Dilemas Frente Epidemias Estratégicas: Vacinar ou Reiniciar?

Vilc Rufino[1,2], **Daniel Menasché**[2], **Italo Cunha**[3], **Cabral Lima**[2], **Leandro P. de Aguiar**[4]

[1]Marinha do Brasil, [2] DCC/PPGI/UFRJ, [3] DCC/UFMG, [4] Siemens Corporate Research

Abstract. *Network epidemics are ubiquitous. As botnets evolve, they compromise additional users to join DDoS attack campaigns. Such users face a dilemma with respect to which countermeasures to take: hard (e.g., vaccination), soft (e.g., rebooting and rejuvenation) or no countermeasures at all. To tackle this dilemma, one option is to leverage insights from analytical models. Our key contribution consists of novel results on the steady state solution of epidemic models wherein the attacker is strategic and has a finite attack budget. To this aim, the most probable states of the model are analyzed, and are used to derive closed form expressions that approximate the steady state probability of infection of a node. Then, model's insights are contrasted against simulations. The simulations qualitatively support the observations of the model and extend the analysis allowing general distributions to the times between the events.*

Resumo. *Epidemias de rede são ubíquas. Todos os dias, sistemas são comprometidos por* botnets *e participam de campanhas de ataque DDoS. Assim, os usuários enfrentam um dilema com relação a quais contramedidas tomar: duras (por exemplo, vacinação), suaves (por exemplo, reinicialização e rejuvenescimento) ou nenhuma contramedida. Para resolver esse dilema, uma opção é tomar proveito de modelos analíticos. Neste artigo, apresentamos novos resultados sobre a solução em estado estacionário de modelos epidêmicos em que o atacante é estratégico e tem uma capacidade de ataque finita. Para tanto, analisamos os estados mais prováveis do modelo, indicando suas propriedades; apresentamos fórmulas que aproximam a probabilidade de infecção e contrastamos os* insights *do modelo com simulações. As simulações suportam qualitativamente as observações do modelo e estendem a análise permitindo distribuições gerais para os tempos entre os eventos.*

1. Introdução

Motivação Epidemias em rede de computadores são ubíquas. Todos os dias sistemas são comprometidos por códigos maliciosos que executam ações sem consentimento do seu dono legítimo (*botnets*), os quais participam de ataques de negação de serviço (*DDoS*). Tais ataques originam-se de centrais de comando e controle e fazem uso de dispositivos que são comprometidos a partir de infecções endógenas (i.e., entre vizinhos na rede local) ou exógenas (i.e., a partir de um dispositivo de uma rede remota).

Desafios Dentre os desafios enfrentados pelos administradores de sistemas, destacamos o dilema entre vacinar seus dispositivos (e.g., aplicando *patches*) ou esperar e reiniciar certos processos, ou o sistema como um todo, de tempos em tempos (e.g., para fazer o rejuvenencimento do mesmo). Embora a vacinação seja mais efetiva, ela pode envolver efeitos colaterais que indisponibilizem o sistema por um longo tempo e isso pode ser inviável, e.g., sistemas de controle industrial (ICS). Para lidar com o *tradeoff*

entre aplicar contramedidas mais fortes ou suaves e os possíveis custos associados a uma invasão, é fundamental ter um melhor entendimento sobre a probabilidade de infecção dos nós da rede. Entretanto, ainda existem muitas questões em aberto no que concerne a caracterização da probabilidade de infecção frente a atacantes estratégicos.

Objetivo e metodologia Neste artigo, nosso objetivo é caracterizar a probabilidade de infecção de nós de uma rede, em função das taxas de infecção endógena e exógena. Para tal, focamos em modelos analíticos, e buscamos soluções aproximadas, o método de Newton fornece uma equação iterativa que desenrolada duas vezes obtém-se uma boa aproximação. Em seguida, usamos simulações para entender o que ocorre quando relaxamos as premissas dos modelo.

Lacunas no estado da arte Existe uma ampla literatura sobre epidemias em redes de computadores, cuja base matemática remonta às epidemias biológicas. Embora as epidemias em redes de computadores e biológicas tenham semelhanças entre si, elas também possuem importantes distinções. Dentre as distinções destacamos o fato de que o atacante da rede de computadores é estratégico, e pode varrer a rede completa na busca por nós vulneráveis. Modelos matemáticos levando em conta este tipo de comportamento são escassos, e não é de nosso conhecimento pesquisa anterior que tenha derivado fórmulas fechadas para a probabilidade de infecção de nós neste cenário.

Contribuições: (i) **Análise do modelo analítico de epidemias** estendemos a análise do modelo analítico de epidemias proposto em [Rufino et al. 2018], indicando formas práticas de parametrizá-lo e analisando seus estados mais prováveis; (ii) **Fórmulas para probabilidade de infecção** obtemos, via método de Newton, fórmula iterativa que desenrolada duas vezes aproxima a probabilidade de contaminação. As fórmulas são simples e dependem apenas dos parâmetros do sistema; (iii) **Simulação** executamos simulações e verificamos que o comportamento capturado pelo modelo analítico é também observado no sistema simulado. Em particular, as simulações levam em conta nós que entram e saem da rede assim como tempos entre eventos gerais (e.g., determinísticos), enquanto que o modelo analítico assume que todos os tempos entre eventos são exponencialmente distribuídos.

Organização o restante deste artigo está organizado da seguinte forma. A Seção 2 descreve o sistema em questão, seguido pelo modelo na Seção 3. Apresentamos fórmulas que aproximam o modelo na Seção 4 e simulações na Seção 5. Finalmente, trabalhos relacionados e conclusão vêm nas Seções 6 e 7.

2. Descrição do sistema

Nesta seção, descrevemos o comportamento de um código malicioso real (Mirai). Destacamos alguns pontos do sistema que serão analisados e observados no restante do artigo.

Ciclo de operação do código malicioso *Mirai Botnet* O Mirai é um código malicioso que ficou conhecido por executar o maior ataque DDoS conhecido até 2016 [Krebs 2016]. As principais etapas de um ataque do *Mirai Botnet* são descritas a seguir: (i) **Varredura**: Os dispositivos contaminados buscam por vítimas vulneráveis na rede local. *Essa é a chamada contaminação endógena.* Além disso, alguns dispositivos também enviam assincronamente pacotes TCP SYN para endereços IPv4 pseudoaleatórios. *Essa é a chamada contaminação exógena (em geral, entre dispositivos de redes distintas).* Caso

encontre uma vítima, passa-se para uma fase de tentativa de autenticação por força bruta; (ii) **Relatório**: Após o primeiro sucesso de autenticação, o *bot* envia as credenciais da vítima para um Servidor de Relatórios, sob controle do atacante; (iii) **Despacho**: por meio de um processo separado, o Servidor de Carregamento, usando as informações colhidas ou diretamente fornecidas pelo *botMaster*, se autentica nos dispositivos vítimas e carrega o programa do mirai, de acordo com a arquitetura identificada. A vítima passa a ser um novo bot sob controle do atacante (*botMaster*); (iv) **Comandos**: O atacante, por meio de um servidor com uma interface de comando e controle, envia comandos a serem executados pelos bots (dispositivos que executam o código malicioso); (v) **Retransmissão**: O servidor de comando e controle retransmite os comandos para os dispositivos controlados (os bots) que foram selecionados e estejam ativos; (vi) **Execução**: Com os comandos recebidos são executados pelos *bots* conforme as instruções do *botMaster*.

