UDC 519.218.3

A Model for Network Virus Protection Based on Regenerative Process

Yu. Grishunina*, L. Manita*

* *National Research University Higher School of Economics, Moscow Institute of Electronics and Mathematics, Russia*

Abstract. We construct a mathematical model of antivirus protection of local area networks. The model belongs to the class of regenerative processes. To protect the network from the external attacks of viruses and the spread of viruses within the network we apply two methods: updating antivirus signatures and reinstallings of operating systems (OS). Operating systems are reinstalled in the case of failure of any of the computers (non-scheduled emergent reinstalling) or at scheduled time moments. We consider a maximization problem of an average unit income. The cumulative distribution function (CDF) of the scheduled intervals between complete OS reinstallings is considered as a control. We prove that the optimal CDF has to be degenerate, i.e., it is localized at a point τ.

Keywords: regenerative process, average unit profit, Laplace transform, virus propagation.

1. Introduction

The choice of strategy of antivirus protection is one of the key factors that determine effectiveness of the functioning of the local networks. Damage or loss of information that is commercial or state secret can lead to million losses and security threats. One of the main methods of solving the problems of the qualitative estimation of the possible risks and simulations of the virus propagation is mathematical modeling. Since propagation of computer viruses happens similarly as propagation of an epidemics in the population, epidemiological methods and terminology are widely used. To take random effects into account, stochastic epidemiological models were developed for modeling the propagation of viruses in a computer network (see, e.g., [1]- [4]). In these models it is assumed that the spread of viruses occurs only as a result of the interaction of infected and uninfected nodes within the network and does not take into account external virus attacks. In this paper we consider a stochastic model of virus protection of local area networks on the assumption that any computer can become infected due to the following two reasons: external attacks of viruses and spreading within network. We propose approach to optimization of the strategy of the antivirus protection is based on the fact that nowadays the only way to guarantee extermination of the viruses in the network is OS reinstall (the full system regeneration). Existence of regeneration points allows to construct a mathematical model based on a regenerating stochastic process. Our research continues the investigation of the problem of antivirus protection strategy discussed in [5].

2. Functioning of the local network with antivirus protection

We consider a local network (LAN) consisting of N computers (nodes). The antivirus protection is realized in two ways: by updating antivirus signatures and by reinstalling the OS. A node calls the update base at independent exponentially distributed with parameter β random time moments. The durations of the update install are independent exponential random variables (i.e.r.v.) of parameter γ. This type of antivirus protection does not guarantee the extermination of the viruses; if the node is infected, then it becomes "healthy" after the update with probability p_0 and remains infected with probability $(1 - p_0)$. OS reinstall happens on all the nodes of the network at the same moment of time and all the computers become "healthy" after it, i.e. the whole network regenerates. The expected reinstall time equals T. The decision about next scheduled OS reinstall is taken at the moment of the system regeneration according to a CDF $G(t)$. During the call for the signature update or OS reinstall the node is not working so no infection by viruses can take place.

Infection of a running node can happen in 2 ways: during successful virus attacks from outside the network or by interacting with infected nodes within the network. External virus attacks occur at time intervals that are i.e.r.v. of parameter λ. For any pair of nodes intervals between communications are i.e.r.v. of parameter α. We divide the viruses in 3 groups based on the damage. Type I viruses cause hidden damage: corrupt, destroy or transmit information. Type II viruses cause explicit system failures, so continuing work is impossible. Type III viruses combine features of both abovementionned types: they can cause hidden damage and system failures. We assume that moments of failure caused by infection by viruses of the types II and III are i.e.r.v. of parameter μ. Let p_1, p_2, p_3 be portions of the virus attacks of corresponding types: $p_1+p_2+p_3 = 1$.

The network functions as follows. At the initial moment of time the decision about the scheduled OS reinstall is taken according to the CDF $G(t)$. If there are no node failures until that moment, then the OS is reinstalled. If there is a node failure, then the reinstall happens at the failure moment. After OS reinstall the network is completely regenerated, and the decision about the next OS reinstall is taken accordingly to $G(t)$. The network work restarts at that moment.

The network makes profit from the income of every functioning node minus the antivirus protection expenses and the losses caused by failures. The profit is determined by the following parameters: c_0 – profit of one node per unit of time; c_1 – hidden damage caused by viruses per unit of time; c_2 – cost of the OS reinstall per unit of time; c_3 – cost of the new antivirus software installed during OS reinstall.

