definition of piP');
end


healthyderiv1=simplify(simple(diff(...
    healthyprob,sr4)));


samplingcostderiv1=...
    simple(diff(sr4*healthyprob,sr4))+...
    simple(diff(sr4*pprob,sr4))

samplingcostderiv2=simple(...
    diff(samplingcostderiv1,sr4));
healthyderiv2=simple(diff(healthyderiv1,sr4));

pretty(simple(cv * (-healthyderiv2)+...
    cs * samplingcostderiv2))

secondderivtotalcost=cv * (-healthyderiv2)+...
    cs * samplingcostderiv2;

% next, we find the roots of the second
% derivate of the total cost

simpleroots2=simple(solve( ...
    secondderivtotalcost==0,sr4));

% the solution simpleroots2 is a 1 x 1 vector
% (meaning that the second derivative of C
% admits a unique root)


derivtotalcost=simple(-healthyderiv1 * cv +...
    samplingcostderiv1* cs)


% unfortunately, the derivative of C with
% respect to r_4 is too complicate
% so we cannot find its real roots directly
% (there are 3 roots, that admit
% closed form expressions,  but they
% are not simple, and it is not easy
% to identify which ones are imaginary)

simpleroots=simple(solve(...
    derivtotalcost==0,sr4))
pretty(simpleroots)


% therefore, we compute the following limit
% instead
mylimit=limit(derivtotalcost,sr4,inf)
% mylimit = c_s r_3 / (k beta)
```

Note that the last limit is reported and used in the proof of Theorem 2.