Os nossos modelos analíticos e de simulação focam nas contaminações dos dispositivos, ou seja, na fase de varredura (envolvendo infecções endógenas e exógenas). Em particular, assumimos que a varredura de endereços IPv4 pseudo-aleatórios consome recursos de banda, e que por isso a taxa de contaminações exógenas é limitada pelo poder do atacante *em função do modelo de adversário.*

Modelo de adversário O adversário é o usuário que tem controle sobre o *malware*. O adversário tem capacidade de reconhecer, após análise, se determinado sistema é vulnerável, mas não é capaz de distinguir, de antemão, se determinado sistema está, ou não, contaminado. Consideramos que o adversário tem uma capacidade média de contaminação exógena de Λ contaminações por unidade de tempo. Essa capacidade está limitada pela taxa agregada de varredura e análise de IPs de todos os nós que compõem a *botnet*. Seja N o número de nós vulneráveis na rede (ou seja, N é o número de nós que decidem não se vacinar). *No caso mais simples, supomos que a capacidade de contaminação exógena do adversário é dividida pelos nós vulneráveis presentes na rede, e que cada um é alvo de uma varredura exógena a uma taxa de Λ/N tentativas de contaminação por unidade de tempo.*

Contramedidas As contramedidas visam evitar ou minimizar os danos provocados por um ataque. A aplicação de uma contramedida está associada a um custo, o que leva o decisor a uma escolha entre as seguintes ações. (i) **Vacinar:** Consideramos a vacinação como a aplicação de correção ao software ou dispositivo que o torne imune aos ataques cibernéticos. Esta ação está associada a custos elevados, e normalmente escolhida quando o risco de contaminação é elevado e represente perdas muito grandes. (ii) **Esperar e eventualmente reiniciar (de forma proativa):** A reinicialização consiste em retornar o dispositivo ou software ao seu estado original, removendo possíveis códigos maliciosos que tenham sido instalados. Por sua simplicidade e efetividade, é uma opção de baixo custo, porém mantém o dispositivo ou software vulnerável à re-contaminação. (iii) **Esperar e reiniciar (de forma reativa):** A decisão de simplesmente não se fazer nada é adotada quando o risco de contaminação é baixo. Alguns decisores só tomam ações corretivas quando a contaminação efetivamente causa prejuízo na produção ou nos lucros. Claramente, ações tardias podem representar grandes perdas. *A seguir, apresentamos um modelo epidêmico para caracterizar a probabilidade de infecção de um nó. Iremos então ilustrar um possível uso do modelo para guiar o processo de tomada de tomada de decisões sobre contramedidas, levando em conta o estado da população.*

3. Modelo epidêmico

A seguir, analisamos o modelo epidemiológico SIS multiplicativo, conforme apresentado em [Rufino et al. 2018]. As Seções 3.1 e 3.2 descrevem o modelo. A Seção 3.3 apresenta resultados sobre o estado mais provável, seguida pela busca dos valores ótimos para parametrizar o modelo, relacionando tais valores com o estado mais provável.

variável	descrição	valor de referência
M	tamanho total da população	100
N	população passível de infecção	M
γ	fator de infecção endógena (por aresta)	1.09
λ	fator de infecção exógena (por nó)	Λ/N
Λ	fator de infecção exógena total	10
A	matriz de adjacência (conexões)	completa
d	número de nós vizinhos infectados	-
$\lambda\gamma^d$	taxa de infecção (por nó)	-
μ	taxa de recuperação	1
$\pi(\boldsymbol{x})$	probabilidade do estado $\boldsymbol{x}$	-
i	número de nós infectados na rede	-
ρ	probabilidade de um nó escolhido ao acaso estar contaminado	-

Tabela 1. Notação do modelo analítico e valores de referência.

3.1. Descrição do modelo

Consideramos uma população finita contendo M nós, dos quais, N decidiram não vacinar, e portanto possuem uma vulnerabilidade que pode ser explorada. Cada um desses N nós pode assumir os estados de suscetível (S ou 0) ou infectado (I ou 1).

Um nó infectado pode ser recuperado, passando do estado I para o estado S após um tempo exponencialmente distribuído com média $1/\mu$. Um nó suscetível pode ser infectado por um atacante externo (infecção exógena) ou por um ataque interno (infecção endógena) de um vizinho na rede. Seja d o número de vizinhos infectados. Seja γ a taxa de infecção endógena por vizinho e seja λ a taxa de infecção exógena por nó, $\lambda = \Lambda/N$. Assumimos que o tempo entre infecções é exponencialmente distribuído, com taxa $\gamma^d\lambda$. Ou seja, assumimos contribuições multiplicativas das taxas de infecções endógenas e exógenas.

Seja $\boldsymbol{x}$ um estado possível da rede, entre todos os estados possíveis $\mathcal{X}$. O estado é um vetor N dimensional, $\boldsymbol{x} \in \{0,1\}^N$, $\boldsymbol{x} = (x_1, x_2, \cdots, x_k, \cdots, x_{N-1}, x_N)$, onde $x_k \in \{0,1\}$. A dinâmica do sistema é caracterizada por um processo Markoviano contínuo, homogêneo temporal, irredutível e de estados finitos. Cada estado da rede corresponde a um estado no processo Markoviano. Além disso, o nosso processo Markoviano é reversível, conforme [Kelly 1979].

