Contaminação Epidêmica em Redes: Imunidade Coletiva e Suas Implicações Frente a Atacantes Estratégicos

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

[1]DCTIM - Marinha do Brasil, Centro, Rio de Janeiro - RJ

[2]PPGI - Universidade Federal do Rio de Janeiro, Ilha do Fundão, Rio de Janeiro - RJ

[3]DCC - Universidade Federal de Minas Gerais, Pampulha, Belo Horizonte, MG

[4]Siemens Corporate Research, Princeton, NJ

Abstract. *Herd immunity, one of the most fundamental concepts in network epidemics, occurs when a large fraction of the population is immune against a virus or malware. The few individuals who have not taken countermeasures against the threat are assumed to have very low chances of infection, as they are indirectly protected by the rest of the population. Although very fundamental, herd immunity does not account for strategic attackers scanning the network for vulnerable nodes. In face of such attackers, extant nodes who linger vulnerable in the network become easy targets. The greater the proportion of individuals in a community who are immune, the higher the probability that the attacker will choose those few who are not immune. In this paper, we propose an analytical model which allows us to capture the impact of countermeasures against attackers when both in-network as well as exogenous infections coexist. Using the proposed model, we show that a diverse set of potential attacks produces non-trivial equilibria, some of which go counter to herd immunity, where nodes are more prone to adopt countermeasures when the remainder of the nodes has decided to do so.*

Resumo. *Imunidade coletiva, um dos conceitos fundamentais associados à contaminações em redes, ocorre quando grande parte da população está imune a um ataque oriundo de uma classe de códigos maliciosos. A menor parcela de indivíduos que permanecem sem adotar contramedidas tende a estar mais protegida de ataques, por efeito da redução da contaminação epidêmica. Contudo, esta condição não leva em consideração a capacidade do atacante de estrategicamente buscar por indivíduos vulneráveis, se tornando alvos fáceis. Neste artigo, propomos um modelo analítico que nos permite capturar o impacto das contramedidas face a epidemias e atacantes estratégicos. Usando o modelo propostos, podemos verificar a existência de um equilíbrio não trivial entre o custo de uma imunização e a probabilidade de estar infectado em função do número de suscetíveis.*

1. Introdução

Motivação Modelos epidêmicos são largamente estudados há décadas. Provenientes da biologia, e posteriormente também usados no estudo da propagação de vírus em redes de computadores [Murray 1988, Kephart and White 1991], tais modelos visam capturar

como um vírus se espalha em uma rede, seus impactos e o efeito das contramedidas para mitigar ameaças. As contramedidas geralmente passam por imunizar indivíduos vulneráveis. Quando a parcela da população imunizada atinge um limiar percentual significativo dentro da rede, evita-se a contaminação de indivíduos que ainda não foram imunizados [Libster 2017]. Tal fenômeno é conhecido como imunidade coletiva (do inglês, *herd immunity*), e está relacionado a externalidades positivas decorrentes da imunização.

Desafios e limitações do estado da arte A análise do custo-benefício de um programa de vacinação leva em conta o número de agentes (nós) contaminados e imunes e consequentemente o risco de contaminação de agentes suscetíveis. Numa população onde poucos indivíduos são suscetíveis, estes possuem menor incentivo em investir na vacinação. Esta trata-se de uma decisão racional, que leva em conta indivíduos que seguem um comportamento de *evitar a multidão* (*avoid the crowd*), e que podem seguramente ignorar a vacina (evitando seus custos e efeitos colaterais). Contudo, esta análise ignora a possibilidade de ocorrer uma contaminação exógena, na qual sistemas computacionais podem ser contaminados por adversários estratégicos. Atualmente, é possível que um adversário estratégico consiga localizar sistemas suscetíveis (vulneráveis) em poucas horas, por meio de eficientes varreduras de todo o espaço de endereços IPv4 da Internet [Durumeric et al. 2013]. Ainda que hajam poucos indivíduos suscetíveis, neste caso o risco de contaminação é alto. O risco aumenta para os poucos indivíduos ainda não vacinados. Há um incentivo para que estes apliquem as vacinas, e a decisão racional é de *seguir a multidão* (*follow the crowd*). Não é de nosso conhecimento nenhum trabalho na literatura de epidemias que tenha considerado o efeito de atacantes estratégicos com poder para varrer todo o endereço de IPs da Internet, na busca por nós vulneráveis. O presente trabalho visa preencher esta lacuna.

Objetivos e metodologia Neste trabalho, estabelecemos políticas ótimas de aplicação de contramedidas, levando em conta seus custos e benefícios. Ao considerarmos os benefícios da aplicação de uma contramedida para um determinado nó, levamos em consideração o fato de que estes dependem das estratégias adotadas pelos demais participantes. Por meio de um modelo analítico, determinamos condições para existência de equilíbrios, nos quais os nós não possuem incentivos para modificar suas estratégias de contramedidas.

Nossos achados também são embasados por resultados empíricos sobre a aplicação de *patches*, observados experimentalmente na rede Shodan. Avaliando o comportamento da população de usuários que expõem dispositivos de controle industrial na Internet, podemos observar que para certos dispositivos a população se comporta num padrão condizente com o *follow the crowd*, enquanto que para outros dispositivos o comportamento assemelha-se a *avoid the crowd*.

Contribuições Podemos resumir nossas contribuições conforme a seguir:

Identificação do comportamento de usuários reais Avaliando dados advindos da plataforma Shodan, identificamos que em cenários reais, para inúmeros dispositivos, conjuntos de usuários se comportam de forma similar ao *follow the crowd*. Em

outas palavras, para tais dispositivos, quanto mais usuários aplicam uma contramedida (*patch*), maior a chance de um novo usuário também aplicar tal contramedida. Esta observação serve como motivador adicional para estudar-se tal regime, que vai ao encontro da literatura padrão de epidemiologia, na qual presume-se que o comportamento padrão é do tipo *avoid the crowd*.