We consider a problem of finding a CDF $G(t)$ such that the average profit made by the network per time unit is maximal for networks working for long enough time.

3. Mathematical model

The mathematical model of the described LAN is a stochastic process $X(t) = (\xi(t), \eta(t), r(t))$, where $\xi(t)$ is the number of infected but working nodes at moment t, $\eta(t)$ and $r(t)$ – number of noninfected and infected nodes respectively, on which the updates are installed at the moment t. Obviously, the state space $\mathcal{X}$ of the process $X(t)$ is defined as follows: $\mathcal{X} = \mathcal{X}_0 \cup \{\mathcal{R}\}$, where $\mathcal{X}_0 = \left\{(i,j,k) : i,j,k = \overline{0,N},\, i+j+k \le N\right\}$. Here the event $\{X(t) = \mathcal{R}\}$ means that the OS reinstall happens at moment t. Since at the moment of the OS reinstall the network regenerates completely and restarts, the moments of the end of the OS reinstall (the moments of the transition of the process $X(t)$ from the state $\mathcal{R}$ to $(0,0,0)$) are regeneration points of the stochastic process $X(t)$.

The regeneration period consists of 2 intervals: time until the next scheduled or emergency (in case of failure of one of the nodes) OS reinstall and the time of the reinstall itself. Let Z be the duration of the regeneration period, τ – time between the update time and the beginning of the scheduled OS reinstall, $G(t) = P(\tau < t)$, Y – time between the update time and an emergency OS reinstall. Then

$$\mathsf{E}\, Z = \mathsf{E}\, \min(\tau, Y) + T = \int_0^\infty \mathsf{E}\, \min(t, Y)\, dG(t) + T$$

Let $Q_{i,j,k}(t)$ be the mean time before OS reinstall under condition that $X(t)$ starts from the state (i,j,k) and at the moment t the OS reinstall is scheduled. Note that $\mathsf{E}\, \min(t, Y) = Q_{0,0,0}(t)$. By virtue of the total expectation formula we get a system of integral convolution-like equations with respect to $Q_{i,j,k}(t)$:

$$\begin{aligned} Q_{0,0,0}(t) \;=\; & te^{-N(\lambda+\beta)t} + \int_0^t Ne^{-N(\lambda+\beta)x}\Big[\lambda(x + Q_{1,0,0}(t-x)) \\ & + \beta(x + Q_{0,1,0}(t-x))\Big]\, dx \end{aligned}$$

$$\begin{aligned} Q_{i,j,k}(t) = {} & te^{-\Lambda_{i,j,k}t} + \int_0^t x(p_2+p_3)i\mu e^{-\Lambda_{i,j,k}x}\, dx + \\ & + \int_0^t \Big[\Big((N-(i+j+k))\lambda + i\alpha\tfrac{N-(i+j+k)}{N-(1+j+k)}\Big)(x + Q_{i+1,j,k}(t-x)) \\ & + i\beta(x + Q_{i-1,j,k+1}(t-x)) + k\gamma p_0(x + Q_{i,j,k-1}(t-x)) \\ & + k\gamma(1-p_0)(x + Q_{i+1,j,k-1}(t-x)) + j\gamma(x + Q_{i,j-1,k}(t-x)) \end{aligned}$$

$$+ (N-(i+j+k))\beta(x+Q_{i,j+1,k}(t-x))\Big] e^{-\Lambda_{i,j,k}x}\,dx,$$

where

$$\Lambda_{i,j,k} = (p_2+p_3)i\mu+(N-(i+j+k))\lambda+i\alpha\tfrac{N-(i+j+k)}{N-(1+j+k)}+(N-(j+k))\beta+(j+k)\gamma$$

Remark 1. $Q_{0,0,0}(t)$ does not depend on $G(t)$.

By $R(Z)$ denote the profit made by the network over one regeneration period. $R(Z)$ consists of incomes of single nodes minus the damage caused by viruses on the interval between regeneration till the OS reinstall, OS reinstall costs. Let $R_{i,j,k}(t)$ be a mean profit of the network from the initial moment till the start of the OS reinstall under condition that X starts from (i,j,k) and time till next scheduled OS reinstall equals t. Then

$$\mathsf{E}\,R(Z) = \int_0^\infty R_{0,0,0}(t)\,dG(t) - c_2T - c_3.$$

We get a system of integral convolution-like equations with respect to $R_{i,j,k}(t)$ that is similar to the system for $Q_{i,j,k}(t)$.