Consideramos uma topologia totalmente conectada, onde todos os nós estão conectados entre si. A probabilidade do estado $\boldsymbol{x}$ é dada por $\pi(\boldsymbol{x})$, derivada em [Zhang et al. 2017],

$$\pi(\boldsymbol{x}) \;=\; \frac{\tilde{\pi}(\boldsymbol{x})}{Z} \tag{1}$$

onde

$$\tilde{\pi}(\boldsymbol{x}) = \left(\frac{\lambda}{\mu}\right)^{1^T\boldsymbol{x}} \gamma^{\boldsymbol{x}^T A\boldsymbol{x}/2}\,, \boldsymbol{x} \in \mathcal{X}, \qquad Z = \sum_{\boldsymbol{x}\in\mathcal{X}} \tilde{\pi}(\boldsymbol{x}). \tag{2}$$

A Tabela 1 resume a notação. Tomando proveito da simetria do problema, e com certo abuso de notação, seja $\pi(\iota)$ a probabilidade de haver ι nós infectados:

$$\pi(\iota) = \frac{\tilde{\pi}(\iota)}{Z}, \qquad \tilde{\pi}(\iota) = \binom{N}{\iota} \left(\frac{\lambda(N)}{\mu}\right)^{\iota} \gamma^{\iota(\iota-1)/2}, \quad \iota = 0, \ldots, N. \tag{3}$$

O valor esperado do número de nós infectados é

$$E(I) = \sum_{\iota=0}^{N} \iota \frac{\tilde{\pi}(\iota)}{Z} = N\rho(N), \quad \rho(N) = \frac{1}{N} \sum_{\iota=0}^{N} \iota \frac{\tilde{\pi}(\iota)}{Z} \tag{4}$$

onde $\rho(N)$ é a probabilidade de infecção de um nó escolhido aleatoriamente.

3.2. Modelo binomial

A análise direta das equações acima é complexa, por envolver um termo quadrático no expoente de γ em (3). Para simplificar a análise, consideramos uma solução aproximada. Para tal, definimos $\hat{\rho}(N) \approx \rho(N)$ e $\hat{\pi}(\iota) \approx \tilde{\pi}(\iota)$:

$$\hat{\rho}(N) = \frac{1}{N} \sum_{\iota=0}^{N} \iota \frac{\hat{\pi}(\iota)}{\hat{Z}}, \quad \hat{Z} = \sum_{\iota=0}^{N} \hat{\pi}(\iota), \quad \hat{\pi}(\iota) = \binom{N}{\iota} \left(\frac{\lambda(N)}{\mu} \gamma^{N^\star}\right)^{\iota} \tag{5}$$

onde $N^\star(N)$ é uma função crescente de N, que denotamos simplesmente por $N^\star$ para simplificar a notação. Nos referimos ao modelo proposto para aproximar a solução do modelo original como *modelo binomial*, por fazermos uso do binômio de Newton na demonstração do resultado a seguir.

Lema 3.1. *No modelo binomial, temos que:*

$$\hat{\rho}(N) = \frac{1}{1 + \mu/(\lambda(N)\gamma^{N^\star})} \tag{6}$$

A demonstração do lema acima consta em [Rufino et al. 2018], notando o pequeno ajuste de notação na nova versão do resultado, refletindo a nova definição de $N^\star$. Como discutido na Seção 3.3 a seguir, $N^\star$ é adequadamente parametrizado como $N^\star = (N - 1)\hat{\rho}(N)$. Simplificando a notação, removendo as dependências com relação a N,

$$\hat{\rho} = \frac{1}{1 + \mu/(\lambda\gamma^{(N-1)\hat{\rho}})} \tag{7}$$

A equação acima dá origem a um problema de ponto fixo, analisado na Seção 4.1.

3.3. Buscando valor ótimo para $N^\star$

A seguir, buscamos o valor ótimo de $N^\star$ em função de N. Para tal, nos aproveitamos de um resultado recentemente derivado em [Zhang and Moura 2018] sobre os estados mais prováveis do modelo epidemiológico aqui discutido. Em [Zhang and Moura 2018], os autores discutem o cenário no qual a taxa de infecção exógena, λ, é constante, independente de N. Reproduzimos o principal resultado de [Zhang and Moura 2018], tendo em vista que ele nos traz *insights* sobre o valor ótimo de $N^\star$.

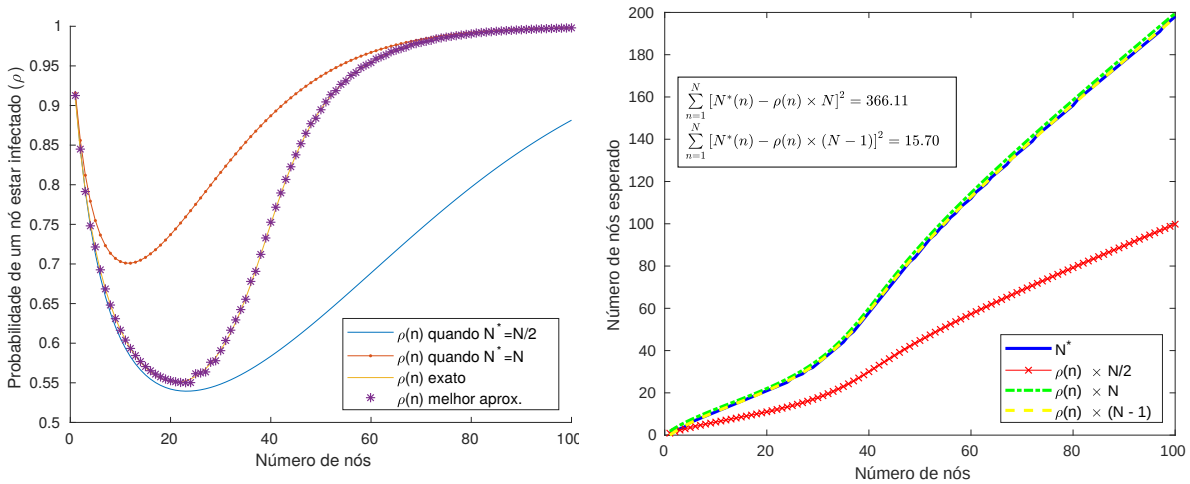
Seja $\boldsymbol{x}^\star$ a *configuração mais provável* do sistema, $\boldsymbol{x}^\star = [x_1^\star, x_2^\star, \ldots, x_N^\star]$, onde $\boldsymbol{x}^\star = \underset{\boldsymbol{x}\in\mathcal{X}}{\arg\max}\ \pi(\boldsymbol{x})$. Se $\tilde{\pi}(\boldsymbol{x}^\star) \gg \tilde{\pi}(\boldsymbol{x})\ , \forall \boldsymbol{x} \in \mathcal{X}/\boldsymbol{x}^\star$, então

$$P(x_k = 1) \approx \frac{1}{1 + \mu/\left(\lambda\gamma^{m_k^\star}\right)}, \qquad \text{onde } m_k^\star = \sum_{j=1}^{N} a_{kj} x_j^\star. \tag{8}$$

O resultado acima decorre do fato de que a probabilidade do estado $\boldsymbol{x}$ pode ser expressa como $\pi(\boldsymbol{x}) = e^{H(\boldsymbol{x})}$, onde $H(\boldsymbol{x}) = 1^T\mathbf{x}\ \log(\lambda/\mu) + \left(\mathbf{x}^T A\mathbf{x}\ \log\gamma\right)/2$. Notando então a relação entre $\pi(\boldsymbol{x})$ e a distribuição de Gibbs, o resultado segue.