Modelo analítico Propomos um modelo analítico capaz de capturar fatores positivos e negativos associados as contramedidas de segurança computacional em um jogo de defesa e ataque. O modelo proposto é simples, tratável, enquanto ainda tem poder expressivo para produzir *tradeoffs* relacionados à vacinação de nós em rede.

Equilíbrio Apresentamos uma investigação numérica de equilíbrio no sistema, indicando que para dois valores extremos de regime a probabilidade de infecção cresce monotonicamente como função do tamanho da população suscetível (vulnerável), correspondendo ao comportamento de seguir a multidão. Com o sistema sob alguns ajustes, a probabilidade de infecção primeiro diminui e depois cresce em função do tamanho da população suscetível.

Organização O restante deste artigo é organizado da seguinte forma. Na Seção 2 apresentamos uma descrição dos aspectos relevantes do sistema sobre o qual se propagam as epidemias em consideração. Avaliamos o comportamento de usuários reais com relação a aplicação de contramedidas na Seção 3. Tais comportamentos motivam aspectos do modelo analítico apresentado na Seção 4, seguido por sua análise na Seção 5. Trabalhos relacionados e a conclusão seguem nas Seções 6 e 7.

2. Descrição do sistema

2.1. *Malware*: infecções exógenas e endógenas no mundo real

WannaCry é um código malicioso que ficou conhecido em 12 de maio de 2017 por ser um *ransomware* (sequestra arquivos de usuários e exige resgate) que em apenas um dia havia atingido 230.000 usuários infectados em mais de 150 países [Lee 2017]. As vulnerabilidades exploradas no protocolo SMBv2 haviam sido divulgadas e corrigidas pela Microsoft ainda em março. Em [Lee 2017] questiona-se como é possível que uma ameaça explorando um protocolo típico de redes locais tenha se espalhado tão rapidamente. O *WannaCry* é um código malicioso que possui uma alta taxa de contaminação endógena, mas bastante limitada a capacidade de contaminação entre redes. Em geral, o código entrava nas redes locais por meio de anexos em emails falsos (contaminações exógenas). *Neste trabalho, propomos um modelo analítico que visa capturar o impacto de infecções exógenas na propagação de epidemias em sistemas computacionais.*

Mirai é outro código malicioso que também se espalhou rapidamente pela Internet. Porém, diferente do *WannaCry*, seu alvo eram dispositivos que pudessem estar com configurações inadequadas ou mesmo de fábrica. Estes foram usados como fonte de ataques de Negação de Serviço Distribuídos (DDoS), por meio de um controle centralizado (*botmaster*). A análise de [Antonakakis et al. 2017] revela que a estrutura do código fonte possui uma parte do código, executado nas vítimas, que busca por novos alvos e realiza comandos do *botmaster*; ao encontrar uma vítima, são testados combinações conhecidas de usuário e senhas, e caso haja sucesso, a informação é passada para um servidor, que de

forma assíncrona usa o login e senha para carregar o código apropriado de acordo com a arquitetura do dispositivo. Como o código não é residente, uma simples reinicialização do dispositivo pode fazer com que o código executado na vítima seja descarregado, e o servidor que serve ao atacante deve recontaminar as vítimas novamente. *Este comportamento é similar ao do modelo SIS, considerado neste trabalho.*

Após ter conhecimento de quais dispositivos estão vulneráveis, a capacidade do atacante se limitará a capacidade do servidor de carregar o código nas vítimas e de manter a rede operacional. Como não há contaminação de vítimas para vítimas, podemos supor que a taxa de contaminação endógena é pequena, mas não nula pois a descoberta de novas vítimas ainda se dá pelas vítimas. A taxa de contaminação exógena, por outro lado, é um elemento chave do sistemas, e é limitada pela capacidade do carregador do código malicioso injetá-lo nas vítimas identificadas. *O impacto de tal taxa de contaminação exógena é um dos objetos de estudo deste trabalho.*

2.2. Poder do atacante

Consideramos um adversário ao sistema (atacante que possui controle de um *malware*) que possui um poder computacional limitado. Sua capacidade média de contaminação por unidade de tempo é denotada por Λ. No caso mais simples, esta capacidade de ataque é voltada de forma uniforme entre os nós suscetíveis existentes no sistema.

2.3. Contramedidas

Contramedidas moderadas A vacinação é uma contramedida importante para evitar a disseminação epidêmica. Em sistemas computacionais existem formas de vacinações de efeito moderado e de efeito total. Vacinação de efeitos moderados são aquelas realizadas por meio de atualizações de sistemas e anti-vírus baseados em assinatura, aos quais possuem a necessidade de constantes renovações, diárias ou semanais ("jogo do gato e rato"). Tão logo as atualizações são disponibilizadas, *hackers* usam técnicas de despistamento, modificando o código e o comportamento dos programas maliciosos. Assim, promovem-se várias gerações de um mesmo programa malicioso. As novas versões de sistemas de proteção (anti-vírus) precisam lidar com as evoluções de códigos maliciosos, caracterizando-se assim um processo de contaminação epidêmica tipicamente caracterizado como SIS (Susceptible Infected Susceptible). *Segundo este modelo, adotado neste trabalho, os nós alternam entre os estados de suscetível e infectado.*

