Vilc Queupe Rufino

Modelos Analíticos de Segurança Cibernética

Rio de Janeiro, Brasil

2019, v-1.0.0

Vilc Queupe Rufino

Modelos Analíticos de Segurança Cibernética

Qualificação de Doutorado a ser submetida à banca de Corpo Docente selecionada pelo Programa de Pós-Graduação em Informática, Instituto Tércio Pacitti de Aplicações e Pesquisas Computacionais, instituto de Matemática da Universidade Federal do Rio de Janeiro. Área de Concentração Informática.

Universidade Federal do Rio de Janeiro – UFRJ

Instituto de Matemática

Instituto Tércio Pacitti de Aplicações e Pesquisas Computacionais

Programa de Pós-Graduação em Informática

D.Sc. Daniel Sadoc Menasché

D.Sc. Josefino Cabral Melo Lima

Rio de Janeiro, Brasil

2019, v-1.0.0

Vilc Queupe Rufino
Modelos Analíticos de Segurança Cibernética/ Vilc Queupe Rufino. – Rio de Janeiro, Brasil, 2019, v-1.0.0-
0p. : il. (algumas color.) ; 30 cm.

D.Sc. Daniel Sadoc Menasché

Tese (Doutorado) – Universidade Federal do Rio de Janeiro – UFRJ
Instituto de Matemática
Instituto Tércio Pacitti de Aplicações e Pesquisas Computacionais
Programa de Pós-Graduação em Informática, 2019, v-1.0.0.

1. Epidemias. 2. Segurança da Informação. 3. Atacante estratégico. I. Orientador D.Sc. Daniel Sadoc Menasché. II. Universidade Federal do Rio de Janeiro. III. Programa de Pós-Graduação em Informática. IV. Modelos Analíticos de Segurança Cibernética

ERRATA

Folha	Linha	Onde se lê	Leia-se

Vilc Queupe Rufino

Modelos Analíticos de Segurança Cibernética

Qualificação de Doutorado a ser submetida à banca de Corpo Docente selecionada pelo Programa de Pós-Graduação em Informática, Instituto Tércio Pacitti de Aplicações e Pesquisas Computacionais, instituto de Matemática da Universidade Federal do Rio de Janeiro. Área de Concentração Informática.

Trabalho aprovado. Rio de Janeiro, Brasil, 18 de abril de 2019:

D.Sc. Daniel Sadoc Menasché
Orientador

D.Sc. Josefino Cabral Melo Lima
Coorientador

D.Sc. Claudio Miceli de Farias
PPGI-UFRJ

D.Sc.Edmundo Albuquerque de Souza e Silva
PESC-UFRJ

D.Sc Alberto Avritzer
ESULABSOLUTIONS

Rio de Janeiro, Brasil
2019, v-1.0.0

Este trabalho é dedicado às crianças adultas que,
quando pequenas, sonharam em se tornar cientistas.

AGRADECIMENTOS

Os agradecimentos principais são direcionados à Gerald Weber, Miguel Frasson, Leslie H. Watter, Bruno Parente Lima, Flávio de Vasconcellos Corrêa, Otavio Real Salvador, Renato Machnievscz[1] e todos aqueles que contribuíram para que a produção de trabalhos acadêmicos conforme as normas ABNT com LaTeX fosse possível.

Agradecimentos especiais são direcionados ao Centro de Pesquisa em Arquitetura da Informação[2] da Universidade de Brasília (CPAI), ao grupo de usuários *latex-br*[3] e aos novos voluntários do grupo *abnTeX2*[4] que contribuíram e que ainda contribuirão para a evolução do abnTeX2.

[1] Os nomes dos integrantes do primeiro projeto abnTeX foram extraídos de http://codigolivre.org.br/projects/abntex/

[2] http://www.cpai.unb.br/

[3] http://groups.google.com/group/latex-br

[4] http://groups.google.com/group/abntex2 e http://www.abntex.net.br/

"Não vos amoldeis às estruturas deste mundo,
mas transformai-vos pela renovação da mente,
a fim de distinguir qual é a vontade de Deus:
o que é bom, o que Lhe é agradável, o que é perfeito.
(Bíblia Sagrada, Romanos 12, 2)

RESUMO

Todos os dias sistemas são comprometidos por códigos maliciosos e participam de campanhas de ataques a sistemas computacionais. 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 trabalho, apresentamos uma proposta para carcterização do estado estacionário de modelos epidêmicos em que o atacante é estratégico e tem uma capacidade de ataque finita. Para tanto, é analisado os estados mais prováveis do modelo, indicando suas propriedades; apresenta-se fórmulas fechadas que aproximam a probabilidade de infecção e contrastamos os *insights* do modelo com simulações. Simulações suportam qualitativamente as observações do modelo epidêmico e estendem a análise permitindo distribuições gerais.

Palavras-chave: epidemia, contaminação em redes, segurança da informação.

ABSTRACT

Every day systems are compromised by malicious code and participate in campaigns to attack computer systems. Thus, users face a dilemma as to which countermeasures to take: harsh (for example, vaccination), mild (for example, restart and rejuvenation) or no countermeasures. To resolve this dilemma, one option is to take advantage of analytical models. In this work, we present a proposal to characterize the steady state of epidemic models in which the attacker is strategic and has a finite attack capacity. For that, the most probable states of the model are analyzed, indicating their properties; closed formulas are presented that approximate the probability of infection and we contrast the *insights* of the model with simulations. Simulations qualitatively support the observations of the epidemic model and extend the analysis by allowing general distributions.