Note que $m_k^\star$ representa o números de vizinhos de k contaminados no estado mais provável. Comparando (8) e (6), vemos que $m_k^\star$ está diretamente relacionado a $N^\star$. Seja n o número de vizinhos de um nó típico. Ao considerarmos um grafo completo, o número de vizinhos de cada nó é $n = N - 1$, e o número médio de vizinhos contaminados é ρn. Substituindo $N^\star$ em (6) por $m_k^\star$, obtemos então a expressão (8). Tal argumento sugere que o valor ótimo de $N^\star$ é dado por $N^\star = \hat{\rho}n$. A seguir, ilustramos numericamente tal fato.

A Figura 1 ilustra como a probabilidade de infecção em função do número de nós não vacinados na rede, N. Consideramos $\lambda = \Lambda/N$, $\Lambda = 10$, $\mu = 1$ e $\gamma = 1.09$. O valor de $\hat{\rho}$ obtido via aproximação, quando selecionamos o melhor valor de $N^\star$, é muito próximo do valor de ρ exato. A Figura 1(a) mostra também que quando fazemos $N^\star = N$ e $N^\star = N/2$ obtemos limites superiores e inferiores para a probabilidade de contaminação. Para avaliar a qualidade da aproximação $N^\star = \rho(N-1)$, a Figura 1(b) mostra o valor ótimo de $N^\star$, em função de N, comparado com $N/2$, ρN e $\rho(N-1)$. Tanto ρN quanto $\rho(N-1)$ apresentam excelentes aproximações. Calculando a soma dos erros quadráticos (em destaque na figura), podemos identificar que de fato $\rho(N-1)$ é uma melhor aproximação, corroborando os resultados derivados nessa seção.



(a) $\rho(n)$ exato e usando aproximações para $N^\star$ (b) $N^\star$ ótimo em função de N

Figura 1. Validação da solução aproximada pelo modelo binomial ($\gamma = 1,09$). Buscando uma melhor aproximação de $N^\star$

4. Fórmulas via método de Newton

A seguir indicamos como usar o método de Newton para achar fórmulas aproximadas para (6). Pelo fato de ρ aparecer tanto do lado direito quanto do lado esquerdo de (6), a equação não é passível de solução exata em fórmula fechada. Ao invés de buscar por soluções exatas, buscamos então por aproximações. Para tal, vamos considerar as seguintes funções auxiliares,

$$f(\rho) = \rho\left(1+\frac{\mu}{\lambda}\gamma^{-\rho n}\right) - 1 = \rho + \rho\frac{\mu}{\lambda}\gamma^{-\rho n} - 1 \quad (9)$$

$$\frac{\partial f(\rho)}{\partial \rho} = f'(\rho) = 1 + \frac{\mu}{\lambda}\gamma^{-\rho n}\left(1 - \rho n \ln\gamma\right) \quad (10)$$

$$\frac{\partial^2 f(\rho)}{\partial^2 \rho} = f''(\rho) = g\frac{\mu \ln\gamma}{\lambda}\gamma^{-\rho n}\left(\rho(\ln\gamma) - 2\right) \quad (11)$$

Encontrar a solução ρ para (7) é equivalente a encontrar as raízes (i.e., os zeros) de (9).

A iteração do método de Newton, adaptada ao nosso cenário, é dada por,

$$\rho_{i+1} = \rho_i - \frac{f(\rho_i)}{f'(\rho_i)} = \frac{\lambda - \mu\gamma^{-\rho_i n}\left(\rho_i^2 n \ln\gamma\right)}{\lambda - \mu\gamma^{-\rho_i n}\left(\rho_i n \ln\gamma - 1\right)} \quad (12)$$

Destacamos que $f(0) = -1$ e $f(1) = \frac{\mu}{\lambda\gamma} > 0$, onde $\mu, \lambda > 0$ e $\gamma > 1$. Além disso, $f(0) = -1$ e $f'(0) = \mu/(\lambda\gamma)$, assim como $f''(0) = \mu\ln\gamma/\lambda$. Desta forma, se $\gamma > 1$, então $f(0)f''(0) > 0$ (f e f'' têm o mesmo sinal). Pelo teorema de Darboux [Darboux 1869], iniciando com $\rho_0 = 0$, o método de Newton converge sem ultrapassar (*overshoot*) a solução.

4.1. Obtendo fórmulas aproximadas

Usando a abordagem descrita acima, obtemos fórmulas para uma aproximação da probabilidade de um nó estar infectado. Numericamente, identificamos que considerar duas iterações do método de Newton é suficiente para obter boas aproximações.

A condição inicial do método de Newton tem um papel importante no resultado. Consideramos então duas condições iniciais extremas. Seja ρ_0 a condição inicial do método. Considerando $\rho_0 = 0$ e $\rho_0 = 1$, obtemos duas aproximações para a probabilidade de contaminação. Na próxima seção, apresentamos uma heurística simples para determinar quando adotar uma condição inicial ou a outra. Na Seção 4.2, indicamos numericamente que as aproximações junto com a heurística capturam o comportamento da probabilidade de infecção.

Seja $\rho_i^{(0)}$ a probabilidade de infecção aproximada após i iterações do método de Newton, com condição inicial $\rho_0 = 0$. Então,

$$\rho_0^{(0)} = 0, \quad \rho_1^{(0)} = \frac{\lambda}{\lambda+\mu}$$

$$\rho_2^{(0)} = \frac{\lambda - \mu\gamma^{-\rho_1 n}\left(\rho_1^2 n\ln\gamma\right)}{\lambda - \mu\gamma^{-\rho_1 n}\left(\rho_1 n\ln\gamma - 1\right)} = \frac{\lambda - \mu\gamma^{-\left(\frac{\lambda}{\lambda+\mu}\right)n}\left(\left(\frac{\lambda}{\lambda+\mu}\right)^2 n\ln\gamma\right)}{\lambda - \mu\gamma^{-\left(\frac{\lambda}{\lambda+\mu}\right)n}\left(\left(\frac{\lambda}{\lambda+\mu}\right) n\ln\gamma - 1\right)}$$