Contramedidas rigorosas Algumas contramedidas contra códigos maliciosos envolvem tratamentos mais rigorosos, tais como a desconexão de nós da rede, instalação de sistemas operacionais mais modernos e sistemas de anti-vírus de efeito total. Estes últimos detectam mais eficientemente os códigos maliciosos, porém possuem um custo de manutenção, poder computacional e capacidade de memória maiores que os anti-vírus de efeito moderado. *Para todos os propósitos práticos, neste trabalho os nós que aplicam algum tipo de contramedida rigorosa são considerados imunes, e são removidos da população de interesse. O tamanho da população, portanto, é igual ao número de nós que não aplicaram contramedidas ou que aplicaram contramedidas moderadas.*

2.4. Dilema da atualização (*patch management*)

Toda aplicação de um *patch* de correção ou a atualização de um sistema computacional possui um custo. Muitas vezes a simples paralisação do sistema computacional pode gerar perdas consideráveis. Além disso, uma atualização pode representar uma mudança de tecnologia, que pode representar a substituição de todo um parque instalado e investimentos elevados. Os modelos de epidemias podem auxiliar essa decisão, fornecendo a probabilidade e o valor esperado de um agente ou rede ser contaminada. Neste trabalho, focamos no impacto da decisão de um agente sobre a decisão dos demais.

Evitando a multidão Essa é a abordagem clássica, e se assemelha ao modelo biológico. Se todos os indivíduos foram vacinados, o risco da epidemia diminui, pois há poucos indivíduos que podem ser contaminados, diminuíndo o contágio. Portanto no caso da maioria dos indivíduos aplicarem a atualização (*patch*), a decisão de fazer a atualização é desincentivada.

Seguindo a multidão Outra abordagem é semelhante a brincadeira infantil de *pique-esconde*, no qual quanto mais indivíduos se salvaram, maiores as chance dos indivíduos que ainda não foram salvos serem descobertos e perderem o jogo. Este é o caso onde existe um adversário inserido no sistema em busca de sistemas vulneráveis para contaminação. Portanto, se o poder do atacante é finito e a maioria dos indivíduos aplicaram a atualização, a decisão de também fazer a atualização é incentivada.

3. Avaliação do comportamento real de usuários: identificando *follow the crowd* e *avoid the crowd* na Internet

A seguir, visamos identificar na Internet comportamentos que se assemelhem a padrões do tipo *seguir* ou *evitar* a multidão. Para tal, usamos dados reportados em [Wang et al. 2017] sobre a aplicação de *patches* por parte dos usuários. De acordo com [Wang et al. 2017], em uma análise de mais de 64 mil amostras de dispositivos de sistemas de controle industrial, menos de 30% desses sistemas são atualizados para versões imunes a ataques, dentro de um prazo de 60 dias desde a descoberta da vulnerabilidade.

Seleção dos dados Adotamos os seguintes critérios para selecionar os sistemas cujos dados são relevantes para nossas análises: 1) selecionamos sistemas que possuem pelo menos duas versões amostradas; 2) dado que os sistemas podem ter várias versões, para cada sistema escolhemos as versões mais populares, ou seja, as que estavam presentes em mais dispositivos durante o histórico de medições; 3) excluímos os sistemas que não tiveram ao menos dez dispositivos ativos em uma única medição em qualquer uma das versões mais populares; 4) para facilitar e padronizar a caracterização dos sistemas selecionados, numeramos sequencialmente o número das versões, de modo que os valores menores são os mais antigos e valores maiores correspondem aos mais modernos.

Caracterizando o comportamento Os típicos comportamentos de *seguir* ou *evitar* a multidão foram caracterizados conforme a adoção de versões mais modernas ou mais antigas. O comportamento de *seguir* (resp., *evitar*) a multidão ocorre quando o pico de usuários da versão mais moderna (resp., mais antiga) **ocorre após** do pico de usuários da

versão mais antiga (resp., mais moderna), mostrando que há um típico comportamento para atualizar (resp., ignorar) as versões mais recentes do sistema, que se propaga pela população.

A Tabela 1 apresenta o resultado da observação do comportamento de *seguir* a multidão (S), *evitar* a multidão (E) e um comportamento indefinido (I), onde não podemos caracterizar nenhum comportamento típico. As duas primeiras colunas se referem respectivamente ao índice da versão mais popular (1º pop) e ao dia no qual ocorreu a medição com maior número de sistemas utilizando a respectiva versão (topo 1º) [1] e as duas próximas colunas referem-se a segunda versão mais popular. A coluna intitulada S/E/I indica o comportamento típico da população. Para fazer a classificação do comportamento, definimos o índice ι=(topo 1º- topo 2º) / (topo 1º+ topo 2º). O caso $\iota > 0,01$ (resp., $\iota < -0.01$) corresponde ao comportamento de *seguir* (resp., *evitar*) a multidão. Valores entre $-0,01$ e $0,01$ são associados a comportamentos indefinidos.