Keywords: latex. epidemic, network contamination, information security.

RÉSUMÉ

Chaque jour, les systèmes sont compromis par un code malveillant et participent à des campagnes pour attaquer les systèmes informatiques. Ainsi, les utilisateurs sont confrontés à un dilemme quant aux contre-mesures à prendre: sévères (par exemple, vaccination), douces (par exemple, redémarrage et rajeunissement) ou aucune contre-mesure. Pour résoudre ce dilemme, une option consiste à tirer parti des modèles analytiques. Dans ce travail, nous présentons une proposition pour caractériser l'état stationnaire des modèles épidémiques dans lesquels l'attaquant est stratégique et a une capacité d'attaque finie. Pour cela, les états les plus probables du modèle sont analysés, indiquant leurs propriétés; des formules fermées sont présentées qui approchent la probabilité d'infection et nous contrastons les *insights* du modèle avec des simulations. Les simulations soutiennent qualitativement les observations du modèle épidémique et étendent l'analyse en permettant des distributions générales.

Mots-clés: épidémie, contamination du réseau, sécurité de l'information.

LISTA DE ILUSTRAÇÕES

LISTA DE QUADROS

LISTA DE TABELAS

LISTA DE ABREVIATURAS E SIGLAS

ABNT	Associação Brasileira de Normas Técnicas
abnTeX	ABsurdas Normas para TeX

LISTA DE SÍMBOLOS

Γ	Letra grega Gama
Λ	Lambda
ζ	Letra grega minúscula zeta
$\in$	Pertence

SUMÁRIO

1 INTRODUÇÃO

Motivação Epidemias em rede de computadores são onipresentes, todos os dias sistemas são comprometidos por códigos maliciosos que executam ações sem consentimento do seu dono legítimo (*botnets*), os quais participam de atividades criminosas, como a espionagem industrial, sequestro de dados e transmissão de pornografia infantil, como também de atos de guerra, provocados por conflitos de estados ou grupos ideológicos, como ataques de negação de serviço (*DoS*) e transmissão de informações falsas por meio de usuários legítimos. Tais tipos de 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); em todo caso, há um adversário (ou grupo) que intencionalmente faz as contaminações de seus códigos maliciosos (*malwares*) para que possam exercer seu controle remotamente.

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 e riscos que os sistemas assumem por sua estratégia de aplicação de contramedidas.

Objetivo e metodologia Neste trabalho, o objetivo é apresentar a proposta de tese de doutorado para desenvolvimento de modelos analíticos que sirvam de base para a análise de segurança em redes de computadores. Para cumprir tal objetivo, nesta qualificação é proposta a metodologia de: caracterização de comportamento dos nós em rede, em função da probabilidade de infecção, como em (**??**), da resposta resposta dos sistemas distribuídos em função dos ataques, como (**??**), e da medida de relevância de nós em função das métricas estabelecidas (centralidade), que é a proposta de integração dos trabalhos anteriores. Deste modo, a proposta é estabelecer formas de caracterização e estabelecimento de medidas que vão desde a infecção em sua origem, até os efeitos que esses ataques exercem sobre os sistemas computacionais em rede, que auxiliem a tomada de decisão e relevância do investimento em contramedidas eficazes de acordo com as ameaças e seus possíveis efeitos.

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 pode ser estratégico, com alguma capacidade limitada, o qual 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. Quando se trata de sistemas distribuídos, heterogêneos, tal como as nuvens, as informações disponíveis sobre os sistemas hospedados são superficiais, e neste cenário é possível coexistir ameaças e vulnerabilidades. Portanto a pergunta que desejamos responder é: poderia a configuração (capacidades de processamento, armazenamento, comunicação, estratégia de contramedidas) fornecida por um provedor de nuvem ou centro de dados, auxiliar na segurança dos sistemas computacionais? E quais são as métricas necessárias, para a tomada de decisão?

Contribuições: (*i*) **Análise do modelo analítico de epidemias** análise do modelo analítico de epidemias proposto em (**??**), indicando forma de parametrizá-lo e analisando seus estados mais prováveis; (*ii*) **Fórmulas fechadas para probabilidade de infecção** obtemos, via método de Newton, fórmulas fechadas para aproximar 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. (*iv*) **Caracterização da resposta de sistemas distribuídos** por meio de modelos de filas M/G/K, apresentamos a resposta de sistemas distribuídos para diferentes configurações e regimes de trabalho, inclusive sob ataques de negação de serviço por meio do consumo dos recursos disponíveis.

Organização o restante deste texto está organizado da seguinte forma. O Capítulo **??** descreve o sistema em questão, seguido pela caracterização de usuários reais na Capítulo **??**, e o modelo na Capítulo **??** com algumas fórmulas fechadas. Apresentamos simulações na Capítulo **??**. Finalmente, trabalhos relacionados e conclusão vêm nos Capítulos **??** e **??**.

2 DESCRIÇÃO DO SISTEMA

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

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 (**??**). As vulnerabilidades exploradas no protocolo SMBv2 haviam sido divulgadas e corrigidas pela Microsoft ainda em março. Em (**??**) 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 (**??**) 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

2.3.1 Contramedidas moderadas (Reinicialiação)

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 são 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.*

2.3.2 Contramedidas rigorosas (Vacina)

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.

2.3.3 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.*

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 impoacto 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.