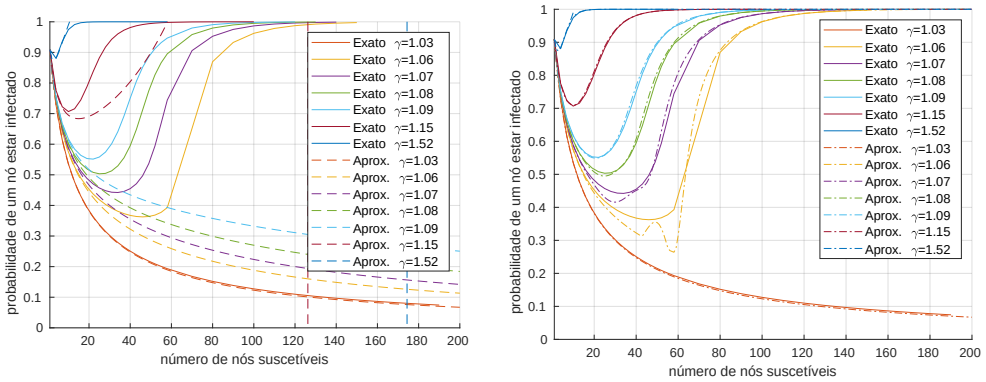
Analogamente, seja $\rho_i^{(1)}$ a probabilidade de infecção aproximada após i iterações do método de Newton, com condição inicial $\rho_0 = 1$. Então,

$$\rho_2^{(1)} = \frac{\lambda - \mu\gamma^{-\left(\frac{\lambda-\mu\gamma^{-n}(n\ln\gamma)}{\lambda-\mu\gamma^{-n}(n\ln\gamma-1)}\right)n}\left(\left(\frac{\lambda-\mu\gamma^{-n}(n\ln\gamma)}{\lambda-\mu\gamma^{-n}(n\ln\gamma-1)}\right)^2 n\ln\gamma\right)}{\lambda - \mu\gamma^{-\left(\frac{\lambda-\mu\gamma^{-n}(n\ln\gamma)}{\lambda-\mu\gamma^{-n}(n\ln\gamma-1)}\right)n}\left(\left(\frac{\lambda-\mu\gamma^{-n}(n\ln\gamma)}{\lambda-\mu\gamma^{-n}(n\ln\gamma-1)}\right) n\ln\gamma - 1\right)} \tag{13}$$

A fórmula fechada com $\rho_0 = 0$ é bem mais simples que com $\rho_0 = 1$. Conforme iremos indicar nas seções a seguir, para muitos cenários a primeira aproximação, mais simples, é suficiente.

4.2. Heurística para determinação da condição inicial

A seguir consideramos uma heurística para determinar a condição inicial ótima da iteração de Newton descrita na seção anterior. Para tal, ilustramos o comportamento da aproximação quando $\rho_0 = 0$ na Figura 2(a) e $\rho_0 = 1$ na Figura 2(b) exceto a curva para $\gamma = 1.03$, usando os valores referência na Tabela 1. Na medida em que N aumenta, a condição inicial $\rho_0 = 1$ tende a produzir melhores aproximações. Entretanto, para valores pequenos de γ (e.g., $\gamma = 1,03$) é preciso utilizar a condição inicial $\rho_0 = 0$ mesmo para valores grandes de N.



(a) $\rho_0 = 0$ (b) ρ_0 dado por (15)

Figura 2. Probabilidade de infecção, calculada usando método de Newton com (a) condição inicial $\rho_0 = 0$ e (b) heurística para condição inicial.

Dependendo da condição inicial, o método de Newton pode convergir para valores maiores que 1 ou menores que 0. Como ilustrado na Figura 2(a), para $\gamma = 1,52$ e $1,15$. Portanto, nossa heurística para determinar a inicialização parte da definição das seguintes quantidades auxiliares adicionais,

$$\bar{\rho}_2^{(z)}(N) = \begin{cases} \rho_2^{(z)}(N), & \text{se } 0 \le \rho_2^{(z)}(N) \le 1 \text{ e } \bar{\rho}_2^{(z)}(N-1) \neq -\infty, \\ -\infty, & \text{caso contrário.} \end{cases} \tag{14}$$

onde z é a condição inicial, $0 \le z \le 1$. Segundo (14), se o método de Newton convergir para valores além do domínio de interesse para determinada condição inicial, tal condição é descartada daí em diante. Em (14) deixamos explícita a dependência de ρ com relação a $N = n+1$ (na Seção 4.1 tal dependência foi mantida implícita). Motivado pela discussão acima, nossa heurística é então dada por,

$$\bar{\rho}(N) = \max(\bar{\rho}_2^{(0)}(N), \bar{\rho}_2^{(1)}(N)) \tag{15}$$

A Figura 2(b) ilustra a qualidade das aproximações obtidas por meio da heurística de inicialização. Para os cenários em consideração a heurística foi capaz de determinar boas escolhas para inicializar os parâmetros.

4.3. Vacinar, reiniciar ou esperar?

Tendo em vista a solução com fórmulas aproximadas descrita nesta seção, podemos gerar curvas como aquela apresentada na Figura 2 de forma bem eficiente. Dada esta curva, assumimos então um custo fixo da contramedida mais custosa (vacinar). Para fins de ilustração, consideramos que o custo é dado pela probabilidade de um nó estar infectado. A utilidade do usuário é dada então pela diferença entre a probabilidade de infecção e o custo. Considere, por exemplo, o caso $\gamma = 1,07$ na Figura 2 e o custo igual a $0,5$. Segundo a Figura 2, se o número de nós não vacinados for menor ou igual a 20, a probabilidade de infecção é alta e os nós têm incentivos para se vacinarem (sistema dominado por infecções exógenas). Por outro lado, se o número de nós não vacinados variar entre 20 e 50, o custo de vacinação é superior à probabilidade de infecção. Nesse caso, os nós têm incentivo para não vacinarem-se, e simplesmente esperarem e reiniciarem suas máquinas quando detectarem um ataque (de forma reativa) ou quando avaliarem que o sistema está ocioso (de forma proativa).

Mensagem desta seção Nesta seção, apresentamos fórmulas para estimar a probabilidade de contaminação de nós na rede. As fórmulas fechadas aproximadas podem ser usadas para guiar a tomada de decisão com relação a contramedidas (e.g., vacinar, reiniciar ou esperar, conforme discutido acima). Na seção a seguir, indicamos por meio de simulações que de fato os regimes discutidos acima, nos quais diferentes contramedidas são adotadas, também são observados nos cenários simulados.

5. Simulação

Nesta seção apresentamos o simulador[1] de propagação de *malware* construído para reproduzir o comportamento descrito na Seção 2. Nossos objetivos são (i) ilustrar o comportamento do sistema sob condições diferentes do modelo analítico, e.g., assumindo que os nós podem entrar e sair da rede, e que os tempos entre eventos não são necessariamente exponenciais e (ii) comparar os resultados do modelo analítico contra simulações. Nosso simulador é flexível e permite a avaliação de *malware* com diferentes padrões de comportamento, que descrevemos a seguir.

