Multiple Cyber-Threats Containment via Kendall's Birth-Death-Immigration Model

Vincenzo Matta, Mario Di Mauro, Maurizio Longo
University of Salerno, I-84084, Fisciano (SA), Italy
Email: {vmatta, mdimauro, longo}@unisa.it

Alfonso Farina
Selex-ES (retired), Visiting Professor UCL, Rome, Italy
Email: alfonso.farina@outlook.it

***Abstract*—This work examines the problem of modeling and containing multiple cyber-threats that propagate across multiple subnets of a data network. With regard to threat modeling, we propose to employ the Birth-Death-Immigration (BDI) model pioneered by Kendall in his seminal work of 1948 [1]. With regard to threat containment, assuming that a certain resource budget is available to mitigate the threats, we illustrate how the notable properties of the BDI model can be exploited to provide the optimal resource allocation across the attacked subnets.**

I. Introduction and Motivation

In our digital era, an ever-increasing number of activities rely on the exchange of information that takes place across a network of interconnected entities. While the network attributes enable appealing opportunities, when a network is in place there is always a theoretical possibility that a cyber-threat could rise. As a matter of fact, networks are natively exposed to the rapid propagation of malicious threats (e.g., a worm, a malware, or a virus), whose dynamics are often regulated by cascade mechanism resembling those governing the spread of epidemic diseases. A recent instance of this type is *WannaCry*, the malware performing a famous worldwide network attack launched in May 2017 [2]. The chameleonic nature of cyber-threats makes them capable to renew themselves, exploiting the new vulnerabilities of the cyber systems that arise as the result of technology advances. In order to contain the threats, the defender must put in place proper countermeasures, including: strategies for threat identification; analysis of the threat propagation algorithm; production of software aimed at defeating the threat (e.g., security patches to the operating systems, update of the anti-viruses). All these defense strategies are costly (e.g., in terms of time, working/computational power), and it is demanding to deliver these countermeasures in a timely manner, i.e., at the early stage of threat propagation. In addition, the defender is typically faced with multiple subnets that experience simultaneously different types of attacks, whose relative importance must be accurately judged in view of the overall containment task. This work focuses on two fundamental issues: i) providing an analytical model for threat propagation, aimed at characterizing the power of the threats associated to the different subnets; ii) establishing some optimal recipe to allocate the resources available to mitigate threats' propagation.

A. Related Work

This work belongs to the field of signal processing for cyber-security, which addresses topics as: identifying patterns of intrusions and data falsification [3], [4]; detecting anomalous activities in the network traffic [5], [6]; discovering the route of clandestine information flows across the network [7]; identifying a maliciously camouflaged source under an adversarial perspective [8]; locating the sources of Distributed-Denial-of-Service attacks [9]–[11]; containing the spread of cyber-threats over networks, which is the focus of this work.

Significant commonalities emerge between the mechanism of threat propagation and other propagative phenomena, such as epidemics, or population growth. For this reason, several mathematical models borrowed from other disciplines (e.g., natural and social sciences, biology, medical science [12]–[14]) have been applied also in the cyber-security domain — see [15] for a comprehensive account. Several studies have shown, by theoretical considerations as well as by experimental verification, that epidemic models are able to capture many important aspects of the cyber-threat dynamics [16], [17]. The primal, *deterministic* epidemic models [12] have been subsequently generalized to *stochastic* models, in order to face the the high and unpredictable variability of several complex network applications. With regard to *stochastic* threat propagation, a relevant paradigm is furnished by *queueing theory* [18], where nodes "arrive" (i.e., they get sick) following a certain arrival process, while nodes "depart" (i.e., they are cured) according to a certain departure process [19]. Two main features emerge from the relevant technical literature regarding threat propagation over networks: *i)* due to the lack of knowledge about the network details, cyber-threat propagation must be often described in terms of some summary/average parameters, according a "thermodynamic" approach that enables a satisfying description at a *macroscopic* level; *ii)* the number of infected nodes grows exponentially in the early stage of infection before the curing process becomes effective [20], which motivates the successful use of stochastic *branching-type* processes, namely, the Galton-Watson process [21], and the Fibonacci model [22]. These models are *discrete-time* models, where the time for a host to get infected is modeled as a *deterministic* time. In this work we shall focus instead on a *continuous-time* model, with the additional flexibility of working with *random* times.