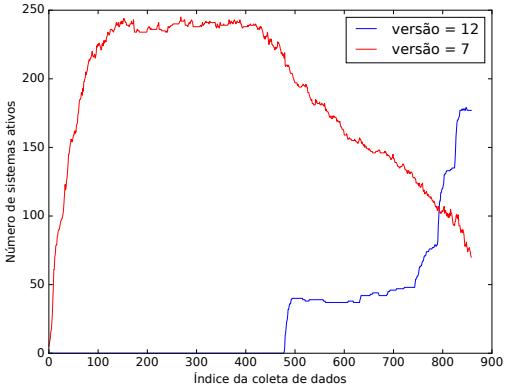
1º pop	topo 1º	2º pop	topo 2º	S/E/I	Nome do sistema (arquivo)
11	847	6	268	S	Allegro RomPager
2	857	1	854	I	AMX NetLinx A
35	833	23	850	E	Apache httpd
25	450	21	191	S	AVM FRITZ!Box Fon WLAN 7170 SIP
5	831	2	847	I	Boa HTTPd
11	156	4	857	E	Dropbear sshd
6	757	4	755	I	Lantronix MSS100 serial interface fingerd
13	448	8	311	S	lighttpd
7	849	5	848	I	Microsoft IIS httpd
12	850	6	833	S	Microsoft SQL Server
2	558	1	567	I	MoxaHttp
45	852	36	833	S	MySQL
36	837	8	848	I	nginx
33	852	28	852	I	OpenSSH
17	857	6	164	S	ProFTPD
7	513	5	521	I	Schneider BMX NOE 0100
5	532	2	525	I	Schneider BMX P34 2020
8	533	7	533	I	Schneider Electric SAS TSXETY4103
8	569	6	572	I	Siemens BACnet Field Panel
3	564	1	547	S	Siemens PXG3
9	1105	8	1096	I	Siemens SIMATIC IM151
5	1119	4	1117	I	Siemens SIMATIC S7 1200
39	2097	38	2097	I	Siemens SIMATIC S7 300
18	838	1	840	I	Tridium Niagara httpd
4	845	3	847	I	Virata-EmWeb
5	841	4	839	I	vxTarget ftpd
7	858	5	858	I	VxWorks ftpd
4	856	3	845	I	WindWeb

Tabela 1. Comportamento *seguir* ou *evitar* a multidão, observado em populações de usuários de sistemas reais de controle industrial conectados à Internet. A maioria dos produtos possui comportamento típico de *seguir* a multidão, o que vai ao encontro de *evitar* a multidão, típico de sistemas biológicos.

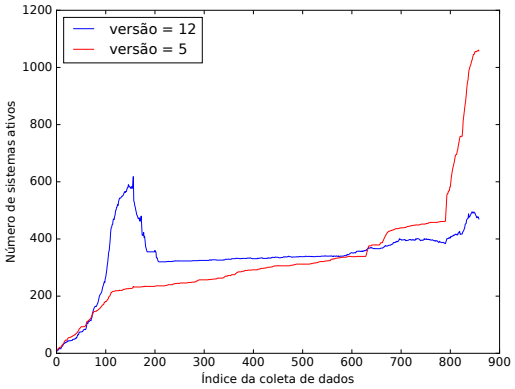
A Figura 1 ilustra o número de dispositivos adotando cada uma das versões de um determinado produto, ao longo do tempo, para dois produtos distintos (Allegro RomPage e Dropbear sshd). A população do Allegro RomPage possui comportamento típico de *seguir* a multidão (a versão mais nova substitui a versão mais antiga). Já o Dropbear sshd possui população com comportamento compatível com *evitar* a multidão. O número de dispositivos com versão antiga cresce em conjunto com o número de dispositivos adotando a versão mais nova do produto. Este último comportamento deve-se, por exemplo,

[1] Os dias são medidos com relação à data da primeira coleta.

a novas instalações do produto, que muitas vezes podem vir embarcados com versões antigas do firmware. Mesmo que o comportamento de *evitar* a multidão não esteja sendo tomado de forma consciente e estratégica, os seus impactos são os mesmos que aqueles observados numa população em que indivíduos param de aplicar uma vacina (contramedida), por considerar que a ameaça é desprezível. Eventualmente, a população pode se ver em face a uma epidemia de um vírus que se pensava erradicado.



a) Seguir a multidão



b) Evitar a multidão

Figura 1. a) Comportamento de seguir a multidão : sistema Allegro RomPage. b)Comportamento de evitar a multidão: sistema Dropbear sshd

4. Modelo epidêmico

Nesta seção apresentamos o modelo a ser usado no decorrer do presente artigo.

4.1. Visão geral

Na presente seção apresentamos um exemplo para ilustrar algumas propriedades importantes do modelo proposto. A seção serve para apreciar intuitivamente algumas das características do modelo. Os parâmetros apresentados possuem fins ilustrativos, de modo a capturar e evidenciar as propiedades de interesse. O modelo será apresentado na próxima seção em maiores detalhes.

Assumimos que o atacante pode identificar nós vulneráveis, mas não pode distingui-los entre suscetíveis e infectados. Além disso, quando existem N nós vulneráveis na rede, a taxa de ataque por nó é $\lambda = \Lambda/N$.

A Figura 2 ilustra graficamente como a probabilidade de infecção varia em função do tamanho da população de nós vulneráveis em diferentes regimes. Quando há um domínio da infecção endógena, o sistema se comporta como a curva mais acima do gráfico, onde a probabilidade de infecção é elevada, independente da quantidade de nós vulneráveis. Todos os incentivos são para que cada indivíduo aplique a vacina (contramedida mais custosa); quando a taxa de infecção endógena atinge valores moderados (próximo do nulo multiplicativo), a probabilidade de infecção passa a depender do número de nós vulneráveis. Nesse caso, na medida em que o número de indivíduos na rede aumenta, verificamos inicialmente um domínio da infecção exógena, e posteriormente da endógena. Para atacantes de capacidade limitada, o aumento do número de nós