Configuração do simulador O simulador construído permite verificar o comportamento e dinâmica de uma rede, sob a perspectiva de um ataque de código malicioso tipo *Botnet Mirai*, contando com um atacante estratégico, o *Botmaster*. As contaminações se dão pelo processo de varredura, autenticação e infecção descrito na Seção 2. As tentativas de autenticação podem falhar porque o dispositivo alvo é seguro ou porque já foi infectado. Em ambos os casos, o dispositivo alvo não responde à tentativa de autenticação.

O *botMaster* busca por *hosts* vulneráveis e troca mensagens para realizar a infecção. Caso a latência seja superior a um *timeout* determinado, a contaminação falha. A taxa de contaminação do *botMaster* é fixa, independente do número de nós na rede. No modelo, tal taxa corresponde ao parâmetro Λ. A taxa de contaminação exógena

[1]Código disponível em `https://github.com/TopologyMapping/miraisim/`.

por *host* é $\lambda = \Lambda/N$. Cada *host* contaminado torna-se um *bot*, que pode iniciar o processo de contaminação de todos os *hosts* vulneráveis alcançáveis. Tal contaminação endógena começa por uma autenticação na vítima, seguida pelo processo de tentativa de infecção. Os parâmetros do simulador com seus valores de referência estão listados na Tabela 2.

PARAM.	DESCRIÇÃO	REFERÊNCIA
Tamanho da rede		
M	Total de indivíduos	10 a 500
$N_p = N/M$	Proporção de indivíduos vulneráveis (não vacinados)	100%
Comportamento dos dispositivos		
$\mathcal{D}_{\text{on}}$	Distribuição do período ligado (up-time)	Exponencial
$P_{\mathcal{D}_{\text{on}}}$	Parâmetros que definem a distribuição	Média $\tau = 65$
τ	do período ligado (Média, Variância, ...) - Exp(τ)	unid. tempo
$\mathcal{D}_{\text{off}}$	Distribuição do período desligado (down-time)	Exponencial
$P_{\mathcal{D}_{\text{off}}}$	Parâmetros que definem a distribuição	Média de $0,1$
	do período desligado (Média, Variância, ...)	unid. tempo
Latência fim-a-fim		
l_{min} e l_{max}	Latência fim-a-fim mínima e máxima,	$0,01$ e $0,4$
	sendo latência uniformemente distribuída	
T	*Timeout*, tempo máximo de conexão	$2,0$
m_{auth}	Mensagens em uma tentativa de autenticação	7
m_{infect}	Mensagens em uma tentativa de infecção	700
Comportamento do malware		
$\mathcal{B}_{\text{exe}}$	Modelo de execução do *bot*	"BroadcastBot"
$\beta_{\mathcal{B}_{\text{exe}}}$	Parâmetro do modelo de execução do *bot*	Taxa de contaminação 5×10^{-5}
$\mathcal{M}_{\text{exe}}$	Modelo de execução do *botMaster*	"UnicastBot"
$\alpha_{\mathcal{M}_{\text{exe}}}$	Parâmetro do modelo de execução do *botMaster*	Taxa de contaminação 2×10^{-2}

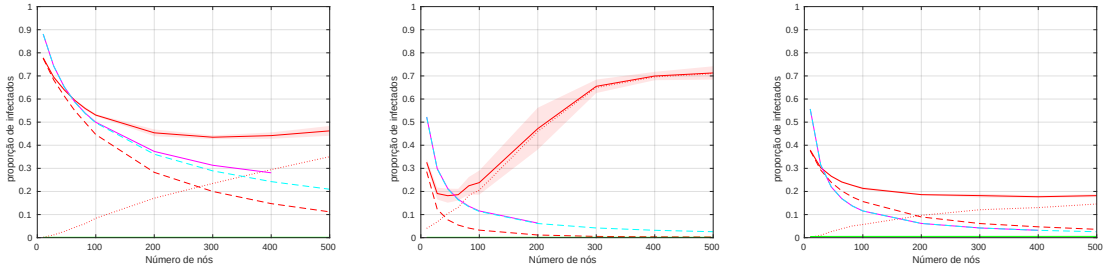
Tabela 2. Parâmetros do simulador e valores de referência.

Modelo analítico e simulação A Fig. 3 ilustra a probabilidade de infecção segundo o modelo analítico proposto (tanto solução exata quanto aproximada, nas curvas magenta contínuo e ciano tracejado, respectivamente). O modelo captura qualitativamente o comportamento da simulação, indicando que no regime inicial, quando o número de nós na rede é pequeno, o sistema é dominado por infecções exógenas. Na medida em que o número de nós na rede aumenta, a probabilidade de infecção primeiro diminui e depois aumenta, atingindo o segundo regime no qual o sistema é dominado por infecções endógenas.

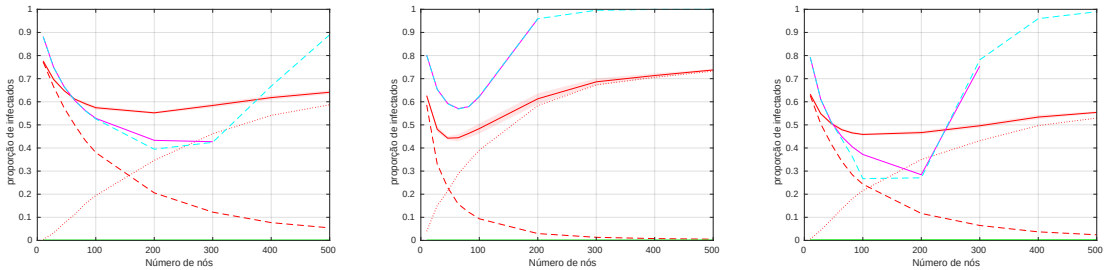
Resultados de simulação ainda na Figura 3, é apresentado a proporção média da população vulnerável infectada (vermelho) e desligada (verde) em função do números de nós vulneráveis (N). O tempo que cada nó passa desligado foi selecionado um valor muito pequeno em relação aos intervalos analisados, pois não é previsto no modelo analítico quando o nó está desligado. Cada curva reflete a média de oito rodadas de simulações. As linhas pontilhadas e tracejadas representam o número de nós infectados de forma endógena e exógena, respectivamente. Somando os valores correspondentes a estas duas linhas, obtemos a fração de nós infectados (linha vermelha).