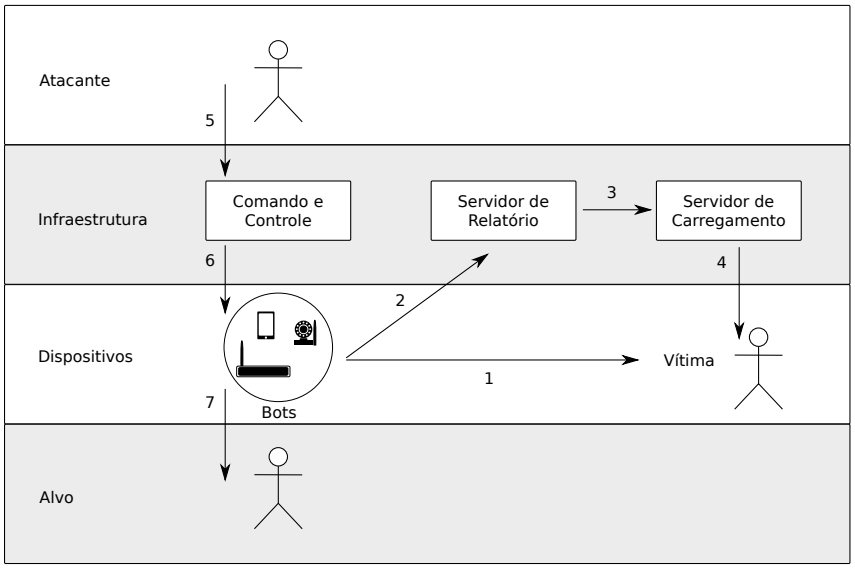


Figura 1 – Operação do Mirai.

2.5 Ciclo de operação do código malicioso *Mirai Botnet*

O Mirai é um um código malicioso que ficou conhecido por executar o maior ataques de negação de serviço conhecido até 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.*

3 EVIDÊNCIAS DOS COMPORTAMENTOS SEGUIR OU EVITAR A MULTIDÃO

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 (**??**) sobre a aplicação de *patches* por parte dos usuários. De acordo com (**??**), 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.

3.1 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.

3.2 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 **??** apresenta o resultado da observação do comportamento de *seguir* a multidão (S), *evitar* a multidão (N) 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 (1ºpop) e ao dia no qual ocorreu a medição com maior número de sistemas utilizando a respectiva versão (topo 1º) [1] e as duas

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

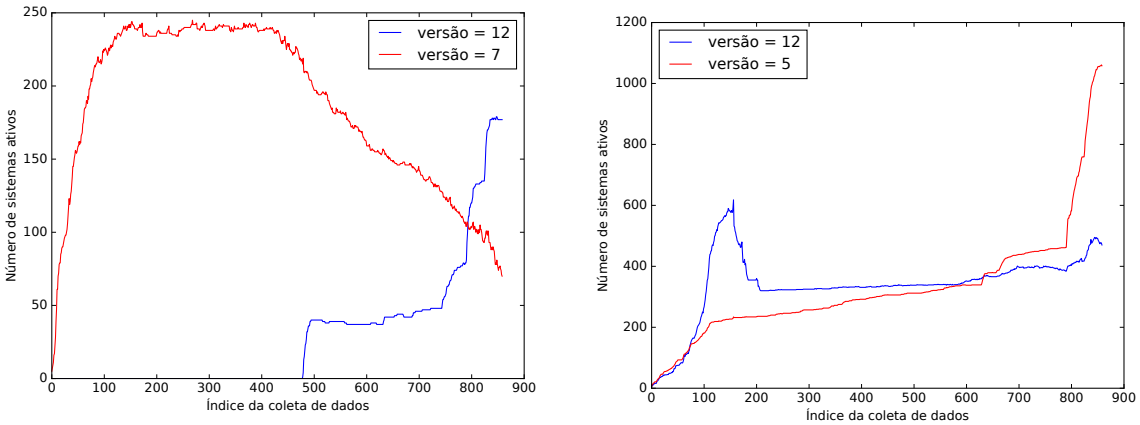
próximas colunas referem-se a segunda versão mais popular. A coluna intitulada S/E/I indica o comportamento típico da população. Para fazer a classificação do comportamento, definimos o índice ι=(topo 1º- topo 2º) / (topo 1º+ topo 2º). O caso $\iota > 0,01$ (resp., $\iota < -0.01$) corresponde ao comportamento de *seguir* (resp., *evitar*) a multidão. Valores entre $-0,01$ e $0,01$ são associados a comportamentos indefinidos.

1ºpop	topo 1º	2ºpop	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 **??** 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, 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 2 – a) Comportamento de seguir a multidão : sistema Allegro RomPage. b)Comportamento de evitar a multidão: sistema Dropbear sshd

4 MODELO EPIDÊMICO

A seguir, analisamos o modelo epidemiológico SIS multiplicativo, conforme apresentado em (**??**). As Seções **??** e **??** descrevem o modelo. A Seção **??** 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.

Tabela 2 – Tabela de notação

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	-

4.1 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 (**??**).