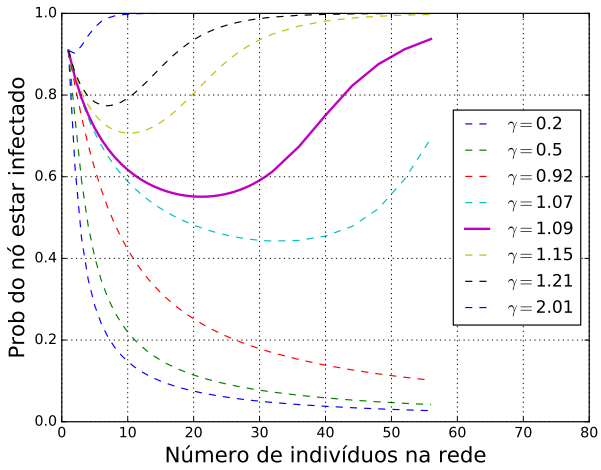


Figura 2. Probabilidade de um nó estar infectado em função do tamanho da população de vulneráveis.

vulneráveis faz com que diminua-se a probabilidade de um determinado nó estar infectado. Os indivíduos são incentivados a aplicar a vacina quando o tamanho da população é extremamente pequeno, a não aplicá-la quando o tamanho da população é um pouco maior, e novamente a aplicá-la quando a população é grande.

Quando a infecção endógena é $\gamma = 0,2$, a probabilidade de infecção apenas diminui com o aumento de nós vulneráveis, como mostrado nas curvas mais abaixo; quando a taxa de contaminação endógena aumenta, na medida em que o número de nós vulneráveis aumenta eles novamente passam a ter incentivo para aplicar a vacina. Assim, há uma transição do comportamento de seguir para evitar a multidão.

variável	descriçao	valor de referência
N	população pronta para infecção	-
M	tamanho total da população	56
γ	taxa de infecção endógena (por aresta)	1.09
λ	taxa de infecção exógena (por nó)	Λ/N
Λ	taxa 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 2. Tabela de notação

Estendendo um pouco mais a análise do gráfico da Figura 2, destacamos a curva que possui a contaminação endógena $\gamma = 1,09$, linha contínua do gráfico; considerando como um custo limite para aplicar a vacina uma probabilidade de contaminação igual a $0,6$ podemos verificar na Figura 3 dois pontos de equilíbrio: o primeiro instável entre 11 e 12 nós vulneráveis, e outro estável entre 30 e 31 nós vulneráveis. A probabilidade de contaminação é mínima quando o número de nós vulneráveis é igual a 21, atingindo

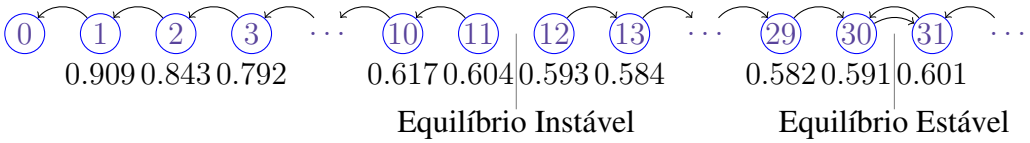


Figura 3. Dinâmica do número de nós na vulneráveis (suscetíveis ou infectados). Considerando uma taxa de contaminação endógena $\gamma = 1.09$ e custo da contramedida igual a 0.6 obtemos dois pontos de equilíbrio. O equilíbrio dinâmico instável (resp., estável) conta com um número de nós na rede que varia entre 11 e 12 (resp., 30 e 31).

o valor de $0,5510$. Segundo o modelo proposto, para alcançar-se este ótimo global são necessários incentivos adicionais por parte de entidades regulatórias e/ou empresas de seguros [Grossklags et al. 2008].

4.2. O modelo multiplicativo de contaminação (*scaled* SIS)

Consideramos uma população finita contendo M nós, dos quais, um total de N nós não aplicaram nenhuma forma de vacina, e portanto possuem uma vulnerabilidade que pode ser explorada por um atacante. Os $M-N$ nós, aplicaram alguma vacina e consideramo-os imunes.

Cada um desses N nós podem assumir os estados de suscetível (S ou 0) ou infectado (I ou 1). Um nó infectado pode ser recuperado para o estado suscetível após um período médio $1/\mu$ distribuído exponencialmente.

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 de um determinado nó, podemos assumir que a taxa de infecção é exponencialmente dependente de d, isto é, a contaminação endógena ocorre a uma taxa γ^d. O efeito da contaminação exógena é assumido como multiplicativo, o qual ocorre a uma taxa λ por nó. Logo, a taxa total de infecção é $\lambda\gamma^d$, assumiremos ainda que o tempo entre as infecções é dado por uma distribuição exponencial.

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].

4.3. Função de utilidade e topologia da rede

Cada nó decide se deve investir na aplicação de uma vacina. O investimento na aplicação de uma vacina possui um custo C, enquanto que a contaminação possui um custo H (usando a terminologia consistente com [Hayel et al. 2014]). Se um determinado nó decide não investir em uma vacina, ele passa estar sujeito ao processo de contaminação epidêmica, podendo ser contaminado. Contudo, após ser contaminado o nó pode se recuperar, sem necessariamente ser imunizado, tal como retorno ao estado inicial, formatação e reinicialização. Após a remoção dos nós que decidiram investir na vacina e estão imunes, a topologia restante da rede determina o processo de dispersão da epidemia.

A topologia da rede é dada por uma matriz de adjacência A_{NN}, onde cada entrada a_{kl} é igual a 1 se o nó x_k e x_L estão contaminados, e zero caso contrário. Como tratamos de conexões não direcionadas, então $a_{kl} = a_{lk}$. Os elementos na diagonal de A são todos iguais a 0.

4.4. Matriz geradora infinitesimal

Seja Q a *matriz infinitesimal* associada ao processo Markoviano. Seja $d_k^{(i)}$ o número de vizinhos infectados ao nó k no estado i, e seja $x_k^{(i)}$ a k-ésima entrada do vetor de estados, no qual o vetor esteja no estado i do processo Markoviano.