Remark 2. $R_{0,0,0}(t)$ does not depend on $G(t)$.

4. Optimal distribution of the intervals between OS reinstalls

Let $S(t)$ be the average profit from the network functioning on the interval $(0;t)$ and $\rho = \lim_{t\to\infty}\frac{S(t)}{t}$. It is know from the regeneration theory [6] that $\rho = \mathsf{E}\,R(Z)/\mathsf{E}\,Z$. Hence

$$\rho = \frac{\int_0^\infty R_{0,0,0}\,(t)\,dG\,(t) - c_2T - c_3}{\int_0^\infty Q_{0,0,0}\,(t)\,dG\,(t) + T}$$

Remarks 1 and 2 imply that the functional ρ is a linear fractional with respect to the distribution $G\,(t)$.

Consider the following optimization problem $(\mathcal{Z}_1)$: $\rho \to \max_{\{G(\cdot)\}}$ where $\{G(\cdot)\}$ is the set of distribution functions such that $G(t) = 0$ as $t \le 0$. Using the theorem about the maximum of a linear fractional functional [6] and remarks 1 and 2, we get the following result:

Theorem. The optimal solution of $(\mathcal{Z}_1)$ (i.e. distribution of τ) is degenerate: $\hat{G}(t) = \begin{cases} 0, & t \le r \\ 1, & t > r \end{cases}$.

Corollary. The problem $(\mathcal{Z}_1)$ is equivalent to the problem $(\mathcal{Z}_2)$:

$$\frac{R_{0,0,0}(r) - c_2T - c_3}{Q_{0,0,0}(r) + T} \longrightarrow \max_{r\ge 0}.$$

For concrete networks the problem $(\mathcal{Z}_2)$ can be effectively resolved. Let $f^*(s) = \int_0^t e^{-st} f(t)\,dt$ denote the Laplace transform (LT) of $f = f(t)$. First we find the LT R^* and Q^* from systems of linear equations obtained by applying the LT to the integral equations for R and Q. Note that $R^*(s)$ and $Q^*(s)$ will be rational functions and there is no problem in concrete situations to invert the LTs and to find R and Q explicitly. Then maximum points of ρ can be found numerically.

5. Conclusions

We construct a mathematical model of virus protection of LAN on the assumption that infection of a running node can happen in two ways: during successful virus attacks from outside the network or by interacting with infected nodes within the network. We consider two methods to protect the network: updating antivirus signatures and OS reinstallings. OS are reinstalled in the case of failure of any of the computers (non-scheduled emergent reinstalling) or at scheduled time moments. The developed model belongs to the class of regenerative processes. We consider a maximization problem of an average unit profit. We prove that the optimal CDF of the scheduled intervals between complete OS reinstallings has to be degenerate.

References

1. *Kephart J., White S.* Directed-graph epidemiological models of computer viruses // Proceedings of the 1991 IEEE Computer Society Symposium on Research in Security and Privacy. — 1991. — P. 343–359.
2. *Mieghem P., Omic J. and Kooij R.* Virus spread in networks // IEEE/ACM Transactions on Networking. — 2009. — Vol. 17, no.1. — P. 1–14.
3. *Li C., Bovenkamp R., Mieghem P.* Susceptible-infected-susceptible model: A comparison of N-intertwined and heterogeneous mean-field approximations // Physical Review E. — 2012. — Vol. 86, no. 2 — Article ID 026116.
4. *Amador J., Artalejo J.* Stochastic modeling of computer virus spreading with warning signals // Journal of the Franklin Institute. — 2013. — V. 350, no. 5. — P. 1112–1138.
5. *Grishunina Yu., Manita L.* Stochastic Models of Virus Propagation in Computer Networks: Algorithms of Protection and Optimization // Lobachevskii Journal of Mathematics. — 2017. — Vol. 38, no. 5. — P. 906–909.
6. *Kashtanov V.A., Medvedev A.I.* Reliability theory of the composite systems. — M.: Fizmatlit, 2009.