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 (**??**),

$$\pi(\boldsymbol{x}) \quad = \quad \frac{\tilde{\pi}(\boldsymbol{x})}{Z} \tag{4.1}$$

onde

$$\tilde{\pi}(\boldsymbol{x}) = \left(\frac{\lambda}{\mu}\right)^{1^T \boldsymbol{x}} \gamma^{\boldsymbol{x}^T A \boldsymbol{x}/2} \ , \ \boldsymbol{x} \in \mathcal{X}, \qquad Z = \sum_{\boldsymbol{x} \in \mathcal{X}} \tilde{\pi}(\boldsymbol{x}). \tag{4.2}$$

A Tabela **??** 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:

$$\pi(\iota) = \frac{\tilde{\pi}(\iota)}{Z}, \qquad \tilde{\pi}(\iota) = \binom{N}{\iota} \left(\frac{\lambda(N)}{\mu}\right)^{\iota} \gamma^{\iota(\iota-1)/2}, \quad \iota = 0, \ldots, N. \tag{4.3}$$

O valor esperado do número de nós infectados é

$$E(I) = \sum_{\iota=0}^{N} \iota \frac{\tilde{\pi}(\iota)}{Z} = N\rho(N), \quad \rho(N) = \frac{1}{N} \sum_{\iota=0}^{N} \iota \frac{\tilde{\pi}(\iota)}{Z} \tag{4.4}$$

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

4.2 Aproximação binomial

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

$$\hat{\rho}(N) = \frac{1}{N} \sum_{\iota=0}^{N} \iota \frac{\hat{\pi}(\iota)}{\hat{Z}}, \quad \hat{Z} = \sum_{\iota=0}^{N} \hat{\pi}(\iota), \quad \hat{\pi}(\iota) = \binom{N}{\iota} \left(\frac{\lambda(N)}{\mu} \gamma^{N^\star}\right)^{\iota} \tag{4.5}$$

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.

Lemma 4.2.1. *No modelo binomial, temos que:*

$$\hat{\rho}(N) = \frac{1}{1 + \mu/(\lambda(N)\gamma^{N^\star})} \tag{4.6}$$

A demonstração do lema acima consta em (**??**), notando o pequeno ajuste de notação na nova versão do resultado, refletindo a nova definição de $N^\star$.

Como discutido na Seção **??** 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,

$$\hat{\rho} = \frac{1}{1 + \mu/(\lambda\gamma^{(N-1)\hat{\rho}})} \tag{4.7}$$

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

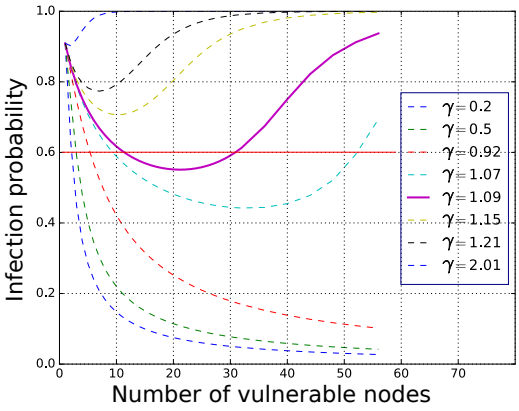


Figura 3 – Equilíbrio em relação as motivações para vacinar ou não vacinar ($\gamma = 1,09$).

4.3 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 (**??**) sobre os estados mais prováveis do modelo epidemiológico aqui discutido. Em (**??**), os autores discutem o cenário no qual a taxa de infecção exógena, λ, é constante, independente de N. Reproduzimos o principal resultado de (**??**), 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 $\boldsymbol{x}^\star = \arg\max_{\boldsymbol{x}\in\mathcal{X}} \pi(\boldsymbol{x})$. Se $\tilde{\pi}(\boldsymbol{x}^\star) \gg \tilde{\pi}(\boldsymbol{x})$, $\forall \boldsymbol{x} \in \mathcal{X}/\boldsymbol{x}^\star$, então

$$P(x_k = 1) \approx \frac{1}{1 + \mu/\left(\lambda\gamma^{m_k^\star}\right)}, \quad \text{, onde } m_k^\star = \sum_{j=1}^{N} a_{kj} x_j^\star. \tag{4.8}$$

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 $H(\boldsymbol{x}) = 1^T\boldsymbol{x}\,\log(\lambda/\mu) + \left(\boldsymbol{x}^T A\boldsymbol{x}\,\log\gamma\right)/2$. 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 (**??**) e (**??**), 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 (**??**) por $m_k^\star$, obtemos então a expressão (**??**). 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 **??** 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 **??**(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 **??**(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.

4.4 Fórmulas fechadas via método de Newton

A seguir indicamos como usar o método de Newton para achar fórmulas fechadas aproximadas para (**??**). Pelo fato de ρ aparecer tanto do lado direito quanto do lado esquerdo de (**??**), 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,

$$f(\rho) = \rho\left(1 + \frac{\mu}{\lambda}\gamma^{-\rho n}\right) - 1 = \rho + \rho\frac{\mu}{\lambda}\gamma^{-\rho n} - 1 \tag{4.9}$$

$$\frac{\partial f(\rho)}{\partial \rho} = f'(\rho) = 1 + \frac{\mu}{\lambda}\gamma^{-\rho n}\left(1 - \rho n \ln \gamma\right) \tag{4.10}$$

$$\frac{\partial^2 f(\rho)}{\partial^2 \rho} = f''(\rho) = g\frac{\mu \ln \gamma}{\lambda}\gamma^{-\rho n}\left(\rho(\ln \gamma) - 2\right) \tag{4.11}$$

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

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

$$\rho_{i+1} = \rho_i - \frac{f(\rho_i)}{f'(\rho_i)} = \frac{\lambda - \mu\gamma^{-\rho_i n}\left(\rho_i^2 n \ln \gamma\right)}{\lambda - \mu\gamma^{-\rho_i n}\left(\rho_i n \ln \gamma - 1\right)} \tag{4.12}$$

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 (**??**), iniciando com $\rho_0 = 0$, o método de Newton converge sem ultrapassar (*overshoot*) a solução.