Em geral, o modelo tende a superestimar a probabilidade de infecção em relação à simulação. Isto deve-se ao fato de que (i) no modelo assumimos que os nós estão

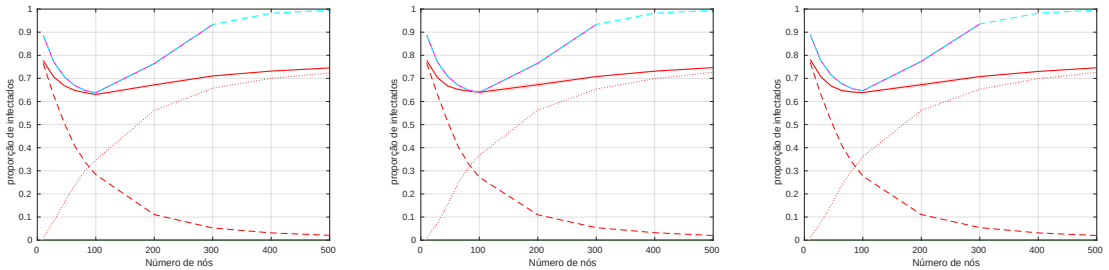
Taxa de infecção endógena | Taxa de Infecção exógena | Tempo médio ligado



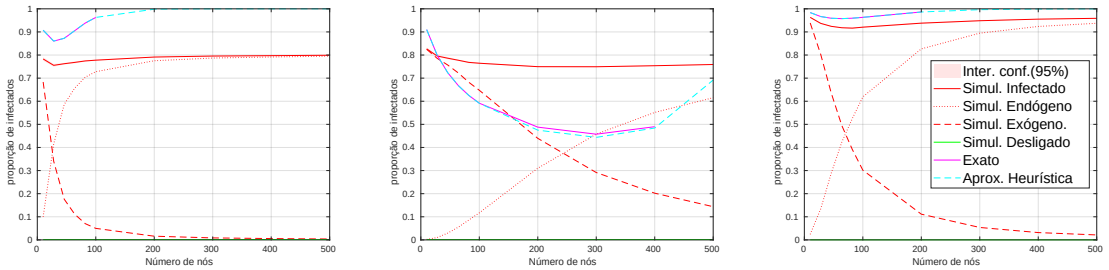
(a) $\alpha = 8 \times 10^{-5}$ $\mu = 21.3801$, $\gamma = 1.0071$ | (b) $\beta = 5 \times 10^{-2}$ $\mu = 154.8886$, $\gamma = 1.0262$ | (c) $\tau = 18$ $\mu = 122.9879$, $\gamma = 1.0061$



(d) $\alpha = 20 \times 10^{-5}$ $\mu = 21.6194$, $\gamma = 1.0092$ | (e) $\beta = 32 \times 10^{-2}$ $\mu = 44.7100$, $\gamma = 1.0262$ | (f) $\tau = 40$ $\mu = 43.5209$, $\gamma = 1.0148$



(g) $\alpha = 50 \times 10^{-5}$ $\mu = 21.4835$, $\gamma = 1.0148$ | (h) $\beta = 200 \times 10^{-2}$ $\mu = 21.4111$, $\gamma = 1.0148$ | (i) $\tau = 65$ $\mu = 20.9333$, $\gamma = 1.0148$



(j) $\alpha = 500 \times 10^{-5}$ $\mu = 20.7848$, $\gamma = 1.0383$ | (k) $\beta = 2000 \times 10^{-2}$ $\mu = 15.7172$, $\gamma = 1.0071$ | (l) $\tau = 260$ $\mu = 2.8819$, $\gamma = 1.0171$

Figura 3. Resultados de simulação para o comportamento da rede totalmente conectada, sob atuação da *Botnet Mirai* na presença de um atacante estratégico. Quando não especificado, o modelo analítico e a simulação foram ajustados respectivamente para: $\Lambda = 1500$, $\alpha = 5 \times 10^{-5}$, $\beta = 2 \times 10^{-2}$ e $\tau = 65$, exceto quando específicado na figura.

sempre ligados, enquanto que na simulação os nós alternam entre ligados e desligados, (ii) o modelo assume contribuições multiplicativas das taxas de infecção, enquanto que a simulação considera contribuições aditivas e (iii) no modelo, todos os tempos entre eventos são exponencialmente distribuídos, enquanto que na simulação a latência na rede é uniforme (tal latência não é levada em conta no modelo). Trabalhos futuros consistem em verificar sob que condições o modelo produz um limite superior para a probabilidade de infecção de fato observada na rede. Note também que embora a aproximação proposta tenha apresentado bons resultados na Figura 2, em alguns cenários da Figura 3 a aproximação distanciou-se do valor exato previsto pelo modelo, e estamos no momento averiguando formas de refinar a aproximação.

Análise de sensitividade Para estudar a sensitividade da probabilidade de infecção em função dos diferentes parâmetros do sistema, mantemos todos os parâmetros fixos e variamos um de cada vez para avaliar seu impacto. Em particular, na primeira, segunda e terceira colunas da Figura 3 variamos a taxa de contaminação endógena (α), taxa de contaminação exógena (β) e tempo médio que o dispositivo permanece ligado (τ). Nas curvas obtidas via simulação indicamos o intervalo de confiança de 95%. As linhas roxa e verde correspondem, respectivamente, à solução exata do modelo (eq. (3)) e à aproximação de Newton, com duas iterações, conforme heurística definida na Seção 4.2. *Cabe ressaltar que o simulador caracteriza detalhadamente o comportamento do Mirai Botnet, enquanto o modelo proposto captura a essência do sistema.*

O sistema passa por dois regimes fundamentais, primeiro sendo dominado por infecções exógenas e depois por infecções endógenas. Em todos os cenários apresentados na Figura 3 observa-se que o sistema passa por dois regimes. Tal fato pode ser constatado focando-se nas linhas pontilhadas e tracejadas, que crescem e diminuem, respectivamente, na medida em que o número de nós no sistema aumenta. Tal comportamento observado em simulações está de acordo com o previsto pelas equações (7) e (12). No primeiro regime, o sistema é dominado por infecções exógenas (linha tracejada acima da linha pontilhada). Na medida em que o número de nós na rede aumenta, as infecções endógenas também passam a exercer papel importante. No segundo regime, o sistema é dominado por infecções endógenas (linha tracejada acima da linha pontilhada). Em nossas simulações, observamos que o número de nós no sistema correspondente ao cruzamento dos gráficos de infecções endógenas e exógenas (cruzamento das curvas pontilhada e tracejada) é igual ou aproximadamente igual a aquele que minimiza a proporção de nós infectados (curva vermelha).

A probabilidade de infecção é mais sensível à taxa de contaminação endógena que à taxa de contaminação exógena. Isto ocorre porque a taxa de infecção endógena é amplificada pelo número de nós infectados na rede, enquanto que a taxa de infecção exógena é limitada pelo *Botmaster*. Um aumento em torno de 60 vezes da taxa de infecção endógena produz os efeitos observados na primeira coluna da Figura 3. Já a taxa de infecção exógena teve um aumento de 400 vezes para observarmos a variação de padrões na segunda coluna da Figura 3.