A entrada na linha i e coluna j de Q, $q_{i,j}$, é dada conforme a seguir:

$$q_{i,j} = \begin{cases} -\sum\limits_{\substack{j=1 \\ j\neq i}}^{2^N} q_{i,j}, & \text{se } i = j \\ \lambda\gamma^{d_k^{(i)}}, & \text{se } x_k^{(i)} = 0, x_k^{(j)} = 1, x_l^{(i)} = x_l^{(j)}, \text{ para } l \neq k \\ \mu, & \text{se } x_k^{(i)} = 1, x_k^{(j)} = 0, x_l^{(i)} = x_l^{(j)}, \text{ para } l \neq k \\ 0, & \text{caso contrário} \end{cases} \tag{1}$$

4.5. Modelo de ameaça

Consideramos um atacante com uma capacidade de ataque igual a Λ infecções por unidade de tempo. Em um ajuste simplista, tal qual a capacidade total dividido pelo número total de nós suscetíveis, obtemos uma taxa de infecção por nó de $\lambda = \Lambda/N$.

Mais genericamente, podemos permitir que Λ seja escrita como uma função de N, e λ assume uma forma mais generica de função dada por Λ e N, temos que $\lambda(\Lambda, N)$ denota a taxa de infecção exógena por nó.

4.6. Distribuição de probabilidades dos estados em regime estacionário

A distribuição de probabilidades dos estados em regime estacionário é dada por [Zhang and Moura 2014, Zhang et al. 2017].

$$\pi(\boldsymbol{x}) \;=\; \frac{\tilde{\pi}(\boldsymbol{x})}{Z} \tag{2}$$

onde

$$\tilde{\pi}(\boldsymbol{x}) = \left(\frac{\lambda}{\mu}\right)^{1^T\boldsymbol{x}} \gamma^{\boldsymbol{x}^T A\boldsymbol{x}/2}\,, \boldsymbol{x} \in \mathcal{X} \tag{3}$$

e $Z = \sum_{\boldsymbol{x}\in\mathcal{X}} \tilde{\pi}(\boldsymbol{x})$.

O número de nós infectados no estado $\boldsymbol{x}$ é dado por $1^T\boldsymbol{x} = \sum\limits_{k=1}^{N} x_k$. O número de ligações (arestas) onde os dois nós vizinhos estão contaminados (arestas contaminadas) é dado por $\frac{1}{2}\boldsymbol{x}^T A\boldsymbol{x} = \frac{1}{2}\sum\limits_{k=1}^{N}\sum\limits_{\substack{l=1 \\ l\neq k}}^{N} x_k x_l a_{kl}$.

5. Análise do modelo

A seguir, focamos em uma topologia totalmente conectada. Neste caso, por simetria temos que

$$\tilde{\pi}(i) = \binom{N}{i} \left(\frac{\lambda(N)}{\mu}\right)^i \gamma^{i(i-1)/2}, \quad i = 0, \ldots, N \tag{4}$$

A probabilidade de infecção de um nó escolhido de forma aleatoria e uniforme é

$$\rho(N) = \frac{1}{N} \sum_{i=0}^{N} i \frac{\tilde{\pi}(i)}{Z} \tag{5}$$

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

$$\hat{\rho}(N) = \frac{1}{N} \sum_{i=0}^{N} i \frac{\hat{\pi}(i)}{\hat{Z}} \tag{6}$$

$$\hat{\pi}(i) = \binom{N}{i} \left(\frac{\lambda(N)}{\mu} \gamma^{N^\star/2}\right)^i \tag{7}$$

$$\hat{Z} = \sum_{i=0}^{N} \hat{\pi}(i) \tag{8}$$

onde $N^\star(N)$ é uma função crescente de N, que denotamos simplesmente por $N^\star$ para simplificar a notação. Nos referimos a o 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 5.1. *No modelo binomial, temos que*

$$\hat{\rho}(N) = \frac{1}{1 + \mu/(\lambda(N)\gamma^{N^\star/2})} \tag{9}$$

Demonstração. O resultado é fruto de manipulações algébricas,

$$\hat{Z}\hat{\rho}(N) = \frac{1}{N} \sum_{i=0}^{N} i \binom{N}{i} \left(\frac{\lambda(N)}{\mu} \gamma^{N^\star/2}\right)^i = \sum_{i=1}^{N} \binom{N-1}{i-1} \left(\frac{\lambda(N)}{\mu} \gamma^{N^\star/2}\right)^i \tag{10}$$

$$= \left(\frac{\lambda(N)}{\mu} \gamma^{N^\star/2}\right) \left(1 + \frac{\lambda(N)}{\mu} \gamma^{N^\star/2}\right)^{N-1} \tag{11}$$

O resultado segue a partir da obtenção, de forma similar, da expressão de $\hat{Z}$. □

O resultado acima pode ser usado, por exemplo, para caracterizarmos os pontos de equilíbrio do sistema.