4.5 Obtendo fórmulas fechadas

Usando a abordagem descrita acima, obtemos fórmulas fechadas 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 **??**, 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,

$$\begin{aligned}
\rho_0^{(0)} &= 0, \quad \rho_1^{(0)} = \frac{\lambda}{\lambda+\mu} \\
\rho_2^{(0)} &= \frac{\lambda - \mu\gamma^{-\rho_1 n}\left(\rho_1^2 n \ln\gamma\right)}{\lambda - \mu\gamma^{-\rho_1 n}\left(\rho_1 n \ln\gamma - 1\right)} = \frac{\lambda - \mu\gamma^{-\left(\frac{\lambda}{\lambda+\mu}\right)n}\left(\left(\frac{\lambda}{\lambda+\mu}\right)^2 n\ln\gamma\right)}{\lambda - \mu\gamma^{-\left(\frac{\lambda}{\lambda+\mu}\right)n}\left(\left(\frac{\lambda}{\lambda+\mu}\right) n\ln\gamma - 1\right)}
\end{aligned}$$

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

$$\rho_2^{(1)} = \frac{\lambda - \mu\gamma^{-\left(\frac{\lambda-\mu\gamma^{-n}(n\ln\gamma)}{\lambda-\mu\gamma^{-n}(n\ln\gamma-1)}\right)n}\left(\left(\frac{\lambda-\mu\gamma^{-n}(n\ln\gamma)}{\lambda-\mu\gamma^{-n}(n\ln\gamma-1)}\right)^2 n\ln\gamma\right)}{\lambda - \mu\gamma^{-\left(\frac{\lambda-\mu\gamma^{-n}(n\ln\gamma)}{\lambda-\mu\gamma^{-n}(n\ln\gamma-1)}\right)n}\left(\left(\frac{\lambda-\mu\gamma^{-n}(n\ln\gamma)}{\lambda-\mu\gamma^{-n}(n\ln\gamma-1)}\right) n\ln\gamma - 1\right)} \tag{4.13}$$

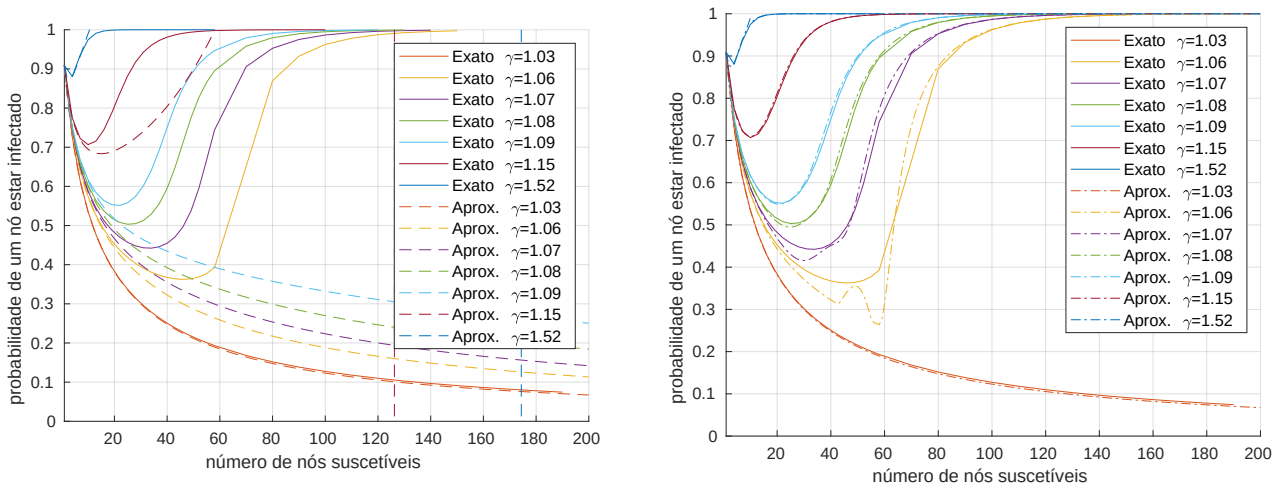
A fórmula fechada com $\rho_0 = 0$ é bem mais simples que com $\rho_0 = 1$. Conforme iremos indicar nas seções a seguir, para muitos cenários a primeira aproximação, mais simples, é suficiente.

4.5.1 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 **??**(a) e $\rho_0 = 1$ na Figura **??**(b) exceto a curva para $\gamma = 1.03$, usando os valores referência na Tabela **??**. 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.

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

$$\bar{\rho}_2^{(z)}(N) = \begin{cases} \rho_2^{(z)}(N), & \text{se } 0 \le \rho_2^{(z)}(N) \le 1 \text{ e } \bar{\rho}_2^{(z)}(N-1) \neq -\infty, \\ -\infty, & \text{caso contrário.} \end{cases} \tag{4.14}$$



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

Figura 4 – Probabilidade de infecção, calculada usando método de Newton com (a) condição inicial $\rho_0 = 0$ e (b) heurística para condição inicial.

onde z é a condição inicial, $0 \leq z \leq 1$. Segundo (**??**), 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 (**??**) deixamos explícita a dependência de ρ com relação a $N = n + 1$ (na Seção **??** tal dependência foi mantida implícita). Motivado pela discussão acima, nossa heurística é então dada por,

$$\bar{\rho}(N) = \max(\bar{\rho}_2^{(0)}(N), \bar{\rho}_2^{(1)}(N)) \tag{4.15}$$

A Figura **??**(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.