O tempo médio que um dispositivo não vacinado permanece ligado (logo, suscetível) é também um fator relevante na simulação. O valor assintótico da fração de nós infectados, por exemplo, depende do tempo médio que um dispositivo permanece ligado, conforme vemos na última coluna da Figura 3. Na medida em que os nós permanecem

mais tempo ligados, a proporção de nós infectados também aumenta. *Simplesmente desligar os sistemas pode ser uma estratégia eficaz para conter epidemias. Entretanto, tal estratégia pode acabar por atender aos anseios do atacante, de causar um ataque de DDoS por indisponibilidade dos sistemas alvos.*

6. Trabalhos Relacionados

A literatura sobre modelos epidemiológicos é vasta, levando em conta aspectos transientes [Ganesh et al. 2005] e estacionários [Keeling and Eames 2005, Tavares et al. 2018], bem como infecções endógenas e exógenas [Zhang et al. 2017, Zhang 2015, Zhang and Moura 2018]. Entretanto, não é de nosso conhecimento nenhum trabalho que tenha analisado modelos analíticos levando em consideração atacantes estratégicos, de capacidade limitada, capazes de causar infecções exógenas, gerando expressões para a probabilidade de infecção dos nós.

Este trabalho é uma extensão de [Rufino et al. 2018], onde propusemos o modelo epidemiológico analisado no presente artigo, dentre as principais contribuições do presente trabalho destacamos três: (i) aproveitando-se de resultados recentes em [Zhang and Moura 2018], derivamos um método iterativo para calcular a probabilidade de infecção dos nós. Em seguida, apresentamos (ii) fórmulas fechadas (iterativa com duas iterações) para aproximar a probabilidade de infecção assim como (iii) resultados de simulação.

A proliferação de *malware* e a formação de grandes *botnets* permitem a execução de ataques DDoS com volume capaz de afetar serviços com grande capacidade [Kolias et al. 2017, York 2016]. O crescimento da Internet das Coisas (IoT) [Peterson 2019], combinado com as vulnerabilidades presentes nestes dispositivos e a dificuldade de atualizá-los criaram um ambiente propício para construção de *botnets* [Angrishi 2017]. As caracterizações e comportamentos observados em *malware* real podem ser utilizados para parametrizar nossos modelo e simulador. Nossos modelos e o simulador são gerais o suficiente para serem aplicados a novos *malwares* que venham a ser identificados e caracterizados.

Vários trabalhos na literatura estudam o comportamento e a evolução de *malwares* [Antonakakis et al. 2017, Marzano et al. 2018]. Existe uma constante evolução dos *malwares* por parte dos atacantes. *Assim, a transição do estado infectado para o estado suscetível considerada neste trabalho pode refletir o fato de que um nó infectado, após aplicar uma contramedida, voltou a se tornar suscetível com relação a novas variantes de um mesmo* malware.

7. Conclusões

Neste trabalho, consideramos a caracterização do processo de propagação de epidemias frente atacantes estratégicos. Em particular, considerando o modelo analítico previamente proposto em [Rufino et al. 2018], propusemos um método iterativo para calcular a probabilidade de infecção dos nós. Comparando os resultados analíticos com resultados de simulações, observamos que o modelo captura o comportamento do sistema mesmo quando consideramos nós intermitentes, bem como tempos entre eventos que não sejam exponencialmente distribuídos. Acreditamos que os resultados apresentados neste artigo constituam um passo no sentido de estabelecer os fundamentos para melhor modelagem

e análise do processo de propagação de epidemias. Tal entendimento é crítico para auxiliar na tomada de decisões, por exemplo, sobre vacinar ou reiniciar sistemas conectados à rede, levando em conta os custos de tais contramedidas e os respectivos riscos da não implementação das mesmas.

Referências

Angrishi, K. (2017). Turning Internet of Things (IoT) into Internet of Vulnerabilities (IoV): IoT Botnets. *CoRR*, abs/1702.03681.

Antonakakis, M., April, T., Bailey, M., Bernhard, M., and et al. (2017). Understanding the Mirai Botnet. In *Proc. of USENIX Security Symposium*.

Darboux, M. (1869). Sur la méthode d'approximation de Newton. In *Nouvelles annales de mathématiques*, volume 8, pages 17–27.

Ganesh, A., Massoulié, L., and Towsley, D. (2005). The effect of network topology on the spread of epidemics. In *INFOCOM*, volume 2, pages 1455–1466. IEEE.

Keeling, M. J. and Eames, K. T. (2005). Networks and epidemic models. *Journal of the Royal Society Interface*, 2(4):295–307.

Kelly, F. P. (1979). *Reversibility and stochastic networks*. John Wiley, New York.

Kolias, C., Kambourakis, G., Stavrou, A., and Voas, J. (2017). DDoS in the IoT: Mirai and other Botnets. *Computer*, 50(7):80–84.

Krebs, B. (2016). Krebsonsecurity hit with record ddos. `https://tinyurl.com/krebs2019`.

Marzano, A., Alexander, D., Fonseca, O. L. H. M., Fazzion, E. C., Hoepers, C., Steding-Jessen, K., Chaves, M. H. P. C., Cunha, Í. S., Guedes, D. O., and Jr., W. M. (2018). The evolution of Bashlite and Mirai IoT botnets. In *Computers and Communications*.

Peterson, D. (2019). ICS security patching: Never, next, now. `https://tinyurl.com/wperf2019b`.

Rufino, V., Menasche, D., Cunha, I., Lima, C., and de Aguiar, L. P. (2018). Contaminação epidêmica em redes: Imunidade coletiva e suas implicações frente a atacantes estratégicos. In *WPerformance*, volume 17. SBC.

Tavares, J., Iacobelli, G., and Figueiredo, D. R. (2018). Simulação escalável de epidemias em redes baseadas em passeios aleatórios. In *WPerformance*, volume 17. SBC.

York, K. (2016). Dyn Statement on 10/21/2016 DDoS Attack. `http://dyn.com/blog/dyn-statement-on-10212016-ddos-attack/`.

Zhang, J. (2015). *Network Process: How Topology Impacts the Dynamics of Epidemics and Cascading Failures*. PhD dissertation, Carnegie Mellon University.

Zhang, J. and Moura, J. M. (2018). Who is more at risk in heterogenous networks? In *Acoustics, Speech and Signal Processing (ICASSP)*, pages 1–5. IEEE.

Zhang, J., Moura, J. M., and Zhang, J. (2017). Contact process with exogenous infection and the scaled sis process. *Journal of Complex Networks*.