B. This work

One of the most elegant stochastic models that capture the aforementioned fundamental features is the Birth-Death-Immigration (BDI) proposed by Kendall in 1948 [1], whose application to the cyber-security domain has been perhaps overlooked so far. It is a *stochastic* model that encodes the main features into three parameters (birth, death, and immigration

ISBN 978-90-827970-1-5

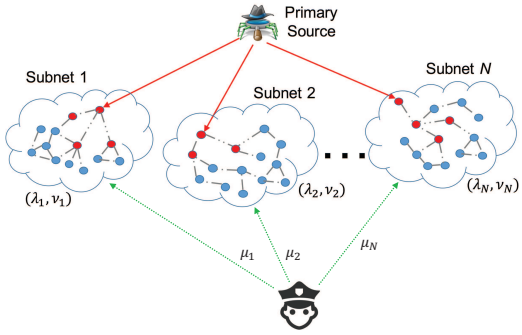

Fig. 1. Representation of the considered scenario. First, *multiple* threats are spreading across N subnets of a data network. Then, the agency employs an optimal allocation policy to distribute the available curing capacity.

rates) with a clear physical meaning. The considered setting is schematically illustrated in Fig. 1. The network is partitioned into N subnets, where each subnet is susceptible to a specific threat, which is disseminated across the individual subnet by one or more attackers. The attackers act as *primary* sources of infection: they explore *continually*, as time elapses, new portions of the network, looking for new vulnerable nodes. When a *vulnerable* node is found: i) it gets "sick" after a certain (random) time; and ii) it becomes a *secondary* source of infection. According to the terminology used in the context of the BDI model, the aforementioned mechanism is characterized by two main quantities (subscript ℓ refers to the ℓ-th subnet): the "immigration" rates $\{\nu_\ell\}$, which represent the number of hosts per unit time infected by the primary source of attack (*external* infection rate); and the "birth" rates $\{\lambda_\ell\}$, which represent the rate of hosts infected by another infected host (*internal* infection rate). The defender must put in place proper countermeasures to contain the multiple threats. When countermeasures are in use, nodes are "cured" at certain (random) times, and the "death" rates $\{\mu_\ell\}$ represent the number of cured hosts per unit time. The unavoidable resource limitations mentioned in the previous section will be abstracted by saying that a total-rate constraint must be imposed, such that the overall curing capacity must fulfill: $\sum_{\ell=1}^{N} \mu_\ell \leq C$. We remark that different types of resources (e.g., human resources, software installation, computing power) might be needed to tackle different threats, and, hence, handling the mapping from resources to curing rates might be a nontrivial task.

II. BDI Formal Model

Let $I(t)$ be the number of infected nodes at time t, let $p(n;t) \triangleq \mathbb{P}[I(t) = n]$, and let further

$$\Psi(x;t) \triangleq \mathbb{E}[e^{xI(t)}] \tag{1}$$

be the Moment Generating Function (MGF) of the number of infected nodes at time t. For the BDI model, it is possible to find a closed-form solution for the MGF (and, then, for the corresponding probability distribution) [14], [23], by describing the time evolution of the MGF through a first-order linear partial differential equation, namely through Eq. (6) further ahead. We start by outlining shortly how such time evolution can be obtained. Following the classic queueing theory paradigm, the infection and the curing processes will be modeled as *memoryless*, which amounts to say that the arrival process is a Poisson process and that the service times are exponentially distributed [15], [19]. Independence across distinct nodes is assumed. Let us consider a *vanishing* time interval of size ϵ. When there are $n-1$ already infected nodes in the system, the infection (i.e., arrival) process is a Poisson process of global rate $\lambda(n) \triangleq (n-1)\lambda + \nu$, which aggregates the internal (i.e., $(n-1)\lambda$) as well as the external (i.e., ν) infection rate components. Using the known properties of Poisson processes, the probability of reaching state n is approximately given by $\lambda(n)\epsilon$. Likewise, when there are $n+1$ infected nodes in the system, and using the known properties of the exponential distribution, the probability of reaching state n is approximately given by $\mu(n)\epsilon$, with $\mu(n) \triangleq (n+1)\mu$. For the same reasons, the probability of reaching state n from state $n \pm k$, with $k > 1$, is an infinitesimal of higher order, and is neglected, finally yielding:

$$\begin{aligned} p(n;t+\epsilon) &= \lambda(n)\epsilon\, p(n-1;t) + \mu(n)\epsilon\, p(n+1;t) \\ &+ [1 - \lambda(n)\epsilon - \mu(n)\epsilon]\, p(n;t). \end{aligned} \tag{2}$$

Dividing by ϵ, and taking the limit as $\epsilon \to 0$, we get the system of classic birth-and-death *master equations* [1]:

$$\begin{aligned} \frac{dp(n;t)}{dt} &= \lambda(n)\, p(n-1;t) - [\lambda(n) + \mu(n)]\, p(n;t) \\ &+ \mu(n)\, p(n+1;t). \end{aligned} \tag{3}$$