4.6 Vacinar, reiniciar ou esperar?

Tendo em vista a solução com fórmulas fechadas descrita nesta seção, podemos gerar curvas como aquela apresentada na Figura **??** 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 **??** e o custo igual a $0,5$. Segundo a Figura **??**, se o número de nós não vacinados for menor ou igual a 20, a probabilidade de infecção é alta e os nós tem 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 tem 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 fechadas para estimar a probabilidade de contaminação de nós na rede. As fórmulas fechadas 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 de propagação de *malware* construído para reproduzir o comportamento descrito no Capítulo **??**. 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.

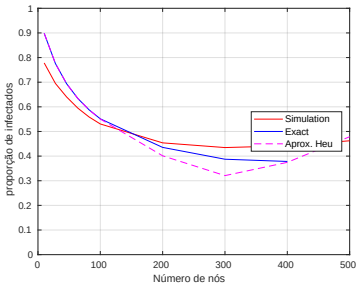
Configuração do simulador O simulador construído permite verificar o comportamento e dinâmica de uma rede, sob a perspectiva de um ataque de código malicioso tipo *Botnet Mirai*, contando com um atacante estratégico, o *Botmaster*. As contaminações se dão pelo processo de varredura, autenticação e infecção descrito no Capítulo **??**. 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 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 **??**.

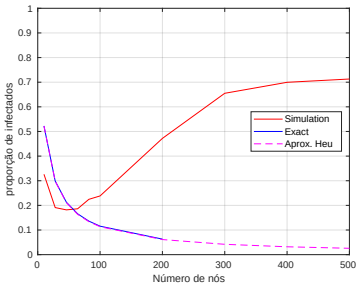
Resultados de simulação Os resultados da simulação estão sumarizados na Figura **??**, onde apresentamos a proporção média da população vulnerável infectada (vermelho) e desligada (verde) em função do números de nós vulneráveis (N). Cada curva reflete a média de oito rodadas de simulações. As linhas 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 infecatdos (linha vermelha).

Modelo analítico e simulação A Fig. **??** ilustra a probabilidade de infecção segundo o modelo analítico proposto (tanto solução exata quanto aproximada, nas curvas rosa e azul, 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

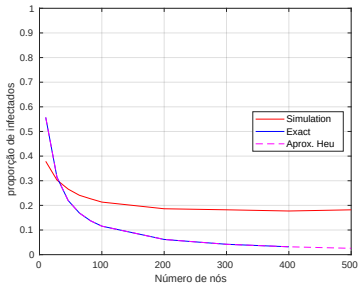
Taxa de infecção endógena | Taxa de Infecção exógena | Tempo médio ligado



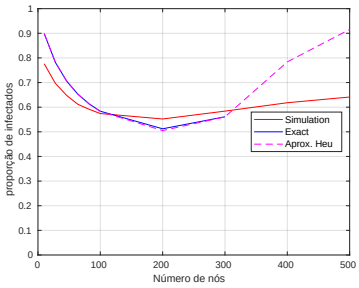
(a) $\alpha = 8 \times 10^{-5}$
$\mu = 17.959$, $\gamma = 1.0071$



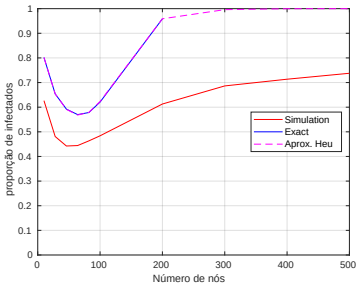
(b) $\beta = 5 \times 10^{-2}$
$\mu = 154.888$, $\gamma = 1.0262$



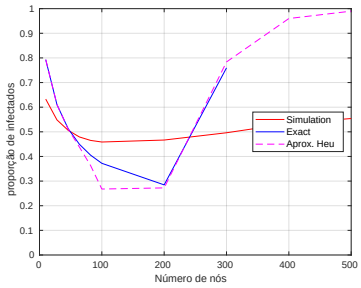
(c) $\tau = 18$
$\mu = 122.987$, $\gamma = 1.0061$



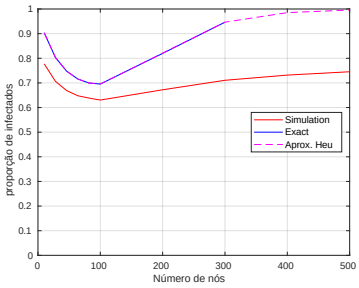
(d) $\alpha = 20 \times 10^{-5}$
$\mu = 18.160$, $\gamma = 1.0092$



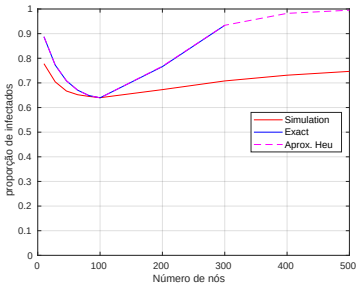
(e) $\beta = 32 \times 10^{-2}$
$\mu = 44.710$, $\gamma = 1.0262$



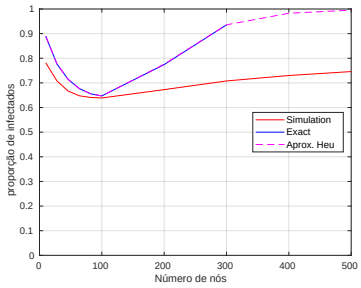
(f) $\tau = 40$
$\mu = 43.521$, $\gamma = 1.0148$