Teorema 5.1. *O modelo binomial admite no máximo dois equilíbrios interiores ao considerarmos um custo de vacinação constante, desde que $\gamma > 1$, $\partial N^\star / \partial N$ seja positivo e não decrescente e $\lambda(N)$ decrescente.*

Demonstração. Seja $\tau(N) = (\lambda(N)/\mu)\gamma^{N^\star/2}$. Então, pelo lema acima, temos que $\partial\hat{\rho}(N)/\partial N = (\partial\tau/\partial N)/(\tau^2(1+1/\tau)^2)$. Claramente, todos os termos de $\partial\hat{\rho}(N)/\partial N$ são positivos, com exceção de $\partial\tau/\partial N$. Temos que

$$\frac{\partial\tau}{\partial N} = \lambda(N)\gamma^{N^\star/2}\left(\frac{1}{2}\log\gamma\frac{\partial N^\star}{\partial N} + \frac{\lambda'(N)}{\lambda(N)}\right) \tag{12}$$

Se $\frac{\partial N^\star}{\partial N}$ for positivo e não decrescente, e $\lambda(N)$ for decrescente, a expressão acima admite um único zero. Assim, a função $\hat{\rho}(N)$ possui no máximo um ponto de mínimo interno, e por isso cruza qualquer linha horizontal em no máximo dois pontos. □

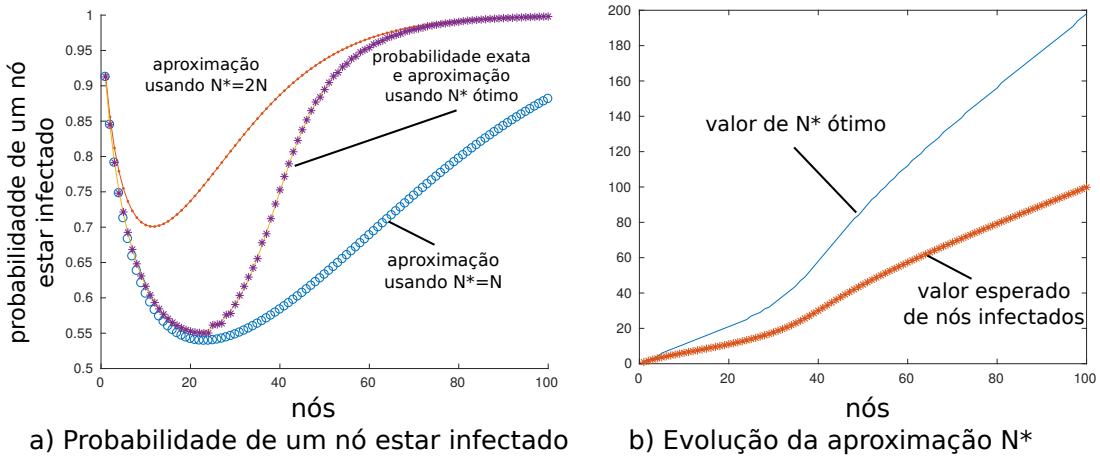
O resultado acima está de acordo com a ilustração apresentada nas Figuras 2 e 3. Segundo a Figura 2, em todos os casos em que $\gamma > 1$, a probabilidade de um nó estar infectado primeiro diminui e depois aumenta, ou simplesmente sempre aumenta. Conforme discutido na Figura 3, a um ponto de mínimo correspondem dois equilíbrios, um estável e um instável.

Caso especial: $\lambda(N) = \Lambda/N$ O modelo proposto é factível de análise em fórmula fechada para vários casos especiais da função $\lambda(N)$. Para fins de ilustração, consideramos o caso especial em que $\lambda(N) = \Lambda/N$. Este caso corresponde a um atacante que tem poder de ataque (*budget*) constante igual a Λ infecções por segundo, e divide esse poder entre os N nós da rede. Nesse caso, $\lambda'(N) = -\Lambda/N^2$. Assumindo para fins de simplificação que $N^\star = N$, temos

$$\frac{\partial}{\partial N}\hat{\rho}(N) = \kappa\left(\frac{1}{2}\log\gamma - \frac{1}{N}\right) \tag{13}$$

onde κ é uma constante positiva. Podemos verificar que $\frac{\partial}{\partial N}\hat{\rho}(N) = 0$ quando $N = (2/\log\gamma)$. No caso de $\gamma = 1.09$ encontramos o valor crítico quando $N \approx 23$, que está de acordo com aquele apresentado na Figura 2.

Note também que podemos obter fórmulas fechadas para os pontos de equilíbrio. Para tal, seja C o custo de aplicação da contramedida. Então, os pontos de equilíbrio interno são pontos tais que $\hat{\rho}(N) - C = 0$. Assumindo para fins de simplificação que $N^\star = N$, temos que os valores de N que satisfazem a equação de equilíbrio são $N = \frac{-2}{\log\gamma}W\left(\frac{\Lambda(C-1)\log\gamma}{2C\gamma^{1/2}}\right)$ onde $W(x)$ é a função de Lambert, que admite dois valores reais, correspondentes aos ramos -1 e 0. No caso de $\Lambda = 10$, $\gamma = 1,09$ e $C = 0.6$, por exemplo, temos os valores de N correspondentes aos ramos -1 e 0 dados por 45,6 e 9,7. O segundo equilíbrio condiz com o resultado da Figura 3, enquanto que o primeiro foi superestimado. Tal fato deve-se à simplificação de que $N^\star = N$, conforme ilustrado na Figura 4, podendo o resultado ser melhor aproximado em refinamentos sucessivos do modelo. A Figura 4(a) mostra a probabilidade de um nó estar infectado, em função do número de nós na rede. A aproximação $N^\star = N$ captura bem o comportamento do sistema antes de $\rho(N)$ atingir o seu ponto de mínimo. Depois deste ponto, é necessário ajustar $N^\star$ para o seu valor ótimo, que é crescente e sempre maior que N como ilustrado na Figura 4(b) (satisfazendo os critérios do Teorema 5.1). De forma mais geral, cabe destacar que o fato de a função de Lambert possuir dois ramos reais está de acordo com a constatação de que o sistema admite no máximo dois equilíbrios internos (vide Teorema 5.1).



a) Probabilidade de um nó estar infectado b) Evolução da aproximação N*

Figura 4. Validação da solução aproximada pelo modelo binomial ($\gamma = 1,09$).