If we multiply both sides of the above equation by e^{nx}, and sum over n, after simple algebraic manipulations we obtain:

$$\begin{aligned} &\sum_{n=0}^{\infty} \frac{dp(n;t)}{dt} e^{nx} = \frac{d}{dt} \sum_{n=0}^{\infty} p(n;t) e^{nx} = \frac{\partial \Psi(x;t)}{\partial t} \\ &= \lambda e^{x} \frac{\partial \Psi(x;t)}{\partial x} + \nu e^{x} \Psi(x;t) - (\lambda + \mu) \frac{\partial \Psi(x;t)}{\partial x} \\ &\qquad - \nu \Psi(x;t) + \mu e^{-x} \frac{\partial \Psi(x;t)}{\partial x}, \end{aligned} \tag{4}$$

which, grouping the terms, corresponds to the first-order linear partial differential equation in (6). Solving this equation provides us the MGF, from which it is then possible to retrieve the probability distribution of the number of infected nodes. The next property summarizes this important known result — see [14], [23] for a detailed proof.

Property 1 (Statistical characterization of $I(t)$):

1) Let:

$$a(x) \triangleq [\lambda(1 - e^{x}) + \mu(1 - e^{-x})], \quad b(x) \triangleq \nu(e^{x} - 1). \tag{5}$$

Then, the moment generating function of $I(t)$ obeys the following first-order linear partial differential equation:

$$\frac{\partial \Psi}{\partial t} + a(x) \frac{\partial \Psi}{\partial x} = b(x)\, \Psi. \tag{6}$$

2) Let us introduce the following normalized quantities:

$$\Delta \triangleq \lambda - \mu, \qquad \rho \triangleq \lambda/\mu, \qquad \eta \triangleq \nu/\lambda, \tag{7}$$

and let also:

$$\pi_t \triangleq \frac{e^{\Delta t} - 1}{e^{\Delta t} - 1/\rho}, \quad q_t \triangleq \frac{e^{\Delta t} - \rho}{e^{\Delta t} - 1}. \tag{8}$$

ISBN 978-90-827970-1-5

Then, the moment generating function that solves (6) is defined in the range $x < \ln(1/\pi_t)$, and is equal to:

$$\Psi(x;t) = \left(\frac{1-\pi_t}{1-\pi_t\, e^x}\right)^{\eta+n_0} \left(\frac{1-q_t\, e^x}{1-q_t}\right)^{n_0}, \qquad (9)$$

where n_0 is the initial number of infected nodes.[1]

□

A. Asymptotic (large t) regimes

The detailed statistical characterization offered in Property 1 enables a powerful study of the asymptotic behavior of the BDI process, which will be useful in our setting. Some known asymptotic results will be collected in Properties 2 and 3 further ahead. We introduce preliminarily some notation. A negative binomial random variable of parameters $r > 0$ and $0 < p < 1$ is denoted by $\mathscr{N}_b(r,p)$, with probability mass function [24]:

$$p_{\text{nb}}(n) = \binom{n+r-1}{n}(1-p)^r p^n, \quad n = 0,1,\ldots \qquad (10)$$

A unit-scale gamma random variable with shape parameter $r > 0$ is denoted by $\mathscr{G}(r)$, with probability density function [24]:

$$f_{\text{G}}(z) = \frac{1}{\Gamma(r)} z^{r-1}\, e^{-z}, \quad z > 0, \qquad (11)$$

where $\Gamma(\cdot)$ is the (complete) gamma function. Finally, let $\mathscr{Y}(r,s,m)$, with $r > 0$, $s > 1$, and $m \in \{0,1,\ldots\}$, be a random variable with moment generating function:

$$\Psi_{\text{Y}}(x) = \left(\frac{1}{1-xs/(s-1)}\right)^{r+m} \left(1-\frac{x}{s-1}\right)^m, \qquad (12)$$

defined for $x < 1 - 1/s$.

We have the following result.

Property 2 (Asymptotic regimes of operation):

$$\begin{aligned} I(t) &\xrightarrow[t\to\infty]{\text{d}} \mathscr{N}_b(\eta,\rho), && \text{if } \rho < 1, \\ \frac{I(t)}{\lambda t} &\xrightarrow[t\to\infty]{\text{d}} \mathscr{G}(\eta), && \text{if } \rho = 1, \\ I(t)\, e^{-\Delta t} &\xrightarrow[t\to\infty]{\text{d}} \mathscr{Y}(\eta,\rho,n_0), && \text{if } \rho > 1 \end{aligned} \qquad (13)$$

with $\xrightarrow[t\to\infty]{\text{d}}$ denoting convergence in distribution as $t \to \infty$ [25].