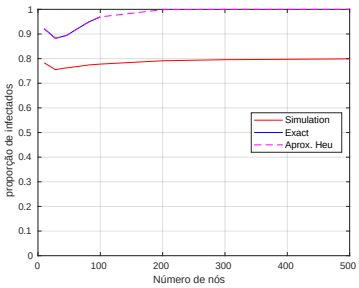
(g) $\alpha = 50 \times 10^{-5}$
$\mu = 18.046$, $\gamma = 1.0148$



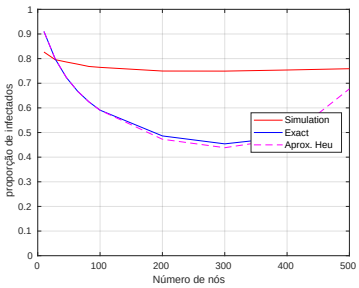
(h) $\beta = 200 \times 10^{-2}$
$\mu = 15.717$, $\gamma = 1.0071$



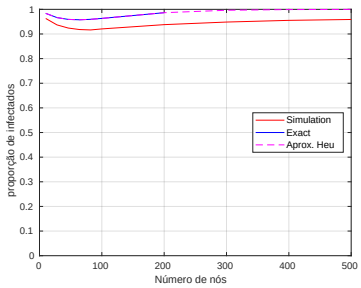
(i) $\tau = 65$
$\mu = 2.882$, $\gamma = 1.0171$



(j) $\alpha = 500 \times 10^{-5}$
$\mu = 17.459$, $\gamma = 1.0383$



(k) $\beta = 2000 \times 10^{-2}$
$\mu = 21.411$, $\gamma = 1.0148$



(l) $\tau = 260$
$\mu = 20.933$, $\gamma = 1.0148$

Figura 5 – Resultados de simulação para o comportamento da rede totalmente conectada, sob atuação da *Botnet Mirai* na presença de um atacante estratégico. Modelo analítico da Seção **??** parametrizado com $\Lambda = 1500$ e valores adicionais indicados na figura.

Tabela 3 – Parâmetros do simulador e valores de referência.

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 do período ligado (Média, Variância, ...)	Média de 65 unid. tempo
$\mathcal{D}_{\text{off}}$	Distribuição do período desligado (down-time)	Exponencial
$P_{\mathcal{D}_{\text{off}}}$	Parâmetros que definem a distribuição do período desligado (Média, Variância, ...)	Média de $0,1$ unid. tempo
Latência fim-a-fim		
$l_{\min}$ e $l_{\max}$	Latência fim-a-fim mínima e máxima, sendo latência uniformemente distribuída	$0,01$ e $0,4$
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}

segundo regime no qual o sistema é dominado por infecções endógenas.

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 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 **??**, em alguns cenários da Figura **??** 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 **??** 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. (**??**)) e à aproximação de Newton, com duas iterações, conforme heurística definida na Seção **??**. *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 endógenas e depois por infecções exógenas. Em todos os cenários apresentados na Figura **??** 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 (**??**) e (**??**). 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 **??**. Já a taxa de infecção exógena teve um aumento de 400 vezes para observarmos a variação de padrões na segunda coluna da Figura **??**.

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 **??**. 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 ANÁLISE DO MODELO

6.1 Análise do modelo

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

$$\tilde{\pi}(i) = \binom{N}{i} \left(\frac{\lambda(N)}{\mu}\right)^i \gamma^{i(i-1)/2}, \quad i = 0, \ldots, N \tag{6.1}$$

A probabilidade de infecção de um nó aleatório escolhido de uma distribuição uniforme dos nós:

$$\rho(N) = \frac{1}{N} \sum_{i=0}^{N} i \frac{\tilde{\pi}(i)}{Z} \tag{6.2}$$

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 $\hat{\pi}(i) \approx \tilde{\pi}(i)$,

$$\hat{\rho}(N) = \frac{1}{N} \sum_{i=0}^{N} i \frac{\hat{\pi}(i)}{\hat{Z}} \tag{6.3}$$

$$\hat{\pi}(i) = \binom{N}{i} \left(\frac{\lambda(N)}{\mu} \gamma^{N^\star/2}\right)^i \tag{6.4}$$

$$\hat{Z} = \sum_{i=0}^{N} \hat{\pi}(i) \tag{6.5}$$

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 *aproximação binomial*, por fazermos uso do binômio de Newton na demonstração do resultado a seguir:

Lemma 6.1.1. *No modelo binomial, temos que*

$$\hat{\rho}(N) = \frac{1}{1 + \mu/(\lambda(N)\gamma^{N^\star/2})} \tag{6.6}$$

Demonstração. O resultado é fruto de manipulações algébricas,

$$\hat{Z}\hat{\rho}(N) = \frac{1}{N} \sum_{i=0}^{N} i \binom{N}{i} \left(\frac{\lambda(N)}{\mu} \gamma^{N^\star/2}\right)^i = \sum_{i=1}^{N} \binom{N-1}{i-1} \left(\frac{\lambda(N)}{\mu} \gamma^{N^\star/2}\right)^i \tag{6.7}$$

$$= \left(\frac{\lambda(N)}{\mu} \gamma^{N^\star/2}\right) \left(1 + \frac{\lambda(N)}{\mu} \gamma^{N^\star/2}\right)^{N-1} \tag{6.8}$$

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.