6. Trabalhos relacionados

A literatura sobre epidemias é vasta. Entretanto, não é de nosso conhecimento nenhum trabalho anterior que tenha avaliado externalidade negativas resultantes da aplicação de contramedidas, usando um modelo de disseminação de epidemias como o SIS.

Em [Maillé et al. 2011], por exemplo, os autores discutem as motivações financeiras e econômicas por trás dos interesses dos hackers na elaboração de *malware*. Os autores partem do princípio de que quanto mais difundida uma vulnerabilidade, maior o estímulo para tomar proveito da mesma. Neste trabalho, entretanto, focamos no cenário oposto, que consiste em um atacante que busca promover ataques direcionados, se aproveitando de vulnerabilidades que foram sanadas pela maioria da população mas que ainda contem com vítimas fáceis.

Embora os modelos clássicos epidemiológicos sejam usados para explicar comportamentos biológicos observáveis na natureza, a propagação de programas maliciosos (*malwares*) em rede de computadores precisa capturar as contaminações intencionais e direcionadas. Neste artigo, uma solução usada para capturar essa característica foi adicionar ao modelo biológico mais uma forma de contaminação, que chamamos de contaminação exógena, tal como em [Altman et al. 2014, Zhang and Moura 2015, Van Mieghem and Cator 2012].

Em [Grossklags et al. 2008], os autores consideram externalidades positivas, por exemplo, advindas da instalação de *honey pots* em uma rede. Tal externalidade é similar em essência a ideia de que quanto mais nós suscetíveis, menores as chances de um nó escolhido de forma uniforme e aleatoria ser infectado. Embora o presente trabalho tenha semelhanças em termos de suas conclusões com [Maillé et al. 2011], os autores deste último negligenciaram aspectos epidemiológicos, que são o foco do presente estudo.

7. Conclusões

O campo da epidemiologia demonstra que quanto maior o número de indivíduos imunes, menor é a chance de que uma contaminação se espalhe na rede. Quando há muitos indivíduos que aplicaram uma contramedida, os indivíduo restantes têm menos incetivo

para aplicar a contramedida (evitando a multidão). Contudo, com a caracterização de um atacante externo com capacidade limitada, argumentamos que é possível que o número menor de indivíduos que não aplicaram a contramedida torne-os alvos preferidos dos atacantes externos. Isso faz com que os poucos indivíduos que não aplicaram algum tipo de contramedida sejam incentivados a fazê-lo (seguindo a multidão). Para capturar tais efeitos, propusemos um modelo analítico, e analisamos os seus pontos de equilíbrio. Numericamente, indicamos os regimes que caracterizam situações de evitar e seguir a multidão, e evidenciamos que tais regimes de fato ocorrem na prática, usando dados reais providos pela plataforma Shodan sobre o comportamento de *patching* de usuários na Internet.

Referências

Altman, E., Avritzer, A., El-Azouzi, R., Menasche, D. S., and de Aguiar, L. P. (2014). Rejuvenation and the spread of epidemics in general topologies. In *ISSREW*, pages 414–419. IEEE.

Antonakakis, M., April, T., Bailey, M., Bernhard, M., Bursztein, E., Cochran, J., Durumeric, Z., Halderman, J. A., Invernizzi, L., Kallitsis, M., et al. (2017). Understanding the mirai botnet. In *USENIX Security Symposium*.

Durumeric, Z., Wustrow, E., and Halderman, J. A. (2013). Zmap: Fast internet-wide scanning and its security applications. In *Usenix Security*, volume 2013.

Grossklags, J., Christin, N., and Chuang, J. (2008). Secure or insure?: a game-theoretic analysis of information security games. In *WWW*, pages 209–218. ACM.

Hayel, Y., Trajanovski, S., Altman, E., Wang, H., and Van Mieghem, P. (2014). Complete game-theoretic characterization of sis epidemics protection strategies. In *Decision and Control (CDC), 2014 IEEE 53rd Annual Conference on*, pages 1179–1184. IEEE.

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

Kephart, J. O. and White, S. R. (1991). Directed-graph epidemiological models of computer viruses. In *Research in Security and Privacy*, pages 343–359. IEEE.

Lee, B. (2017). Ransomware: Unlocking the lucrative criminal business model. Technical report, Palo Alto Networks. `https://www.paloaltonetworks.com/`.

Libster, R. (2017). The power of herd immunity. TED Talk.

Maillé, P., Reichl, P., and Tuffin, B. (2011). Interplay between security providers, consumers, and attackers: A weighted congestion game approach. In *GameSec*, pages 67–86. Springer.

Murray, W. H. (1988). The application of epidemiology to computer viruses. *Computers & Security*, 7(2):139–145.

Van Mieghem, P. and Cator, E. (2012). Epidemics in networks with nodal self-infection and the epidemic threshold. *Physical Review E*, 86(1):016116.

Wang, B., Li, X., de Aguiar, L. P., Menasche, D. S., and Shafiq, Z. (2017). Characterizing and modeling patching practices of industrial control systems. *POMACS*, 1(1):18.

Zhang, J. and Moura, J. M. (2014). Diffusion in social networks as sis epidemics: Beyond full mixing and complete graphs. *IEEE Journal of Selected Topics in Signal Processing*, 8(4):537–551.

Zhang, J. and Moura, J. M. (2015). Contact process with exogenous infection and the scaled sis process. *arXiv preprint arXiv:1507.00396*.

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