□

We see from Property 2 that three possible regimes exist. The first one ($\rho < 1$) is a *stable, subcritical* regime, where the probability distribution of the number of infected nodes $I(t)$ approaches, as time elapses, a negative binomial distribution of parameters η and ρ. The second one ($\rho = 1$) is an *unstable, critical* regime, where the probability distribution of $I(t)$ scaled by λt, converges to a gamma distribution with unit scale parameter and with shape parameter equal to η. Loosely speaking, under this regime the number of infected nodes increases *linearly* with time. The third regime ($\rho > 1$) is a *strongly unstable, supercritical* regime, where the probability distribution of $I(t)$ scaled by $e^{\Delta t}$, converges to the distribution of the random variable $\mathscr{Y}(\eta,\rho,n_0)$. The number of infected nodes here increases *exponentially* with time.

[1]For $\rho = 1$, Eq. (9) holds true if $\pi_t = \frac{\lambda t}{\lambda t+1}$ and $q_t = \frac{\lambda t-1}{\lambda t}$, which is obtained from (8) by setting $\rho = \lambda/(\lambda-\Delta)$ and taking the limit as $\Delta \to 0$.

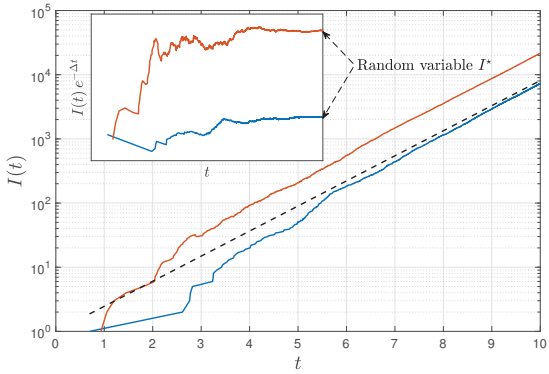

Fig. 2. Time evolution of the number of infected nodes under the strongly unstable regime. Two realizations (blue and red curves) of the random process $I(t)$ are displayed, along with the function $e^{\Delta t}$ (black dashed curve). In the inset, the normalized process $I(t)e^{-\Delta t}$ is depicted. The two points marked by the arrows correspond (approximately, since the time horizon is finite) to two distinct realizations of the limiting random variable, $I^\star$.

B. The Strongly Unstable Regime ($\rho > 1$)

The strongly unstable regime is of interest whenever the curing rate is smaller than the infection rate, a situation frequently met at the early stages of the infection propagation. Under this regime, the pertinent ($\rho > 1$) result in Property 2 can be refined by establishing that the number of infected nodes diverges *almost surely* at an exponential rate equal to Δ.

Property 3 (Almost sure exponential divergence of $I(t)$ under the strongly unstable regime): The scaled process $I(t)e^{-\Delta t}$ converges almost surely to a limiting random variable $I^\star$ distributed as $\mathscr{Y}(\eta,\rho,n_0)$:

$$I(t)e^{-\Delta t} \xrightarrow[t\to\infty]{\text{a.s.}} I^\star \sim \mathscr{Y}(\eta,\rho,n_0), \quad (\rho > 1). \qquad (14)$$

□

The result in Property 3 deserves special attention in our setting, for the following reasons. Since almost sure convergence takes place *over the sample paths*, Property 3 reveals that, even if the process $I(t)$ is random, (almost) all realizations of $I(t)$ will share the same behavior:

$$I(t) \approx I^\star\, e^{\Delta t} \qquad \text{for large } t, \qquad (15)$$

i.e., they eventually increase exponentially fast with rate Δ. In Fig. 2 we illustrate such behavior by depicting two realizations of the process (case $\rho > 1$) along with the theoretical exponential curve $e^{\Delta t}$ (dashed line). We see that all these realizations stay nearly parallel to the theoretical exponential curve, which matches perfectly (14) and (15). The different heights of the curves correspond to the different realization of the *random* multiplying constant, $I^\star$. The latter behavior is magnified in the inset of Fig. 2, where $I(t)e^{-\Delta t}$ is depicted.

III. Optimal Resource Allocation for Threat Mitigation

We are now in the position of tackling the problem of resource allocation for threats containment. We will carry out the analysis by assuming that the *infection vectors* $\boldsymbol{\lambda}$ and $\boldsymbol{\nu}$ are perfectly known. The case of unknown parameters can be dealt with by designing proper estimators that leverage the properties of the BDI model, as detailed in [26].