Theorem 6.1.1. *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

$$\frac{\partial\tau}{\partial N} = \lambda(N)\gamma^{N^\star/2}\left(\frac{1}{2}\log\gamma\frac{\partial N^\star}{\partial N} + \frac{\lambda'(N)}{\lambda(N)}\right) \tag{6.9}$$

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 **??** e **??**. Segundo a Figura **??**, 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 **??**, a um ponto de mínimo correspondem dois equilíbrios, um estável e um instável.

Caso especial: $\lambda(N) = \Lambda/N$

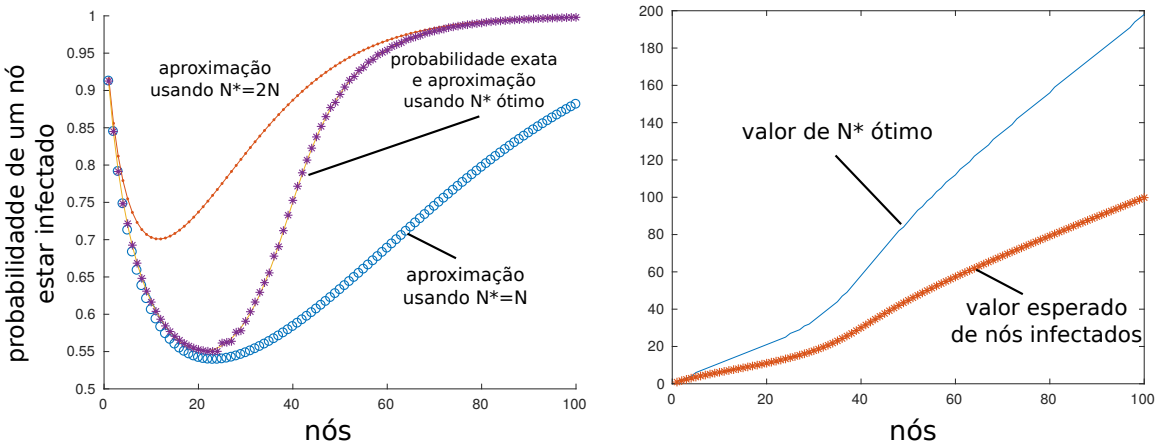
O modelo proposto é factível de análise em fórmula fechada para vários casos especiais da função $\lambda(N)$. Para fins de ilustração, consideramos o caso especial em que $\lambda(N) = \Lambda/N$. Este caso corresponde a um atacante que tem poder de ataque (*budget*) constante igual a Λ infecções por 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

$$\frac{\partial}{\partial N}\hat{\rho}(N) = \kappa\left(\frac{1}{2}\log\gamma - \frac{1}{N}\right) \tag{6.10}$$

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 **??**.

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 **??**, enquanto que o primeiro foi superestimado. Tal fato deve-se à simplificação de que $N^\star = N$, conforme ilustrado na Figura **??**, podendo o resultado ser melhor aproximado em refinamentos sucessivos do modelo. A Figura **??**(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 **??**(b) (satisfazendo os critérios do Teorema **??**). 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 **??**).



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

Figura 6 – Validação da solução aproximada pelo modelo binomial ($\gamma = 1,09$).

7 TRABALHOS RELACIONADOS

A literatura sobre modelos epidemiológicos é vasta, levando em conta aspectos transientes (**??**) e estacionários (**????**), bem como infecções endógenas e exógenas (**??????**). 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 (**??**). Em (**??**) propusemos o modelo epidemiológico analisado no presente artigo. Dentre as principais contribuições do presente trabalho, que não constavam em (**??**), destacamos três: (*i*) aproveitando-se de resultados recentes apresentados em (**??**), derivamos um método iterativo para calcular a probabilidade de infecção dos nós. Em seguida, apresentamos (*ii*) fórmulas fechadas 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 distribuídos de negação de serviço com volume capaz de afetar serviços com grande capacidade (**????**). O crescimento da Internet das Coisas (IoT) (**??**), combinado com as vulnerabilidades presentes nestes dispositivos e a dificuldade de atualizá-los criaram um ambiente propício para construção de *botnets* (**??**). 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* (**????**). 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.

8 CRONOGRAMA DE TRABALHO

O restante do trabalho será desenvolvido em cotutela com a *Université d'Avignon*, onde nossos resultados serão incrementados com desenvolvimento de formulações matemáticas que auxiliem a caracterização de medidas de centralidade em redes, com interesse em segurança da informação.

Atividades em conjunto com a "Université d'Avignon":

1. **2019 - MAI:** Revisão das definições e metodologias de centralidades em redes;
2. **2019 - JUN:** Desenvolvimento das formulações matemáticas de interesse para o problema de Segurança da Informação ;
3. **2019 - JUL** Implementação da primeira solução, usando a metodologia desenvolvida em **??**;
4. **2019 - AGO** Validação dos resultados com o mundo real, usando-se dados públicos ou dados simulados;
5. **2019 - SET** Submissão de trabalho científico.
6. **2019 - OUT** Revisar os experimentos com possíveis dados locais, possivelmente com a integração dos trabalhos;
7. **2019 - NOV e DEZ** Realização de experimentos computacionais;
8. **2020 - JAN** Submissão de trabalho científico;
9. **2020 - FEV e MAR** Escrita de Tese e revisão pelos orientadores;
10. **2020 - ABR** Submissão de Tese;
11. **2020 - MAI** Defesa de Tese.

9 CONCLUSÃO

Neste trabalho, consideramos a caracterização do processo de propagação de epidemias frente atacantes estratégicos. Em particular, considerando o modelo analítico previamente proposto em (**????**), 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.