ISBN 978-90-827970-1-5 © EURASIP 2018

The curing rate that will be allocated to the ℓ-th subnet will be denoted by μ_ℓ, and the global curing rate that can be delivered, $\sum_{\ell=1}^N \mu_\ell$, is limited to a maximum curing capacity, C. Ideally, one would like to solve the optimization problem:

$$\min_{\boldsymbol{\mu}} \sum_{\ell=1}^N I_\ell(t) \quad \text{s.t.} \ \sum_{\ell=1}^N \mu_\ell \leq C. \tag{16}$$

However, the functions $I_\ell(t)$ are *random* processes, and, hence, we must choose a proper cost function that is amenable to optimization. We must distinguish two regimes for the optimization, determined by the available system capacity C.

The first regime corresponds to the case that $\lambda_{\text{tot}} \triangleq \sum_{\ell=1}^N \lambda_\ell > C$. Clearly, under this regime, at least for one ℓ we should have $\lambda_\ell > \mu_\ell$. Let us accordingly define:

$$\Delta_{\max} \triangleq \max_{\ell=1,2,\ldots,N} (\lambda_\ell - \mu_\ell) > 0. \tag{17}$$

As it can be seen from the next proposition, the quantity $\Delta_{\max}$ represents the *exponential rate of propagation* associated to the *overall* number of infected nodes, *in the presence of countermeasures*, and when the available curing *capacity is smaller than the internal infection rate*.

Proposition 1 (Exponential divergence of the overall number of infected nodes): If $\lambda_\ell > \mu_\ell$ for all $\ell = 1, 2, \ldots, N$, then, almost surely, the overall number of infected nodes diverges exponentially with exponent equal to $\Delta_{\max}$:

$$\sum_{\ell=1}^N I_\ell(t) e^{-\Delta_{\max} t} \xrightarrow[t\to\infty]{\text{a.s.}} I^\star_{\text{sum}}. \tag{18}$$

The random variable $I^\star_{\text{sum}}$ has the following MGF:

$$\prod_{\ell \in \mathcal{S}} \left(\frac{1}{1 - x\rho_\ell/(\rho_\ell - 1)} \right)^{\eta_\ell + n_{0,\ell}} \left(1 - \frac{x}{\rho_\ell - 1} \right)^{n_{0,\ell}} \tag{19}$$

for $x < 1 - (\min_{\ell \in \mathcal{S}} \rho_\ell)^{-1}$, where

$$\mathcal{S} \triangleq \{\ell \in [1, N] : \lambda_\ell - \mu_\ell = \Delta_{\max}\}, \tag{20}$$

and where η_ℓ, ρ_ℓ and $n_{0,\ell}$ are the parameters pertaining to the ℓ-th subnet. In addition, if $\lambda_\ell \leq \mu_\ell$ for some ℓ, the convergence in (18) holds in probability.

Proof: In view of (20), $\lambda_\ell - \mu_\ell = \Delta_{\max}$ for all $\ell \in \mathcal{S}$. Thus, straight application of Property 3 yields:

$$I_\ell(t) e^{-\Delta_{\max} t} \xrightarrow[t\to\infty]{\text{a.s.}} I^\star_\ell \sim \mathscr{Y}(\eta_\ell, \rho_\ell, n_{0,\ell}), \quad \forall \ell \in \mathcal{S}. \tag{21}$$

Let us now show that $I_\ell(t) e^{-\Delta_{\max} t}$ converges to zero in probability when $\ell \notin \mathcal{S}$, namely that:

$$I_\ell(t) e^{-\Delta_{\max} t} \xrightarrow[t\to\infty]{\text{p}} 0, \qquad \forall \ell \notin \mathcal{S}. \tag{22}$$

To this aim, we will make repeated use of the following result regarding stochastic convergence [25]:

$$Y(t) \xrightarrow[t\to\infty]{\text{d}} Y, \ \text{ and } \ f(t) \xrightarrow[t\to\infty]{} 0, \quad \Longrightarrow f(t)Y(t) \xrightarrow[t\to\infty]{\text{p}} 0. \tag{23}$$

For the case $\rho_\ell < 1$, we know that $I_\ell(t)$ converges in distribution in view of Property 2. Setting $Y(t) = I_\ell(t)$ and $f(t) = e^{-\Delta_{\max} t}$ into (23), we conclude that (22) holds true because $e^{-\Delta_{\max} t}$ vanishes as t goes to infinity. If $\rho_\ell = 1$, then $(\lambda_\ell t)^{-1} I_\ell(t)$ converges in distribution in view of Property 2. Setting $Y(t) = (\lambda_\ell t)^{-1} I_\ell(t)$ and $f(t) = (\lambda_\ell t) e^{-\Delta_{\max} t}$ into (23), we conclude that (22) holds true because $(\lambda_\ell t) e^{-\Delta_{\max} t}$ vanishes. Finally, if $\rho_\ell > 1$, then $I_\ell(t) e^{-\Delta_\ell t}$ converges in distribution in view of Property 2. Setting $Y(t) = I_\ell(t) e^{-\Delta_\ell t}$ and $f(t) = e^{(\Delta_\ell - \Delta_{\max}) t}$ into (23), we conclude that (22) holds true because $e^{(\Delta_\ell - \Delta_{\max}) t}$ vanishes for all $\ell \notin \mathcal{S}$.

Combining (21) and (22), we have in fact proved that the convergence in (18) holds in probability, provided that we set $I^\star_{\text{sum}} = \sum_{\ell \in \mathcal{S}} I^\star_\ell$. It is straightforward to show that such convergence can be strengthened to a.s. convergence when $\lambda_\ell > \mu_\ell$ for all $\ell = 1, 2, \ldots, N$.

Claim (19) now follows because: the processes corresponding to different subnets are statistically independent; the MGF of the sum of independent variables is given by the product of the MGFs of the variables; in view of (21) the MGF of $I^\star_\ell$ is given by (12), with the choices: $r = \eta_\ell$, $s = \rho_\ell$, and $m = n_{0,\ell}$. ■

The main consequence of Proposition 1 is that, in the case $\lambda_{\text{tot}} > C$, the asymptotic behavior of the number of infected nodes across the N subnets is determined by the largest exponent, namely $\sum_{\ell=1}^N I_\ell(t) \approx I^\star_{\text{sum}} e^{\Delta_{\max} t}$. It is thus meaningful to focus on minimizing the exponent, which amounts to reformulate (16) as:

$$\min_{\boldsymbol{\mu}} \max_{\ell \in [1,N]} (\lambda_\ell - \mu_\ell) \quad \text{s.t.} \ \sum_{\ell=1}^N \mu_\ell \leq C, \tag{24}$$

It is easy to show that the sought minimizer can be obtained through the following reverse water-filling solution [27]:

$$\mu^\star_\ell = \max(0, \lambda_\ell - \gamma) \qquad (\ell = 1, 2, \ldots, N) \tag{25}$$

with γ being chosen so as to meet the constraint $\sum_{\ell=1} \mu_\ell = C$.

We now switch to examine the most favorable case where $\lambda_{\text{tot}} < C$. Now it is clearly possible to meet the desirable requirement $\lambda_\ell < \mu_\ell$ *for all* $\ell = 1, 2, \ldots, N$, which prevents from exponential divergence in *all* subnets. From a more technical perspective, we see from Property 2 that in this case we cannot rely on minimization at the exponent. An alternative and reasonable choice is to minimize the expected number of infected nodes. To this aim, we mention that the convergence expressed by the first equation in (13) is in fact implied by the stronger result that the MGF of $I(t)$ converges to the MGF of $\mathscr{N}_b(\eta, \rho)$. Since $i)$ convergence of the MGF implies convergence of moments [25], and $ii)$ the expectation of $\mathscr{N}_b(\eta, \rho)$ is equal to $\frac{\eta\rho}{1-\rho} = \frac{\nu}{\mu - \lambda}$ [24], we can write:

$$\mathbb{E}\left[\sum_{\ell=1}^N I_\ell(t) \right] \xrightarrow[t\to\infty]{} \sum_{\ell=1}^N \frac{\nu_\ell}{\mu_\ell - \lambda_\ell}. \tag{26}$$

We shall accordingly focus on the optimization problem:

$$\min_{\boldsymbol{\mu}} \sum_{\ell=1}^N \frac{\nu_\ell}{\mu_\ell - \lambda_\ell} \quad \text{s.t.} \ \sum_{\ell=1}^N \mu_\ell \leq C \tag{27}$$

with $\mu_\ell > \lambda_\ell$ for all $\ell = 1, 2, \ldots, N$. Such a problem can be solved by the method of Lagrange multipliers. We introduce the Lagrangian $J(\boldsymbol{\mu}) = \sum_{\ell=1}^N, \frac{\nu_\ell}{\mu_\ell - \lambda_\ell} + \beta \sum_{\ell=1}^N \mu_\ell$, where β

ISBN 978-90-827970-1-5 © EURASIP 2018

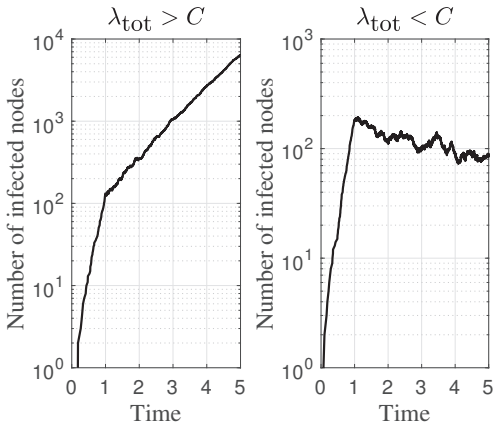

Fig. 3. Number of infected nodes spreading across $N = 3$ subnets. The vectors of internal and external infection rates are $\boldsymbol{\lambda} = [5, 4, 3.9]$ and $\boldsymbol{\nu} = [2.5, 4, 7.8]$, respectively. When $C = 0.8\,\lambda_{\text{tot}}$ (leftmost panel), the number of infected nodes grows exponentially and the countermeasures are not sufficient to contain such undesired behavior. When $C = 1.1\,\lambda_{\text{tot}}$ (rightmost panel), the optimization procedure is able to guarantee stability of the threat growth.

is the Lagrange multiplier. Taking the partial derivative with respect to the k-th component, we get:

$$\frac{\partial J}{\partial \mu_k} = -\frac{\nu_k}{(\mu_k - \lambda_k)^2} + \beta = 0 \Rightarrow \mu_k = \lambda_k + \sqrt{\nu_k/\beta}. \quad (28)$$

Imposing the constraint with equality yields the optimal $\mu_k^\star$:

$$\mu_k^\star = \lambda_k + \delta C \frac{\sqrt{\nu_k}}{\sum_{\ell=1}^{N} \sqrt{\nu_\ell}} \qquad (k = 1, 2, \ldots, N) \qquad (29)$$

where $\delta C = C - \lambda_{\text{tot}}$.

IV. Simulation Results

We now illustrate an application of the aforementioned optimization procedure. A more detailed analysis of the values assumed by the various parameters in a real setting, and a careful comparison with semi-realistic threat-propagation models, are presented in [26].

In Fig. 3, the overall number of infected nodes spreading across $N = 3$ subnets is displayed as a function of time for two cases, namely, $\lambda_{\text{tot}} > C$, and $\lambda_{\text{tot}} < C$. The vector of internal infection rates is $\boldsymbol{\lambda} = [5, 4, 3.9]$, whereas the vector of external infection rates is $\boldsymbol{\nu} = [2.5, 4, 7.8]$. The leftmost panel refers to the less favorable scenario where $C = 0.8\,\lambda_{\text{tot}}$, namely, the countermeasures cannot be sufficient to contain the exponential propagation of the threat. In particular, the number of infected nodes grows exponentially (a logarithmic scale is adopted) up to about one fifth of the time window; afterwards, a sharp slope variation occurs as a consequence of the applied countermeasures following the optimization procedure.

In contrast, the rightmost panel addresses the advantageous situation where $C = 1.1\,\lambda_{\text{tot}}$. In this case, the optimization is performed by exploiting (29), and the final result is that the threat containment is effective in preventing the undesired exponential behavior.

References

[1] D. G. Kendall, "On some modes of population growth leading to R. A. Fisher's logarithmic series distribution," *Biometrika*, vol. 35, no. 1/2, pp. 6–15, May 1948.

[2] "What you need to know about the WannaCry Ransomware." https://www.symantec.com/connect/blogs/what-you-need-know-about-wannacry-ransomware.

[3] S. Marano, V. Matta, and L. Tong, "Distributed detection in the presence of Byzantine attacks," *IEEE Trans. Signal Process.*, vol. 57, no. 1, pp. 16–29, Jan. 2009.

[4] B. Kailkhura, S. Brahma, B. Dulek, Y. S Han, and P. Varshney, "Distributed detection in tree networks: Byzantines and mitigation techniques," *IEEE Trans. Inf. Forensics Security*, vol. 10, no. 7, pp. 1499–1512, Jul. 2015.

[5] M. Mardani, G. Mateos, and G. B. Giannakis, "Dynamic anomalography: tracking network anomalies via sparsity and low rank," *IEEE J. Sel. Topics Signal Process.*, vol. 7, no. 1, pp. 50–66, Feb. 2013.

[6] S. Fortunati, F. Gini, M. Greco, A. Farina, A. Graziano, and S. Giompapa, "An improvement of the state-of-the-art Covariance-based Methods for Statistical Anomaly Detection Algorithms", *Signal, Image and Video Processing*, vol. 10, no. 4, pp. 687–694, Apr. 2016.

[7] S. Marano, V. Matta, T. He, and L. Tong, "The embedding capacity of information flows under renewal traffic," *IEEE Trans. Inf. Theory*, vol. 59, no. 3, pp. 1724–1739, Mar. 2013.

[8] M. Barni and B. Tondi, "The source identification game: an information theoretic perspective," *IEEE Trans. Inf. Forensics Security*, vol. 8, no. 3, pp. 450–463, Mar. 2013.

[9] V. Matta, M. Di Mauro, and M. Longo, "DDoS attacks with randomized traffic innovation: Botnet identification challenges and strategies," *IEEE Trans. Inf. Forensics Security*, vol. 12, no. 8, pp. 1844–1859, Aug. 2017.

[10] V. Matta, M. Di Mauro, and M. Longo, "Botnet identification in randomized DDoS attacks," in *Proc. EUSIPCO*, Budapest, Hungary, Aug./Sep. 2016, pp. 2260–2264.

[11] V. Matta, M. Di Mauro, and M. Longo, "Botnet identification in multi-clustered DDoS attacks," in *Proc. EUSIPCO*, Kos island, Greece, Aug./Sep. 2017, pp. 2235–2239.

[12] W. O. Kermack, A. G. McKendrick, "A contribution to the mathematical theory of epidemics" in *Proc. Roy. Soc. Lond. Series A*, vol. 115, no. 772, pp. 700–721, Aug. 1927.

[13] P. J. Costa, *Applied Mathematics for the Analysis of Biomedical Data*, John Wiley & Sons, New York, 2017.

[14] N. T. J. Bailey, *The Elements of Stochastic Processes with Applications to the Natural Sciences*, John Wiley & Sons, New York, 1964.

[15] V. Karyotis and M. H. R. Khouzani, *Malware Diffusion Models for Modern Complex Networks. Theory and Applications*, Morgan Kaufmann, 2016.

[16] C. C. Zou, W. Gong, D. Towsley, and L. Gao, "The monitoring and early detection of Internet worms," *IEEE/ACM Trans. Netw.*, vol. 13, no. 5, pp. 961–974, Oct. 2005.

[17] M. H. R. Khouzani, S. Sarkar, and E. Altman, "Maximum damage malware attack in mobile wireless networks," *IEEE/ACM Trans. Networking*, vol. 20, no. 5, pp. 1347–1360, Oct. 2012.

[18] L. Kleinrock, *Queueing Systems. Volume I: Theory*, John Wiley & Sons, New York, 1975.

[19] V. Karyotis and S. Papavassiliou, "Macroscopic malware propagation dynamics for complex networks with churn," *IEEE Commun. Lett.*, vol. 19, no. 4, pp. 577–580, Apr. 2015.

[20] S. Staniford, V. Paxson, and N. Weaver, "How to 0wn the Internet in your spare time," in *Proc. USENIX Security Symposium*, Berkeley, CA, USA, Aug. 2002, pp. 149–167.

[21] S. H. Sellke, N. B. Shroff and S. Bagchi, "Modeling and automated containment of worms," *IEEE Trans. Depend. Secure Comput.*, vol. 5, no. 2, pp. 71–86, Apr.-Jun. 2008.

[22] A. Farina, C. Fantacci, and M. Frasca, "Stochastic filtering of a random Fibonacci sequence: Theory and applications," *Signal Processing*, vol. 104, pp. 212–224, Apr. 2014.

[23] R. N. Nucho, "Transient behavior of the Kendall birth-death process — Applications to capacity expansion for special services," *The Bell System Technical Journal*, vol. 60, no. 1, pp. 57–87, Jan. 1981.

[24] N. L. Johnson, S. Kotz, and N. Balakrishnan, *Continuous Univariate Distributions, Volume 1*, John Wiley & Sons, New York, 1994.

[25] A. Gut, *Probability: A Graduate Course*, 2nd ed. Springer, New York, 2013.

[26] V. Matta, M. Di Mauro, M. Longo, and A. Farina "Cyber-Threat Mitigation Exploiting the Birth-Death-Immigration Model," *IEEE Trans. Inf. Forensics Security*, doi: 10.1109/TIFS.2018.2838084.

[27] T. Cover and J. Thomas, *Elements of Information Theory*, 2nd ed. John Wiley & Sons, New York, 2006.

ISBN 978-90-827970-1-5